The Role of Group Boundaries in Human Altruism and Anti-Social Behavior

Stephan Meier,[1*] David Huffman,[2*] Matthias Sutter,[3] Lorenz Goette[4†]

[1]Graduate School of Business, Columbia University, 3022 Broadway, New York, NY 10027, USA
[2]Department of Economics, Swarthmore College, 500 College Avenue, Swarthmore, PA 19081, USA
[3]Department of Public Finance, University of Innsbruck, Universitaetsstrasse 15, 6020 Innsbruck, Austria
[4]Department of Economics, University of Lausanne, Internef, 1015 Lausanne-Dorigny, Switzerland

[*]These authors contributed equally to this work.
[†]To whom correspondence should be addressed; E-mail: lorenz.goette@unil.ch.

Exchange between humans can be facilitated by altruistic cooperation, and the willingness to punish those who act selfishly. But in some societies punishment is directed at cooperators as well as defectors, in a way that harms efficiency and is more akin to outright hostility and aggression (1). This raises the fundamental question of how such different patterns could have emerged. One possible explanation involves social group boundaries, and the economic environment between groups. When there is competition, a group's survival chances may be improved if individuals cooperate within their group, but attack other groups (2). In neutral environments, by contrast, it may be optimal to cooperate and use punishment to enforce a norm of cooperation on behalf of one's group (3; 4; 5). We conduct experiments among platoons in the Swiss Army to examine these conflicting views about punishment. We find stronger cooperation with one's own group, norm enforcement, and no sign of hostil-

ity between groups in neutral environments. When competition is added, we find even stronger cooperation within groups, and a strong pattern of hostile punishment towards other groups. These results help explain how, both, social and anti-social punishment can arise, (6; 7; 8; 9; 1; 10) and shed light on the sources of cooperation and conflict between groups.

Many beneficial exchanges require humans to cooperate and trust each other, even though narrow self-interest may tempt them to act otherwise. Evidence suggests that altruistic cooperation, while measurable, is not very strong on its own (11; 12). But experiments show that humans are willing to expend resources to impose harm on others who act selfishly (13; 14; 15; 16; 12; 3; 17). This altruistic norm enforcement can deter defection and thus has profound implications, potentially explaining the human ability to sustain large-scale cooperation (5). However, the idea that punishment enhances efficiency in exchanges is hotly contended (18; 19; 1; 10; 20). Experiments reveal that in some societies, people will also spend resources to harm cooperators, in what could be interpreted as plain hostility or attacking (1). This anti-social punishment does not enhance efficiency and wastes resources of both the punisher and the punished. A fundamental question is thus: Why would human punishment behavior have developed to play both pro-social and destructive roles, and what determines whether the beneficial or harmful form manifests?

Theory suggests that group boundaries are key to understanding punishment motives as well as altruistic cooperation. For exemple, current evolutionary models based on the idea of (cultural) group selection show that a pattern of altruistic cooperation and punishment of defectors can emerge within groups. These altruistic behaviors can survive because they enhance group fitness, and make groups composed of altruists more likely to survive environmental shocks (21; 22; 5; 20). Crucially, altruism must be parochial, or preferentially directed towards own group members, otherwise altruistic groups lose their relative tness advantage. Anti-social punishment can emerge, however, with the introduction of competition for resources between groups. In this case the seemingly benign trait of altruism can play a surprising role, because enhancing own-group tness is not the only way to win: damaging competitor groups is also a viable strategy. In addition to being even more cooperative within their group, altruists might become hostile towards other groups, and use anti-social punishment as a way to damage outsiders. This hostility could survive because it reinforces the relative tness advantage of groups with altruists (2; 23). Thus, hostility towards other groups in the form of anti-social punishment, and even stronger cooperation within groups, can emerge if there is competition for resources between groups.

In this study, we conduct experiments to test whether the interaction of group boundaries and economic environment can explain altruistic cooperation and the sometimes pro-social and sometimes anti-social manifestations of punishment behavior. Previous studies with real groups are vulnerable to two fundamental confounds. For example, some studies have found hostility between real-life groups even when there is no competition between them (24; 25). But it is difficult to interpret these results, as there are other stark differences between these individuals besides just group affiliation. Studying the effect of competition presents a second, similar challenge (26; 27; 28; 29). Competition may cause group composition to change, attracting particularly aggressive individuals, making it difficult to examine whether competition changes a given individual's preferences. A solution to these confounds is to randomly assign both groups and competition in the lab. In general such studies find in-group favoritism, but do not find significant hostility between groups, and find no specific evidence of what could be interpreted as altruistic punishment (30; 31; 32; 33; 34; 35). Yet, it is unclear whether the lack of hostility generalizes to real social groups, because the minimal-group paradigm used in this literature creates such a weak group manipulation, as pointed out elsewhere (33) (see also 36).

Experimental Design. Our approach avoids these confounds. We exploit random assignment to army platoons as a strong and meaningful manipulation of group membership, and randomize the economic environment in which the interaction takes place. Our subjects are officer candidates in the Joint Officer Training Program of the Swiss Army. They are randomly assigned to platoons and spend four weeks of intense work within these units. Despite the fact that platoons are assigned orthogonally to previous social ties, the members choose to spend most of what little off-duty time they have with members of their newly assigned group, indicating that the platoons quickly form strong within-group social ties (see SI). In the officer training program there is no institutionalized conflict between the groups, thus allowing us to study the effects of group membership on preferences in neutral economic environments.

In week three of the four-week training, we conducted two anonymous, one-shot experiments with the officer candidates. Experiment 1 was a simultaneous prisoners dilemma (PD) game. The players, labeled A1 and A2, were each endowed with 20 points and simultaneously decided whether to keep the points or pass all of them to the other player (4 points equaled CHF 1). Passed points were doubled. The subjects interacted anonymously in the PD game but knew the group membership of the other player which was either a member from their own platoon or a member from another platoon.[1] If both players passed on the points (cooperated), this caused each to get 40. But each player would have done better individually by keeping the points (defecting): he would gain 20 and get 60 in this case, but cause a loss of 40 to the other. Thus, a rational selfish players will keep the points, even though this is inefficient. Thus, passing on the points requires a non-selfish motivation. The game therefore allows us to identify the effect of group boundaries, and environment, on these non-selfish motives.

Experiment 2 was a simultaneous prisoners dilemma game with third-party punishment (17). Two players, labeled A1 and A2, play another PD game as described above. Now a third party - two additional players denoted B1 and B2 - observe the actions of the A-players

[1] After they made their decision in the PD, we also elicited members' beliefs about in- and out-group members' cooperation rates.

in the PD and can then punish either A1 or A2, respectively. The B-players can condition their punishment decision on both A-players actions using the strategy method. They are free to punish defection (norm enforcement) or cooperation (anti-social punishment) of their respective A-player. Each punishment points spent cost 1 point for the B-player and destroyed 3 points for the A-player. Punishment, whether pro- or anti-social, is therefore costly for the third party and a selfish third party will never punish. As punishment requires non-selfish motivation this design allows us to examine how these motivations changes as a function of the group composition and the economic environment. For the remainder of the paper, we refer to the group composition in Experiment 2 from B1s perspective. Thus, A1 always refers to the player that the B-player can punish, while we refer to the other A-player as A2.

We implemented two treatments to investigate the effect of the economic environment. In the *Neutral Group Environment (NG)* (N=228) there was no economic competition among groups. This allows us to test for motives of altruistic cooperation with the in-group, and of parochial norm enforcement, i.e. stronger punishment of defection if the victim is a member of one's own group (5). Our second treatment is the *Competitive Group Environment (CG)* (N=297). We randomized group composition in the same way as in NG, but added a bonus, to be won by the platoon with the highest average payoffs from the A-players' choices. This was implemented both in experiment 1 and experiment 2. Importantly, the bonus was chosen such that it didn't change the incentives of a strictly selfish individual. The bonus was sufficiently small (20 points) that a selfish A-player would still keep the points (see SI). Furthermore, since the recipients of the bonus were determined using only the A-players' payoffs before deducting the punishment points, it left the B-players' material incentives completely unchanged.[2] This treatment adds two important features: First, it increases the group-specific public good, thus prompting theories of altruistic cooperation and punishment to predict stronger cooperation and

[2]In all experiments, before making decisions, subjects answered control questions to verify that they had a complete understanding of the rules in the experiment. See SI for complete instructions.

norm enforcement on behalf of one's groups. Second, it also creates a psychological feeling of competition between groups. If this treatment triggers hostility in punishment towards other groups, it cannot be for strategic reasons: the B-players' choices have no effect on winning the bonus. Thus, anti-social punishment of outsiders would provide evidence that competition generates an impulse to harm other groups, in line with models in which a taste for hostility emerges (2).

Results. Panel A in Figure 1 shows the fraction of individuals cooperating as a function of the group composition and the treatment. In the NG treatment, there is a significant and large increase in cooperation if individuals are paired with someone from their own platoon rather than another platoon: cooperation rates are 18 percentage points higher for within-group interactions than between-group ($p = 0.03$, see Table S2). This is consistent with the prediction of parochialism, because it shows that altruistic cooperation is preferentially directed towards own group members. In the CG treatment, favoritism of the in-group is even more extreme: Cooperation rates are 36 percentage points higher in within-group than between-group interactions ($p < 0.01$, see Table S2). The interaction of the in-group effect with the economic environment shows that the increase in cooperation rates among in-group members is stronger in the competitive environment ($p = 0.02$, see column (3) in Table S2). The results thus support the prediction that in-group favoritism in altruistic cooperation should increase in the presence of competition.[3]

Figure 2 displays the results for punishment. The figure allows us to highlight two distinct motives related to the group membership. By varying the identity of A1, the person who can be punished, we can see if punishment depends on whether A1 was a member of the punisher's own group (blue lines) or another group (grey lines). By varying the identity of A2, the player who

[3]The results on cooperation behavior are also fully reflected in the individuals' beliefs; people report that they expect in-group favoritism in NG, and significantly greater favoritism in CG (see Panel B in Figure 1 and Table S2).

is the potential victim of defection, we can examine if punishment of A1 depends on whether the victim of defection was from the punisher's group (solid lines) or some other group (dashed lines). The figure also distinguishes between whether A1 cooperated or defected.

Panel A displays the results for the NG treatment. There is a clear pattern of norm enforcement in the data: A1 is punished more strongly for defection than cooperation ($p < 0.01$, see Table S5). Punishment of A1 also depends on the identity of A2. If A1 defects, the solid lines (A2 from the punisher's group) are always above the dashed lines (A2 from another group). The effect is significant ($p = 0.05$, see Table S3). It is evident from the figure that the identity of A2 does not matter if A1 cooperates. These results are consistent with the prediction that punishers engage in altruistic punishment in a way that enforces a norm of cooperation toward members of their own group. Notice also that we do not see any evidence of hostility in this treatment. Hostility would imply stronger punishment of an A1 that belongs to another group, regardless of what A1 does. As can be seen in the graph, there is essentially no difference as a function of A1 group affiliation, and regressions reported in the SI confirm that there is no signficant effect.

Turning to the competition treatment, we see that the punishment choices in panel B are starkly different. Most importantly, there is now a clear difference in punishment depending on whether A1 belongs to the punisher's own group or not. Grey lines (A1 is from another group) are clearly above the blue lines (A1 is from the punisher's group). Thus, out-group individuals are punished significantly harder than in-group members, and importantly, this is true no matter whether the individual cooperates or defects (grey lines are above blue lines in both cases). These punishment differences are statistically significant in CG ($p < 0.01$ for cooperation; $p < 0.01$ for defection; see Table S3). Thus, the introduction of competition leads to hostility in punishment behavior. Furthermore, there is no relationship between the identity of A2 and punishment in CG, so the tendency to preferentially punish defection against the in-group is no longer present (see Table S3).

The two different economic environments (neutral and competitive) generated qualitatively different patterns of punishment, as is evident in the figure. A formal statistical test confirms this impression: We clearly reject the hypothesis that the effect of A1's and A2's group affiliation on punishment are the same across the two treatments ($p < 0.01$, see Table S3). In addition, the two panels show that punishment seems to depend less on A1's behavior in CG than in NG: While, in Panel A, punishment is clearly higher when A1 defects, regardless of the group composition, that relationship is almost completely muted in Panel B. A formal test shows that indeed, the difference is signficant ($p < 0.01$, see Table S5) and conditioning of punishment on actions of A1 is weaker in CG. Notably, this is even true when A2 is from the punisher's own group ($p = 0.03$, see Table S5), the case when norm enforcement was strongest in NG.

Discussion: Our results demonstrate that the interaction of group boundaries, and the economic environment surrounding the groups, can generate two starkly different patterns of cooperation and punishment behaviors. In the absence of competition between groups, individuals cooperate more within their group, and use punishment to enforce cooperation norms towards their group. This finding is consonant with evolutionary theories of group selection that show how within-group cooperation and altruistic norm enforcement can emerge jointly (22; 5). The punishment patterns change dramatically when we introduce competition between groups. We find strong out-group hostility, in the form of anti-social punishment of the out-group. Furthermore, the norm enforcement pattern of retaliating for defection against one's own group has completely disappeared in the competitive environment. Thus, punishment here takes on the qualitative pattern of anti-social punishment directed towards the out-group, consistent with the between group hostility and aggression predicted to arise in competitive settings by Choi and Bowles (2; 23). This interpretation is further supported by the increase in in-group cooperation that we find, consistent with stronger altruism within groups in competitive environments.[4] Our finding that anti-social punishment depends on competition between groups fits well with previous evidence that anti-social punishment is especially pronounced in societies with more ”close-knit” social networks, which presumably perceive everyone outside their network as a competitor (1). Importantly, in contrast to earlier studies documenting the pattern of altruistic punishment (13) or anti-social punishment (1; 10), we are able to trigger each by an experimental treatment. Our findings also provide insights into the origins of conflict along group lines, and provide a better understanding of the mechanism leading to hostility in the presence of competition (30; 37; 27; 29).

For illustrative purposes we provided an interpretation based on group selection, but it is important to note that other evolutionary mechanisms could be at work, too. Our setup and results rule out some mechanisms: Models of kin selection predict cooperation only within related individuals, but our groups are not formed according to relatedness (7). Our setup also rules out explanations based on direct reciprocity or networks (8; 9), because interactions were one-shot and anonymous. The results can be interpreted, however, in light of recent findings in evolutionary psychology. These findings suggest that group membership may create cues of reputation concerns even in completely anonymous interactions (38; 39). Therefore, individuals may treat in-group members more favorably, and engage in hostility on behalf of their group, because of a subconscious cue of repeated interactions.

Finally, our results also provide additional evidence that social preferences are endogenous to the economic environment (40; 41). This literature argues that changes in economic environments brings about changes in preferences. But these changes are typically assumed to be slow, e.g., operating through slow-changing norms of cooperation (1). It is noteworthy that we ob-

[4]Surprisingly, we do not find evidence of hostility in cooperation, i.e., a decrease in cooperation with the outgroup. However, the same puzzle has been observed in another context (26). The finding also parallels the results in Herrmann et al., where they find no noticeable differences in cooperation in the absence of punishment between societies that display anti-social or pro-social punishment (1).

serve a particularly strong form of endogenous preferences: Our treatments are between-subject manipulations, and yet we immediately observe starkly different punishment and cooperation strategies. Thus, this evidence also suggests that different motives of social preferences may be dormant in humans, and triggered by different economic environments.

References and Notes

[1] B. Herrmann, C. Thoeni, S. Gachter, *Science* **319**, 1362 (2008).

[2] J. Choi, S. Bowles, *Science* **318**, 636 (2007).

[3] S. Gaechter, E. Renner, M. Sefton, *Science* **322**, 1510 (2008).

[4] O. Gurerk, B. Irlenbusch, B. Rockenbach, The competitive advantage of sanctioning institutions (2006).

[5] R. Boyd, H. Gintis, S. Bowles, P. Richerson, *Proceedings of the National Academy of Sciences* **100**, 3531 (2003).

[6] H. Gintis, *Journal of Theoretical Biology* **206**, 169 (2000).

[7] W. Hamilton, *Journal of Theoretical Biology* **7**, 1 (1964).

[8] R. Trivers, *The Quarterly Review of Biology* **46**, 35 (1971).

[9] M. Nowak, K. Sigmund, *Nature* **437**, 1291 (2005).

[10] K. Abbink, J. Brandts, B. Herrmann, H. Orzen, *American Economic Review* (2009).

[11] E. Fehr, K. Schmidt, *Quarterly Journal of Economics* **114**, 817 (1999).

[12] L. Goette, D. Huffman, *Review of General Psychology* **11**, 142 (2007).

[13] E. Fehr, S. Gaechter, *Nature* **415**, 137 (2002).

[14] E. Ostrom, J. Walker, R. Gardner, *The American Political Science Review* **86**, 404 (1992).

[15] D. De Quervain, *et al.*, *Science* **305**, 1254 (2004).

[16] J. Buckholtz, *et al.*, *Neuron* **60**, 930 (2008).

[17] E. Fehr, U. Fischbacher, *Evolution and Human Behavior* **25**, 63 (2004).

[18] A. Dreber, D. Rand, D. Fudenberg, M. Nowak, *Nature* **452**, 348 (2008).

[19] M. Egas, A. Riedl, *Proceedings of the Royal Society* **275**, 871 (2008).

[20] M. Nowak, *Science* **314**, 1560 (2006).

[21] G. Price, *Nature* **227**, 520 (1970).

[22] J. Henrich, *Journal of Economic Behavior and Organization* **53**, 3 (2004).

[23] S. Bowles, *Science* **324**, 1293 (2009).

[24] H. Bernhard, U. Fischbacher, E. Fehr, *Nature* **442**, 912 (2006).

[25] P. Kollock, *Modelling Rationality, Morality and Evolution*, P. A. Danielson, ed. (Oxford University Press, 1998), pp. 186 – 210.

[26] D. Rand, *et al.*, *Proceedings of the National Academy of Sciences* **106**, 6187 (2009).

[27] V. Esses, L. Jackson, T. Armstrong, *Journal of Social Issues* **54**, 699 (1998).

[28] K. Auletta, *The New York Times Magazine* **24**, 36 (1985).

[29] M. Sherif, O. Harvey, B. J. White, W. Hood, C. Sherif, *Intergroup conflict and cooperation: The Robbers Cave experiment* (University Book Exchange Norman, 1961).

[30] M. Hewstone, M. Rubin, H. Willis, *Annual Review of Psychology* **53**, 575 (2002).

[31] M. Brewer, *Journal of Social Issues* **55**, 429 (1999).

[32] Y. Chen, S. Li, *American Economic Review* **99**, 431 (2009).

[33] N. Halevy, G. Bornstein, L. Sagiv, *Psychological Science* **19**, 405 (2008).

[34] C. Efferson, R. Lalive, E. Fehr, *Science* **321**, 1844 (2008).

[35] T. Yamagishi, T. Kiyonari, *Social Psychology Quarterly* **63**, 116 (2000).

[36] S. Levitt, J. List, *Science* **319**, 909 (2008).

[37] D. H. Posner, *American Political Science Review* **98**, 529 (2004).

[38] K. Haley, D. Fessler, *Evolution and Human Behavior* **26**, 245 (2005).

[39] M. Rigdon, K. Ishii, M. Watabe, S. Kitayama, *Journal of Economic Psychology* (2009).

[40] S. Bowles, *Journal of Economic Literature* **36**, 75 (1998).

[41] S. V. Burks, J. P. Carpenter, L. Goette, *Journal of Economic Behavior and Organization* (2009).

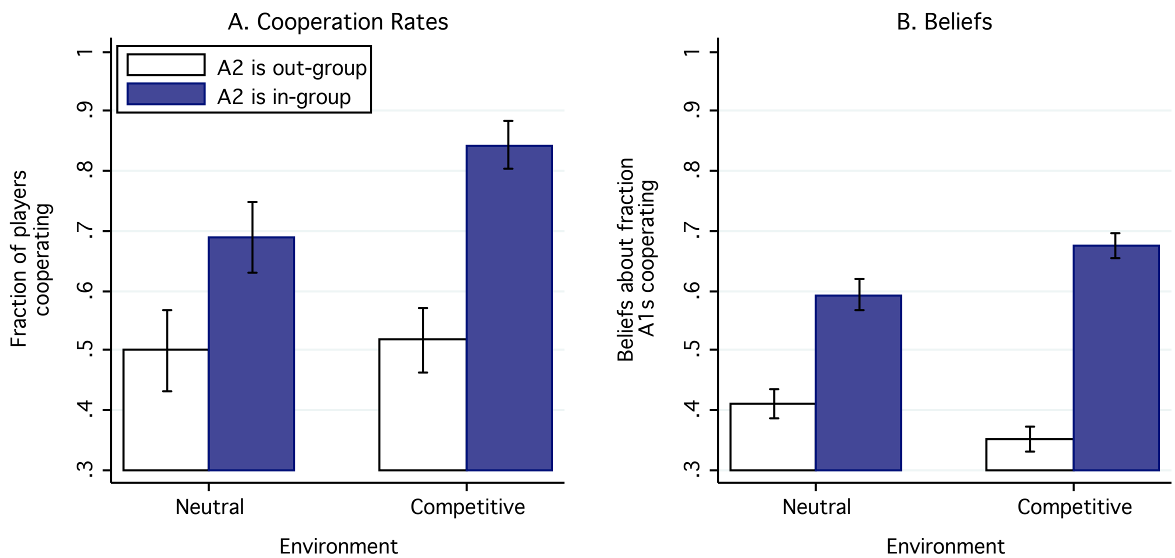


Fig. 1. Cooperation rates and beliefs about cooperation in NG and CT. The bars show standard errors of the mean. White indicates an out-group pairing while blue indicates an in-group pairing. Panel A shows the fraction of A-players passing their endowment to the other player in a simultaneous one-shot PD. Panel B shows beliefs about the fraction of A-players who pass their endowment in the PD.

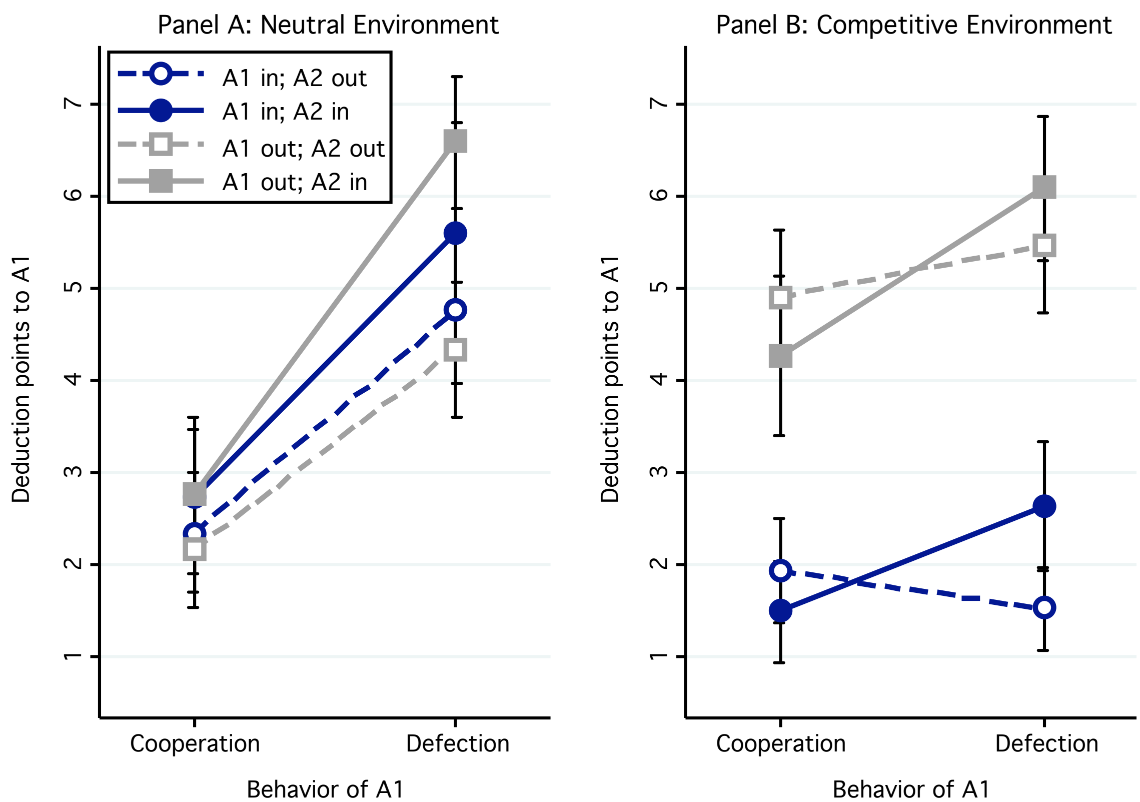


Fig. 2. Deduction points of B-players imposed on A1-players in the case A2-player cooperated. B-players could deduct between 0 and 10 points. Each deduction point costs B-players 1 point and A1-players 3 points. Deduction points were made conditional on whether A1-players cooperated or defected using the strategy method. Error bars show standard errors of the mean. Blue indicate that A1 in the same group as B1, while grey indicate that A1 is another group. Solid lines indicate that A2 is in the same group as B1, while dashed lines indicate that A2 is in another group.

Supporting Online Material for

The role of group boundaries in human altruism and anti-social behavior

Stephan Meier,* David Huffman,* Matthias Sutter, Lorenz Goette†

*These authors contributed equally to this work.

†To whom correspondence should be addressed; E-mail: lorenz.goette@unige.ch.

Contents

1 Materials and Methods

1.1 Random Group Assignment and Off-duty Interactions (Table S1)

All Swiss males are required to perform at least 300 days of military service, beginning with twenty-one weeks of basic training. In week seven, about one fourth are selected to go through ten weeks of officer-candidate training. Of these, one fourth are promoted to officers and continue on to the Joint Officer Training Program (JOTP). Whereas officer-candidate training is specific for each branch of service, and occurs in separate locations, JOTP brings new officers from all branches of service together, to the same location, for four weeks. Officers are randomly assigned to a platoon at the beginning of JOTP, and spend virtually all time during the day with their platoon. Training involves mainly coursework, on principles of security, combat in large military units, logistics, and leadership. At the end of JOTP, the platoons are dissolved and officers are once again sent to separate locations, for further, advanced training specific to each branch of service.

We used the random assignment of candidates to platoons as our manipulation of social groups. Each platoon is identified by a different number. Assignment to platoons is random, and stratified according to the different branches of service. The army assigns platoons orthogonally to any previous social ties among officers with the aim of promoting exchanges of perspectives among different individuals and branches of service (see also 1).

The assignment mechanism is ideal, in several ways, for investigating the impact of group membership on behavior. First, trainees know that platoon composition is designed to be identical and that nobody could choose which platoon to join. Indeed, statistical tests reveal no significant differences in platoon composition, by branch of service, education, or age. Second, there is no competition between the groups (or trainees) for evaluations or other resources. Relative performance evaluations were completed previously, in candidate training. Thus, there is no function of the group assignment, other than to effect the circle of individuals with whom an officer interacts most frequently. Third, social interactions within a platoon are intense. Platoon members spend the whole workday with their group, for the three weeks leading up to our experiments. Despite the fact that platoons are assigned orthogonally to previous social ties, social interactions and ties also arise endogenously within platoons in after-work time: In a questionnaire, officers in our study report to a question on *"How often do you spend off-duty time with members of a) your own platoon or b) the other platoons?"* that they spend significantly more time off-duty with members of their own platoon. This is remarkable in itself, given that 79.8 percent of the trainees know people in other platoons, mostly from earlier stages of their training. Yet, as illustrated in Table S1, they choose to spend most of what little off-duty time they have with members of their newly assigned group. Thus, platoon assignment provides a strong group manipulation.

Table S1: Off-Duty Time Spent per Week

	Own Platoon	Other Platoon
Less than once	4.5%	30.4%
Between 1 and 2 times	45.0%	44.0%
Twice or more	50.5%	25.7%
N	489	491

Note: The two distributions are significantly different (Wilcoxon Signed-Rank test, $p < 0.001$).

1.2 Experiments and Group Conditions (Figure S1)

In the third week of the four-week training period, we conducted two experiments with the officer candidates, to see how random group assignment, and random introduction of competition between groups, affect behavior.

Experiment 1: Cooperation. The game was a simultaneous prisoners' dilemma (PD). The players, labeled A1 and A2, were each endowed with 20 points worth real money (4 points = 1 CHF). They simultaneously decided whether to keep the points or pass all of them to the other player. Passed points were doubled. Thus, if both players passed their points (cooperation), they each got 40 points. However, a selfish player could always do better by keeping the points (defecting), regardless of the other player's decision: Defecting when the other defected would yield 20, whereas cooperating would sacrifice the endowment and yield nothing in return; defecting when the other cooperated would yield 60, the maximum possible individual payoff in the game (while leaving the cooperator with 0). Cooperating thus entails lowering ones own payoff, and improving the payoff of the other player, and is an indicator of non-selfish motives. We use the game as our workhorse for studying how group boundaries, and economic environment, affect non-selfish motives for cooperation.

Experiment 1 involved two conditions in a between-subject design. In all conditions, a subject never learned the individual identity of their partner. In the in-group condition, subjects interacted anonymously, except for being informed that the other player was a member of their platoon. The out-group condition was the same, except subjects were informed that the other player was a member of another platoon. Group affiliation was clearly marked on the decision sheets. These conditions allow us to examine how group assignment affects cooperation.

We also elicited individual's beliefs about in-group and out-group cooperation. Independent of the condition they were in, we asked participants to state both their belief for in- and out-group cooperation. We asked them to predict the percentage of the in- and out-group that would send all of the points (cooperate). They were given an incentive to make their best guess: the knew that their prediction would be compared to the percentage actually observed. If the deviation was less than 10 percent, then they would get one extra point.

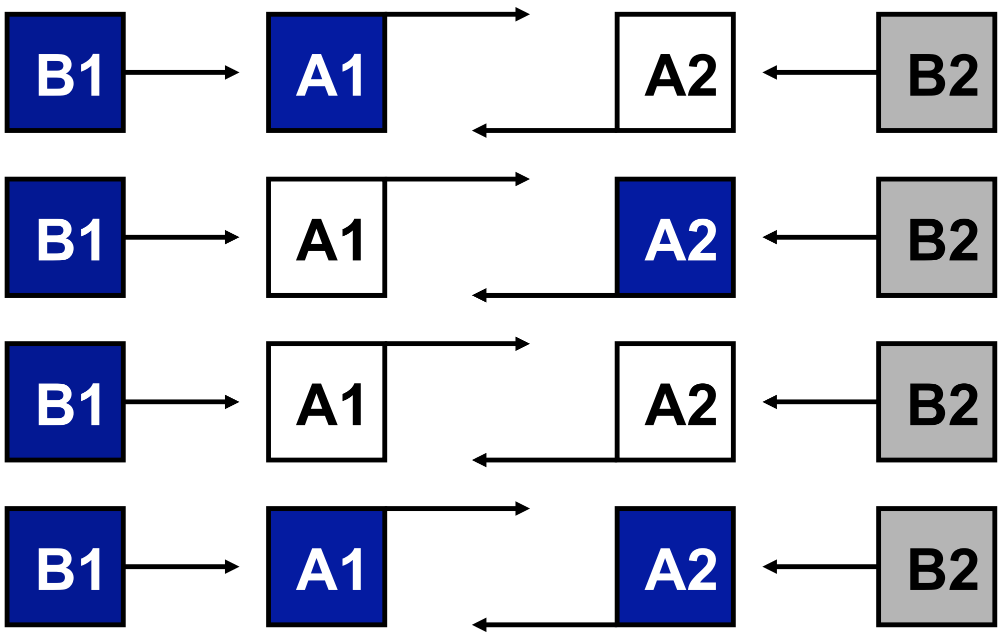
At the very end of the experimental sessions, we conducted a short survey in which we asked participants whether they agreed or disagreed with three statements about trust: 1) *"In general, people can be trusted."*, 2) *"Nowadays, you can't rely on anybody."*, and 3) *"Dealing with strangers, it is better to be cautious before trusting them."*. Participants answered on a 4-point scale (1 "Agree Strongly", 2 "Agree Slightly", 3 "Disagree Slightly", and 4 "Disagree Strongly"). We created an individual variable, $Trust$, by adding the answers to the three question and assigning a 1 for the least amount of trust and 4 the highest amount of trust per question (answers to question 1 are reversed coded). This is used to help capture individual differences in beliefs about trustworthiness, in our statistical analysis.

Experiment 2: Punishment. In Experiment 2, two players A1 and A2 played a PD as in Experiment 1, but we added two additional players, B1 and B2. Each B-player was endowed with 70 points. B1 could assign up to 10 deduction points to A1, and B2 to A2. Each deduction point subtracted three points from the A-player, and cost the B-player one point of his or her endowment. The B-players could condition their choices on the actions of A1 and A2. Thus Experiment 2 incorporated the possibility of third-party punishment (2), and is suited for examining determinants of whether punishment takes the form of norm enforcement (selectively punishing defection) or anti-social punishment (punishing both cooperation and defection). To examine the impact of group membership on norm enforcement, we varied the composition of players in each game in a between-subject design. For the remainder of the paper, we refer to the group composition in Experiment 2 from B1's perspective. Thus, A1 always refers to the player that the B-player can punish, while we refer to the other A-player as A2. The four different group compositions we implemented are shown in Figure S1.

To examine the impact of group membership on norm enforcement, we varied the composition of players in each game in a between-subject design. For the remainder of the paper, we refer to the group composition in Experiment 2 from B1's perspective. Thus, A1 always refers to the player that the B-player can punish, while we refer to the other A-player as A2. The four different group compositions we implemented are shown in Figure S1.

Varying the group membership of A1 allows us to investigate how the group identity of the person being punished matters. We also study how punishment varies with the group affiliation of A2, the person affected by A1s actions. Section 1.6 shows a translation of the instructions for one group composition in the *Neutral Group Environment* treatment.

Figure S1: Third-Party Punishment Game



Note: The game allowed B1 to punish A1, and B2 to punish A2, conditional on the actions and A1 and A2 in a simultaneous prisoners' dilemma game. The dark shading indicates the four possible group combinations for B1, A1, and A2, which were implemented as different treatments (players with the same shading are from the same group). The design deliberately did not vary all possible combinations of B1 and B2 group roles, because of number of observations, so the effect of B2 group identity on B1 behavior is not studied. The pattern of B-player (and A-player) group compositions was identical across the Economic Environments, the *NG* and *CG* treatments.

1.3 Economic Environment Treatments

We used two treatments to analyze the effect of the economic environment on cooperation and punishment behavior, within and between groups.

Neutral Group Environment (NG): In this treatment, we used the randomly assigned groups as our group manipulation and varied the group composition as described above. There was no economic conflict between the groups.

Competitive Group Environment (CG): We added competition to the 'Neutral Group Environment' treatment by offering a bonus to the group that got the highest payoff in the PD stage. The bonus was 20 points for each group member if the group got the highest average payoff in the PD. In case of a tie between two groups, the winning team was randomly determined. Because the bonus was based on average payoffs for pairs playing the PD, and cooperation maximized payoffs for the pair, cooperation facilitated winning the bonus for the group. Importantly, however, the bonus did not change the incentives for a selfish individual: the best strategy for a selfish A-player was still to defect (see below for intuition). Furthermore, in Experiment 2 the bonus was calculated based on the A-player average payoffs, *before* deducting any punishment points imposed by the B-players. B-players (and A-players) knew this. Thus the bonus was irrelevant for the choices of the B-players, regardless of whether they were selfish or altruistic. The rules of the game were made clear in the instructions, and we only began the experiment after control questions verified that all participants understood them.

We added competition to the 'Neutral Group Environment' treatment by offering a bonus to the group which got the highest payoff in the PD stage. The bonus was 20 points for each group member if the groups got the highest average payoff in the PD. In case of a tie between two groups the winning team will be randomly determined. The bonus doesn't change the Nash equilibrium in the two experiments. For the A-players, it is still a Nash equilibrium to defect in the PD (see below for a formal proof). The punishment behavior of the B-players has no influence on the bonus. This feature of the bonus was made clear in the instructions and we checked in the control questions that all participants understood it.

1.4 Nash Equilibrium in the PD in *Competitive Group Environment*

In this section, we explain why it was optimal for selfish A-players to defect in the Competitive Group Environment. The intuition is straightforward: Cooperating never leads to an increased payoff, because cooperating costs 20 points, and the bonus is only 20 points. In fact, our rules for tie-breaking in case two groups have the same number of points imply that individuals always lose money when cooperating, because the bonus is only 10 in expected terms. Thus, adding competition cannot generate an increase in cooperation rates through selfish incentives; an increase in cooperation under competition must reflet an affect working through non-selfish motives. Below, we formalize this intuition, based on the notion of Nash equilibrium from game theory. A Nash equilibrium is simply a combination of strategies such that no player has a (selfish) incentive to change what they are doing. Given the lack of incentive to change, it is a stable, or equilibrium state.

We show that the Nash equilibrium in the game involves all (selfish) A-players defecting.

(i) In our experiment, there are within-group and between-group pairings. Obviously, a selfish player never cooperates with a player from another group, since, on top of costing him 20 points, he may also be pivotal in losing the bonus. Therefore, what remains to be considered are within-group pairings. First consider the case of $K = 2$ groups, denoted X and Y. Now pick an arbitrary collection of strategies in which some individuals cooperate in within-group pairings. We ask whether this strategy can be a Nash equilibrium. Two possible cases can arise: Either one of the groups, say group Y, loses, or the two groups tie.

We first show that groups can never tie with some individuals cooperating.

- Pick an arbitrary member of group k who is cooperating. Since the groups are tied, he wins a bonus with probability 0.5. If he defects, his group will lose for sure. However, defecting saves 20 points, while costing only 10 points in expected bonus. Thus, when two groups are tied, cooperating players have an incentive to defect.

We now show that it is impossible to have a Nash equilibrium in which group Y loses for sure.

- If group Y loses, then it cannot be a Nash equilibrium for anyone in group Y to cooperate. Given the others' strategies, members of Y who cooperate can increase their payoff by 20 points if they defect.

- Given this result, it follows that in group X, at most one player will cooperate. If more than one player in group X cooperated, a player could switch to defection while still winning the bonus, holding the other players' strategies constant.

- However, if one player in X cooperates, the tying rule now implies that the player can defect, and save 20 points, but only lose 10 points in expected bonus (since the two groups now tie).

Thus, the only equilibrium for $K = 2$ groups involves both groups tying, and this equilibrium involves all players defecting.

(iii) The above arguments immediately generalize to $K > 2$ groups. The only difference is that the expected bonus in the case of a tie will be even smaller, $20/m$, where $m \leq K$ is the number of groups tying. Thus, the same reasoning applies.

1.5 Experimental Procedures

The experiment was conducted with paper-and-pencil in a large auditorium. Subjects were ordered into the auditorium and did not know of the experiment in advance. The experiment lasted 45 minutes.

Special care was taken to ensure anonymity. Our anonymity conditions are stronger than the typical social preferences experiment, because not only were subjects never told the identity of their partner (s), they knew that payoffs would be mailed to home addresses ten days after the experiment, so that all participants would only learn the outcome of the experiment after JOTP was over and they were no longer with their group members. These conditions ensured that the experiment was truly one-shot, and that defection was the optimal choice for a selfish individual. For example, subjects did not need to fear reprisal after the experiment if they chose to defect.

Points earned were converted into Swiss Francs (one point = 0.25 CHF) and the subjects earned on average CHF 14.4 (approximately $14). There was no show-up fee.

Overall, 525 subjects participated in the experiments: 228 in the 'Neutral Group Environment' treatment and 297 in the 'Competitive Group Environment' treatment. 281 were assigned the role of A-players and participated in Experiment 1. Half were assigned to the in-group treatment, and half to the out-group treatment. In the few cases in which the groups had an uneven number of A-players, we randomly used the action of some A-players twice to calculate payoffs. The payoff of these players was determined by the decisions associated with the first match. After participating in Experiment 1, these same subjects participated as A-players in Experiment 2. This procedure introduces a possible order effect for the A-players, but choices of the A-players in Experiment 2 are not of interest for our purposes. 244 subjects were assigned the role of B-players. They participated only in Experiment 2, and were assigned to one of four treatments. We elicited B-players' deduction points using the strategy method, i.e., they specified how many points to deduct from their associated A-player for each possible combination of actions by A1 and A2.

1.6 Experimental Instructions

1.6.1 Instructions of Experiment 1 for Player A (Translation)

What is this about?

Two subjects participate in this decision situation. They will be called A1 and A2. Both, A1 and A2, will get an endowment of 20 points. Each participants has to decide between two options:

- **Keep**: The participant keeps his 20 points.
- **Transfer**: The participant transfers his 20 points to the other participants. The transfered points will be doubled.

Each participant has to decide whether to Keep or the Transfer without knowing how the other participant decided. So, the following payoffs can result:

		Payoffs in this case:
Case 1:	A1 keeps the points	A1: 20 points
	A2 keeps the points	A2: 20 points
		Payoffs in this case:
Case 2:	A1 transfers 20 points	A1: 0 points
	A2 keeps the points	A2: 60 points
		Payoffs in this case:
Case 3:	A2 keeps the points	A1: 60 points
	A2 transfers 20 points	A2: 0 points
		Payoffs in this case:
Case 4:	A1 transfers 20 points	A1: 40 points
	A2 transfers 20 points	A2: 40 points

How will you decide?

- You will be in the role of A1.
- Your assigned **participant A2 is from another platoon**.

None of the participants will ever find out to whom he was assigned. We guarantee total anonymity. When all the participants reached a decision, we will calculate the points and the resulting monetary payoffs in the following way:

4 points = CHF1

The amount will be delivered to you by mail.

Everything clear?

Before you decide, answer the following questions. The question make sure that all the participants understand the instructions.

If you have questions, please contact the staff.

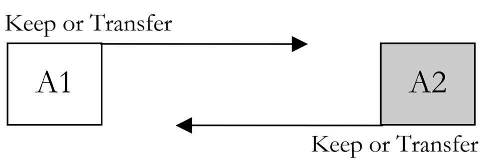
1. A1 and A2 keep their points. Please calculate the resulting points for all participants. State all the steps in getting to the result.
2. A1 and A2 transfer their points. Please calculate the resulting points for all participants. State all the steps in getting to the result.
3. A1 keeps his points and A2 transfer his points. Please calculate the resulting points for all participants. State all the steps in getting to the result.

Please contact the staff when you are done with the questions or if you have questions.

Decision Sheet

- You were assigned the role of **A1**.
- Your assigned **participant A2 is from another platoon**.

In the following figure the participants from the other platoons are shaded.



Please decide which option to pick:

- Keep
- Transfer

Please let the staff know when you decided.

1.6.2 Instructions of Experiment 2 for Player B (Translation)

What is this about?

Four subjects participate in this decision situation. They will be called A1, A2, B1 and B2. The decision situation will have two steps.

Step 1: A1 and A2 will get an endowment of 20 points. Each participants has to decide between two options:

- **Keep**: The participant keeps his 20 points.
- **Transfer**: The participant transfers his 20 points to the other participants. The transfered points will be doubled.

		Payoffs in this case:
Case 1:	A1 keeps the points	A1: 20 points
	A2 keeps the points	A2: 20 points
		Payoffs in this case:
Case 2:	A1 transfers 20 points	A1: 0 points
	A2 keeps the points	A2: 60 points
		Payoffs in this case:
Case 3:	A2 keeps the points	A1: 60 points
	A2 transfers 20 points	A2: 0 points
		Payoffs in this case:
Case 4:	A1 transfers 20 points	A1: 40 points
	A2 transfers 20 points	A2: 40 points

Each participant has to decide whether to Keep or the Transfer without knowing how the other participant decided. So, the following payoffs can result:

Step 2: B1 and B2 will get an endowment of 70 points each and A1 and A2 will get another 10 points each. In Step 2, B1 and B2 can assign deduction points. B1 can assigned deduction points to A1 and B2 can assign deduction points to A2. B1 and B2 can each assign a maximum of 10 deduction points.

Before explaining how B1 and B2 will make their decisions, we will describe how deduction points will change the payoffs. **Each deduction point will reduce the payoff of B by one point and the payoff of A by three points.** For example, if B1 assigns 3 deduction points, this will reduce A1's payoff by 9 points and B1's payoff by 3 points.

B1 and B2 will decide about the assignment of deduction points for each potential case in Step 1. That is, they will decide about assigning deduction points for the following four potential cases in Step 1:

- **Case 1**: A1 and A2 keep their points.
- **Case 2**: A1 transfers his points and A2 keeps his points.
- **Case 3**: A1 keeps his points and A2 transfers his points.
- **Case 4**: A1 and A2 transfer their points.

This will lead to the following payoffs:

How will you decide?

- You will be in the role of **B1**.
- Your assigned **participant A1 is from another platoon**.
- The participant **A2** is **from another platoon**. He got assigned to a participant **B2 from your platoon**.

Payoff of A1 =	Payoff from Step 1 + 10 points from Step 2 - 3*Deduction points from B1
Payoff of A2 =	Payoff from Step 1 + 10 points from Step 2 - 3*Deduction points from B2
Payoff of B1 =	Endowment of 70 points - Deduction points to A1
Payoff of B2 =	Endowment of 70 points - Deduction points to A2

None of the participants will ever find out to whom he was assigned. We guarantee total anonymity. When all the participants reached a decision, we will calculate the points and the resulting monetary payoffs in the following way:

4 points = CHF1

The amount will be delivered to you by mail.

Everything clear?

Before you decide, answer the following questions. The question make sure that all the participants understand the instructions.

If you have questions, please contact the staff.

1. In Step 1, A1 and A2 keep their points. In Step 2, neither B1 nor B2 assign any deduction points. Please calculate the resulting points for all participants. State all the steps in getting to the result.
2. In Step 1, A1 and A2 transfer their points. In Step 2, neither B1 nor B2 assign any deduction points. Please calculate the resulting points for all participants. State all the steps in getting to the result.
3. In Step 1, A1 keeps his points and A2 transfers his points. In Step 2, B1 assigns 2 deduction points and B2 assigns 5 deduction points. Please calculate the resulting points for all participants. State all the steps in getting to the result.
4. In Step 1, A1 transfers his points and A2 transfers his points. In Step 2, B1 assigns 1 deduction points and B2 assigns 4 deduction points. Please calculate the resulting points for all participants. State all the steps in getting to the result.

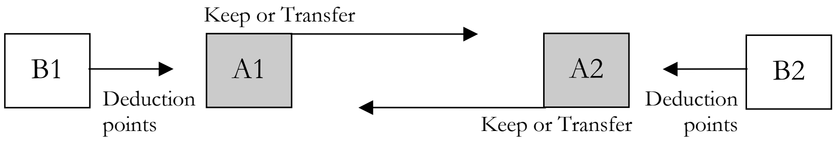
Please contact the staff when you are done with the questions or if you have questions.

Decision Sheet

- You will be in the role of **B1**.
- Your assigned **participant A1 is from another platoon**.

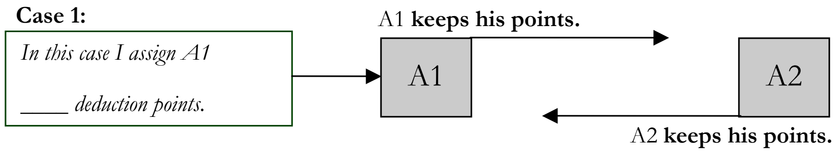
- The participant **A2** is **from another platoon**. He got assigned to a participant **B2 from your platoon**.

In the following figure the participants from the other platoons are shaded.

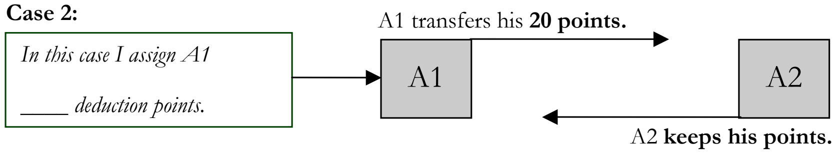


Please decide about the assignment of the deduction points for all possible cases. Only the cases that really happen will determine your payoff and the payoff of the other participants. In each of the cases, you can assign between 0 and 10 deduction points.

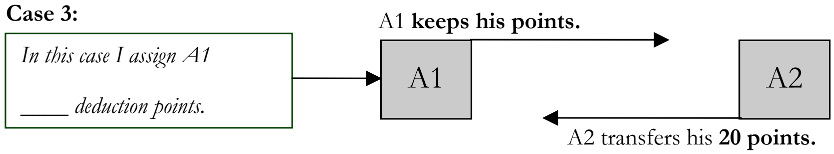
Case 1:



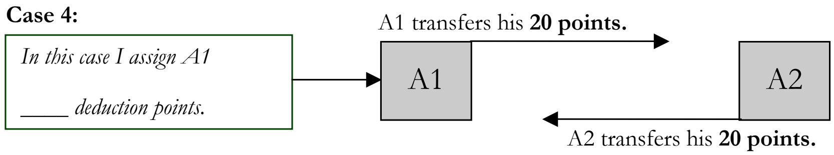
Case 2:



Case 3:



Case 4:



Please let the staff know when you decided.

2 Supporting Empirical Analysis

In the following, we detail the statistical models and empirical results we discuss in the text.

2.1 Table S2: Cooperation

We estimate logit models of the following form

$$coop_i = \alpha + \gamma_0 IG_i + e_i \tag{1}$$

where *coop* is an indicator variable equal to 1 if individual i cooperates, and zero otherwise. IG is an indicator variable equal to 1 if the individual is paired with another individual from his platoon and zero if the other player is from another platoon. The first estimate provides a basic test of whether there is a significant in-group effect. We then add interaction terms for the economic environment:

$$coop_i = \alpha + \gamma_0 IG_i + \gamma_1 IG_i \times CG_i + \delta CG_i + x_i\beta + e_i \tag{2}$$

The indicator variable CG_i is equal to 1 for the 'Competitive Group Environment' treatment, and zero otherwise. In some specifications, we also add control variables x to increase the precision of the estimates. The results are displayed in Table S2. In the text, we refer to the following results

- There is a significant overall in-group effect in column (1).
- In columns (2) and (3), we separate the effect of group membership in the two treatments. In the NG treatment, cooperation is about 20 percentage points higher if the interaction is in-group ($p = 0.03$ in column (2), and $p = 0.05$ in column (3)).
- We also find that the strength of the in-group effect depends on the economic environment. The interaction term between IG and CG shows that the cooperation differential in in-group interactions is about 20 percentage points larger when there is competition ($p = 0.07$

in column (2), and $p = 0.021$ in column (3), where we include an index of trust questions obtained from the questionnaire, explained in Section 1.1).

Table S2: Results for Cooperation Rates and Beliefs about Cooperation

Dependent Variable:	**Cooperation (=1)**			**Beliefs: % cooperating**		
	(1)	(2)	(3)	(4)	(5)	(6)
Ingroup (=1)	0.28***	0.18**	0.17*	0.26***	0.18***	0.18***
	(0.05)	(0.08)	(0.09)	(0.02)	(0.02)	(0.02)
CG (=1)	0.08	0.00	0.01	0.01	-0.06*	-0.06*
	(0.06)	(0.08)	(0.08)	(0.03)	(0.03)	(0.03)
Ingroup×CG		0.18*	0.22**		0.13***	0.13***
		(0.10)	(0.10)		(0.03)	(0.04)
Trust			0.06***			0.03***
			(0.02)			(0.01)
Constant				0.37***	0.41***	0.34***
				(0.02)	(0.02)	(0.04)
(Pseudo)-R^2	0.07	0.08	0.10	0.20	0.21	0.25
N	281	281	267	538	538	515

Notes: In column (1) to (3), marginal effects from logit models. In columns (3) and (4) coefficients from OLS models. The model in columns (4) to (6) uses two observations per individual, therefore standard errors of the estimates in column (4) to (6) are adjusted for clustering on individuals.
Level of significance: $^*: 0.05 \leq p < 0.1$, $^{**}: 0.01 \leq p < 0.05$, $^{***}: p < 0.01$.

2.2 Table S2: Beliefs about Cooperation

Beliefs about cooperation rates as a function of the group composition are elicited within individuals. Similar to above, we estimate the following regressions for beliefs

$$bel_{ik} = \alpha + \gamma_0 IG_k + e_{ik} \tag{3}$$

where bel_{ik} is individual i's belief about the percent of individuals cooperating in the two group configurations: $IG = 0$ if the interaction is between groups ($k = 0$), and $IG = 1$ if the interaction is in-group ($k = 1$). As with cooperation rates, we then add interaction terms for the economic environment:

$$bel_{ik} = \alpha + \gamma_0 IG_k + \gamma_1 IG_k \times CG_i + \delta CG_i + x_i\beta + e_{ik} \tag{4}$$

Because we use two observations per individual, we adjust the errors by clustering on individuals for possible correlations in e_{ik} within individuals. The results are displayed in columns (4) to (6) of Table S2. In the text, we refer to the following results:

- Column (4) shows that there is a strong overall in-group effect in beliefs, of almost the identical magnitude as observed in behavior ($p < 0.01$).
- We then separate the in-group effect in the two environments. Beliefs about cooperation are significantly higher for in-group pairings in NG, about the same magnitude as we find for behavior.
- There is a significant interaction with the economic environment: The in-group differential is 13 percentage points larger in CG than in NG ($p < 0.01$ in both specifications).
- All in-group differentials in beliefs are within a standard deviation of the in-group differentials in cooperation, showing that the individuals had well-calibrated beliefs.

2.3 Tables S3 to S5: Punishment

2.3.1 Table S3: Punishment as a Function of Group Membership (A2 cooperates)

We estimate the following OLS regressions[1]

$$PP_{ik} = \alpha + \gamma_1 I_i(\text{A1 out-group}) + \gamma_2 I_i(\text{A2 in-group}) + e_i \tag{5}$$

where PP are the punishment points that individual i assigns in case k. We include two indicator variables to capture the effect of the group composition on i's punishment; I(A2 in-group) is equal to 1 if player A2 is from the same group as B1 and 0 otherwise and I(A1 out-group) is equal to 1 if player A1 is from another group as B1 and 0 otherwise.

We estimate equation (5) separately for the two cases where A1 cooperates and the two cases where A1 defects, and estimate these again separately for *NG* and *CG*.

Table S3 displays the result for the case in which A2 cooperates. In the text, we refer to the following results from the *NG* treatment:

- In the neutral environment, we find stronger punishment of defection against a member of one's own group, i.e. A2 is an in-group member ($p = 0.05$, column(1)), but no effect of the identity of A1 on punishment ($p = 0.85$, column (1)).
- In the *NG* environment, we find no effects of the group composition on punishment of cooperation (column (3)).

In the text, we refer to the following results from the *CG* treatment:

- In *CG*, we observe a different pattern in punishment. This can be seen in column (2) of table S3. The identity of A2 is no longer significant ($p = 0.201$).
- A1 gets punished more heavily, whether he cooperates or defects, if he is from a different group than the punisher, i.e. A1 is out-group ($p < 0.01$ in columns (2) and (4)).

[1] The results are maintained in tobit regressions.

Table S3: Punishment as a Function of Group Membership

	(1)	(2)	(3)	(4)
Behavior of A1:	A1 defects		A1 cooperates	
Environment:	Neutral	Comp	Neutral	Comp
A1 out-group (γ_1)	0.155	3.742***	–0.099	2.898***
	(0.853)	(0.658)	(0.693)	(0.684)
A2 in-group (γ_2)	1.694**	0.868	0.535	–0.544
	(0.840)	(0.676)	(0.697)	(0.693)
Constant	4.487***	1.636***	2.307***	1.988***
	(0.705)	(0.447)	(0.578)	(0.517)
R^2	0.039	0.203	0.005	0.125
N	111	132	111	132
Tests across equations (environments):				
Test that γ_1 differs	$p < 0.01$		$p < 0.01$	
Test that γ_2 differs	$p = 0.44$		$p = 0.27$	
Test that γ_1 and γ_2 differ	$p < 0.01$		$p < 0.01$	

Note: Dependent variable: # of deduction points. OLS estimates for the cases in which A2 cooperates. Robust standard errors in parentheses. p-values in cross-equation tests are all two-sided.
Level of significance: $^*: 0.05 \leq p < 0.1$, $^{**}: 0.01 \leq p < 0.05$, $^{***}: p < 0.01$.

2.3.2 Table S4: Punishment as a Function of Group Membership

As a robustness test in Table S4 we add the cases in which A2 defects and estimate the following regression

$$PP_{ik} = \alpha + \gamma_1 I_i(\text{A1 out-group}) + \gamma_2 I_i(\text{A2 in-group}) + \delta I_k(\text{A2 defects}) + e_i \tag{6}$$

which adds an indicator variable to equation 5 controlling for A2's choice. The variable equals 1 if A2 defects and 0 otherwise. Since we now use two observations from each individual, we adjust the standard errors for clustering on individuals. The results are qualitatively the same, but slightly weaker.

Table S4: Punishment as a Function of Group Membership

	(1)	(2)	(3)	(4)
Behavior of A1:	A1 defects		A1 cooperates	
Environment:	Neutral	Comp	Neutral	Comp
A1 outgroup (γ_1)	–0.103	3.249***	0.018	1.336***
	(0.661)	(0.549)	(0.524)	(0.470)
A2 ingroup (γ_2)	1.445**	0.308	0.297	–0.485
	(0.645)	(0.558)	(0.534)	(0.469)
A2 defects (=1)	–2.774***	–1.477***	–1.084***	–1.705***
	(0.387)	(0.269)	(0.376)	(0.401)
Constant	4.746***	2.127***	2.345***	2.768***
	(0.622)	(0.420)	(0.460)	(0.438)
R^2	0.141	0.205	0.028	0.091
N	223	264	223	264
Tests across equations (environments):				
Test that γ_1 differs	$p < 0.01$		$p = 0.06$	
Test that γ_2 differs	$p = 0.18$		$p = 0.27$	
Test that γ_1 and γ_2 differ	$p < 0.001$		$p = 0.10$	

Note: Dependent variable: # of deduction points. OLS estimates. Robust standard errors clustered on the individual in parentheses. p-values in cross-equation tests are all two-sided. *Level of significance*: * : $0.05 \leq p < 0.1$, ** : $0.01 \leq p < 0.05$, *** : $p < 0.01$.

2.3.3 Table S5: Differential Punishment of Cooperation and Defection (Norm Enforcement)

In order to examine the differential in punishment between cooperation and defection, we estimate for each treatment the following equations:

$$PP_{ik} = \alpha + \gamma_1 I_i(\text{A1 out-group}) + \gamma_2 I_i(\text{A2 in-group}) + \gamma_3 I_k(\text{A1 defects}) + e_i \qquad (7)$$

$$\begin{aligned} PP_{ik} =& \alpha + \gamma_1 I_i(\text{A1 out-group}) + \gamma_2 I_i(\text{A2 in-group}) \\ &+ \gamma_4 I_k(\text{A1 defects}) \times I_i(\text{A2 out-group}) + \gamma_5 I_k(\text{A1 defects}) \times I_i(\text{A2 in-group}) + e_i \end{aligned} \qquad (8)$$

in which *A1 defects* equals 1 if A1 defects and 0 otherwise. Regression 8 adds two interaction terms for *A1 defects* and whether A2 is an out-group member or an in-group member

The results are displayed in Table S5. In the text, we refer to the following results:

- Defection is more strongly punished than cooperation in NG ($p < 0.01$, column (1)).

Table S5: Norm Enforcement Across Environments

Environment:	Neutral (1)	Comp (2)	Neutral (3)	Comp (4)
A1 out-group (γ_1)	0.028	3.320***	0.028	3.320***
	(0.624)	(0.589)	(0.625)	(0.590)
A2 in-group (γ_2)	1.114*	0.162	0.513	–0.533
	(0.609)	(0.595)	(0.691)	(0.694)
A1 defects (γ_3)	2.838***	0.682**		
	(0.445)	(0.330)		
A1 defects × A2 out-group (γ_4)			3.510***	1.482***
			(0.722)	(0.560)
A1 defects × A2 in-group (γ_5)			2.306***	0.092
			(0.549)	(0.385)
Constant	1.978***	1.471***	2.244***	1.766***
	(0.565)	(0.465)	(0.578)	(0.495)
R^2	0.130	0.161	0.136	0.167
N	222	264	222	264
Tests across equations (environments):				
Test that γ_1 differs	$p < 0.01$		$p < 0.01$	
Test that γ_2 differs	$p = 0.26$		$p = 0.29$	
Test that γ_3 differs	$p < 0.01$			
Test that γ_4 differs			$p = 0.03$	
Test that γ_5 differs			$p < 0.01$	
Test that γ_1, γ_2, and γ_3 differ	$p < 0.01$			
Test that γ_1, γ_2, γ_4, and γ_5 differs			$p < 0.01$	

Note: Dependent variable: # of deduction points. OLS estimates. Robust standard errors clustered on the individual in parentheses.
Level of significance: * : $0.05 \leq p < 0.1$, ** : $0.01 \leq p < 0.05$, *** : $p < 0.01$. p-values in cross-equation tests are all two-sided.

2.3.4 Comparisons across Equations

The bottom panels of Table S3 to S5 also display cross-equation tests. For single-coefficient tests, we calculate

$$z = \frac{\gamma_j - \tilde{\gamma}_j}{\sqrt{\Sigma_{jj} + \tilde{\Sigma}_{jj}}} \tag{9}$$

where γ_j and $\tilde{\gamma}_j$ are the two coefficients of interest from the two equations, and Σ_{jj} and $\tilde{\Sigma}_{jj}$ are the corresponding main diagonal elements in the covariance matrix (because the two coefficients come from two separate equations, their covariance, by construction, is zero). z has a standard normal distribution under the null of no difference. We report two-sided p-values to be conservative. In the case of coefficient vectors, we calculate the analogous test statistic

$$\chi = (\gamma - \tilde{\gamma})(\Sigma + \tilde{\Sigma})^{-1}(\gamma - \tilde{\gamma})' \tag{10}$$

which has a chi-square distribution with k degrees of freedom, where k is the number of variables in γ. In the text, we refer to the following results

- The punishment responds differently to the group affiliations of A1 and A2 in the two treatments ($p < 0.01$, Table S3).
- In treatment CG, A1 is punished more heavily than in NG if he is out-group ($p < 0.01$ for both, cooperation and defection of A1, Table S3).
- In treatment NG, punishment is much more sensitive to whether A1 defected or cooperated ($p < 0.01$, Table S5).
- Even when A2 is an in-group member, punishment of defection (*A1 defects*) is less strong in CG ($p < 0.01$, Table S5).

References

[1] L. Goette, D. Huffman, S. Meier, *American Economic Review* **96**, 212 (2006).

[2] E. Fehr, U. Fischbacher, *Evolution and Human Behavior* **25**, 63 (2004).