THE UNIVERSITY OF WARWICK

September 2012 No.99

Reciprocity towards groups: a laboratory experiment on the causes

David Hugh-Jones, Martin A. Leroch

WORKING PAPER SERIES

Centre for Competitive Advantage in the Global Economy

Department of Economics

CAGE Competitive Advantage in the Global Economy

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Reciprocity towards groups: a laboratory experiment on the causes[1]

Authors: David Hugh-Jones[*], Martin A. Leroch[**]

Affiliations:
[*] CAGE, University of Warwick, D.Hugh-Jones@warwick.ac.uk
[**] University of Mainz, leroch@politik.uni-mainz.de

Abstract: Field studies of conflict report cycles of mutual revenge between groups, often linked to perceptions of intergroup injustice. We test the hypothesis that people are predisposed to reciprocate against groups. In a computerized laboratory experiment, subjects who were harmed by a partner's uncooperative action reacted by harming other members of the partner's group. This *group reciprocity* was only observed when one group was seen to be unfairly advantaged. Our results support a behavioral mechanism leading from perceived injustice to intergroup conflict. We discuss the relevance of group reciprocity to economic and political phenomena including conflict, discrimination and team competition.

1This work was supported by British Academy Small Research Grant SG101553. We thank Astrid Buba, Vittoria Levati, Werner Güth, Eva Steiger, Johannes Weisser, Ro'i Zultan, David Reinstein, Ryan McKay, Eva van den Broek, Shaun Hargreaves Heap, Brian Lickel, Arthur Lupia, Tore Ellingsen, Lorenz Goette, Alicia Melis, Michelle Brock, and seminar participants at the Max Planck Institute of Economics, University of Hamburg, University of Warwick, King's College London and IMEBE, for helpful comments; and the Max Planck Institute ESI group hiwis and administrative staff: Martin Beck, Nadine Erdmann, Adrian Liebtrau, Christian Williges, Christian Streubel, Claudia Zellmann and especially Christoph Göring.

Introduction

Laboratory experiments show that humans are prepared to reciprocate wrongs by harming the offender, even at a cost to themselves (Fehr and Gächter 2000). Field studies of ethnic and communal violence report tit-for-tat processes, with attacks on one group being avenged by attacks on (previously uninvolved) members of the attackers' group (Horowitz 1985, 2001; Chagnon 1988). For example, in Atlanta, 1906, after a spate of newspaper coverage of black attacks on white women, a group of whites went downtown to a black neighborhood and killed 25 black men (Bauerlein 2001). Similarly, after an argument between an Indian Dalit and an upper caste farmer, upper caste villagers attacked 80 Dalit families and subsequently imposed a "social boycott", forcing Dalits to leave the village (Hoff et al. 2011). Such reports suggest that humans reciprocate not only towards individuals, but also *towards entire groups*, even if this means harming innocent people.

This behavior, which we call "group reciprocity", could explain the persistence of intractable conflicts which hinder development in many of the world's poorest states (World Bank 2011). Group reciprocity may also affect the international macroeconomy. Keynes (1922), describing the Treaty of Paris' intentional devastation of the German economy, quoted Thomas Hardy: "Nought remains / But vindictiveness here amid the strong, / And there amid the weak an impotent rage. " Similar motivations might be at play between German and Greek voters in the current European crisis.

We conducted a computerized laboratory experiment to test for group reciprocity, looking both for *direct* group reciprocity in response to actions affecting oneself, and for *indirect* group reciprocity, in response to actions affecting others (Nowak and Sigmund 2005). Individual transgressions do not always lead to retaliation against an entire group, so it is important to learn when and why this

occurs. Our controlled laboratory conditions allow us to test different causal explanations in a way hard to achieve with field research. In particular, many studies suggest that intergroup inequality and injustice lead to intergroup conflict (Tambiah 1996; Horowitz 2001; Cederman et al. 2011). Group reciprocity might provide one mechanism for this causal link: in a context of intergroup unfairness, resentment at one individual's actions may spread towards that person's entire group, and lead to intergroup revenge behavior. We therefore use different treatments to vary the perceived unfairness of intergroup allocations.

Our results support the hypothesis that humans group-reciprocate. We find little evidence for indirect group reciprocity, but strong evidence for direct group reciprocity. We also confirm that there is a link with perceived unfairness. Subjects only group-reciprocated when a large reward had been arbitrarily allocated to one group, and subjects who saw the intergroup allocation as unfair were more likely to group-reciprocate.

The impact of group membership on human behavior was first studied by social psychologists (Sumner 1906; Sherif et al. 1961; Tajfel et al. 1971). More recently economists have taken up the same topic, using laboratory and field experiments. The literature shows some fairly robust results. People cooperate more with in-group members (de Cremer and van Vugt 1999; Goette et al. 2006, Guala et al. 2012). Group members give more to each other, punish each other less for misbehavior and reward each other more for good behavior (Bernhard et al. 2006; Chen and Li 2009; Currarini and Mengel 2012). Moreover, people place a value on their group membership, and prefer to interact with others from their own group (Hargreaves-Heap and Zizzo 2009; Currarini and Mengel 2012).

None of these phenomena, by itself, appears sufficient to explain historically observed episodes of extreme violence against outgroups (Brewer 1999). Indeed, the in-group bias shown in allocating goods disappears when subjects allocate a "bad" such as exposure to aversive noise (Mummendey

et al. 1992). Psychologists have therefore developed broader theories of intergroup emotions to explain how people react to outgroups; these include “vicarious retribution”, a propensity to retaliate against the group (Lickel et al. 2006). Experiments on vicarious retribution thus far have examined only verbal attitudes (Stenstrom et al. 2008), or cannot cleanly distinguish individual-level reciprocity from group reciprocity (Gaertner et al. 2008).

Some economic experiments have examined group-level outcomes of intergroup dynamics (Bornstein et al. 1992, 2003; Abbink and Herrmann 2009). Group reciprocity may play a similar explanatory role here to standard reciprocity in public goods games (Fehr et al. 2000), by providing the individual-level psychological mechanism underlying phenomena such as costly vendettas and increased effort in intergroup Prisoner’s Dilemmas.

The interpretation of experiments on reciprocity remains debated: do they show “strong reciprocity”, i.e. reciprocity as a primitive in the utility function, or only “a misapplication... of a heuristic rule ” derived from settings where intergroup reciprocity may be individually rational (Guala 2012)? Our experiment is compatible with both interpretations. We simply claim that in either case, it provides valuable evidence about human behavioral dispositions.

Experiment design

We ran 15 sessions of 16 subjects each. Each session was divided into two parts: a group quiz and a sequence of 2-player public goods games. In each session, subjects were randomly allocated into four color-coded groups of four: green, orange, purple, and brown. Each subject’s color group was shown on screen throughout the experiment. Payments were shown in Experimental Currency Units (ECU) with 1 ECU = 5 Euro cent.

After the group assignment, groups were given a 10-minute quiz consisting of 20 multiple-choice questions. Answers were chosen individually, but members of each group could communicate with

each other via online chat. (It was forbidden to communicate personal information; chat records show that all subjects followed these instructions.) Each group member's correct answer was rewarded with 10 ECU for the group. A group's earnings were divided equally among its members, and 100 ECU per subject were used as an endowment for the second part of the experiment. To ensure that each subject had at least 100 ECU, minimum group earnings of 400 ECU were implemented. This is purely a framing manipulation to increase subjects' sense that they had "earned" their ECU.

The group quiz had two functions: building group identity, by giving groups a common task and an element of common fate; and creating between-group inequality (Chen and Li 2009). To increase this and to manipulate subjects' perceptions of the fairness of intergroup allocations, we gave one group a 100 ECU bonus. In 3 "winner bonus" sessions, the group with the most points was awarded the bonus. In the remaining 12 "random bonus" sessions, this bonus was instead given to a randomly selected group. In both cases, the instructions explained how the bonus would be awarded, and the group receiving the bonus was announced directly after the quiz. We expected subjects to perceive between-group inequality as less fair when the bonus was awarded randomly.

The quiz was followed by a brief questionnaire. Subjects then played eight repetitions of a two round, two player public goods game. In each repetition, subjects were re-matched into groups of 4 with members of different colors. They could observe all group members' color, as well as their earnings from the quiz. The 4 subjects were then matched into pairs for each of the two rounds.

The basic structure of each repetition is always the same. Subjects played a public goods game with one other subject from the four. They then received feedback about one other subject's play from the four, who we label F. Lastly, they played another public goods game with one other subject from the four: the second round partner, whom we label P.

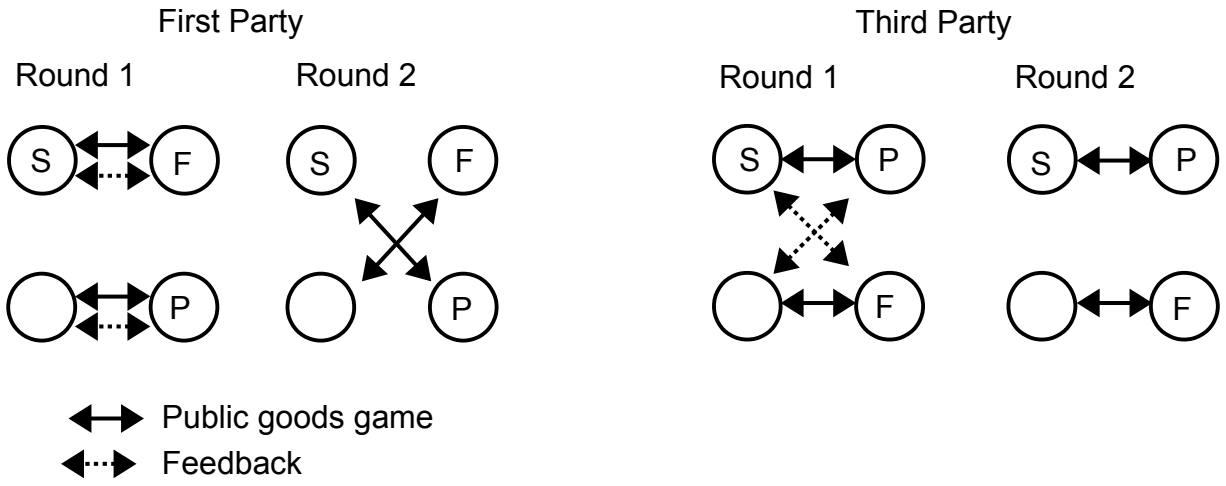


Figure 1: Experiment design. Subject is marked with an S. F denotes the feedback subject, P the second round partner

Different treatments used different matching, illustrated in Figure 1. In "first party" treatments, subjects were paired "horizontally" in the first round and received feedback on their partner's play: i.e., F was the subject's own first round partner. In the second choice, subjects were paired "diagonally". Thus, subjects learnt about how F had played *against them* in the first round, and could then react to this by playing differently against P in the second round. These treatments test for direct or "first party" group reciprocity.

In third party treatments, first choices were made with the horizontal player, but feedback was given (only) on the behavior of the diagonal player. Then subjects played the horizontal player again. Thus, subjects learnt about how F had played *against another participant* in the first round, and could react to this by playing differently against P in the second round. These treatments test for "indirect" or "third party" group reciprocity.

Because conflicts are often modelled as prisoners' dilemmas (Hardin 1995; Fearon and Laitin 1999), we chose to implement a standard linear public goods game in each round: the two subjects shared a fund of 100 ECU, where 50 tokens were taken from each of the two subjects' earnings

from the quiz. Each could then take up to 50 ECU from the fund. ECU remaining in the fund were multiplied by 1.5 and shared equally between the two. Withdrawing tokens from the fund is thus harmful for one's partner because it reduces his earnings from the quiz. In the worst case (subject withdraws nothing, partner withdraws 50 tokens), the subject (after rounding) receives only 36 tokens. The partner would in this case earn the maximum possible of 86 tokens. In case neither subject withdraws anything, both earn 75 tokens. Total earnings from one random repetition were used for payment.

Cross-cutting treatments varied the color group membership of the four subjects, as Figure 2 shows. F's action might influence subjects' play against second round partners from all groups, due to imitation or generalized reciprocity (Dufwenberg et al. 2001; Nowak and Roch 2007). For clean identification of reciprocity towards groups, in *different group* (*same group*) repetitions, F and P were from different groups (the same group). In addition, in *own same group* (*own different group*) repetitions, the "vertical" player was from the subject's own group (another group). In third party repetitions, this allows us to examine the effect of the group membership of the other participant helped or harmed by F: do subjects react more strongly when F's action affects a member of their own group? To avoid a confound with in-group altruism, subjects never shared group membership with the feedback player F, or with either of their partners in the public goods game.

After the second round, subjects learned the choices of all their partners in the repetition, and their total earnings from the repetition. They were then rematched into different groups of 4 for the next repetition. The rematching ensured that all subjects experienced all 8 treatments: {first party, third party} × {same group, different group} × {own same group, own different group}. This "within subjects" design allows more accurate inferences about group reciprocity's individual-level covariates. Treatments were balanced over repetitions.

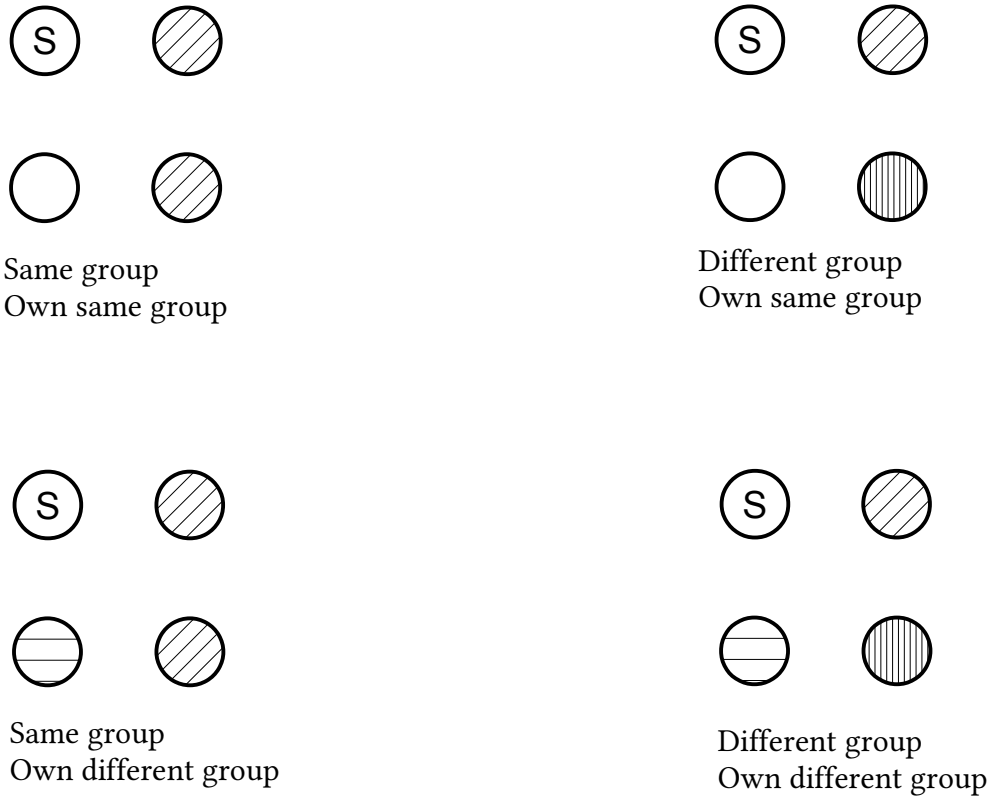


Figure 2: Color treatments. Subject is marked with an S.

After the eight repetitions, subjects answered a questionnaire including demographics, measures of group identity, reactions to other groups, and questions about the experiment. Lastly payments were made privately.

Supplementary Treatments

The feedback player F's choice might influence subjects' play either by changing preferences towards their second round partner P, or by changing expectations about P's play. To investigate this, in repetitions 2 and 7 of sessions 1-10 and 13-15, expectations about P's choice were elicited.

These were incentivized by a payment based on the difference between the guess and the true amount taken by the partner (details in appendix table A2).

In sessions 11-12, among each group of four, one player's second round amount taken was determined randomly, by a computerized draw from the uniform distribution on {0,...,50}. The player and his or her second round partner were informed of this in advance. All players were paid as normal. This allows us to examine the effect of expectations in a different way, by holding S's expectations of P's behavior constant. In order to focus on first party group reciprocity, these sessions only included first party treatments, with two repetitions of each color treatment.

To test the "contact hypothesis" that prejudice can be reduced by cooperation on a common task, in sessions 13-15 the public good games were interrupted after 4 repetitions to play an additional 5 minute quiz containing 10 questions. As before, participants could chat with each other to answer the questions. However, this time, while one color group in each session chatted only within itself, the other 3 color groups were redistributed into new groups of four, so that they chatted with members from different color groups. Color group membership was shown in the chat window. The public goods games were then continued for the 4 final rounds.

Table 1 summarizes our sessions.

Sessions	Bonus	Treatments	Notes
1-3	Winner bonus	First and third party	
4-10	Random bonus	First and third party	
11-12	Random bonus	First party only	Some choices made by computer
13-15	Random bonus	First and third party	Second “intergroup” quiz after 4 repetitions

Table 1: Experimental sessions

The experiment took place in the computer laboratory of the University of Hamburg, using the computer software zTree (Fischbacher 2007). Recruiting took place via ORSEE (Greiner 2004). 240 subjects participated on four separate days. Table 2 shows participants’ descriptive statistics, including demographics. Sessions lasted about an hour. Average earnings per session were 14.48 EUR; the maximum session average was 16.45 EUR and the minimum session average was 13.02 EUR. Individual earnings ranged from 9.40 EUR to 21.85 EUR.

Courses				
	Law	Natural sciences	Social sciences	Economics
	22	17	42	88
	Other	Not a student	No reply	
	66	2	3	
Gender				
	Male	Female	No reply	
	110	128	2	
Native German speaker				
	Yes	No		
	188	52		
Any other participants known to subject				
	Yes	No		
	18	222		
	Min	Max	Mean	Median
Profit (EUR, inc. showup fee)	9.40	21.85	14.48	13.82
Quiz earnings (ECU, inc. bonus)	0	173	48.58	27.5
Age	19	42	24.67	24
First period take	0	50	28.2	33
Second period take	0	50	29.60	35
Quiz score (out of 20)	4	18	12.20	12
Quiz 2 score (out of 10)	1	8	4.04	4

Table 2: Descriptive statistics

Results

To identify group reciprocity, we run regressions of the following form:

$$\textit{Second round take} = \alpha + \beta\, \textit{F take} + \gamma\, \textit{Same group} + \delta\, (\textit{F take} \times \textit{Same group}) + X\theta + \varepsilon. \quad (1)$$

Here, β gives the partial correlation between *F take*, the amount taken by the feedback participant F, and the subject's *Second round take* against his second round partner P, when F and P are in

different groups. The sum $\beta + \delta$ gives this correlation when F and P are in the *Same group*. δ is the extra correlation caused by group membership, i.e., the level of group reciprocity. We expect $\delta > 0$. X is a vector of possible controls.

Our key results are as follows.

Result 1: Subjects showed first party group reciprocity in random bonus sessions, but not in winner bonus sessions.

Result 2: Subjects did not group-reciprocate behavior towards third parties.

Figure 3 plots *F take* against *Second round take* in first party treatments during random bonus sessions. As expected, the slope of *Second round take* on *F take* was about doubled when F and P were from the same group.

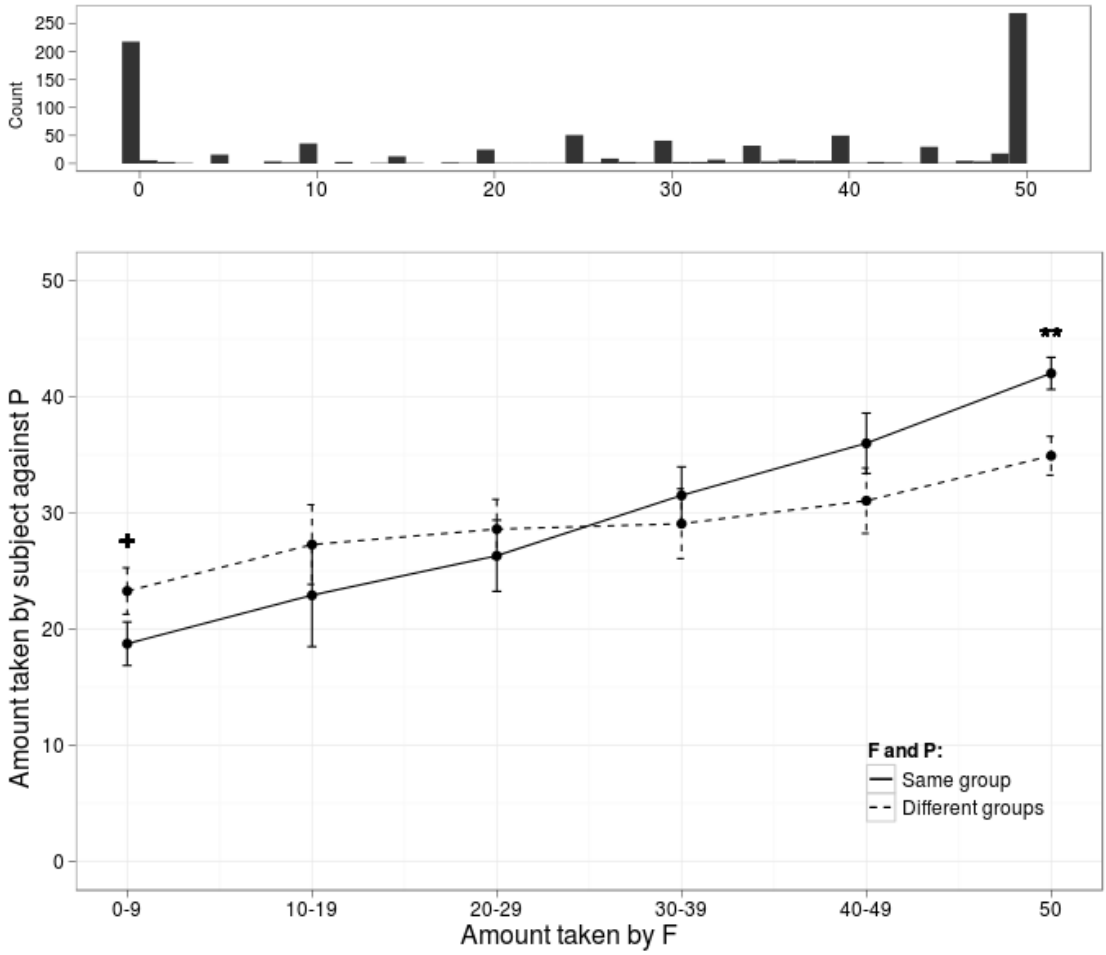


*Figure 3: Mean observed subjects' taking against P, by F's taking, first party treatments, random bonus sessions. Bars show ± 1 s.d. + $p < 0.10$; ** $p < 0.01$. Top plot is histogram of amount taken by F.*

Subjects' choices within a session may not be independent. Therefore, to test for significance, we first calculate values of β and β + δ in (1) separately for each session, and treat them as a single matched pair of observations. This procedure is analogous to running non-parametric statistics on session averages: here, instead of a session average, we are using a partial correlation. Within first party treatments, β + δ was higher than β in 11 out of 12 random bonus sessions, but in no winner bonus sessions. The null hypothesis is that β and β + δ are distributed with the same mean (equivalently, the mean δ is 0). A signed-rank test on the matched pairs rejects this over all sessions at $p = 0.0353$ (two-sided), and over random bonus sessions only at $p = 0.00928$ (two-sided). We

found no evidence for group reciprocity in third party treatments: $\beta + \delta$ was higher than β in 7 out of 10 random bonus sessions and 2 out of 3 winner bonus sessions ($p = 0.216$, two-sided, over all sessions).

Next we examine individual behavior. Table 3 estimates (1) for both first and third party treatments. Column 1 pools data from all sessions, columns 2 and 3 use random and winner bonus sessions respectively. In first party treatments, the δ coefficient on *F take* × *Same group* is positive and highly significant in random bonus sessions (and pooling the data), but not in winner bonus sessions. In third party treatments, δ is small and insignificant. These results are robust to alternative specifications.[2]

2 These include using session-clustered standard errors; adding per individual and per repetition dummies; adding controls for the history of play; and running Tobit regressions to account for the many observations at 0 and 50. The appendix reports these analyses.

	(1)	(2)	(3)
	All sessions	Random bonus sessions	Winner bonus sessions
(Intercept)	22.5	23.1	19.7
	(1.61) ***	(1.8) ***	(3.46) ***
Third party	-2.13	-2.97	1.37
	(1.96)	(2.23)	(4)
First party ✕ F take (β)	0.255	0.223	0.4
	(0.0433) ***	(0.0483) ***	(0.0947) ***
———"——— ✕ Same group (γ)	-4.02	-5.24	1.62
	(1.92) *	(2.16) *	(4.2)
———"——— ✕ F take ✕ Same group (δ)	0.17	0.229	-0.111
	(0.0588) **	(0.0656) ***	(0.13)
Third party ✕ F take (β)	0.282	0.279	0.298
	(0.0486) ***	(0.0558) ***	(0.0977) **
———"——— ✕ Same group (γ)	0.372	0.667	-0.505
	(2.09)	(2.38)	(4.32)
———"——— ✕ F take ✕ Same group (δ)	0.0412	0.0461	0.0213
	(0.0629)	(0.072)	(0.126)
N	1856	1472	384
N indiv.	240	192	48
Adj. R2	0.108	0.111	0.109

*Table 3: Estimates of group reciprocity, first and third party treatments. Dependent variable: amount taken by subjects against P in round 2 of the public goods game. Independent variables are amount taken by F in round 1 (*F take*), whether F and P were in the* Same group*, and the interaction of these variables. Robust standard errors clustered by individual in parentheses. + p < 0.10; * p < 0.05; p < 0.01; *** p < 0.001.*

To find correlates of group reciprocity, we focus on first party treatments in random bonus sessions. First, we want to check our interpretation that the random bonus sessions caused group reciprocity by making subjects feel that the inter-group distribution of income was unfair. Our post-quiz questionnaire included a 1-5 Likert scale “Did you feel that the quiz was fair?” Answers were significantly less positive in random bonus sessions (Goodman and Kruskal’s gamma, $p = 0.072$).

Result 3: Subjects who perceived the quiz as unfair showed more group reciprocity.

Column 1 of Table 4 interacts equation (1) with a dummy variable *Fair*, which is 1 if the subject perceived the quiz as (very fair or) fair and 0 otherwise. The coefficient on *F take* × *Same group* × *Fair* is negative and significant. Subjects who perceived the quiz as unfair were about three-and-a-half times more group-reciprocal (0.362 versus 0.362 - 0.251 = 0.101). Including winner bonus sessions in these regressions gives similar results[3]; controlling for fairness perceptions reduces but does not eliminate the difference between sessions.

3 Available on request.

	(1)	(2)	(3)	(4)
(Intercept)	27.1	3.46	22.3	23
	(2.61) ***	(2.02) +	(1.8) ***	(1.87) ***
F take (β)	0.11	-0.00703	0.244	0.217
	(0.0677)	(0.0511)	(0.0489) ***	(0.05) ***
Same group (γ)	-7.73	-3.22	-4.92	-5.15
	(3.11) *	(1.89) +	(2.19) *	(2.22) *
F take ✕ Same group (δ)	0.362	0.1	0.217	0.224
	(0.0929) ***	(0.0514) +	(0.0663) **	(0.0676) ***
Fair	-7.36	--	--	--
	(3.53) *			
F take ✕ Fair	0.214	--	--	--
	(0.0953) *			
Same group ✕ Fair	4.56	--	--	--
	(4.22)			
F take ✕ Same group ✕ Fair	-0.261	--	--	--
	(0.128) *			
Expectations	--	0.965	--	--
		(0.0574) ***		
Random choice	--	--	13	--
			(6.71) +	
F take ✕ Random choice	--	--	-0.345	--
			(0.189) +	
Same group ✕ Random choice	--	--	-7.01	--
			(7.89)	
F take ✕ Same group ✕ Random choice	--	--	0.225	--
			(0.233)	
Open quiz 2	--	--	--	0.327
				(5.05)
F take ✕ Open quiz 2	--	--	--	0.0832
				(0.154)
Same group ✕ Open quiz 2	--	--	--	-0.367
				(7.54)
F take ✕ Same group ✕ Open quiz 2	--	--	--	0.0101
				(0.21)
Model	Linear	Linear with multiple imputation	Linear	Linear
N	832	832	832	832
N indiv.	192	192	192	192
Adj. R2	0.136	0.479	0.13	0.127

*Table 4: Estimates of group reciprocity, first party treatments, random bonus sessions. Dependent variable: amount taken by subjects against P in round 2 of the public goods game. Independent variables are amount taken by F in round 1 (*F take*), whether F and P were in the* Same group*, and the interaction of these variables, plus further terms. Robust standard errors clustered by individual in parentheses. + p < 0.10; * p < 0.05; p < 0.01; *** p < 0.001.*

Result 4: Group reciprocity remained significant after controlling for subjects' expectations of their second round partner's choices, and also when those choices were made randomly by the computer.

The feedback participant F's actions might change the subject's expectations about other members of F's group, including P. Subjects might then group-reciprocate because of these expectations, e.g. if they wish to match P's expected take, rather than because they directly wished to harm F's group members. This would be a form of statistical discrimination (Arrow 1971), which may be important in explaining real-world group reciprocal behavior. However, capturing it in the laboratory is not very informative, because it is unclear what expectations subjects ought to hold about correlations of behavior among color group members.[4]

To see if group reciprocity was explained by expectations, column 2 of table 4 adds *Expectations* to the basic regression. To increase efficiency, we multiply impute *Expectations* for repetitions where it was not elicited. *Expectations* is highly significant, and the δ coefficient on *F take* × *Same group* is approximately halved. However, it remains weakly significant.[5]

Subjects' stated expectations may be affected by their intended play, for example due to self-justification (Messé and Sivacek 1979). Therefore, we used a further method to control for expectations. In sessions 11-12, some subjects' second round choices were made randomly by a computerized draw from the uniform distribution on {0,...,50}. Their partners knew this, and both partners were paid as normal. Subjects' expectations about P's play ought to be unaffected by F's play in this case.[6] Column 3 of Table 4 interacts equation (1) with a dummy variable *Random choice*, which is 1 when P's choice was random. The coefficient on *F take* × *Same group* × *Random choice* is actually positive, and δ plus this coefficient remains significantly positive

4 In fact, there were no significant correlations within groups. See the appendix.
5 The significance is robust to alternative specifications. See the appendix.
6 We cannot rule out "counterfactual reciprocity", based on what subjects believed P would have done.

(p=0.050). We conclude that group reciprocity was not driven solely by expectations.

Result 5: Intergroup cooperation did not reduce group reciprocity.

The "contact hypothesis", that intergroup contact can reduce prejudice, has a long pedigree within social psychology (Sherif et al. 1961, Pettigrew 1998). Subsequent research has emphasized that contact alone may not be enough; subjects may need to cooperate on a common task (Gaertner et al. 1993). We test whether group reciprocity can be decreased by an episode of cooperation between different groups. In 3 sessions, after 4 repetitions a second quiz took place, in which some subjects could cooperate with members of other groups. Column 4 of Table 4 interacts equation (1) with a dummy *Open quiz 2*, for these subjects in repetitions 5-8. There is no evidence that this prevented group reciprocity: all interaction terms are small and insignificant. These null results are not definitive, since a longer interaction between groups might have broken down group reciprocity more effectively. Nevertheless, they suggest that the tendency to reciprocate actions by outgroup members is not easy to break down.

Another interesting non-result concerns gender. Some evolutionary theories of group identification predict that men and women should possess different group psychology; in particular, men should be more coalition-minded (Sidanius and Pratto 1999). Certainly, men are more directly active in violent intergroup conflicts (Goldstein 2003). However:

Result 6: Men and women showed equal levels of group reciprocity.

Table 5 shows separate estimations of (1) for men and women. The results are strikingly similar. Whatever differentiates men and women's conflict behavior, it is not this aspect of psychology.

The appendix tests the above results in a variety of specifications, and also details some others. High-earning participants' actions appear to have caused more group reciprocity, consistent with a link between inequality and group reciprocity. Subjects with a strong in-group identity may have been more group-reciprocal.

	(1) Females	(2) Males
(Intercept)	24.5	21.5
	(2.39) ***	(2.7) ***
First party ✕ F take	0.221	0.221
	(0.0633) ***	(0.0735) **
———"——— ✕ Same group	-6.19	-4.42
	(2.72) *	(3.43)
———"——— ✕ F take ✕ Same group	0.248	0.223
	(0.0856) **	(0.1) *
Third party	-2.89	-2.8
	(2.92)	(3.41)
———"——— ✕ F take	0.28	0.26
	(0.0717) ***	(0.0866) **
———"——— ✕ Same group	-0.201	0.995
	(3.04)	(3.74)
———"——— ✕ F take ✕ Same group	0.0793	0.0467
	(0.0932)	(0.11)
Model	Linear	Linear
Controls	–	–
N	776	696
N indiv.	101	91
Adj. R2	0.128	0.0992

*Table 5: Estimates of equation (1), random bonus sessions, males and females. Standard errors in parentheses, clustered by individual. + $p < 0.10$; * $p < 0.05$; $p < 0.01$; *** $p < 0.001$.*

Conclusion

Group membership matters to social and economic behavior (Arrow 1998; Akerlof and Kranton

2005, 2010; Sen 2007). In this paper we move beyond static considerations of identity and consider how groups react to each other. Our laboratory experiment allows us to confirm hypotheses from the field in a controlled setting.

Many intractable conflicts are driven by cycles of intergroup revenge, in which uninvolved bystanders are harmed for their fellow group members' supposed actions. Some observers blame not deep intergroup hatreds but self-interested politicians and hired thugs (Brass 1997; Kaufman 2001; cf. Glaeser 2005). Our results, however, show that humans reciprocate towards groups even absent these factors; this psychology could provide the demand for conflict which politicians exploit. We also showed the conditions for the motivation to operate. Uncooperative actions triggered group reciprocity only in unfair contexts, when a randomly selected group received a large reward.

It would be of interest to know how group reciprocity might evolve. One possibility is that it developed as a mechanism for preserving intergroup peace: human ethnic groups mainly live at peace with their neighbors (Fearon and Laitin 1996), whereas e.g. chimpanzee intergroup relations appear violent by default (Wilson and Wrangham 2003). In the context of within-group alliances, primates appear to reciprocate against attackers' kin (Aureli at al. 1992), which suggests that the underlying behavioral propensity may be quite ancient. But reciprocity in the field might also be driven by strategic incentives, with subjects bringing the resulting behavioral heuristic into the laboratory (Chagnon 1988; Hardin 1995; Fearon and Laitin 1996).

We draw consequences from our research for several fields. First, experiments on ethnic heterogeneity have used one-shot interactions to reveal intergroup prejudice and discrimination (Fershtman and Gneezy 2001; Habyarimana et al. 2009). Future work should examine intergroup dynamics: how group members *react* to other groups' behavior (cf. Bornhorst et al. 2009). Second, economists of conflict should consider the possibility of endogenous preference change. For

example, counter-terrorist strategies which might deter a rational terrorist actor (Arce and Sandler 2005) may risk increasing revenge motivations and thereby terrorist recruitment.

Most importantly, there is a complex behavioral connection between perceived intergroup unfairness and conflict. Much more needs to be done to understand how this works: what contexts create breeding grounds for intergroup resentment, and how does this psychology play out in particular environments? We look forward to more research on this topic. Field experiments have examined the effect of violent conflict on preferences (Blattman 2009; Cassar et al. 2011; Voors et al. 2012); further work could explore the behavioral factors behind individuals' selection into conflict. Intergroup dynamics may also be expressed in more routine forms of discrimination, as in Shayo and Zussman's (2011) finding that both Arab and Jewish judges in Israel favour their in-group more after terrorist attacks. Group reciprocity could also affect team production and team competition in firms. For instance, employees may reduce cooperation and "shirk" when they feel that certain groups receive bonuses not available for other employees. In public economics, perceptions that certain groups act corruptly or manipulate the tax system to their own advantage may lead disadvantaged groups to behave non-cooperatively, possibly by evading taxes (cf. Alm and Torgler 2006). Ultimately, understanding intergroup dynamics could help businesses and policy-makers to manage them more productively and to forestall some of their worst effects.

References

Abbink, Klaus, Benedikt Herrmann. 2009. "Pointless Vedettas." http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1468452

Akerlof, George A., and Rachel E. Kranton. 2005. "Identity and the Economics of Organizations." *Journal of Economic Perspectives* 19 (1): 9–32.

Akerlof, George A., and Rachel E. Kranton. 2010. *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being*. Princeton University Press.

Alm, James, and Benno Torgler. 2006. "Cultural differences and tax morale in the United States and Europe." *Journal of Economic Psychology* 27: 224–246

Arce, Daniel G., Todd Sandler. 2005. "Counterterrorism: A Game-Theoretic Analysis." *Journal of Conflict Resolution,* 49(2): 183-200.

Arrow, Kenneth J. 1972. "Some mathematical models of race discrimination in the labor market." *Racial discrimination in economic life*, pp.187–203.

Arrow, Kenneth J. 1998. "What Has Economics to Say About Racial Discrimination?" *The Journal of Economic Perspectives* 12 (2) (April 1): 91–100.

Aureli, Filippo, Roberto Cozzolino, Carla Cordischi, and Stefano Scucchi. 1992. "Kin-Oriented Redirection among Japanese Macaques: An Expression of a Revenge System?" *Animal Behaviour,* 44(2): 283–291.

Baldwin, Kate, John D. Huber. 2010. "Economic Versus Cultural Differences: Forms of Ethnic Diversity and Public Goods Provision," *American Political Science Revview,* 104(4): 644-662.

Bauerlein, Marc. 2001. *Negrophobia: A Race Riot in Atlanta, 1906.* New York: Encounter Books.

Bernhard, Helen, Urs Fischbacher, and Ernst Fehr. 2006. "Parochial Altruism in Humans." *Nature,* 442(7105): 912-915.

Blattman, Christopher. 2009. "From Violence to Voting: War and Political Participation in Uganda." *American Political Science Review,* 103(2): 231–247.

Bornhorst, Fabian, Andrea Ichino, Oliver Kirchkamp, Karl H. Schlag, and Eyal Winter. 2010. "Similarities and Differences When Building Trust: The Role of Cultures." *Experimental Economics* 13 (3): 260–283.

Bornstein, Gary. 1992. "The Free-rider Problem in Intergroup Conflicts over Step-level and Continuous Public Goods." *Journal of Personality and Social Psychology* 62 (4): 597.

Bornstein, Gary. 2003. "Intergroup Conflict: Individual, Group, and Collective Interests."

Personality and Social Psychology Review 7 (2): 129–145.

Brass, Paul R. 1997. *Theft of an Idol: Text and Context in the Representation of Collective Violence.* Princeton, NJ: Princeton University Press.

Cassar, Alessandra, Pauline Grosjean, and Sam Whitt. 2011. "Lost in Transition? How Civil War Violence Can Impair the Foundations for Market Development – The Case of Tajikistan." www.cesifo-group.de/link/vsi11_ec_Cassar.pdf .

Cederman, Lars-Erik, Nils B. Weidmann, and Kristian S. Gleditsch. 2011. "Horizontal Inequalities and Ethno-Nationalist Civil War: A Global Comparison." *American Political Science Review,* 105(3): 478-495.

Chagnon, Napoleon A. 1988. "Life Histories, Blood Revenge, and Warfare in a Tribal Population." *Science,* 239(4843): 985–992.

Chen, Yan and Sherry X. Li. 2009. "Group Identity and Social Preferences." *American Economic Review,* 99(1): 431-457.

Currarini, Sergio and Friederike Mengel. 2012. "Identity, Homophily and In-Group Bias." *FEEM Working Paper* 37.

de Cremer, David and Mark van Vugt. 1999. "Leadership in Social Dilemmas: The Effects of Group Identification on Collective Actions to Provide Public Goods." *Journal of Personality and Social Psychology*, 76(4): 587-599.

Dufwenberg, Martin, Uri Gneezy, Werner Güth, and Eric Van Damme. 2001. "Direct vs indirect reciprocity: an experiment." *Homo Oeconomicus* 18: 19-30.

Ellemers, Naomi, Pauline Kortekaas, and Jaap W. Ouwerkerk. 1999. "Self-Categorisation, Commitment to the Group and Group Self-Esteem as Related but Distinct Aspects of Social Identity." *European Journal of Social Psychology,* 29(2-3): 371-389.

Fearon, James D. and David D. Laitin. 1996. "Explaining Interethnic Cooperation." *American Political Science Review,* 90(4): 715-735.

Fehr, Ernst and Simon Gächter. 2000. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review,* 90(4): 980-994.

Fershtman, Chaim and Uri Gneezy. 2001. "Discrimination in a Segmented Society: An Experimental Approach." *Quarterly Journal of Economics,* 116(1): 351–377.

Fischbacher, Urs. 2007. "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Econics,* 10(2): 171-178.

Fleishman, John A. 1988. "The Effects of Decision Framing and Others' Behavior on Cooperation

in a Social Dilemma." *Journal of Conflict Resolution,* 32(1): 162-180.

Gaertner, Samuel L., John F. Dovidio, Phyllis A. Anastasio, Betty A. Bachman, and Marcy C. Rust. 1993. "The Common Ingroup Identity Model: Recategorization and the Reduction of Intergroup Bias." *European Review of Social Psychology,* 4(1): 1-26.

Gaertner, L., Iuzzini, J., and O'Mara, E. M., 2008. "When rejection by one fosters aggression against many: multiple-victim aggression as a consequence of social rejection and perceived groupness." Journal of Experimental Social Psychology 44: 958-970.

Glaeser, Edward. L. 2005. "The Political Economy of Hatred." *Quarterly Journal of Economics*, 120(1): 45–86.

Goette, Lorenz, David Huffman, and Stephan Meier. 2006. "The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence using Random Assignment to Real Social Groups." *American Economic Review* 96(2): 212-216.

Goldstein, Joshua S. 2003. *War and Gender: How Gender Shapes the War System and Vice Versa.* Cambridge, MA: Cambridge University Press.

Greiner, Ben. 2004. "The Online Recruitment System ORSEE 2.0 – A Guide for the Organization of Experiments in Economics." *University of Cologne Working Papers Series in Economics,* 10.

Guala, Francesco. 2012. "Reciprocity: Weak or strong? What punishment experiments do (and do not) demonstrate." *Behavioral and Brain Sciences*, 35(1): 1.

Habyarimana, James P., Macartan Humphreys, Daniel N. Posner, and Jeremy M. Weinstein. 2009. *Coethnicity: Diversity and the Dilemmas of Collective Action,* New York: Russell Sage Foundation Publications.

Hardin, Russel. 1995. *One For All: The Logic of Group Conflict.* Princeton, NJ: Princeton University Press.

Hargreaves-Heap, Shaun, and Daniel J. Zizzo. 2009. "The Value of Groups." *American Economic Review*, 99(1): 295-323.

Hoff, Karla, Mayuresh Kshetramade, and Ernst Fehr. 2011. "Caste and Punishment: The Legacy of Caste Culture in Norm Enforcement." *The Economic Journal* 121 (556): F449–F475.

Horowitz, Donald N. 1985. *Ethnic Groups in Conflict.* Berkeley: University of California Press.

Horowitz, Donald N. 2001. *The Deadly Ethnic Riot.* Berkeley: University of California Press.

Kaufman, Stuart J. 2001. *Modern Hatreds: The Symbolic Politics of Ethnic War.* Ithaca, NY: Cornell University Press.

Keynes, John M. 1922. *The Economic Consequences of the Peace*. Los Angeles, CA: IndoEuropean.

Lickel, Brian, Norman Miller, Douglas M. Stenstrom, Thomas F. Denson, and Toni Schmader. 2006. "Vicarious Retribution: The Role of Collective Blame in Intergroup Aggression." *Personality and Social Psychology Review,* 10(2): 372-703.

Messé, Lawrence A., and John M. Sivacek. 1979. "Predictions of Others' Responses in a Mixed-Motive Game: Self-Justification or False Consensus?" *Journal of Personality and Social Psychology,* 37(4): 602-607.

Miguel, Edward, and Mary K. Gugerty. 2005. "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya." *Journal of Public Economics,* 89(11-12): 2325-2368.

Mummendey, Amélie, Bernd Simon, Carsten Dietze, Melanie Grünert, Gabi Haeger, Sabine Kessler, Stephan Letgen, and Stefanie Schäferhoff. 1992. "Categorization is Not Enough: Intergroup Discrimination in Negative Outcome Allocations." *Journal of Experimental Social Psychology*, 28: 125-144.

Nowak, Martin A., and Karl Sigmund. 2005. "Evolution of Indirect Reciprocity." *Nature,* 437(4131): 1291-1298.

Nowak, Martin A., and Sébastien Roch. 2007. "Upstream Reciprocity and the Evolution of Gratitude." *Proceedings of the Royal Society B: Biological Sciences,* 274(1610): 605-610.

Pettigrew, Thomas F. 1998. "Intergroup Contact Theory." *Annual Review of Psychology,* 49: 65-85.

Rubin, Donald B. 1976. "Inference and Missing Data." *Biometrika,* 63(3): 581-592.

Shayo, M., and A. Zussman. 2011. "Judicial Ingroup Bias in the Shadow of Terrorism." *The Quarterly Journal of Economics* 126 (3): 1447–1484.

Sherif, Muzafer, O. J. Harvey, B. Jack White, William R. Hood, and Carolyn W. Sherif. 1961. *The Robbers Cave Experiment: Intergroup Conflict and Cooperation.* Scranton, PA: Wesleyan University Press.

Sen, Amartya. 2007. *Identity and Violence: The Illusion of Destiny*. W. W. Norton & Company.

Sidanius, Jim and Felicia Pratto. 1999. *Social Dominance: An Intergroup Theory of Social Hierarchy and Oppression.* Cambridge, MA: Cambridge University Press.

Sumner, William G. 1906. *Folkways: A Study of the Sociological Importance of Usages, Manners, Customs, Mores, and Morals*. Boston: Ginn.

Tajfel, Henri, M. G. Billig, R. P. Bundy, and Claude Flament. 1971. "Social Categorization and Intergroup Behavior." *European Journal of Social Psychology,* 1(2): 149-178.

Tambiah, Stanley J. 1996. *Leveling Crowds: Ethnonationalist Conflicts and Collective Violence in South Asia.* Berkeley: University of California Press.

Wilson, Michael L. and Richard W. Wrangham. 2003. "Intergroup Relations in Chimpanzees." *Annual Review of Anthropology,* 32: 363–392.

World Bank. 2011. *The World Development Report 2011:Conflict, Security, and Development*. Washington D.C.: The World Bank.

Voors, Maarten, Eleonora E.M. Nilesen, Philip Verwimp, Erwin H. Bulte, Robert Lensink, and Daan P. Van Soest. 2012. "Violent Conflict and Behavior: A Field Experiment in Burundi." *American Economic Review,* 102(2): 941-964.

For online publication: Appendix

1. Experimental procedure details
    - Before experiment
    - Quiz questions
    - Eliciting expectations
    - Second quiz
2. Supplementary analysis of main results
    - Expectations
    - Fairness
    - The second quiz and the contact hypothesis
3. Further results
    1. Inequality
    2. Males vs. females
    3. In-group identity
4. Participant instructions

1. Experimental details

Before experiment

After entering the lab, subjects drew a number assigning them to their seats. In 5 sessions, seats were located at desks surrounded by blinds, effectively creating a private cubicle. In the remaining 10 sessions, seats were placed in a separate cabin which could be closed with a curtain. At each seat, subjects found a copy of the instructions and a paper receipt slip to be used for their payment next to the computer screen. After all subjects took their place, instructions were read out loud. (The information provided in the instructions covered the structure of the experiment, a complete description of the games to be played and payment structures. An English translation of the instructions can be found below.)

Quiz questions

The questions in the quiz were chosen to provoke communication, i.e. we expected not everybody to individually be able to answer correctly, but possibly so within the group. The average number of chat entries per group were 81, i.e., roughly, each subject made one chat entry per question. In order to avoid frustration, simpler questions were also included. On average, 61% of the questions were answered correctly. The English translation of the questions is shown in Table A1.

1. Which term does not denote a whirlwind? a) Landspout b) Hurricane c) Typhoon d) Blizzard	11. In which Hitchcock movey can the main character not stand to see the color red? a) Marnie b) Vertigo c) North by Northwest d) The Birds
2. In the fairytale of the Brothers Grimm, the sister of Snow White is... a) Rapunzel b) Rose Red c) Snow Drop d) Pitch Black	12. Before taking up his position as President of the Federal Republic of Germany, Horst Köhler, was head of the a) IWF b) WTO c) World Bank d) European Central Bank
3. What is the official language in Brasil? a) Spanish b) Portuguese c) English d) French	13. In which region are the remains of the antic city of Pergamon to be found? a) Cyprus b) Crete c) Turkey d) Italy
4. What is a circular note of the Pope called? a) Encyclical b) Encyclopedia c) Enzyme d) Encephalitis	14. Which Indian movie was awarded the “Oscar” in 2009 for best movie? a) Slumdance Billionaire b) Squaredance Millionaire c) Slumdog Millionaire d) Hotdog Billionaire
5. What is the correct term for a direct hook-up of positive and negative poles of a voltage supply? a) Short b) Direct connection c) Short cut d) Arc of suspense	15. Which German Basketball star recently won his first title in the american professional league NBA? a) Dirk Bach [A well-known German Comedian] b) Dirk Nowitzki c) Dirk Bauermann [Head Coach of German National Team Basketball] d) Dirk Niebel [German Minister for Foreign Aid]
6. Which religion does not respect guardian angels? a) Christianity b) Judaism c) Islam d) Buddhism	16. In which city does the UNESCO have its headquarters? a) London b) New York c) Paris d) Brussels
7. Which oil is used for the production of cosmetics? a) Palm oil b) Heavy oil c) Penetrating oil d) Waste oil	17. With whom did Sir Edmund Hillary first reach the summit of Mount Everest? a) Nasreddin Hodsha b) Nursay Pimsorn c) Tensing Norgay d) Anrindranath Singh
8. On which island did Christoph Columbus land first during his Atlantic journey in 1492? a) Jamaica b) Cuba c) San Salvador d) Hispaniola	18. Who was never emperor of the Holy Roman Empire of German Nations? a) Louis the Bavarian b) Henry the Lion c) Otto the Great d) Frederick Barbarossa
9. What is the name of the traditional horse race in the city of Siena? a) Palio b) Fermo c) Lugo d) Legnano	19. What is the name of a famous sexual therapist? [Known in Germany from the most popular teenage magazine] a) Dr. Spring b) Dr. Sommer c) Dr. Fall d) Dr. Winter
10. Which term does not denote a Greek style of columns? a) Dorian b) Corinthian c) Ionian d) Cretan	20. Against which city did Bonn succeed in the elections on where the capital of Germany should be located in 1949? a) West-Berlin b) Frankfurt / Main c) Kassel d) Hannover

Table A1: Quiz questions and answers

Directly after the quiz, subjects answered a brief questionnaire containing three questions. We asked these early to ensure that behavior in the public goods games did not affect subjects' answers. Answers could be given by selecting a point on a 5-point Likert scale. Subjects were asked:

1) How much did you enjoy the quiz? ("Not at all" = 1, "Very much" = 5)
2) How well do you think your group performed in the quiz? ("Very badly" = 1, "Very well" = 5)
3) How fair did you think the quiz was? ("Not at all fair" = 1, "Very fair" = 5)

A final questionnaire ended the experiment. The questionnaire included questions on gender, age, major subject, native language, whether any of the other subjects were personally acquainted, and the possibility to comment on the experiment. It also included open questions on what subjects believed the experiment to be about and a brief explanation of their choices of the course of the experiment. Mostly, subjects thought the experiment to be about the influence of income on decision behavior, the correlation of certain behaviors with intelligence, or social dilemmas. Only 3 of the 240 subjects mentioned objectives related to group reciprocity ("treat colors differently", "memorize behavior according to color", "do decisions of un-encountered others affect my behavior"). To get indicators of ingroup identity and fairness we added three statements which subjects could agree/disagree with on a 7-point Likert scale:

- I felt committed to my group
- I think the experiment was fair towards all (color) groups
- I enjoyed the experiment

Perceptions of other groups were also elicited by Likert scales. Statements read as follows (each statement was given 4 times, once for each color group):

- Members of the green/orange/purple/brown group worked together well as a team
- Members of the green/orange/purple/brown group behaved fairly towards other groups

Following the questionnaire, subjects were given information on which round was chosen for payment and their total earnings. Subjects were called up individually by seat number (after the respective previous subject had left the lab) and signed a receipt slip for the money.

Eliciting expectations

In sessions 1-10 and 13-15, subjects were asked about their expectations on the choice of the matched partner in order to check whether individuals' choices were influenced by the expectations on the matched partner's behavior. So as to keep the experiment simple and fast-moving, expectations were only elicited in repetitions 2 and 7. Expectations were incentivized in that the difference between the indicated expectation and the true amount withdrawn by the partner determined the size of an extra bonus to be gained. Table A2 illustrates how the possible boni were determined.

Difference	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50
Bonus	10 ECU	9 ECU	8 ECU	7 ECU	6 ECU	5 ECU	4 ECU	3 ECU	2 ECU	1 ECU	0 ECU

Table A2: Boni for expectations elicitation

Second quiz

Sessions 13-15 contained a second quiz after the first 4 repetitions of the public goods game, in which some individuals could cooperate with members of other groups. On average, approximately 126 chat entries were made per session, i.e. 32 per group or 3.2 chat entries per question. The English translations of the additional quiz questions are shown in Table A3.

1. A classic silent movie by Sergey Eisenstein is titled “The Battleship... a) Potemkin b) Rasputin c) Iljushin d) Putin	6. Most Nepalese share the religion of a) Islam b) Buddhism c) Hinduism d) Christianity
2. Which concordat ended the conflict of investitures in 1122? a) Concordat of Konstanz b) Concordat of Worms c) Concordat of Mainz d) Concordat of Speyer	7. What is the name of the ancient God of Fire? a) Cronus b) Helios c) Hephaistos d) Eos
3. What is acrophobia? a) Fear of heights b) Fear of spiders c) Fear of aggression d) Fear of pain	8. Which return did the GDR provide for the Billion D-Mark credit of the FRG? a) Deinstallation of spring guns on the border b) Release of political prisoners c) Exit permits for all occupiers of the embassy d) Reduction of the number of strategic missiles
4. Which country does not have a nuclear power plant? a) Norway b) Switzerland c) Belgium d) Netherlands	5. Who wrote “Nathan the Wise”? a) Goethe b) Kleist c) Schiller d) Lessing
5. Which German city currently has less than 1 million inhabitants? a) Cologne b) Munich c) Berlin d) Frankfurt / Main	6. Who was elected president of the Czech Republic in 2003? a) Vaclav Havel b) Vaclav Klaus c) Wladimir Spidla d) Wladimir Putin

Table A3: Quiz 2 questions and answers

2. Supplementary analyses of main results

We first explore the basic patterns in our data. Figure A1 shows the distribution of amounts taken. There are many observations of both 0, the minimum possible take, and 50, the maximum, which means that tobit regressions may be appropriate. Figure A2 shows mean amounts taken over all sessions, in each of the 8 repetitions. As is typical for a public goods game, amounts taken increased over the course of the experiment.

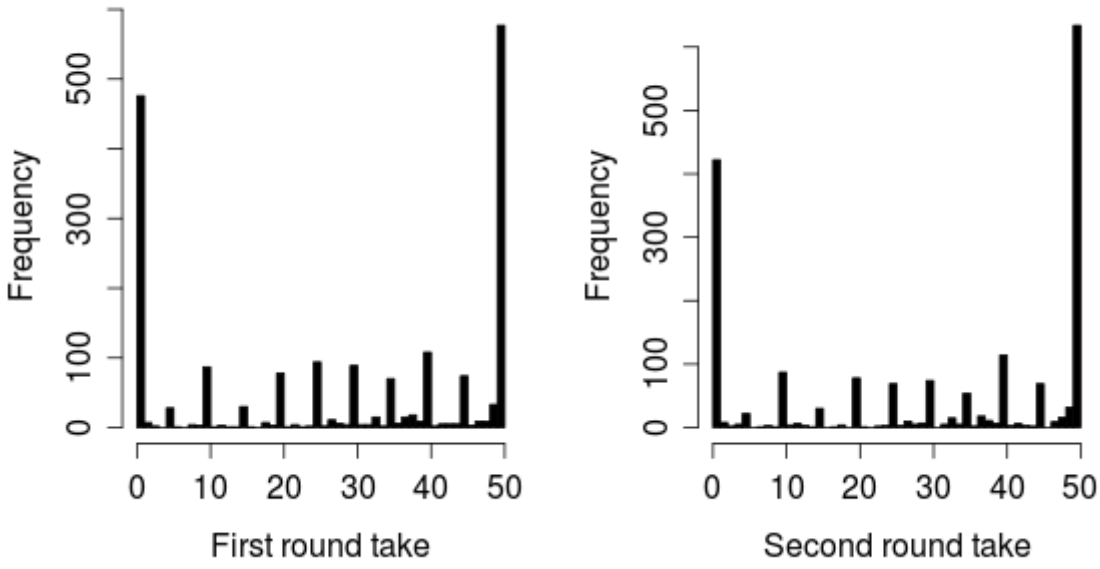


Figure A1: Histograms of amounts taken

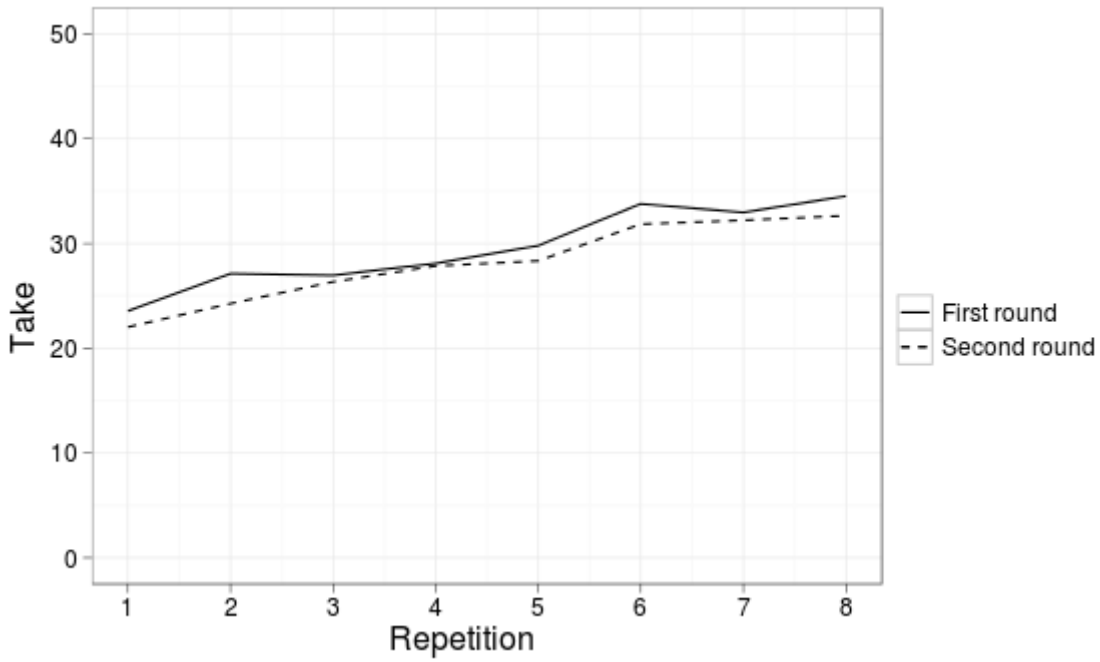


Figure A2: amount taken, repetitions 1-8

Subjects' behavior within a single session may not be independent, because subjects may influence each other, either through their play over the repeated rounds with rematching, or via chat during the quiz. We use two alternative strategies to deal with this. First, we treat each session as a single independent observation and run non-parametric tests on session-level statistics. Second, we run regressions on individual decisions, cluster standard errors by individual, and include covariates in order to control explicitly for the history of play. An alternative approach would be to cluster standard errors at the session level. We believe that errors are likely to be most strongly correlated within individuals, at least when individual-level fixed effects are not included; in any case, the coefficient on *F take* × *Same group* remains highly significant if we cluster at session level (results available on request).

At session level, Figures A3 and A4 show the coefficient on *F take* × *Same group* in equation (1), estimated by simple linear regression without controls for each session, separately for first and third party treatments. We use Wilcoxon signed-rank tests to test whether the values of these coefficients

are significantly different from 0 (equivalently, whether the coefficients on *F take* are significantly larger when *Same group* is 1: thus, the test uses matched pairs of observations from each session for *Same group* = 1 and *Same group* = 0). Table A4 shows two-sided p-values for first and third party treatments in winner bonus and random bonus sessions. It also shows tests pooling all sessions. In third party treatments, δ is never significantly different from 0, while in first party treatments δ is significantly different from 0, whether we pool all sessions or examine random bonus sessions alone.

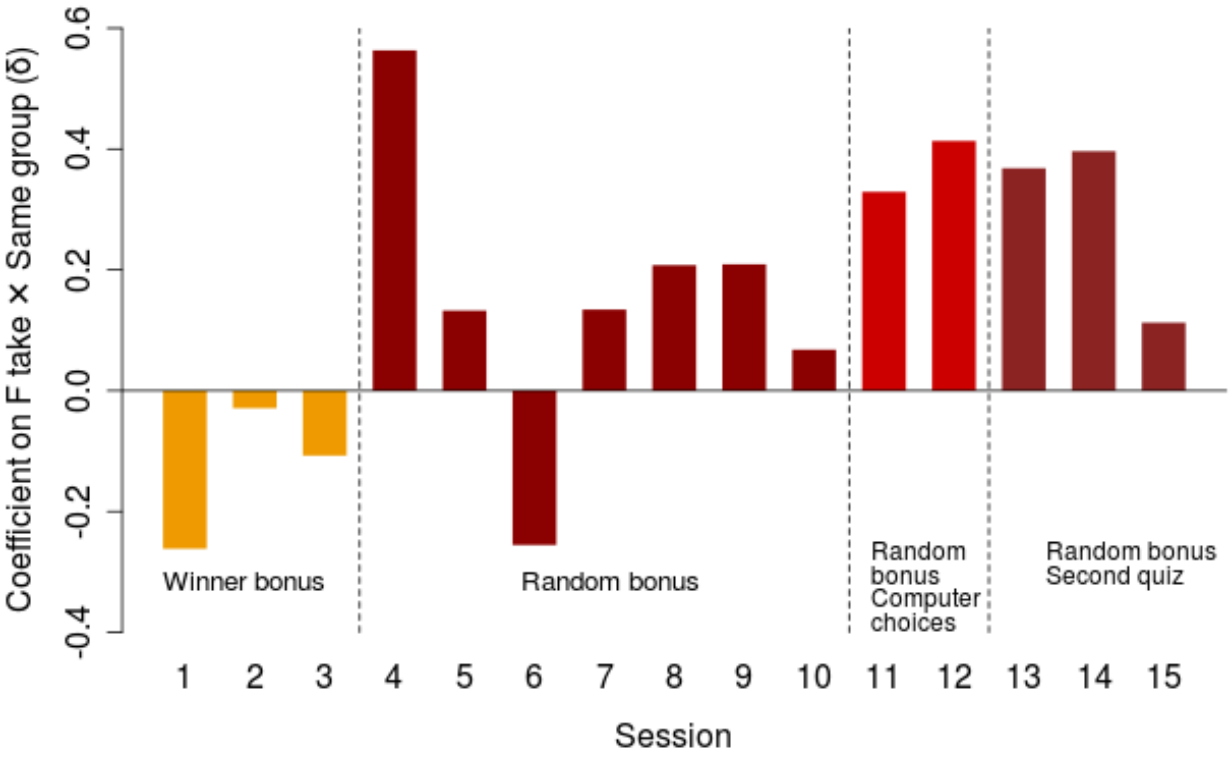


Figure A3: Per-session estimates of δ, first party treatments

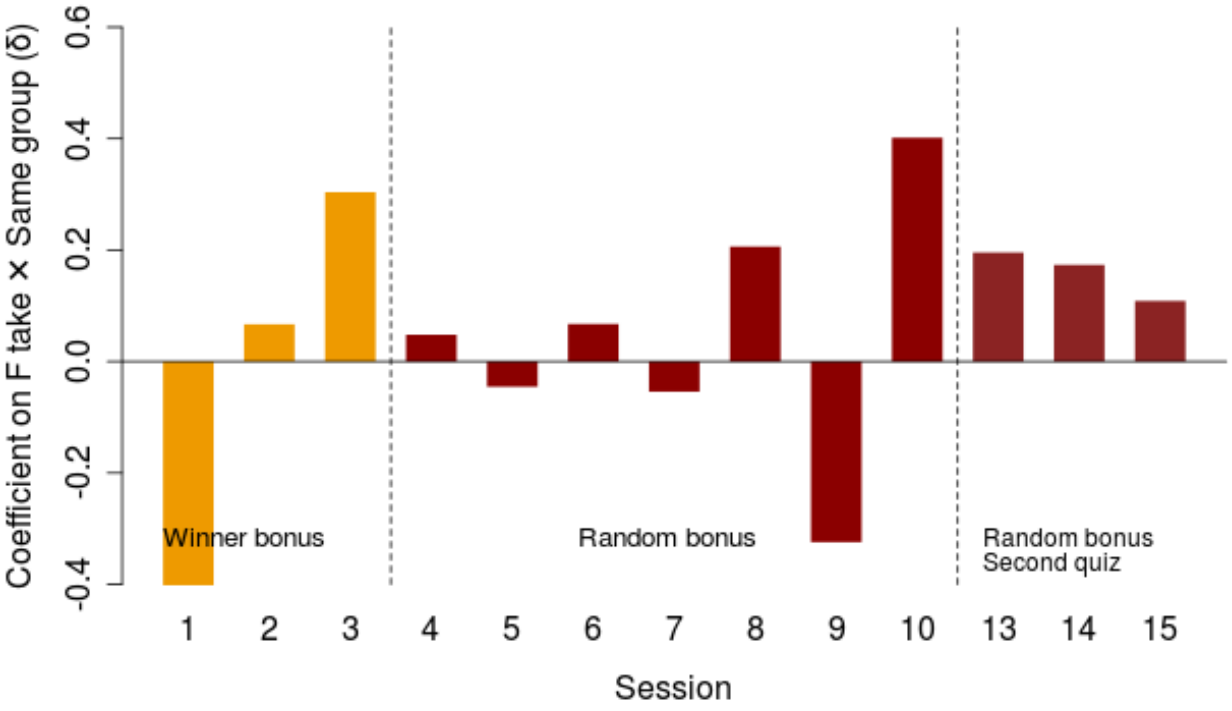


Figure A4: Per-session estimates of δ, third party treatments

	Sessions		
Treatments	Winner bonus	Random bonus	All sessions
First party	0.25	0.00928 **	0.0353 *
Third party	1	0.16	0.216

*Table A4 : Wilcoxon-Mann-Whitney tests of δ=0. Two sided p-values are shown. + p < 0.10; * p < 0.05; ** p< 0.01; *** p < 0.001.*

Next, we take individual decisions as the unit of observation. We would like to know if we can pool our different session types for analysis. As a first gauge of this, we examine the δ coefficient on *F take* × *Same group* in first party treatments. Table A5 estimates equation (1) separately for each

type of session. In first party treatments, δ is not significantly different between the different sub-types of random bonus session (sessions 4-10 vs 11-12: p=0.171; 4-10 vs 13-15: p=0.395, 11-12 vs 13-15: p=0.622), but it is significantly different between the winner bonus sessions and sessions 11-12 (p=0.0146) and sessions 13-15 (p=0.0395), and narrowly misses significance at the 10% level between the winner bonus sessions and sessions 4-10 (p=0.107). This suggests that we can pool the random bonus sessions without undue violence to the data. In third party treatments, δ is always insignificant and does not vary significantly between session types.

	(1) Sessions 1-3	(2) Sessions 4-10	(3) Sessions 11-12	(4) Sessions 13-15
(Intercept)	19.7	21.9	27.2	22
	(3.46) ***	(2.45) ***	(3.76) ***	(3.45) ***
First party ✕ F take	0.4	0.25	0.102	0.271
	(0.0947) ***	(0.0656) ***	(0.0972)	(0.0994) **
———"——— ✕ Same group	1.62	-2.17	-10.9	-7.54
	(4.2)	(2.92)	(4.96) *	(4.11) +
———"——— ✕ F take ✕ Same group	-0.111	0.151	0.381	0.285
	(0.13)	(0.0879) +	(0.146) **	(0.131) *
Third party	1.37	-0.333	–	-4.87
	(4)	(2.78)		(3.81)
———"——— ✕ F take	0.298	0.244	–	0.35
	(0.0977) **	(0.0671) ***		(0.104) ***
———"——— ✕ Same group	-0.505	1.18	–	-1.09
	(4.32)	(2.82)		(4.45)
———"——— ✕ F take ✕ Same group	0.0213	0.0312	–	0.0865
	(0.126)	(0.0851)		(0.137)
Model	Linear	Linear	Linear	Linear
Controls	–	–	–	–
N	384	896	192	384
N indiv.	48	112	32	48
Adj. R2	0.109	0.0915	0.102	0.169

*Table A5: Estimates of equation (1) for each session type. Standard errors in parentheses, clustered by individual. + p < 0.10; * p < 0.05; p < 0.01; *** p < 0.001.*

Next, we pool data from all the random bonus sessions, and test equation (1) in various specifications. Column 1 of Table A6 reports an OLS regression without controls. Column 2 adds in subject and repetition dummies (in econometric language, individual and time fixed effects). The subject dummies control for individual-level variation in baseline propensity to take: in effect, this regression uses the correlation between changes in *F take* and *Same group* and changes in individuals' second round take. The time dummies control for changes in average cooperation over

time. Using fixed effects has advantages and disadvantages. Clearly there is much heterogeneity between individuals, and this might be correlated with the propensity to group-reciprocate; fixed effects mitigate this worry. On the other hand, using fixed effects throws away the information contained in differences between individuals, which can be substantial. Column 3 removes the fixed effects, but adds a vector of controls reflecting the history of play experienced by each individual. Specifically, we add the amount taken by the subject's partners in each round of the previous repetition, as well as the amount taken by the subject's feedback participant (F) in the previous repetition. (Further lags proved insignificant.) We also add *Last take same group*, which is the amount taken by the subject's last partner (in any round and repetition) who was in the same group as the current second-round partner P. The lagged variables in general have the expected sign, but are not strongly significant. However, the coefficient on *Last take same group* is large and highly significant in both first and third party treatments. This suggests to us that even in third party treatments, subjects were directly group-reciprocating the previous behavior of others towards them. The identification of group reciprocity is not clean, however, because subjects cannot be sure that the last partner was a different person from the current P. In later tables, we use the same history variables, but omit their coefficients for the sake of clarity.

The δ coefficient on *F take* × *Same group* remains almost unchanged in all these specifications, and is highly significant throughout for first party treatments. The final column estimates (1) using a tobit functional form, censoring the data at 0 and 50. Here the δ coefficient is larger, because the tobit specification treats the censored observations as less informative.

	(1)	(2)	(3)	(4)
(Intercept)	23.1	–	15.1	20.5
	(1.8) ***		(2.62) ***	(3.44) ***
First party ✕ F take	0.223	0.0277	0.135	0.459
	(0.0483) ***	(0.0368)	(0.0518) **	(0.103) ***
———"——— ✕ Same group	-5.24	-4.05	-6.85	-13.2
	(2.16) *	(1.73) *	(2.51) **	(5.12) *
———"——— ✕ F take ✕ Same group	0.229	0.187	0.257	0.597
	(0.0656) ***	(0.051) ***	(0.0725) ***	(0.152) ***
Third party	-2.97	-2.58	-3.04	-7.68
	(2.23)	(1.87)	(3.3)	(5.45)
———"——— ✕ F take	0.279	0.0859	0.146	0.637
	(0.0558) ***	(0.0417) *	(0.0569) *	(0.12) ***
———"——— ✕ Same group	0.667	1.19	-1.05	0.843
	(2.38)	(1.95)	(2.71)	(5.67)
———"——— ✕ F take ✕ Same group	0.0461	0.00627	0.123	0.121
	(0.072)	(0.0565)	(0.0774)	(0.165)
First party ✕ Lag round 1 take	–	–	0.139	–
			(0.0576) *	
———"——— ✕ Lag round 2 take	–	–	0.0406	–
			(0.0466)	
———"——— ✕ Lag F take	–	–	0.0861	–
			(0.056)	
———"——— ✕ Last take same group	–	–	0.142	–
			(0.0436) **	
Third party ✕ Lag round 1 take	–	–	0.0284	–
			(0.0552)	
———"——— ✕ Lag round 2 take	–	–	0.0771	–
			(0.0501)	
———"——— ✕ Lag F take	–	–	0.146	–
			(0.0516) **	
———"——— ✕ Last take same group	–	–	0.201	–
			(0.0477) ***	
Model	Linear	Linear	Linear	Tobit
Controls	–	Twoway FE	History	–
N	1472	1472	1072	1472
N indiv.	192	192	192	192
Adj. R2	0.111	0.0279	0.218	–
LogLik	–	–	–	-3880

*Table A6: Estimates of equation (1), random bonus sessions. Standard errors in parentheses, clustered by individual for columns 1-3. + p < 0.10; * p < 0.05; p < 0.01; *** p < 0.001.*

We include three further checks of the basic result. First, we split our data up by repetition, analyzing repetitions 1-4 separately from repetitions 5-8. Some behaviors, such as contributions to public goods games, change over repetitions in repeated experiments. Similarly, it could be that group reciprocity is only observed early on, and less so or not at all in later repetitions as subjects gain experience. If so, group reciprocity would not be an "equilibrium phenomenon", although out-of-equilibrium behavior might still be interesting and important. Columns 1-2 of Table A7 estimate equation (1) separately for early and late repetitions. If anything results are stronger in later repetitions, although the difference between the (first party) δ coefficients is not significant (p=0.496).

	(1)	(2)
	Repetitions 1-4	Repetitions 5-8
(Intercept)	23.1	23.5
	(2.43) ***	(2.35) ***
First party × F take	0.17	0.257
	(0.0717) *	(0.0638) ***
———"——— × Same group	-4.6	-6
	(3.01)	(3.31) +
———"——— × F take × Same group	0.174	0.268
	(0.0979) +	(0.0929) **
Third party	-7.3	0.877
	(3.19) *	(3.09)
———"——— × F take	0.384	0.183
	(0.081) ***	(0.0717) *
———"——— × Same group	3.57	0.394
	(3.31)	(3.76)
———"——— × F take × Same group	-0.151	0.123
	(0.109)	(0.102)
Model	Linear	Linear
Controls	–	–
N	736	736
N indiv.	192	192
Adj. R2	0.0851	0.127

*Table A7: Estimates of equation (1), random bonus sessions, early and late repetitions. Standard errors in parentheses, clustered by individual. + $p < 0.10$; * $p < 0.05$; $p < 0.01$; *** $p < 0.001$.*

The second check is to split up decisions by whether the decision-taking subject herself was in the same group as the other subject on her side (i.e., whether the *Own same group* dummy was 1 or 0). In particular, this might make a difference for indirect group reciprocity: subjects may group-reciprocate against behavior towards somebody else in their group, but not behavior towards someone not in their group. However, direct group reciprocity could also be affected, for example if feelings of group identity get stronger when 2 subjects of one group face 2 subjects of another group. Table A8 shows the results. Again, the split makes little difference, either for first or third

party treatments, and the difference between δ coefficients is not significant (first party, p=0.347; third party, p=0.21).

	(1)	(2)
	Own same group	Own different group
(Intercept)	23.3	22.8
	(2.39) ***	(2.32) ***
First party ✕ F take	0.189	0.26
	(0.0665) **	(0.0671) ***
———"——— ✕ Same group	-6.55	-3.85
	(2.95) *	(3.05)
———"——— ✕ F take ✕ Same group	0.288	0.166
	(0.0899) **	(0.0944) +
Third party	-3.01	-2.9
	(3.04)	(3.29)
———"——— ✕ F take	0.256	0.303
	(0.0771) ***	(0.0772) ***
———"——— ✕ Same group	-1.84	3.26
	(3.29)	(3.41)
———"——— ✕ F take ✕ Same group	0.134	-0.0465
	(0.0997)	(0.104)
Model	Linear	Linear
Controls	–	–
N	736	736
N indiv.	192	192
Adj. R2	0.122	0.102

Table A8: Estimates of equation (1), random bonus sessions, Own same group *1 or 0. Standard errors in parentheses, clustered by individual. + p < 0.10; * p < 0.05; p < 0.01; *** p < 0.001.*

Lastly, we replace the linear specification by a set of 6 dummies, 1 each for values of *F take* between 0-9, 10-19, 20-29, 30-39, 40-49 and 50. These are then interacted with *Same group.*The field literature has many more examples of negative group reciprocity – returning evil for evil – than positive group reciprocity, so we expected to see larger and more significant interaction terms at high values of *F take*. This is broadly confirmed in Table A9, which shows the results, using the same four specifications as Table A6.

	(1)	(2)	(3)	(4)
(Intercept)	23.3	–	15.2	21.4
	(1.91) ***		(2.69) ***	(3.58) ***
F take 10-19	4	-1.24	4.47	7.92
	(3.79)	(3.15)	(4.9)	(10.4)
F take 20-29	5.33	-0.15	1.26	9.32
	(3.28)	(2.65)	(4.42)	(8.53)
F take 30-39	5.8	0.739	3.65	8.93
	(3.53)	(2.89)	(4.71)	(7.95)
F take 40-49	7.78	1.09	3.63	15.4
	(3.22) *	(2.58)	(3.46)	(6.86) *
F take 50	11.7	0.805	7.6	24.2
	(2.54) ***	(2.05)	(2.69) **	(5.43) ***
F take 0-9 × Same group	-4.54	-4.52	-7.72	-11.8
	(2.33) +	(1.92) *	(2.72) **	(5.42) *
F take 10-19 × Same group	-4.37	-3.05	-1.77	-4.18
	(5.02)	(4.1)	(6.25)	(13.5)
F take 20-29 × Same group	-2.3	1.56	2.58	-4.27
	(4.1)	(3.28)	(5.21)	(10.5)
F take 30-39 × Same group	2.44	3.12	2.05	5.6
	(3.9)	(3.21)	(4.99)	(9.36)
F take 40-49 × Same group	4.95	3.77	6.58	7.41
	(3.61)	(2.93)	(3.87) +	(8.35)
F take 50 × Same group	7.1	5.55	4.82	22.5
	(2.26) **	(1.92) **	(2.52) +	(5.91) ***
Model	Linear	Linear	Linear	Tobit
Controls	–	Twoway FE	History	–
N	832	832	592	832
N indiv.	192	192	192	192
Adj. R2	0.128	0.0398	0.229	–
LogLik	–	–	–	-2170

*Table A9: Estimates of equation (1), random bonus sessions, first party treatments, dummies for F take. Standard errors in parentheses, clustered by individual for columns 1-3. + $p < 0.10$; * $p < 0.05$; $p < 0.01$; *** $p < 0.001$*

Expectations

There are two possible explanations of the group reciprocal behavior we observe. First, F's action may affect S's *preferences*, broadly defined, for example by making her feel more or less altruistic towards P. Second, F's action may shift S's *expectations* about P's behavior, without necessarily affecting her preferences. For example, if S wishes to take exactly as much as P, and assumes that participants from the same group take similar amounts, then S's taking behavior will be affected by *F take* when F and P are from the same group. We approach this question in two ways: first by eliciting expectations, second by looking at S's behavior when she knows that P's take is determined randomly by a computer, out of P's control.

The variable *Expectations* gives S's expectations of P's second round take, in repetitions 2 and 7, elicited as described above. If we re-estimate equation (1) with this variable added, both the δ coefficient and the β coefficient on *F take* become insignificant in all treatments (estimations not shown). However, re-estimating equation (1) without *Expectations*, but on repetitions 2 and 7 alone, also shows these coefficients becoming insignificant, so the smaller dataset may be to blame rather than the effect of controlling for *Expectations*. One way to resolve this is to multiply impute the *Expectations* variable for other repetitions. Since the missingness of *Expectations* is wholly determined by the repetition number, it is "missing at random" with respect to the value of *Expectations* once repetitions are controlled for. This means that multiple imputation will be unbiased (Rubin 1976). Table A10 shows the results, using the same specifications as Table A6. *Expectations* is highly significantly correlated with subjects' second round take. Indeed in the basic OLS regression, the coefficient is not significantly different from unity, fitting the story that subjects wish to match the amount their partners take. Also, the value of δ is about halved in first party treatments, although it remains weakly significant. This strongly suggests that shifts in expectations help to explain group reciprocal behavior.

	(1)	(2)	(3)	(4)
(Intercept)	3.46	–	3.22	-17.4
	(2) +		(2.06)	(3.99) ***
First party ✕ F take	-0.00703	-0.0494	-0.00837	-0.0287
	(0.0511)	(0.0401)	(0.0514)	(0.103)
———"——— ✕ Same group	-3.22	-3.5	-3.18	-7.16
	(1.89) +	(1.56) *	(1.83) +	(4.04) +
———"——— ✕ F take ✕ Same group	0.1	0.125	0.0995	0.239
	(0.0515) +	(0.0449) **	(0.0504) *	(0.111) *
———"——— ✕ Expectations	0.965	0.627	0.963	1.87
	(0.0572) ***	(0.0448) ***	(0.0654) ***	(0.121) ***
Third party	-0.947	-1.23	-1.61	-1.19
	(2.11)	(1.87)	(2.16)	(4.96)
———"——— ✕ F take	0.0174	-0.00835	0.0083	0.0378
	(0.0465)	(0.0388)	(0.0475)	(0.099)
———"——— ✕ Same group	0.991	0.656	0.978	1.32
	(1.97)	(1.68)	(1.95)	(4.26)
———"——— ✕ F take ✕ Same group	-0.0107	0.00421	-0.0179	-0.00807
	(0.057)	(0.0483)	(0.0553)	(0.121)
———"——— ✕ Expectations	0.965	0.617	0.94	1.84
	(0.0506) ***	(0.0544) ***	(0.0571) ***	(0.102) ***
Model	Linear	Linear	Linear	Tobit
Controls	–	Twoway FE	History	–
N	1472	1472	1072	1472
N indiv.	192	192	192	192
Adj. R2	0.479	0.237	0.481	–

Table A10: Estimates of equation (1), random bonus sessions, adding Expectations *and multiply imputing in repetitions 1, 3-6, 8. 10 imputations. Standard errors in parentheses, clustered by individual for columns 1-3. + p < 0.10; * p < 0.05; p < 0.01; *** p < 0.001.*

We now look directly at whether F's behavior affects subjects' expectations of P's take. To do this we take *Expectations* as the dependent variable, without imputing, and examine the effect of *F take*, crossed with *Same group*. Table A11 shows the results. We skip the fixed effects specification

because the data from repetitions 2 and 7 are not enough for a credible estimate of per-individual effects. *F take* has a significant positive effect on subjects' expectations. This effect appears larger when F and P are in the same group, but the difference is not significant.

	(1)	(2)	(3)
(Intercept)	19.5	5.37	17.1
	(2.57) ***	(3.67)	(3.41) ***
First party ✕ F take	0.344	0.332	0.454
	(0.0724) ***	(0.0737) ***	(0.0993) ***
———"——— ✕ Same group	-4.11	0.602	-6.56
	(3.57)	(3.87)	(4.71)
———"——— ✕ F take ✕ Same group	0.148	0.00937	0.226
	(0.104)	(0.11)	(0.139)
Third party	-1.88	5.81	-1.75
	(3.65)	(5.75)	(4.57)
———"——— ✕ F take	0.313	0.224	0.398
	(0.072) ***	(0.0933) *	(0.0881) ***
———"——— ✕ Same group	-2.76	-4.34	-3.01
	(3.36)	(4.47)	(4.39)
———"——— ✕ F take ✕ Same group	0.0894	0.128	0.128
	(0.101)	(0.119)	(0.124)
Model	Linear	Linear	Tobit
Controls	–	History	–
N	416	296	416
N indiv.	208	207	240
Adj. R2	0.222	0.362	–
LogLik	–	–	-1450

Table A11: Determinants of Expectations*, random bonus sessions, repetitions 2 and 7. Standard errors in parentheses, clustered by individual for columns 1-3. + p < 0.10; * p < 0.05; p < 0.01; *** p < 0.001.*

Are these expectations justified, i.e is F's and P's actual behavior correlated? Table A12 examines this by regressing P's second round take on F's first round take (equivalently, regressing subjects' second round take on the first round take of the other subject on their "side"). Column 1 shows a

small but significant correlation between the two, which does not change significantly when P and F share group membership. However, column 2 shows that this result does not survive adding in time and session dummies, and these seem the most likely explanation for the basic correlation. We found similar null results within each session type (regressions not shown). Nevertheless, subjects might find F's take a useful heuristic for estimating P's take, since they can observe it, but cannot directly observe session effects or repetition effects. F and P's group membership does not appear to be a useful heuristic in this way, since it does not affect the correlation between F and P's behavior.

	(1)	(2)
(Intercept)	26.5	–
	(1.51) ***	
F take (β)	0.122	0.0317
	(0.0381) **	(0.0358)
Same group (γ)	0.493	0.491
	(1.73)	(1.64)
F take × Same group (δ)	-0.0427	-0.0434
	(0.0529)	(0.0502)
Model	Linear	Linear
Controls	–	Time, session FE
N	1856	1856
N indiv.	240	240
Adj. R2	0.0109	0.0689

*Table A12: Regression of P's second round take on F's first round take, all sessions. Standard errors in parentheses, clustered by individual. + $p < 0.10$; * $p < 0.05$; $p < 0.01$; *** $p < 0.001$.*

To sum up these analyses, the effect of group reciprocity is reduced when we control for subjects' expectations. This suggests that expectations are part of the story. There is only weak evidence that subjects take group membership into account when forming expectations; doing so would be incorrect, because the behavior of F and P is not more correlated when they share group

membership (and is not correlated at all after controlling for repetition and session).

As stated in the main text, subjects who intend to take a large amount from the common resource in the second round may report that they expect their partners to take large amounts, due to a self-justification or false consensus effect (Messé and Sivacek 1979); this would bias our estimate of the causal effect of expectations. Therefore, we used an alternative approach to examine behavior when expectations were irrelevant. In 64 of the choices of sessions 11-12, subjects were informed between the rounds that their second round partner's choice would be made by the computer drawing a random number from 0 to 50. All subjects, including those whose choice was made by the computer, were still paid as normal from these rounds. Thus, S's expectations about P's choice ought not to be affected by F's take in round 1. We can use these decisions to examine how the amount taken by F affects the amount given to P, independently of any change in expectations.

Table A13 estimates equation (1) separately for when P's choice would be made by a human (*Human Choice*) and by the computer (*Random Choice*). The coefficient on *F take* × *Same group* is always positive for both values of *Random Choice*, and the difference between coefficients is never significant. In the fixed effects regression, the coefficients on *F take* × *Same group* are not significant. We attribute this to the small sample size and the inefficiency of fixed effects; in any case, the coefficient is substantially larger when *Random Choice*=1. Overall, while expectations seem to play a role in group reciprocity, they are not the whole story.

	(1)	(2)	(3)	(4)
(Intercept)	24	–	19.6	25.2
	(4.24) ***		(5.81) ***	(8.14) **
Human Choice ✕ F take	0.184	0.187	0.118	0.301
	(0.114)	(0.0956) +	(0.129)	(0.235)
Human Choice ✕ Same group	-13.6	-2.91	-15.2	-31.8
	(6.12) *	(5.27)	(6.96) *	(13.9) *
Human Choice ✕ F take ✕ Same group	0.432	0.1	0.44	1.03
	(0.178) *	(0.15)	(0.193) *	(0.393) **
Random Choice	11.3	13.1	7.18	19.2
	(7.22)	(6.2) *	(8.1)	(15.5)
Random Choice ✕ F take	-0.101	-0.194	-0.104	-0.186
	(0.174)	(0.142)	(0.177)	(0.369)
Random Choice ✕ Same group	-11.9	-9.44	-12.2	-20.3
	(7.95)	(7.17)	(9.23)	(17.5)
Random Choice ✕ F take ✕ Same group	0.442	0.307	0.516	0.896
	(0.225) +	(0.199)	(0.244) *	(0.519) +
Model	Linear	Linear	Linear	Tobit
Controls	–	Twoway FE	History	–
N	192	192	143	192
N indiv.	32	32	32	32
Adj. R2	0.125	0.055	0.183	–
LogLik	–	–	–	-520

Table A13: Estimates of equation (1), sessions 11-12, first party treatments, interacted with Random Choice *and* Human Choice*. Standard errors in parentheses, clustered by individual for columns 1-3. + p < 0.10; * p < 0.05; p < 0.01; *** p < 0.001.*

Fairness

Next, we examine perceptions of fairness. The pre-questionnaire contained the question "Did you think the quiz was fair?" Responses were on a 1-5 Likert scale. We created a *Fair* dummy which is 1 if subjects chose 4 or 5 and 0 if subjects chose 1, 2 or 3. Columns 1-4 of Table A14 estimate equation (1) for random bonus sessions, adding an interaction with *Fair*. For clarity, third party treatments are omitted. The interaction of *Fair* with our basic variable of interest *F take × Same*

group is negative, and at least weakly significant in all specifications except column 3 when history is included. Further analysis (available on request) shows that this loss of significance comes from the lower *N* due to the history variables, rather than from the addition of the variables themselves. The sum of the coefficients on *F take* × *Same group* and *F take* × *Same group* × *Fair* is not significantly different from 0 in any specification. It appears that group reciprocity is driven by those subjects who perceived the quiz as unfair.

	(1)	(2)	(3)	(4)
(Intercept)	27.1	–	19.3	30.2
	(2.61) ***		(3.46) ***	(5.1) ***
F take (β)	0.11	-0.0363	0.044	0.182
	(0.0677)	(0.055)	(0.0744)	(0.144)
Same group (γ)	-7.73	-5.39	-9.17	-18.9
	(3.11) *	(2.57) *	(3.75) *	(7.36) *
F take ✕ Same group (δ)	0.362	0.271	0.342	0.945
	(0.0929) ***	(0.0758) ***	(0.105) **	(0.214) ***
Fair	-7.36	–	-6.86	-17.6
	(3.53) *		(3.92) +	(6.81) *
F take ✕ Fair	0.214	0.134	0.162	0.524
	(0.0953) *	(0.0771) +	(0.104)	(0.201) **
Same group ✕ Fair	4.56	2.73	3.92	11
	(4.22)	(3.52)	(4.96)	(10)
F take ✕ Same group ✕ Fair	-0.261	-0.174	-0.152	-0.699
	(0.128) *	(0.105) +	(0.142)	(0.297) *
Model	Linear	Linear	Linear	Tobit
Controls	–	Twoway FE	History	–
N	832	832	592	832
N indiv.	192	192	192	192
Adj. R2	0.136	0.0404	0.232	–
LogLik	–	–	–	-2170

Table A14: Estimates of equation (1), random bonus sessions, first party treatments, interacted with Fair *(perceptions of fairness). Standard errors in parentheses, clustered by individual for columns 1-3. + p < 0.10; * p < 0.05; p < 0.01; *** p < 0.001.*

The second quiz and the contact hypothesis

Here we confirm the analysis of the contact hypothesis from the main text. Recall that *Open quiz 2* is 1 if the subject had previously chatted with members of other group, 0 otherwise. Here, in contrast to the main text, we also include an *Early* dummy for repetitions 1-4. This avoids a confound with time, since *Open quiz 2* only takes the value 1 in repetitions 5-8. Table A15 shows the standard regressions, interacting (1) with both *Early* and *Open quiz 2*. The interaction *F take* × *Same group* × *Open quiz 2* is only significant in the regression with history controls; this is likely to be due to the reduced N, since in other regressions the coefficient is not only insignificant but substantively small. Table A16 repeats the exercise, but replaces *Open quiz 2* with a dummy *Quiz 2 met P*, which is 1 if the subject had previously interacted with participants of P's color during the second quiz. Thus it tests whether group reciprocity is specifically reduced towards members of other groups with whom the subject cooperated. Again, the coefficient on *F take* × *Same group* × *Quiz 2 met P* is not significant and does not have a constant sign.

	(1)	(2)	(3)	(4)
(Intercept)	23.5	–	14.3	20.6
	(2.64) ***		(3.23) ***	(5.85) ***
F take (β)	0.25	0.0159	0.198	0.549
	(0.0702) ***	(0.0581)	(0.0708) **	(0.16) ***
Same group (γ)	-6.07	-6.41	-6.29	-13.8
	(3.71)	(3.06) *	(3.87)	(8.94)
F take ✕ Same group (δ)	0.272	0.235	0.223	0.703
	(0.103) **	(0.0853) **	(0.106) *	(0.246) **
Early	-0.442	–	5.71	1.21
	(3.4)		(3.95)	(7.76)
F take ✕ Early	-0.0804	0.0218	-0.259	-0.251
	(0.101)	(0.0801)	(0.112) *	(0.223)
Same group ✕ Early	1.47	4.46	-4.69	0.828
	(4.94)	(3.91)	(5.5)	(11.3)
F take ✕ Same group ✕ Early	-0.0976	-0.0884	0.155	-0.228
	(0.146)	(0.115)	(0.159)	(0.328)
Open quiz 2	-0.125	3.32	-7.18	-1.33
	(5.41)	(4.94)	(5.3)	(12.3)
F take ✕ Open quiz 2	0.0494	0.0572	0.218	0.106
	(0.162)	(0.14)	(0.158)	(0.374)
Same group ✕ Open quiz 2	0.561	-0.184	6.99	2.92
	(8.03)	(6.84)	(7.89)	(19.6)
F take ✕ Same group ✕ Open quiz 2	-0.0377	-0.0292	-0.145	-0.173
	(0.223)	(0.191)	(0.216)	(0.554)
Model	Linear	Linear	Linear	Tobit
Controls	–	Twoway FE	History	–
N	832	832	592	832
N indiv.	192	192	192	192
Adj. R2	0.136	0.0408	0.241	–
LogLik	–	–	–	-2170

Table A15: Estimates of equation (1), random bonus sessions, first party treatments, interacted with Open quiz 2*. Standard errors in parentheses, clustered by individual for columns 1-3. + p < 0.10; * p < 0.05; p < 0.01; *** p < 0.001.*

	(1)	(2)	(3)	(4)
(Intercept)	24.6	–	14.1	22.6
	(2.41) ***		(3.05) ***	(5.39) ***
F take (β)	0.23	0.104	0.21	0.506
	(0.0654) ***	(0.0435) *	(0.066) **	(0.15) ***
Same group (γ)	-7.4	-3.91	-6.42	-15.6
	(3.48) *	(2.5)	(3.62) +	(8.36) +
F take ✕ Same group (δ)	0.293	0.172	0.227	0.721
	(0.0962) **	(0.0727) *	(0.0987) *	(0.229) **
Early	-1.55	–	5.96	-0.844
	(3.24)		(3.82)	(7.4)
F take ✕ Early	-0.0603	-0.131	-0.27	-0.208
	(0.0969)	(0.0434) **	(0.109) *	(0.216)
Same group ✕ Early	2.8	-1.03	-4.57	2.6
	(4.74)	(2.66)	(5.3)	(10.8)
F take ✕ Same group ✕ Early	-0.119	0.0447	0.152	-0.246
	(0.14)	(0.0896)	(0.153)	(0.315)
Quiz 2 met P	-11.8	4.83	-12.3	-25.2
	(7.55)	(6.1)	(6.83) +	(18.2)
F take ✕ Quiz 2 met P	0.279	-0.168	0.314	0.639
	(0.228)	(0.187)	(0.219)	(0.538)
Same group ✕ Quiz 2 met P	14.3	-2	14.4	25.4
	(10.9)	(9.17)	(10)	(27)
F take ✕ Same group ✕ Quiz 2 met P	-0.204	0.193	-0.264	-0.466
	(0.316)	(0.274)	(0.295)	(0.812)
Model	Linear	Linear	Linear	Tobit
Controls	–	Twoway FE	History	–
N	832	832	592	832
N indiv.	192	192	192	192
Adj. R2	0.14	0.0707	0.243	–
LogLik	–	–	–	-2170

*Table A16: Estimates of equation (1) interacted with Quiz 2 met P, random bonus sessions, first party treatments. Standard errors in parentheses, clustered by individual for columns 1-3. + p < 0.10; * p < 0.05; p < 0.01; *** p < 0.001.*

3. Further results

Inequality

In this section we report some statistical analyses of further topics, which the main paper only briefly summarizes. Between-group inequality is closely related to unfairness, and it may be an important channel leading from ethnic diversity to conflict (Cederman et al. 2011) and inefficient public good provision (Baldwin and Huber 2010). In our experiment, subjects could observe the quiz earnings of F and P. We therefore examine whether high earning groups were more likely to elicit group reciprocity. Table A17 interacts equation (1) with F's quiz earnings. The interaction *F take* × *Same group* × *F earnings* is positive and significant except under the fixed effects specification. So, there is some evidence that high earning groups indeed received more group reciprocity.

	(1)	(2)	(3)	(4)
(Intercept)	21.6	–	12.8	16.8
	(2.41) ***		(3.08) ***	(5.13) **
F take (β)	0.314	0.0702	0.251	0.673
	(0.0663) ***	(0.0538)	(0.0708) ***	(0.152) ***
Same group (γ)	-2.62	-4.14	-2.4	-6.18
	(3.06)	(2.5) +	(3.54)	(7.39)
F take ✕ Same group (δ)	0.083	0.122	0.0698	0.205
	(0.0909)	(0.0736) +	(0.101)	(0.217)
F earnings	0.0308	-0.0121	0.0539	0.0726
	(0.034)	(0.0271)	(0.0401)	(0.0739)
F take ✕ F earnings	-0.00186	-0.000818	-0.00238	-0.00434
	(0.00095) +	(0.000765)	(0.00109) *	(0.00214) *
Same group ✕ F earnings	-0.0527	0.00392	-0.0921	-0.134
	(0.0459)	(0.0364)	(0.0542) +	(0.105)
F take ✕ Same group ✕ F earnings	0.003	0.00132	0.00384	0.00783
	(0.0013) *	(0.00105)	(0.00149) *	(0.00314) *
Model	Linear	Linear	Linear	Tobit
Controls	–	Twoway FE	History	–
N	832	832	592	832
N indiv.	192	192	192	192
Adj. R2	0.132	0.0422	0.235	–
LogLik	–	–	–	-2170

Table A17: Estimates of equation (1) interacted with F earnings*, random bonus sessions, first party treatments. Standard errors in parentheses, clustered by individual for columns 1-3. + p < 0.10; * p < 0.05; p < 0.01; *** p < 0.001.*

In-group identity

Social Identity Theory (SIT) is the longest-established psychological theory of group behavior (Tajfel et al. 1971). It holds that individuals support their group, and discriminate against other groups, in order to protect the self-esteem and security they derive from their social identity as a group member. If we treat group reciprocity as a form of group discrimination, SIT would predict

that subjects with the strongest attachment to their own group will be most likely to group-reciprocate. Our final questionnaire included a set of questions derived from standard measures of group identity (Ellemers et al. 1999). We sum answers to these (on a 1-7 Likert scale) to create the variable *Group ID*. Table A18 interacts equation (1) with Group ID. The interaction term *F take* × *Same group* × *Group ID* is significant in only two out of four specifications, although it always has the expected sign.

	(1)	(2)	(3)	(4)
(Intercept)	23.2	–	15.9	21
	(1.72) ***		(2.52) ***	(3.56) ***
F take (β)	0.217	0.0322	0.134	0.437
	(0.0466) ***	(0.0384)	(0.0505) **	(0.104) ***
Same group (γ)	-4.1	-3.49	-5.87	-10.7
	(2.09) +	(1.75) *	(2.45) *	(5.16) *
F take × Same group (δ)	0.202	0.179	0.236	0.529
	(0.0633) **	(0.0525) ***	(0.071) ***	(0.151) ***
Group ID	-1.02	–	-0.354	-2.58
	(0.709)		(0.75)	(1.49) +
F take × Group ID	-0.00908	-0.02	-0.0101	-0.00176
	(0.0197)	(0.0163)	(0.0211)	(0.0454)
Same group × Group ID	-2.26	-1.9	-2.45	-3.81
	(0.916) *	(0.767) *	(1.05) *	(2.23) +
F take × Same group × Group ID	0.0593	0.0572	0.0497	0.0844
	(0.0277) *	(0.0231) *	(0.0307)	(0.0651)
Model	Linear	Linear	Linear	Tobit
Controls	–	Twoway FE	History	–
N	832	832	592	832
N indiv.	192	192	192	192
Adj. R2	0.162	0.0431	0.249	–
LogLik	–	–	–	-2160

Table A18: Estimates of equation (1), random bonus sessions, first party treatments, interacted with Group ID. *Standard errors in parentheses, clustered by individual for columns 1-3. + p < 0.10; * p < 0.05; p < 0.01; *** p < 0.001.*

4. Participant Instructions

Translated from German.

Welcome to the Experimental Laboratory!

Please turn off your mobile phones and all other electronic devices now!

You are participating in an economic experiment. Depending on your own choices and choices of participants matched with you, you will be able to earn a non-trivial amount of money. For this, it is important that you read carefully and understand the following instructions.

During the experiment communication among the participants is strictly prohibited, except in situations in which it is explicitly allowed! In case you have any questions, please raise your hand and an experimenter will come to you to answer your question.

Non-compliance with this rule and with the instructions of the experimenters will lead to an exclusion from the experiment and all payments!

Your decisions during the experiment will be made anonymously. Only the experimenter will get to know your identity, but your decisions will not be relatable to your identity.

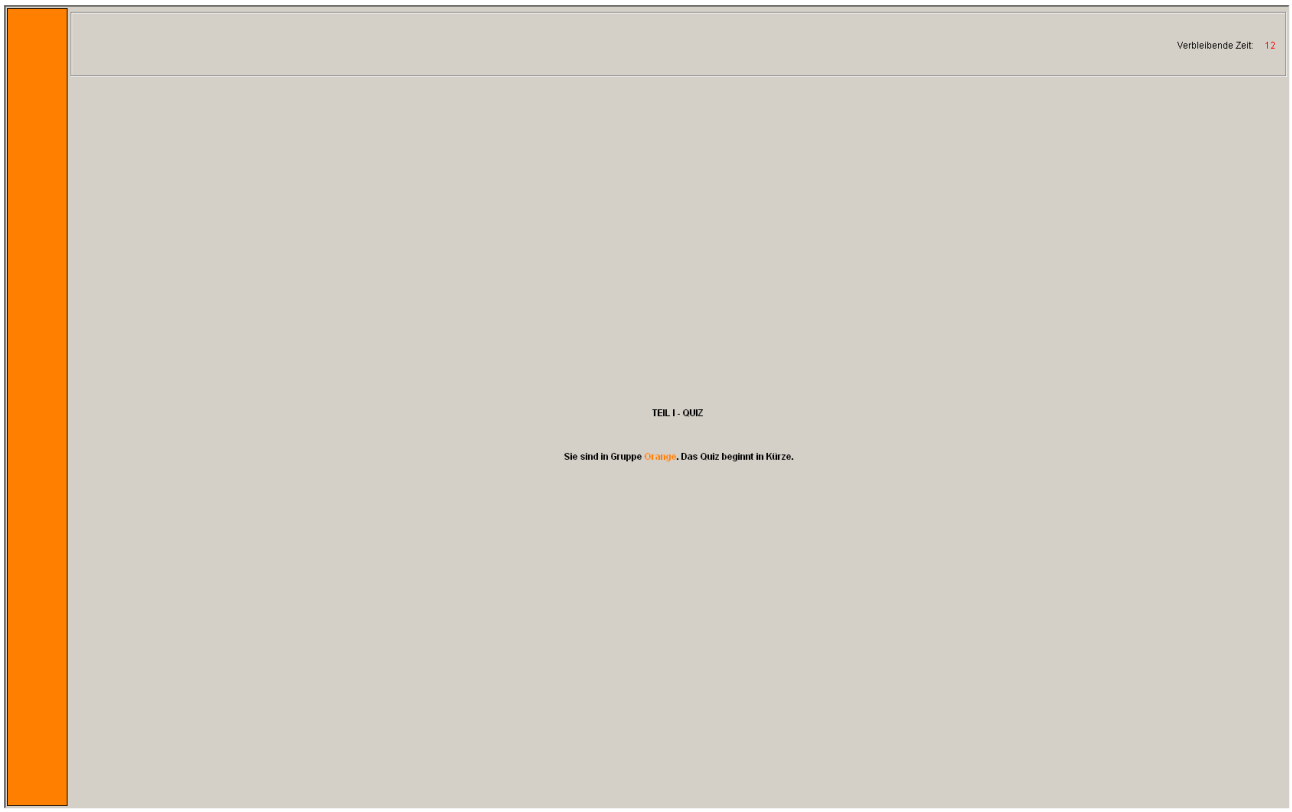
You will receive EUR 5 as show up fee. Additional payments will depend on your decisions and those of participants matched with you. Earnings will during the experiment be entitled Expected Currency Units, abbreviated as ECU. The total amount of ECU you gained during the experiment will be converted at the end, where

1 ECU = 5 CENT (0.05 EUR)

and be paid in cash.

In this experiment, you and the other participants will randomly be assigned one of 4 groups: green,

orange, purple, and brown. Each group consists of 4 participants. The color of your group will be shown to you onscreen. During all parts and rounds of the experiment, you will see the time remaining for a choice in the upper right corner of the screen. The screen at the beginning of the experiment will look similar to the following (please note that no decision has to be made on this first screen):



[THE TEXT ON THE SCREEN READS AS FOLLOWS:
UPPER RIGHT CORNER: “Remaining time”

MAIN BODY: “PART I – QUIZ

You are in group orange. The quiz will begin shortly.”]

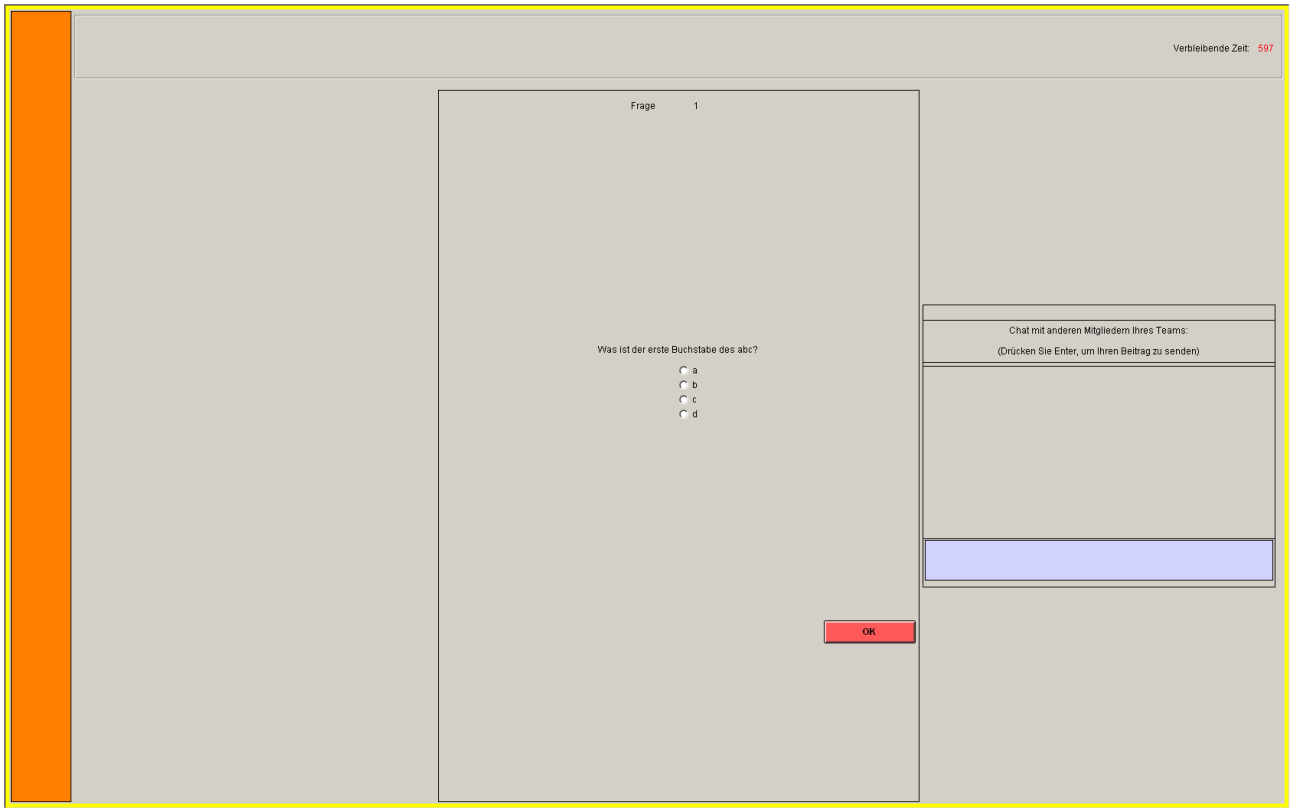
The experiment consists of 2 parts.

PART I

Each group will play a quiz. The aim of the quiz is to answer different questions about general

knowledge. On the whole, there are 20 questions with 4 possible answers each. Each question can be answered individually, but you have the option to exchange information with the other members of your group via computer chat. This is the only allowed form of communication over the course of the experiment.

The screen will look similar to the following. In the blue field on the right hand side you can enter your chat contributions. In the grey field above it you can read the contributions of the other members of your group.



[THE TEXT ON THE SCREEN READS AS FOLLOWS:

CENTER BOX:

“What is the first letter of the abc?”

RIGHT BOX:

“Chat with other members of your team:

(Please press enter to send your contribution)]

For each correct answer, you will receive one point. On the whole you have 10 minutes to answer all questions. Questions not or wrongly answered will be evaluated with 0 points. The computer will add all points of each group's members.

Your group will receive 10 ECU for each of the 80 possible points. (4 participants per group x 20 correct answers = 80 points maximum.)

Example:

Assume the participants of a group answer 20, 15, 14, and 11 questions correctly. The overall number of points of this group will then be 20+15+14+11=60. The group will then receive 10x60=600 ECU.

There is however a minimum earning of 400 ECU per group, i.e. Should your group for instance achieve 0 points in the quiz, you will nevertheless receive 400 ECU.

The earnings of the group will be distributed as follows:

- 400 ECU will be subtracted from the overall earnings and divided equally among the group members. These will serve as endowment for part 2 of the experiment. Therefore, every participant will begin part 2 with 100 ECU.
- All remaining ECU will also be divided equally among the group members (partial amounts will be rounded up) and constitute your earnings of part 1.

Example continued:

Let's take the 600 ECU of the example group. 400 ECU will be used for part 2 of the experiment. The remaining 200 ECU will be divided equally among the participants: Each participant gets 200/4=50 ECU(=2.50 EUR) from part 1.

After the quiz, we ask you to briefly answer three questions to the quiz.

Bonus ECU

[IN WINNER SETTING]: The group achieving the most points in the quiz will receive a bonus of 100 ECU per participant.

[IN RANDOM SETTING]: A randomly selected group will receive a bonus of 100 ECU per participant.

PART II

The second part of the experiment consists of 8 rounds, in each of which you will have to make 2 decisions. At the end of the experiment, *one* of these 8 rounds will be *randomly* selected for your final payment. The earnings of this round only will be added to your earnings of part 1.

In each round you will be interacting with 3 other participants. These can be of any of the different groups – green, orange, purple, brown. After each round the composition of each group will be changed, such that you need not interact with the same participants. In each round participants will be given a number ranging from 1 to 4, which may change from round to round.

In each round, each of the 4 participants will have to make 2 decisions. After each decision, you will receive feedback on *one* of the other participants' choice.

[IN SESSIONS 11-12 THIS PARAGRAPH WAS CHANGED TO:

In each round, principally each participant will have to make 2 decisions. In some rounds the computer may make some of the decisions instead of the respective participant on a random basis.]

DESCRIPTION OF A ROUND

During each round the current participants are shown onscreen. For each participant you will be given information about his group membership and the amount he received in part 1.

DECISION 1

Participants 1 to 4 will be paired. Participants 1 and 2 are paired, and participants 3 and 4.

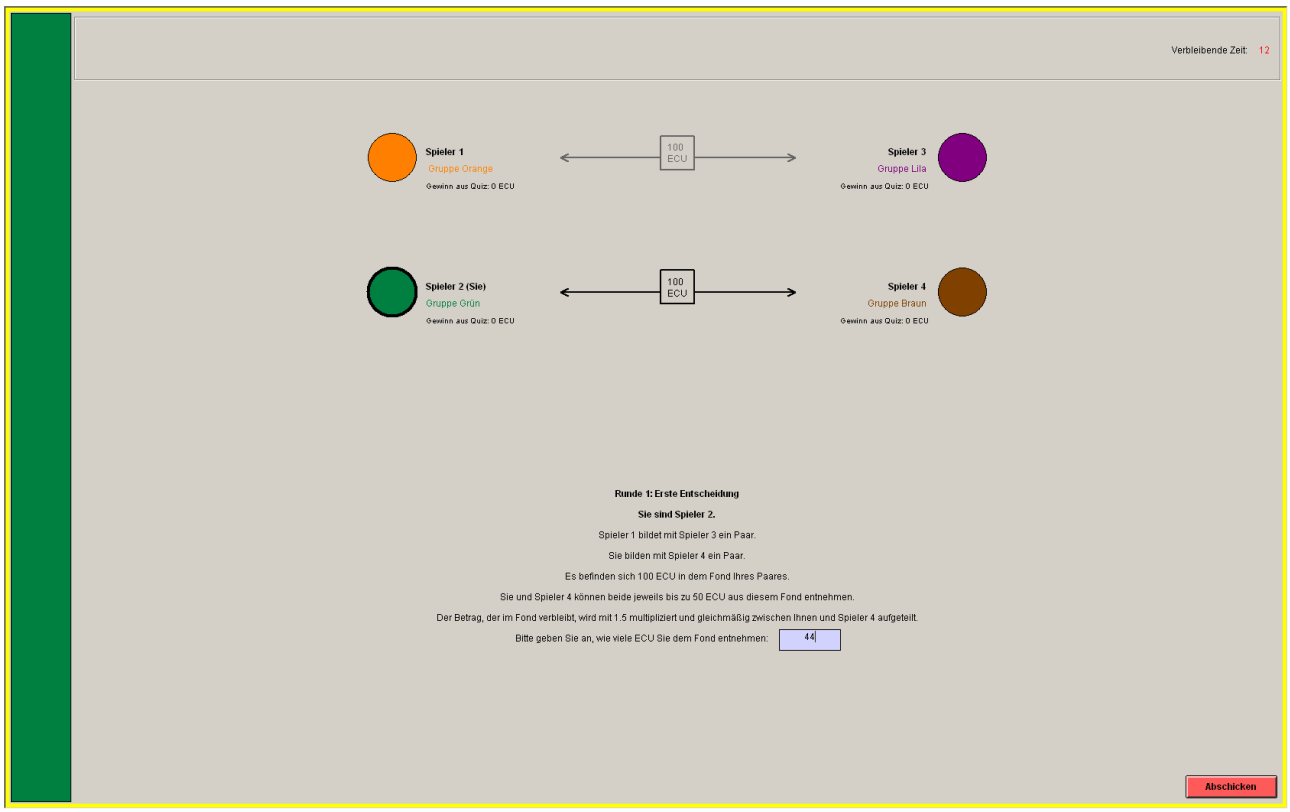
50 ECU of each of a pair's participants' earnings in part 1 will be put into a joint fund. The fund will thus contain 100 ECU.

Each participant of the pair may afterwards choose how much to withdraw from the fund. Both participants can simultaneously withdraw an amount between 0 and 50 ECU. Every ECU remaining in the fund will be multiplied by 1.5 and divided equally between the two (partial amount will be rounded up.)

Example:

Assume that participant 1 had withdrawn 33 ECU and participant 2 had withdrawn 37 ECU (the number mentioned were randomly chosen by a computer.) Then the fund would consist of 100-33-37=30 ECU. These are multiplied by 1.5 yielding 45, and distributed equally among the two participants in the pair. After rounding, each will receive 23 ECU. Participant 1 will receive 33+23=56, participant 2 37+23=60 ECU from this decision.

The screen will look similar to the following (please note that the numbers displayed were randomly drawn):



[THE TEXT ON THE SCREEN READS AS FOLLOWS:

TEXT NEXT TO EACH CIRCLE:

“Player (No.)

Group (color)

Earnings from quiz: 0 ECU”

TEXT AT BOTTOM:

“Round 1: First Decision

You are player 2.

Player 1 is paired with player 3.

You are paired with player 4.

There are 100 ECU in your pair's fund.

You and player 4 can both withdraw up to 50 ECU from the fund.

The amount remaining in the fund will be multiplied by 1.5 and divided equally among you and player 4.

Please indicate how many ECU you withdraw from the fund."]

DECISION 1 FEEDBACK

After decision 1 every participant will be informed about exactly one choice of one of the 4 participants. This can be the decision of the direct partner in the pair, it can also be the decision of a participant of the other pair. The decision will be announced as follows (please again note that the displayed numbers were randomly drawn):

[THE TEXT ON THE SCREEN READS AS FOLLOWS:

“Round 1: Feedback Decision 1

You were player 2.

Player 1 was paired with player 3.

You were paired with player 4.

Player 4 withdrew 41 ECU from your and his fund.

(The other participants may receive feedback from another decision than you do.)]

DECISION 2

Decision 2 equals decision 1. You are again matched with the same 3 other participants in a choice situation and be paired with one of them. But this can now be another participant than in decision 1. Again 50 ECU of each of two participants will be put into a fund, and both participants choose how much they withdraw from the fund. The amounts remaining in the fund will again be multiplied by 1.5 and divided equally among the two participants in the pair.

[IN SESSIONS 11-12 THIS PARAGRAPH WAS CHANGED TO:

In some rounds, some players' second decision may be made randomly by the computer. This will be shown onscreen by the words “random decision” next to the player. In these cases, the computer will automatically draw a number of ECU to take from 0 to 50. Each number from 0 to 50 is equally likely to be chosen. The player and his pair will be paid according to this decision, in the usual way.]

DECISION 2 FEEDBACK

After the second decision, the decisions of you and the participants you were matched with will be shown to you in a summary. The summary will show you how much you and the respective other participant withdrew from the joint fund and the amount of ECU overall achieved in this round. (Please again note that only one of the overall 8 rounds will be randomly selected for payment.)

The summary completes the round. In case another round follows, you will be matched with other participants and the procedure from decision 1 onwards begins anew. The screen of a summary will look similar to the following: (Please again note that the shown amounts are randomly drawn)

[THE TEXT ON THE SCREEN READS AS FOLLOWS:

ROW 1: “Decision 1 / You withdrew

Player 1 withdrew

Remaining in fund

x 1.5

your share (50%)

Share player 1 (50%)”

LAST ROW: “In total”]

[IN SESSIONS 11-12 THE FOLLOWING PART WAS DROPPED]

OTHER QUESTIONS

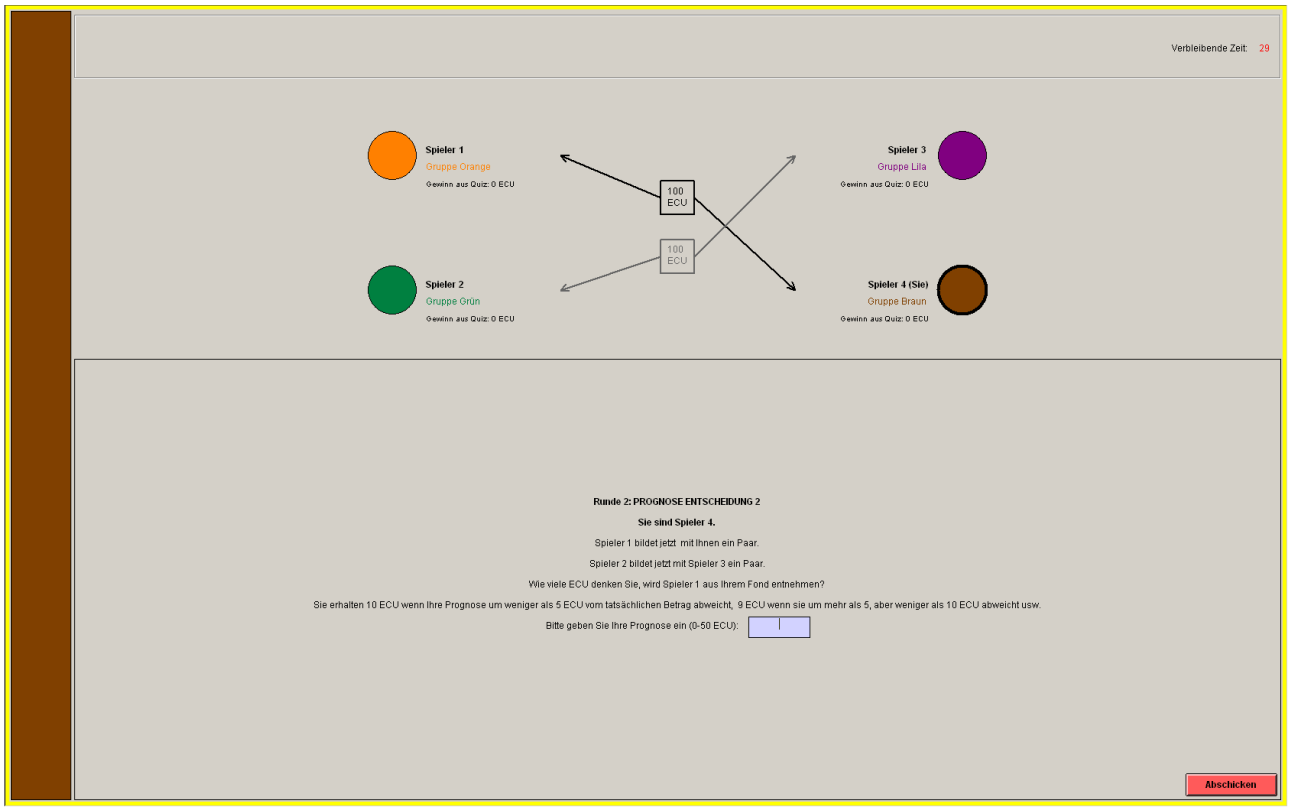
During some rounds you may be asked about your expectations concerning the amount the participant assigned to your pair will withdraw from the fund. If this is the case, you will get the possibility to earn a bonus. The amount of the bonus will be calculated according to the difference of your expectation and the de facto amount the participant withdrew. This difference will be divided into “intervals of five”. If the difference is less than 5 ECU, you will receive 10 ECU bonus. If the difference is between 5 and 9 ECU, you will receive a bonus of 9 ECU, if it is between 10 and 14 ECU you will receive a bonus of 8 and so on, up to a bonus of 1 ECU if the difference is between 45 and 49 ECU, and no bonus if the difference is 50 ECU. The possible boni are summarized in the following table:

Difference	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50
Bonus (ECU)	10	9	8	7	6	5	4	3	2	1	0

Example:

If you believe the other participant will withdraw 45 ECU from the fund, but he only withdrew 27 ECU, the difference between your expectation and the true amount is 45-27=18 ECU. 18 is in the interval between 15 and 19. You will thus receive a bonus of 7 ECU.

The screen for an expectation will look similar to the following:



[THE TEXT ON THE SCREEN READS AS FOLLOWS:

“Round 2: Expectation decision 2

You are player 4.

Player 1 is paired with you.

Player 2 is paired with player 3.

How many ECU do you think player 1 will withdraw?

You receive 10 ECU if the difference between your expectation and the true withdrawal is less than 5 ECU, 9 ECU if it is more than 5 but less than 10 ECU etc.

Please indicate your expectation (0-50 ECU):"]

After the experiment there will be a brief questionnaire. Please answer these questions truthfully.

PAYMENT

After the end of the experiment the round of part 2 chosen for payment will be announced. You will then be called separately and payed privately. Additionally to the ECU earnings of the experiment you will receive 5 EUR show up fee. You will have to sign a receipt for the money, but this receipt will in no way be associated with your choices during the experiment. After receiving your payment, please exit the lab silently.

Do you have any questions regarding the experiment?