Intergroup revenge: a laboratory experiment on the causes[1]

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

Affiliations:
[*] University of Essex, dhughj@essex.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. Which motivations account for such behavior is, however, not clear. We test the hypothesis that people are predisposed to reciprocate against groups. In a 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 as unfairly advantaged. Our results support a behavioral mechanism leading from perceived injustice to intergroup conflict. We discuss the relevance of group reciprocity to political and economic 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, Catherine Scacco, Rafael Hortalla-Vallve, Nadine Chlass, Matthew Braham and seminar participants at the Max Planck Institute of Economics, University of Hamburg, University of Warwick, King's College London, University of East Anglia, NYU-CESS, IMEBE, and ESA 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

Violent intergroup conflict is a constant of human behavior, and one puzzling feature is its apparent irrationality. Here are two examples. After an argument between an Indian Dalit and an upper caste farmer, upper caste villagers attacked 80 Dalit families (Hoff et al. 2011). In Atlanta, 1906, after newspaper allegations of black attacks on white women, a group of whites went downtown to a black neighborhood and killed 25 black men (Bauerlein 2001). In both examples, innocent people were made to suffer for the real or supposed crimes of others. Very many field studies of ethnic and communal violence report similar tit-for-tat processes, with harm to members of one group being avenged by attacks on previously uninvolved coethnics of the original attackers (Horowitz 1985, 2001; Chagnon 1988).

To understand this behaviour, we suggest looking to the psychology of reciprocity. Laboratory experiments show that humans are prepared to reciprocate wrongs by harming the offender, even at a cost to themselves (Fehr and Gächter 2000; Falk and Fischbacher 2006). A natural extension, which would account for the episodes described above, is that humans reciprocate not only towards individuals, but also *towards entire groups*. We label retaliation against a perpetrator's group members *group reciprocity*.

Group reciprocity might underlie a range of important phenomena. First, it could explain the persistence of intractable conflicts, which hinder economic development in many of the world's poorest states (World Bank 2011). Second, it may affect routine social and economic life. For example, on days after terrorist bombings in Israel, Jewish (Arab) judges become more likely to favor Jewish (Arab) plaintiffs in their decisions, and Israeli Arabs face higher prices for used cars (Shayo and Zussman 2011, Zussman 2012). Consumers buy fewer products from countries which they see as politically antagonistic (Klein et al. 1999, Leong et al. 2008). Third, group reciprocity may affect international politics and macroeconomics. German voters are unwilling to bail out "the Greeks" because of "their" previous transgressions against norms of fiscal rectitude, and Greeks have reacted with similar group resentment. Revenge appears to have played an important part in twentieth-century history. The Treaty of Paris' intentional devastation of the German economy, for example, led Keynes (1922) to quote Thomas Hardy's play *The Dynasts*: "Nought remains / But vindictiveness here amid the strong, / And there amid the weak an impotent rage."

Whilst these stories suggest the social importance of intergroup revenge, they are clearly not conclusive evidence. There are many potential explanations of the Versailles treaty. Similarly, the field studies above cannot distinguish group reciprocity from in-group bias under external threat (Sherif et al. 1961). There may also be reputational or other strategic considerations if people are rewarded for defending their coethnics (Chagnon 1988). To identify our phenomenon cleanly, we need the control and replicability afforded by a laboratory setting. We therefore ran a computerized laboratory experiment to test for group reciprocity.

We emphasize that we do not believe episodes of violent conflict can be recreated in the lab. Instead, we aim to test a theory of human behavior which might explain them, and other phenomena, by reproducing a key underlying psychological mechanism. Doing this in a replicable experimental paradigm would open the door to exploration of why, how and when the mechanism operates. This may then help us understand the evolution of real violent conflicts.

Our experimental design has this goal in mind. Qualitative studies and cross-country regressions show an association between intergroup inequality and injustice, and intergroup conflict (Tambiah 1996; Horowitz 2001; Cederman et al. 2011). Group reciprocity might provide a causal explanation for this 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 measure group reciprocity across different treatments which vary the perceived unfairness of intergroup allocations.

Experimental work on groups has shown some fairly robust results. People value group membership, and prefer to interact with others from their own group (Hargreaves-Heap and Zizzo 2009; Currarini and Mengel 2012). They 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 and reward each other more (Bernhard et al. 2006; Chen and Li 2009; Currarini and Mengel 2012). Shared group identity may also provide a simple coordination mechanism for individuals (Chen and Chen 2011).

None of these phenomena appears sufficient to explain extreme prejudice and violence against outgroups (Brewer 1999). Indeed, in-group bias disappears when subjects allocate a "bad" such as exposure to aversive noise (Mummendey et al. 1992). As a result, psychologists have developed broader theories of intergroup emotions, including "vicarious retribution", a propensity to retaliate against groups (Lickel et al. 2006). Experiments on vicarious retribution have so far examined only verbal attitudes (Stenstrom et al. 2008), or cannot cleanly distinguish individual-level reciprocity from group reciprocity (Gaertner et al. 2008). Economists have run experiments on "generalized reciprocity", in which subjects reciprocate by being nicer or nastier to other people in general (Dufwenberg et al. 2001, Greiner and Levatti 2005, Stanca 2009). This phenomenon could be supported by many mechanisms, such as learning, imitation, or changes in mood. To our knowledge, ours is the first experiment that can identify reciprocation against specific groups.

Some economic experiments have examined group-level phenomena, such as intergroup contests and vendettas (Bornstein et al. 1992, 2003; Abbink and Herrmann 2009). Like individual reciprocity in public goods games (Fehr and Gächter 2000), group reciprocity may provide the underlying psychological mechanism that helps to explain these collective outcomes.

Theory

To motivate our experimental design, we present a simple theoretical framework. Consider a player i who interacts in sequence with two other agents. Let i's utility be given by

$$u(p, k_1, k_2) = V(p) + (\mu + \nu I_{SG})k_2 + (\beta + \delta I_{SG})k_1 k_2 \quad (1)$$

where p is own material payoff, $V(\cdot)$ is weakly concave, k_1 is the first agent's kindness to i and k_2 is i's kindness to the second agent. I_{SG} is an indicator taking the value 1 when the two other agents share group membership, and 0 otherwise. Agent *i*'s kindness to agent *j* can be measured as $(p_j - p_{min})/(p_{max} - p_{min})$ where p_j is *j*'s actual material payoff, and p_{max} and p_{min} are his maximum and minimum material payoffs over the set of *i*'s possible actions, taking *j*'s own action as given. The second term in this equation, $\mu + \nu I_{SG}$, represents altruism. The third term, $\beta + \delta I_{SG}$, represents reciprocity. Both terms can vary with the other agents' group membership. What interests us is δ, which measures the increase in reciprocity when the first and second agent are in the same group: in other words, the level of group reciprocity. We expect $\delta > 0$. The β coefficient measures baseline generalized reciprocity when the two agents are not in the same group.

In the episodes described above, revenge was taken not by previous victims themselves, but by others acting on their behalf. Our experiment tests both for *direct* group reciprocity, in response to actions affecting oneself, and for *indirect* or *third party* group reciprocity, in response to actions affecting others (Nowak and Sigmund 2005). Indirect group reciprocity can be analysed in the framework above by letting k_1 be the first agent's kindness to another player, as observed by *i*.

Design

The core of our experiment consisted of 8 repetitions of a two-round linear public goods game. The basic structure of each repetition is always the same. Subjects played a public goods game with one other subject, from a set of four subjects. They then received feedback about exactly one subject's play from the four. Lastly, they played another public goods game with one other subject from the four: the second round partner. After the second round, subjects learned the choices of both their partners, and their total earnings from the repetition.

In each public goods game, both subjects shared a fund of 100 Experimental Currency Units (ECU) with 1 ECU = 5 Euro cent. 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. Formally, subject *i*'s earnings p^i are given by

$$p^i = T^i + (100 - T^i - T^j) \times 0.75 \qquad (2)$$

when s/he withdraws T^i tokens and his or her partner withdraws T^j tokens. (2) clearly shows that withdrawing more tokens materially harms the partner and helps oneself. Total earnings from both rounds of one random repetition were used for payment.

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 own partner's play, labelled F in the figure. 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 their second round partner, labelled P in the figure. These treatments test for direct or "first party" group reciprocity.

In *third party* treatments, first choices were made with the horizontal player, but subjects received feedback only about the diagonal player, who had played against the subject marked O. Subjects did not learn how their own partner had played. Then subjects played the horizontal player again (here labelled P). 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.

F's play could affect subjects' play towards P for many reasons. We are specifically interested in when subjects play differently towards P because F and P are in the same group. To identify this, cross-cutting treatments varied the color group membership of the four subjects, as shown in Figure 2. In *same group* (*different group*) treatments, F and P were from the same (different) groups. By comparing these treatments, we can identify the effect of F and P's shared group membership, i.e. δ in equation (1).

In addition, in *own same group* (*own different group*) treatments, the "vertical" player O was from the subject's own group (a different group). In third party repetitions, this allows us to examine whether 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. Notice also that our design makes it clear that F and P are different people. There is no possibility that subjects are harming a whole group in order to reciprocate individually against one person within it, as there is in, e.g., Stenstrom et al. (2008) and Gaertner et al. (2008).

Between repetitions subjects were rematched into different sets of 4. The rematching ensured that all subjects experienced all 8 treatments over the 8 repetitions: {first party, third party} × {same group, different group} × {own same group, own different group}.[2] This "within subjects" design allows more accurate inferences about group reciprocity's individual-level covariates. Treatments were balanced over repetitions.

Procedure

We ran 15 sessions of 16 subjects each. Each session was divided into two parts: a group quiz, followed by the 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 color assignment, color 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; 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 endowment 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

2 Sessions 11 and 12 used only the first party treatment.

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. The instructions always explained how the bonus would be awarded, and the group receiving the bonus was announced directly after the quiz.

Random allocations are *ex ante* fair. However, in the context of a real effort task, we expected subjects to perceive random allocations as less fair than allocations based on "merit" in task performance (confer Cherry et al. 2002). We confirm this expectation below.

The quiz was followed by a brief questionnaire. Subjects then played eight repetitions of the public goods game, as described above. In each repetition, all 4 subjects' colors, as well as their earnings from the quiz, were displayed on screen. After the public goods game, subjects answered a questionnaire including demographics, measures of group identity, reactions to other groups, and questions about the experiment. Lastly payments were made privately. Certain sessions contained minor variations to this design, which we describe later in the text.

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 1 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.

Results

Our estimation is based on equation (1). Define T_1 to be the amount taken by the first round participant, and T_2 to be the amount *i* takes when playing against the second round partner. It is easy to see that $k_1 = (1 - T_1/50)$ and $k_2 = (1 - T_2/50)$. For the moment, we assume that expectations about the second round partner's take are unaffected by the first round partner's behaviour. As shown in the appendix, using a quadratic for *i*'s material welfare *V(p)*, and taking the first order condition, gives a linear form for T_2:

$$T_2 = \alpha + \beta T_1 + \gamma I_{SG} + \delta I_{SG} T_1 \quad (3)$$

where β and δ are the coefficients on reciprocity defined in (1) above. Based on this, we estimate individual *i*'s second round choice in round *t* as:

$$Second\ round\ take_{i,t} = \alpha + \beta\ F\ take_{i,t} + \gamma\ Same\ group_{i,t} + \delta\ (F\ take_{i,t} \times Same\ group_{i,t}) + X_{i,t}\theta + \varepsilon_{i,t}. \quad (*)$$

Thus, β gives the partial correlation between *F take* (i.e. T_1), the amount taken by the feedback participant F, and the subject's *Second round take* (T_2) 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* ($I_{SG} = 1$). *X* is a possible vector of controls.

Our first key result is:

Result 1: There is significant evidence for first party group reciprocity aggregating over all sessions, and within random bonus sessions, but not within winner bonus sessions.

To give an initial sense of Result 1, Figure 3 plots *F take* against *Second round take* in first party treatments during random bonus sessions. The slope of *Second round take* on *F take* was about doubled when F and P were from the same group.

Subjects' choices within a session may not be independent. As a conservative test of significance, we first calculated values of β and $\beta + \delta$ in (*) 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, $\beta + \delta$ 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 $\beta + \delta$ 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).

Our second key result is a negative one.

Result 2: Subjects showed no third party group reciprocity in either winner or random bonus sessions.

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).

Results 1 and 2 are confirmed in a regression using individual observations. Table 2 estimates equation (*) 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. Regressions include period and individual dummies, to control for individual heterogeneity and changes in behavior over time. When these are accounted for, β is not significantly different from zero, in other words, there is no evidence for generalized reciprocity beyond the boundaries of a color group. By contrast, in first party treatments, the δ coefficient on *F take* × *Same group* is always 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.[3] There is also no evidence for third party group reciprocity in *own same group* repetitions alone, as shown in the appendix.

Why did we not find evidence for third party group reciprocity? One possible explanation is that

3These include using session-clustered standard errors; removing the fixed effects; 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 for all regressions presented in the main paper.

people are more likely to react to other groups as units, than to see themselves as group members, with ties and obligations to their fellows. Our armchair intuition is that many quite selfish people nevertheless treat outgroups as coherent, potentially hostile actors. With a stronger group treatment, group reciprocity and in-group altruism could combine to create third party group reciprocity. A second possibility is that our experimental design encouraged subjects to focus on what was being done to them directly, even when this was not observable.

Next, we look for possible correlates of group reciprocity, focusing 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?" Figure 4 shows the distribution of answers, by session type. Answers were 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 3 interacts equation (*) 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 at the 10% level, and the summed coefficient is not significantly different from zero. Subjects who perceived the quiz as unfair were about two-and-a-half times more group-reciprocal (0.271 versus 0.271 – 0.174 = 0.097). Including winner bonus sessions in these regressions gives similar results[4]; controlling for fairness perceptions reduces but does not eliminate the difference between sessions.

Result 4: Group reciprocity remained significant after controlling for subjects' expectations of their

4 Available on request.

second round partner's choices.

The feedback participant F's actions might change the subject's expectations about other members of F's group, including P. Subjects might then behave differently because of these expectations, either because of a different expected material payoff, or 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 1972), 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.[5]

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. 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.[6] *Expectations* is highly significant, and the δ coefficient on *F take* × *Same group* shrinks by about a third. However, it remains significant. In other words, both expectations and raw preferences seem to be playing a role.

Because subjects' stated expectations may be affected by their intended play, for example due to self-justification (Messé and Sivacek 1979), we used a further method to control for expectations. In sessions 11 and 12, in each repetition, one player's second round take was determined randomly, by a computerized draw from the uniform distribution on {0,...,50}.[7] Their second round partners knew this, and both partners were paid as normal from the decisions. In this case, subjects' expectations about P's play ought to be unaffected by F's play.[8] Thus, we can examine the effect of expectations

5 In fact, there were no significant correlations within groups and expectations were not significantly affected by other players' group memberships. See the appendix.
6 The appendix describes this procedure in detail.
7 In order to gather more information, these sessions only included first party treatments, with two repetitions of each treatment.
8 We cannot rule out "counterfactual reciprocity", based on what subjects believed P would have done.

in a different way, by holding S's expectations of P's behavior constant. Column 3 of Table 4 interacts equation (*) 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, although the summed coefficient is imprecisely estimated. Overall, it appears that group reciprocity cannot be explained by changes in beliefs alone.

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 sessions 13-15, after 4 repetitions a second quiz took place, in which some subjects interacted and cooperated with members of other color groups. Column 4 of Table 4 interacts equation (*) 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 null result concerns gender. Some evolutionary theories 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.

We interacted the basic regression in column 2 of Table 3 with a gender dummy. None of the interactions was significant at the 5% level, and an F test could not reject the null of zero coefficients on all interaction terms ($p = 0.12$). 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.

Conclusion

Group membership matters to social and economic behavior (Arrow 1998; Akerlof and Kranton 2000, 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). The results here show that even absent these factors, humans may be led to take revenge upon groups. But the motivation only became active under certain conditions: in contexts when rewards were allocated indiscriminately, rather than upon the basis of merit. And subjects who perceived this as unfair were more group-reciprocal.

Experiment participants did not reciprocate for actions affecting other people, even other in-group members. Many real world examples of conflict seem to involve reciprocation by third parties, and this phenomenon may be important, because it creates conditions where conflict can spread very fast. So, third party reciprocity needs more empirical study. Researchers could use stronger treatments, or home-grown identities, to increase identification with the in-group.

Previous experiments on groups, including both lab-grown groups and real ethnic groups, have typically used one-shot interactions to reveal intergroup prejudice and discrimination (Fershtman and Gneezy 2001; Habyarimana et al. 2009; cf. Bornhorst et al. 2010). The underlying model is static: subjects begin an interaction with predetermined (perhaps experimentally induced) in-group biases, which they then implement. We believe that the science could progress by focusing more on intergroup dynamics like those in this paper. For instance, history has shown us that apparently peaceful and integrated multicultural societies like Bosnia can descend swiftly into brutal ethnic civil wars. Static models of prejudice, and experiments designed to measure it, seem less helpful in explaining this process than the dynamics of how groups react to one another (cf. Whitt and Wilson 2007). Similarly, rational-actor models have been better at explaining why politicians whip up outgroup hatred than why the public responds (cf. Glaeser 2005). Modelling group reciprocity might explain how politicians can create conflict by provoking small hatred-producing episodes.

Understanding how group reciprocity might evolve could shed light on the evolution of intergroup motivations more generally (cf. Choi and Bowles 2007). 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 humans may also learn group reciprocity as an appropriate response to strategic incentives, then apply the resulting behavioral heuristic in inappropriate settings, including the laboratory (Chagnon 1988; Hardin 1995; Fearon and Laitin 1996).

Lastly, we know little about the behavioral connection between perceived intergroup unfairness and conflict. Much more needs to be done to understand how what contexts create breeding grounds for intergroup resentment, and how this psychology plays out in particular environments. Consider affirmative action. Does it soothe intergroup resentment by remedying past injustices? Or does it risk leading to retaliation from the now-disadvantaged group, and a further cycle of injustice? These questions may be as important as affirmative action's direct effect on incentives. 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); similarly, the introduction of "unfair" policies may lead to the breakdown of cooperation between different groups. Ultimately, understanding intergroup dynamics could help 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. 2000. “Economics and Identity”, *Quarterly Journal of Economics* 115(3): 715-753.

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

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.

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.

Brewer, Marilyn B. 1999. The psychology of prejudice: Ingroup love and outgroup hate?. *Journal of Social Issues* 55 (3): 429-444.

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, Roy and Yan Chen. 2011. "The Potential of Social Identity for Equilibrium Selection." *American Economic Review*, 101(6): 2562-2589.

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

Cherry, Todd L., Peter Frykblom, and Jason F. Shogren. 2002. "Hardnose the Dictator." *American Economic Review* 92(4): 1218-1221.

Choi, Jung-Kyoo and Samuel Bowles. 2007. "The coevolution of parochial altruism and war." S*cience*, 318.5850: 636-640.

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.

Falk, Armin and Urs Fischbacher. 2006. "A Theory of Reciprocity." *Games and Economic Behavior,* 54: 293-315.

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.

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, Lowell, Jonathan Iuzzini, and Erin M. O'Mara. 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.

Greiner, Ben and Vittoria M. Levatti. 2005. "Indirect Reciprocity in Cyclical Networks. An Experimental Study." *Journal of Economic Psychology* 26: 711-731.

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.

Honaker, James, Gary King, and Matthew Blackwell. 2012. "Amelia II: A Program for Missing Data." *R package*.

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.

Klein, Jill Gabrielle, and Richard Ettensoe. 1999. “Consumer Animosity and Consumer Ethnocentrism.” *Journal of International Consumer Marketing* 11 (4): 5–24.

Leong, Siew Meng, Joseph A. Cote, Swee Hoon Ang, Soo Jiuan Tan, Kwon Jung, Ah Keng Kau, and Chanthika Pornpitakpan. 2008. “Understanding Consumer Animosity in an International Crisis: Nature, Antecedents, and Consequences.” *Journal of International Business Studies* 39(6): 996–1009.

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.

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.

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.

Stanca, Luca. 2009. “Measuring Indirect Reciprocity: Whose Back Do We Scratch?” *Journal of Economic Psychology* 30: 190 – 202.

Stenstrom, Douglas M., Brian Lickel, Thomas F. Denson, and Norman Miller. 2008. The Roles of Ingroup Identification and Outgroup Entitativity in Intergroup Retribution. *Personality and Social Psychology Bulletin, 34:* 1570-1582.

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.

Whitt, S., and R. K Wilson. 2007. "The Dictator Game, Fairness and Ethnicity in Postwar Bosnia." *American Journal of Political Science* 51(3): 655–68.

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

Zussman, Asaf. 2012. "Ethnic Discrimination: Lessons from the Israeli Online Market For Used Cars." Working paper.

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 1: Descriptive statistics.

	(1)	(2)	(3)
	All sessions	Random bonus sessions	Winner bonus sessions
Third party	-2.08	-2.58	2.13
	(1.66)	(1.87)	(3.59)
First party × F take (β)	0.0336	0.0277	0.0606
	(0.0329)	(0.0368)	(0.073)
———"——— × Same group (γ)	-3.96	-4.05	-2.11
	(1.54) *	(1.73) *	(3.57)
———"——— × F take × Same group (δ)	0.174	0.187	0.109
	(0.0457) ***	(0.051) ***	(0.105)
Third party × F take (β)	0.0781	0.0859	0.0133
	(0.0368) *	(0.0417) *	(0.0797)
———"——— × Same group (γ)	-0.0389	1.19	-6.84
	(1.73)	(1.95)	(3.75) +
———"——— × F take × Same group (δ)	0.0399	0.00627	0.19
	(0.0497)	(0.0565)	(0.105) +
Model	Linear	Linear	Linear
Controls	Period and indiv. FE	Period and indiv. FE	Period and indiv. FE
N	1856	1472	384
N indiv.	240	192	48
Adj. R2	0.027	0.0279	0.0314

*Table 2: 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, plus per-period and per-individual fixed effects. Robust standard errors clustered by individual in parentheses. + p < 0.10; * p < 0.05; p < 0.01; *** p < 0.001. Total N less than 1920 because 64 decisions were made by computer in sessions 11 and 12.*

	(1)	(2)	(3)	(4)
F take (β)	-0.0363	-0.0526	0.0433	0.0284
	(0.055)	(0.0414)	(0.0399)	(0.04)
Same group (γ)	-5.39	-3.57	-3.7	-3.6
	(2.57) *	(1.62) *	(1.82) *	(1.82) *
F take × Same group (δ)	0.271	0.129	0.18	0.178
	(0.0758) ***	(0.0468) **	(0.0542) ***	(0.0546) **
F take × Fair	0.134	--	--	--
	(0.0771) +			
Same group × Fair	2.73	--	--	--
	(3.52)			
F take × Same group × Fair	-0.174	--	--	--
	(0.105) +			
Expectations	--	0.628	--	--
		(0.0535) ***		
Random choice	--	--	8.41	--
			(5.17)	
F take × Random choice	--	--	-0.206	--
			(0.143)	
Same group × Random choice	--	--	-4.47	--
			(6.45)	
F take × Same group × Random choice	--	--	0.0947	--
			(0.19)	
Open quiz 2	--	--	--	4.1
				(4.7)
F take × Open quiz 2	--	--	--	0.0457
				(0.133)
Same group × Open quiz 2	--	--	--	-2.9
				(6.38)
F take × Same group × Open quiz 2	--	--	--	0.0271
				(0.179)
Model	Linear	Linear with multiple imputation	Linear	Linear
Controls	Period and indiv. FE	Period and indiv. FE	Period and indiv. FE	Period and indiv. FE
N	832	832	832	832
N indiv.	192	192	192	192
Adj. R2	0.0404	0.217	0.0387	0.0393

*Table 3: 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 per-period and per-individual fixed effects and further terms. Robust standard errors clustered by individual in parentheses. + p < 0.10; * p < 0.05; p < 0.01; *** p < 0.001.*

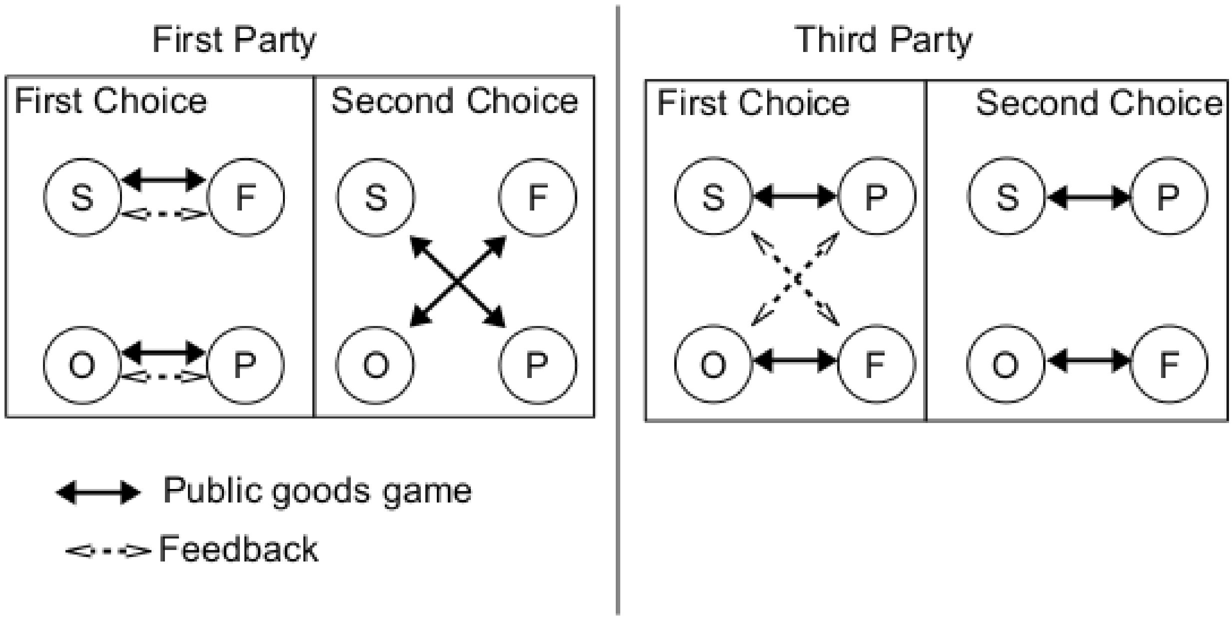

Figure 1: Experiment design. Subject is marked with an S. F denotes the feedback subject, P the second round partner, O the other player. S learns how F played, then reacts in play against P

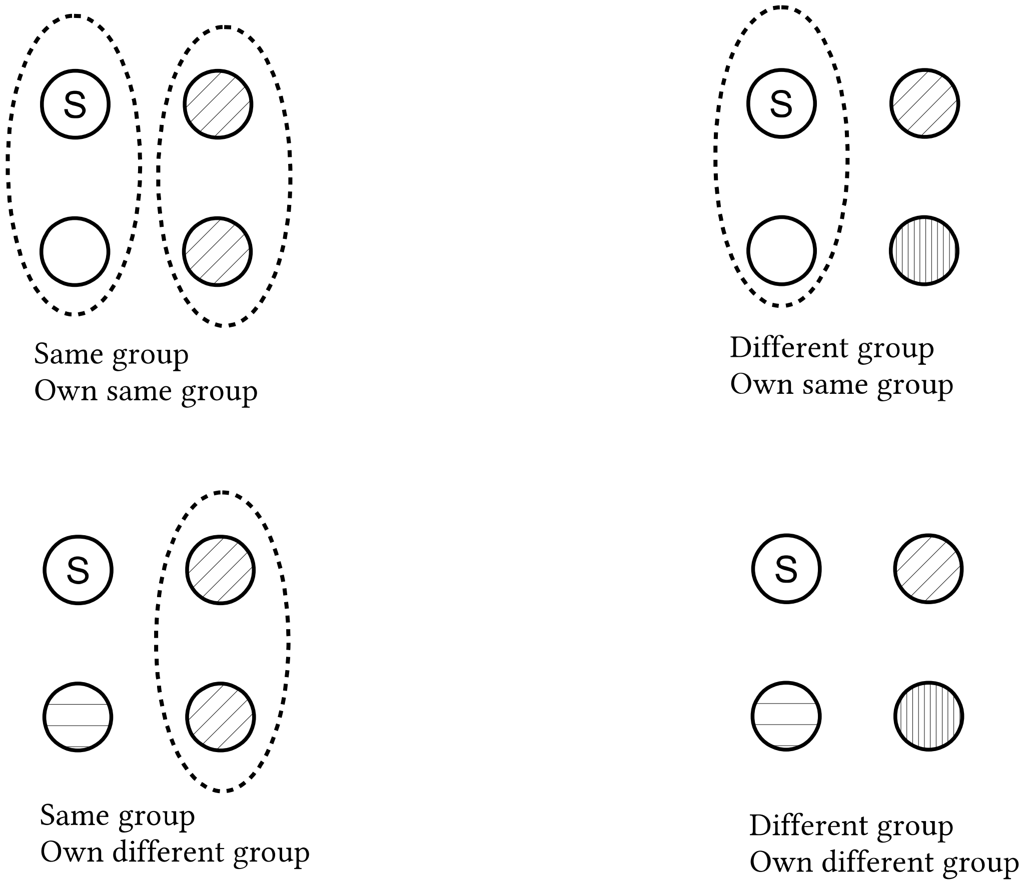

Figure 2: Color treatments. Subject is marked with an S.

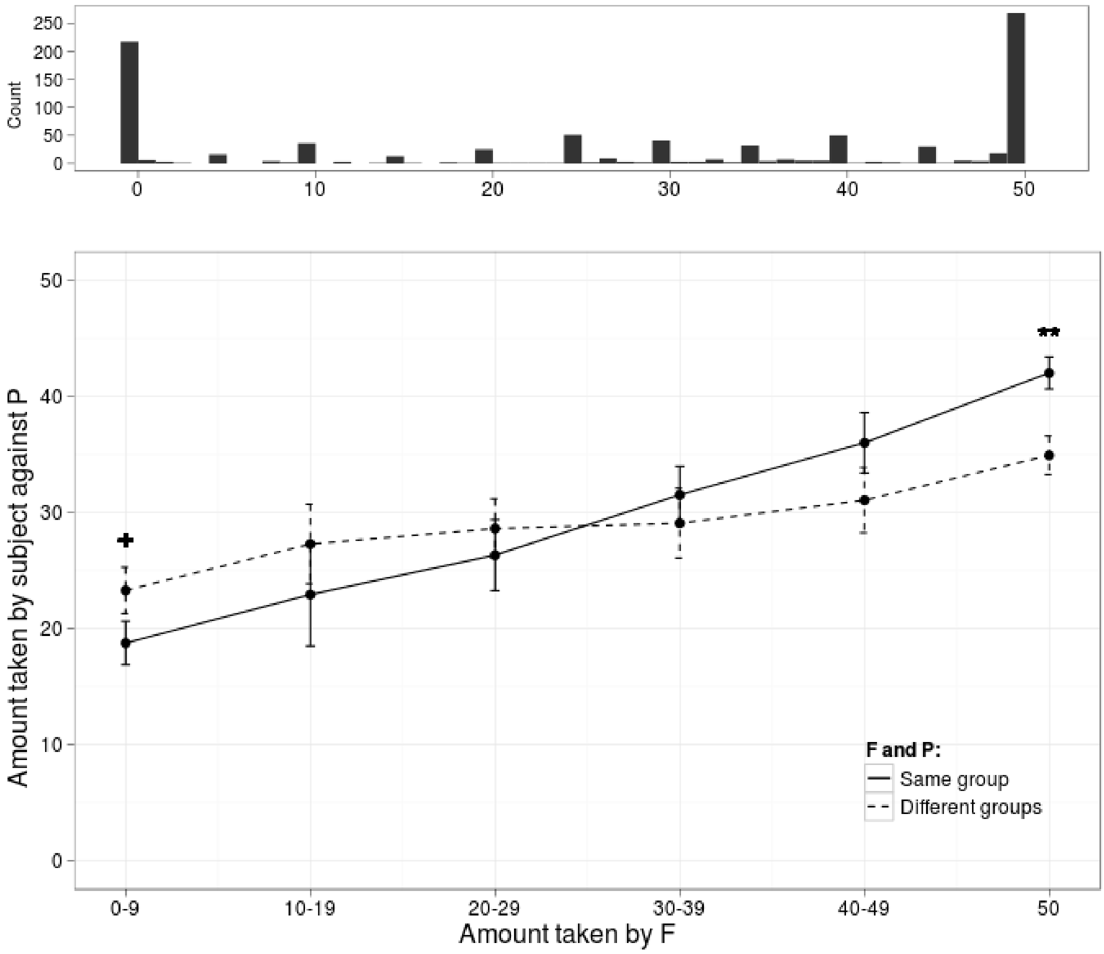

*Figure 3: Mean observed subjects' taking against P, by F's taking, first party treatments, random bonus sessions. Bars show ± 1 s.d. + $p < 0.10$; ** $p < 0.01$. Top plot is histogram of amount taken by F.*

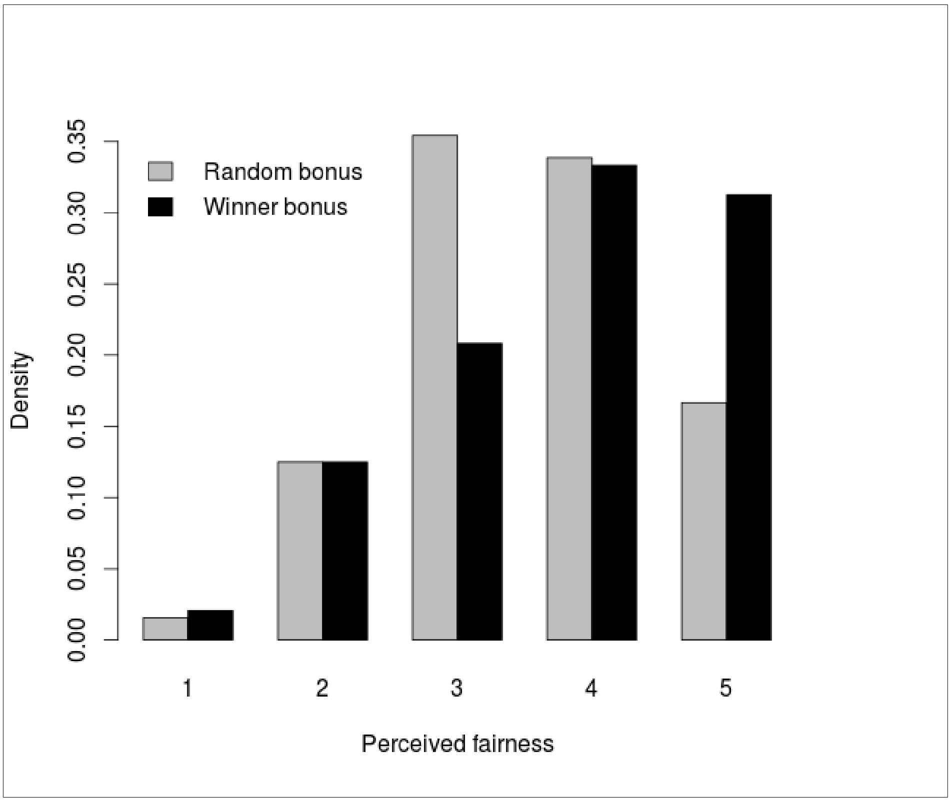

Figure 4: Perceptions of fairness in random and winner bonus sessions (1 = "very unfair", 5 = "very fair")