Group Selection with Imperfect Separation — An Experiment *

Veronika Grimm[†] Friederike Mengel[‡]

University of Cologne University of Alicante

November 3, 2006

Abstract

We experimentally investigate the effect of imperfect separation of groups on group selection and cooperation in a standard prisoner's dilemma environment. Subjects can repeatedly choose between two groups, where in one of them an institutionalized norm fosters cooperation. The degree of separation of the two groups is varied between treatments. We find that both, the share of agents that choose into the normative group and the share of agents that cooperate, rise monotonously with the degree of group separation. We moreover find positive feedback effects in the sense that in treatments with a higher degree of group separation (where cooperative equilibria are played) significantly more subjects claim to be in favor of norm enforcement.

Keywords: Experiments, Cooperation, Group Selection, Social Norms, Population Viscosity.
JEL classification: C70, C73, Z13.

---

*We thank Axel Ockenfels as well as seminar participants at Cologne, Amsterdam, and Magdeburg for helpful comments and suggestions, and Felix Lamoroux and Rene Cyranek for most valuable research assistance. Financial support by the *Deutsche Forschungsgemeinschaft,* the *Instituto Valenciano de Investigaciones Económicas (IVIE)* and the *Spanish Ministery of Education and Science* (grant SEJ 2004-02172) is gratefully acknowledged.

[†]University of Cologne, Staatswissenschaftliches Seminar, Albertus-Magnus-Platz, D-50923 Köln, Germany, vgrimm@uni-koeln.de

[‡]University of Alicante, Departamento de Fundamentos del Análisis Económico, Campus San Vicente del Raspeig, 03071 Alicante, Spain, friederike@merlin.fae.ua.es

1 Introduction

One of the biggest challenges in social sciences, as well as in biology is to explain pro-social behavior in social dilemma situations. The paradigmatic model of such a situation is the prisoner's dilemma game. In this game agents can choose between a cooperative action and a deficient action. Defection is a dominant strategy for both players (thus individually rational), but both would be better off if they jointly cooperated. People are often observed to cooperate in such situations although they could achieve a higher payoff choosing defection, independently of what their opponent does. This phenomenon has been termed the puzzle of pro-sociality.

Cultural Group Selection is an often advanced explanation for the survival of pro-sociality in social dilemma situations. The basic idea is that if there are two separated cultural groups, one in which agents cooperate (because they share a social norm to do so or because they are genetically programmed) and one, in which agents defect, the agents in the cooperative group will earn higher payoffs and thus have higher evolutionary fitness. This will lead to a proliferation of the cooperative trait or norm.[1] Typically though cultural groups are not perfectly separated. However, as long as agents of the cooperative group are matched with a sufficiently high probability among themselves cooperation can still survive. The intuition simply is that if cooperators interact relatively often with other cooperators they will receive the high payoff of joint cooperation a lot of times, whereas defectors will get the low payoff of joint defection a lot of times. Populations in which there is an increased probability of interacting with agents of one's own type (cultural group) are called *viscous populations*.[2]

In this experiment we investigate the effect of population viscosity (imperfect separation of groups) on group selection and cooperation in a standard prisoner's dilemma environment. Subjects can repeatedly choose between two groups, where in one of them an institutionalized norm fosters cooperation. The degree of separation of the two groups is varied between treatments. We find that participants choose into the "normative" group if and only if the degree of separation between groups is high enough. Also while agents in the normative group cooperate if and only if the degree of separation is high enough, agents in the other group almost never cooperate. Finally both, the share of agents that choose into the normative group and the share of agents that cooperate, rise monotonously with the degree of population viscosity (the degree of separation between groups). Population viscosity - sometimes also referred to as assortative matching - seems a powerful mechanism to sustain cooperation.

We also tackle the difficult question of eliciting social norms and attitudes towards norm enforcement in order to gain additional insight into the relation between population viscosity, norms, norm enforcement and cooperation. To these ends we gave participants a questionnaire specially designed to elicit their cooperative types, attitudes towards norm enforcement and normative principles at the end of the ex-

[1]See Mitteldorf and Wilson (2000), Boyd and Richerson (2005), Richerson, Boyd and Henrich (2003) or Wilson and Sober (1994) among others.

[2]See Myerson, Pollock and Swinkels (1991) or Mitteldorf and Wilson (2000) among others.

periment.[3] While we find no significant differences between cooperator types among treatments, we find that in treatments with high separation (where cooperative equilibria are played) significantly more subjects claim to be in favor of norm enforcement.[4] Note that these results jointly imply that the participants' answers in the questionnaire were not mainly driven by dissonance avoidance. There seem to be feedback effects between the equilibrium of the different treatments and attitudes towards norm enforcement. High viscosity seems to favor norm enforcement.

The experimental literature has already demonstrated the existence of different cooperative types. Fischbacher, Fehr and Gächter (2001) find that in their experiment roughly 50% of all subjects can be classified as conditional cooperators, 30% are flat defectors and only very few cooperate always.[5] Their two principal types (conditional cooperators and defectors) are also the main types we find in our questionnaire.

Only very recently experimental economics has started to focus on the relation between interaction structures and cooperation. Coricelli, Fehr and Fellner (2004) or Page, Putterman and Unel (2005) are examples of studies in which agents can endogenously choose interaction partners.[6]

The effect of punishment institutions has been investigated by Kosfeld and Riedl (2004) among others. Guererk, Irlenbusch and Rockenbach (2006) show that subjects learn to choose into a group where a punishment mechanism is at place.[7] Our results in the full separation treatment confirm theirs.

All the above studies deal only with the case of perfect separation of groups. To our knowledge this study is the first to analyze the case of imperfect separation. Another novelty in our experiment is the questionnaire that allows us to gain knowledge about variables such as attitudes towards normative criteria and norm enforcement and their relation to behavior in the experiment.

The paper is organized as follows. In Section 2 we present the theoretical model underlying our study. In Section 3 the hypotheses derived from this model are summarized and Section 4 describes the experimental design. The results from the experiment are presented and discussed in Section 5. Section 6 is devoted to a more general discussion of social norms in experiments and points to some promising directions for future research. Section 7 concludes.

[3] We are aware of the problems of using questionnaire data. We see it as a useful method and as a very first step to tackle the difficult question of measuring social norms in experiments without distorting the participants decisions. See also the discussion in Section 6.

[4] If such feedback effects do indeed exist many interesting theoretical predictions arise (See Benabou and Tirole (2005), Mengel (2006), Lindbeck, Nyberg and Weibull (1999) or Traxler (2005)). See also the discussion in section 6.

[5] See also Fischbacher and Gächter (2006) or Brandts and Schram (2001).

[6] See also Ones and Putterman (2006) or the literature on network experiments reviewed in Falk and Kosfeld (2003).

[7] See also Goette, Huffman and Meier (2006) for a field study on these issues.

2 The Model

2.1 The Basic Game

The game we study is a standard (symmetric) Prisoner's dilemma game, in which agents can either cooperate (C) or defect (D). The payoffs are given by the following matrix

	C	D
C	a	b
D	c	d

Table 1: Payoff Matrix of the Prisoner's Dilemma

with $c > a > d > b$. The cooperative outcome is efficient whenever $a > \frac{b+c}{2}$. In the experiment we use the following parametrization

$$a = 400; b = 50; c = 550; d = 200. \quad (1)$$

If agents are randomly matched to play this game the unique prediction is mutual defection and thus a payoff of d ($= 200$) for both. What happens if there is a cultural group that shares a social norm to cooperate (play C) in the Prisoner's Dilemma? Clearly if agents in this group would exclusively interact among themselves they would receive a payoff of a ($= 400$) and would be better off than agents that do not adhere to the norm. These latter agents interacting only among themselves would receive a payoff of d ($= 200$). Group Selection argues that in this case the social norm would slowly spread and finally take over the population. The reason is that as agents adhering to the social norm receive higher payoffs, they have higher evolutionary fitness and will be selected according to many standard evolutionary dynamics.[8] Whenever cultural groups are not perfectly isolated though, defectors can sometimes exploit norm-adherers. Our model is designed to study group selection and cooperation in the case where cultural groups are *not* perfectly separated.[9]

2.2 Cultural Groups

There are two cultural groups: A and B. Agents in group A share a norm to cooperate. This norm is institutionalized in group A through a punishment institution. More precisely, whenever a member of group A defects in an interaction with another A member she incurs a payoff loss of γ ($= 200$). One can think of this as a social disapproval mechanism. While in the experiment punishment is material, in real

[8]See Boyd and Richerson (2005), Mitteldorf and Wilson (2000) or Wilson and Sober (1994) among others.

[9]Our baseline model of group selection in viscous populations builds on Mengel's (2006) model of cultural transmission of social norms.

life this term will typically correspond to either a psychological payoff loss or to anticipation of a material loss in the future.[10] Thus, whereas a member of group B faces the payoff matrix given in table 1, the relevant payoff matrix for a group A member is given by

	C	D
C	a	b
D	$c - \gamma\delta_{AA}$	$d - \gamma\delta_{AA}$

Table 2: Payoff Matrix Group A

where δ_{AA} takes on the value $\delta_{AA} = 1$ if both interaction partners are group A members. Obviously, for a group B member $\delta_{AA} = 0$ independently of whom he interacts with.

In our model, group-membership defines an agent's type. At all times agents have incomplete information about the type of their match. When choosing an action in the bilateral game they have to estimate the type of their match from the distribution of types in the economy and from their knowledge about the matching technology described below. Clearly, for a group B member defection is a dominant strategy. For a group A member, whether cooperation or defection is optimal depends on the relative size of the two groups and on the degree of separation of the two groups or the matching technology.

	A	B
A	$1 - p_B x$	$p_B x$
B	$p_A x$	$1 - p_A x$

Table 3: Matching Probabilities

Matching takes place randomly in a viscous population, the latter meaning that individuals have a tendency to interact more often with individuals that are of the same type. The degree of viscosity is measured by the parameter $x \in [0, 1]$. $x = 1$ corresponds to the case of random matching. $x = 0$ means that the population is fully viscous, implying that agents interact with probability 1 with agents of the same group and never with agents from another group. This is the case typically considered in models of group selection. In a viscous society with parameter x, if p_A is the share of agents of type A (members of group A) the probability for any one of them to interact with a B type is $(1 - p_A)x = p_B x$ and the probability to interact with a member of group A is $(1 - (1 - p_A)x) = 1 - p_B x$. Obviously if the society is fully viscous ($x = 0$), agents only interact with agents of their own cultural group. In other words, in this case cultural groups are separated. The matching probabilities are summarized in table 3.

[10] See Section 6 for further discussion of this issue.

2.3 Cultural Equilibria

We assume that materially successful groups attract agents and proliferate. Denote p_A the share of agents in group A and assume that p_A evolves as follows:

$$\dot{p}_A = p_A(1 - p_A)[\Pi_A - \Pi_B], \tag{2}$$

where Π_A and Π_B are the average payoffs of group A and group B members.

Remark 1 *Equation (2) is essentially the well known replicator dynamics used in biology and evolutionary game theory.*[11] *It can follow for example from a process in which expected payoff maximizing agents can decide periodically to change their group and adopt the corresponding payoff matrix. Alternatively one could think of agents as being reinforcement learners, that are endowed with attractions for the two groups. A "good" payoff experience in any group reinforces the agents attraction for that group. Such a process can also be approximated by (2). Both are possible models of how group selection operates in our experiment.* [12] *While the first one presumes a rather high degree of rationality of the agents, in the reinforcement model agents hardly need to know anything about the game played. In particular, they do not need to understand the incentive structure, nor calculate best response functions, nor they need to know that there is strategic interaction.*

Let us call a *cultural equilibrium* a share p_A together with an action choice in the bilateral game, such that (i) the action choice is a Nash equilibrium given x and p_A and (ii) p_A is an asymptotically stable equilibrium of (2). Then, the theoretical prediction can be summarized as follows:

Proposition 1

(i) If $x < \frac{1}{4}$ the globally stable cultural equilibrium has $p_A^* = 1$ with all players cooperating.

(ii) If $x \in [\frac{1}{4}, \frac{4}{7}]$ there are two locally stable cultural equilibria: $p_A^* = 1$ with all players cooperating and $p_A^* = 0$ with all players defecting.

(iii) If $x > \frac{4}{7}$ the globally stable cultural equilibrium has $p_A^* = 0$ with all players defecting.[13]

Proof   Appendix A. □

[11]For good introductions to concepts from evolutionary game theory see Weibull (1995) or Vega-Redondo (1996).

[12]Equation (2) could also be derived from models of payoff-biased imitation and cultural transmission of social norms, see Mengel (2006). Deriving the dynamics from such a model would lead to the following equation: $\dot{p}_A = p_A(1 - p_A)x[\Pi_A - \Pi_B]$, where the viscosity factor x represents a change of time scale. If there is little interaction between groups (small x) cultural transmission is slow.

[13]Note that whereas in the theory outlined above there is a continuum of agents this is obviously not the case in the experiment. Proposition 1 is derived for the discrete case of our experiment.

In the experiment we implemented four treatments corresponding to the viscosity parameters $x = 0, \frac{1}{3}, \frac{2}{3}, 1$. The theoretical predictions for these treatments are summarized in the following section.

3 Hypotheses from the Theory

Group Selection In treatments $x = 0$ and $x = \frac{1}{3}$ all subjects should join group A and cooperate. In treatments $x = \frac{2}{3}$ and $x = 1$ all subjects should join group B and defect.

Cooperation Group B members should always defect. Whether cooperation or defection is optimal for group A members depends on (i) the relative size of the two groups, p_A, and on (ii) the degree of separation of the two groups, x. An equilibrium where subjects in group A cooperate exists if and only if viscosity is sufficiently high (i.e. in treatments $x = 0$ and $x = \frac{1}{3}$).

Profits Average profits in the population should be equal to 400 in treatments $x = 0$ and $x = \frac{1}{3}$ and equal to 200 in treatments $x = \frac{2}{3}$ and $x = 1$. Group A members should have higher (lower) profits than group B members in treatments $x = 0$ and $x = \frac{1}{3}$ ($x = \frac{2}{3}$ and $x = 1$).[14]

Rate of Convergence and Learning Dynamics Both approaches (the evolutionary model and the reinforcement learning model) predict that learning is fastest for $x = 0$, slower for $x = 1$ and slowest for the intermediate x–values.

Agents are expected to switch from group A to group B after observing a high share p_A or a high rate of defection in treatments $x = \frac{2}{3}$ and $x = 1$. Experiencing punishment can also induce switching into group B.

Agents are expected to switch from group B to group A after observing a high share p_A or a high rate of defection in treatments $x = 0$ and $x = \frac{1}{3}$.

4 The Experimental Design

The experiment was conducted in four sessions in May, 2006. A total of 128 students (32 per session) were recruited among the student population of the University of Cologne — mainly undergraduate students with no (or very little) prior exposure to game theory.

In order to answer our research questions we implemented four different treatments that differed in the degree of viscosity, x, as defined in section 2. We chose the values $x \in \{0, \frac{1}{3}, \frac{2}{3}, 1\}$. One population consisted of 8 subjects. The members of a population

[14] If this were not the case, an equilibrium where agents select into group A (B) in treatments $x = 0$ and $x = \frac{1}{3}$ ($x = \frac{2}{3}$ and $x = 1$) could not be stable under (2).

were initially randomly assigned to groups A and B in equal proportions. In the first four rounds, each subject played the game described in section 2 with an interaction partner who was assigned randomly according to the matching technology. From round 5 on, each round had two stages. At the first stage, two of the eight subjects could decide to either join the other group, or to stay in their own group. Each subject could make this decision every fourth round. At the second stage of each round, subjects played the (modified) prisoner's dilemma game as given by (2) with an interaction partner who was assigned randomly according to the matching technology. Prior to playing the game they were informed about (a) the percentage of subjects in group A and B, and (b) their individual probability to meet a group–A and group–B member, respectively.

Since in our experiment the population was necessarily finite, one-to-one matching was not feasible for matching technologies with $x \neq 1$ (i.e. in three out of four treatments). Instead, we first realized a random draw with the probabilities given in table 3 to decide whether a subject's interaction partner was from group A or B. Then the interaction partner played the actions "cooperate" or "defect" with probabilities that corresponded to the proportions with which those actions were played in the respective group (in that round). In the unlikely event that only one subject remained in a group (either A or B) and the first random draw determined that he had to play against an member of his own group, the subject's interaction partner was preprogrammed to play the equilibrium strategy.[15] After each of the 100 rounds, subjects were informed of whether their interaction partner belonged to group A or B, his action, and their own monetary payoffs.

At the end of the experiment (after all 100 rounds were finished) we had the participants answer a questionnaire designed to elicit their attitudes towards cooperation, the normative principles their decisions were guided by, and their attitudes towards norm enforcement.

The four experimental sessions were computerized.[16] Written instructions were distributed at the beginning of the experiment.[17] Each session took approximately 120 minutes (including reading the instructions, answering a post-experimental questionnaire and receiving payments). Subjects participating in the experiment received 2.50 Euros just to show up. On average subjects earned Euro 15.25 Euros (all included).

[15] The subjects were informed that the interaction partner would play optimally given the situation in this case.

[16] The experiment was programmed and conducted with the software z-Tree (Fischbacher 1999).

[17] The instructions for $x = 0$, translated into English, can be found in the Appendix. Instructions for the remaining treatments are available upon request.

5 Results

5.1 Prologue: Attitudes Towards Cooperation

Let us first report some results from the post–experimental questionnaire that we will later on refer to in our analysis of the data.

Cooperator Types In the post–experimental questionnaire we asked the subjects whether — in a one shot game — they would cooperate if 0 (25, 50, 75, 100) per cent of the other subjects cooperated. From the answers we identify four "cooperator types": (1) flat defectors (who always defect), (2) altruists (who always cooperate), (3) conditional cooperators (who cooperate if and only if the share of cooperators is sufficiently high) and (4) hump shaped (who cooperate if and only if the share of cooperators is intermediate and defect otherwise). Table 4 reports the results. The attitude of subjects towards cooperation seems to vary somewhat across treatments but the differences are not significant.[18]

Cooperator Type	Treatment $x=0$	$x=\frac{1}{3}$	$x=\frac{2}{3}$	$x=1$	Overall
flat defectors	.19	.41	.41	.28	.32
altruists	.06	–	–	.03	.02
conditional cooperators	.28	.37	.28	.34	.32
hump shaped	.31	.19	.22	.19	.23
none of the others	.16	.03	.09	.16	.11

Table 4: Cooperator Types.

Norm Enforcement A second observation from the questionnaire that we would like to report are the subjects' attitudes towards norm enforcement. We asked subjects the question whether they think that choosing defection in the stage game should be punished. In treatments with a high degree of group separation ($x = 0$ and $x = \frac{1}{3}$), the majority of subjects was in favor of punishment, while in those treatments where group separation was low ($x = \frac{2}{3}$ and $x = 1$), the majority was against punishment of defectors (using random matching as a benchmark the difference is significant at the 5% (1%) level for $x = 0$ ($x = \frac{1}{3}$)). Those answers approximately reflect what the subjects experienced in the experiment. As we have shown in section 2 punishment as

[18] We ran logit regressions on the questionnaire data for the following binary variables: $defector =$ 1 if an agent is classified as defector and zero otherwise, $cooperator = 1$ if an agent is classified as either altruist ($altruist$) or conditional cooperator ($conditionalC$) and $humpshaped = 1$ if an agent is classified as hump-shaped. Estimating the coefficients of treatment dummies gives no significant effects.

implemented in our design induces cooperation for a high degree of group separation whereas it fails to do so for a low degree of separation between groups. Also in the experiment cooperation obtained for $x = 0$ and $x = \frac{1}{3}$ and defection was the predominant pattern for $x = \frac{2}{3}$ and $x = 1$.

	Treatment				
Norm Enforcer?	$x = 0$	$x = \frac{1}{3}$	$x = \frac{2}{3}$	$x = 1$	Overall
no	.44	.34	.53	.69	.50
yes	.56	.66	.47	.31	.50

Table 5: Attitudes Towards Norm Enforcement.

We conclude that there seem to be feedback effects from the treatment variable (x) to the subjects attitudes towards norm enforcement. Furthermore we find that participants that claim to be conditional cooperators in the questionnaire are in favor of norm enforcement significantly (at the 1% level) more often in all treatments.[19]

We will come back to the questionnaire results in section 6. Now let us turn to the analysis of our experimental data. We proceed as follows: We first report the evidence on group selection. Then, we show the effects group selection had on cooperation rates in the two groups (A and B), overall cooperation rates and profits (again, per group and overall) in the population. Finally we report some evidence on the speed of convergence and on learning dynamics.

5.2 Group Selection

Figure 1 impressively illustrates the effect population viscosity has on group selection. While in the perfectly separated population (treatment $x = 0$) almost all subjects join group A, the share of subjects that are in group A decreases as viscosity decreases. However, even under random matching, group A does not entirely disappear.[20]

Statistical analysis yields the following results, which are all significant at the 1% level:[21]

[19]We regress the variable *dpunish* (= 1 if the answer is yes and zero if no) on treatment variables and cooperator types (as well as interaction terms) using logit regressions.

[20]This can in part be explained by the incentives of our experimental design. In treatment $x = 0$, since a single group B member was automatically matched with a preprogrammed defector, there was a strict incentive to join group A in order to realize the gains from mutual cooperation. In treatment $x = 1$, however, if all other subjects were in group B, a subject had the same expected payoff from a given action if it remained in group A (since under random matching it could not meet another group A member and therefore, punishment was impossible).

[21]We ran random effect logit regressions for the variable *groupA* (= 1 if an agent is in group A and zero otherwise). The independent variables were *cooperator*, *dpunish*, the treatment dummies and interaction terms.

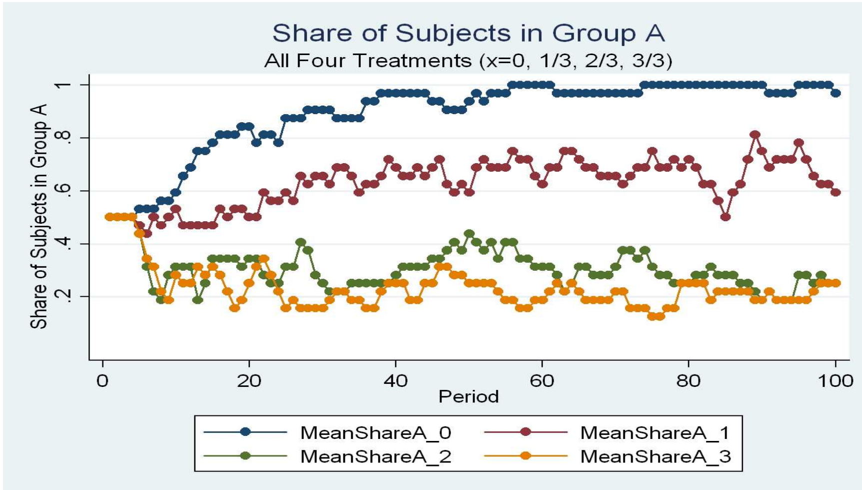

Figure 1: The Share of Subjects in Group A (per treatment).

Result 1 (Group Selection)

(i) The share of subjects in group A is higher, the more viscous the population is.

(ii) "Conditional cooperators" are more likely to be in group A in treatments $x = 0$ and $x = \frac{1}{3}$ and more likely to be in group B in treatments $x = \frac{2}{3}$ and $x = 1$.

(iii) "Norm enforcers" are more likely to be in group A in treatments $x = 0$ and $x = \frac{1}{3}$.

These results have very intuitive interpretations. Conditional Cooperators cooperate whenever matched with high probability with other cooperators. Consequently in treatments $x = 0$ and $x = \frac{1}{3}$ they cooperate and choose group A, whereas in treatments $x = \frac{2}{3}$ and $x = 1$ they defect as the environment is characterized by defection, but then it is optimal to join group B. It also makes perfect sense that norm enforcers choose into the group (A) where the norm is enforced. On the other hand experiencing "successful" norm enforcement in treatments $x = 0$ and $x = \frac{1}{3}$ (where most subjects are in group A) possibly leads agents to regard norm enforcement a sensible thing to do. These are the feedback effects already mentioned in the previous section.[22]

5.3 Cooperation

As figure 2 illustrates, the shares of cooperating subjects in the population evolves in line with the share of subjects in group A (compare figure 1). Analyzing cooperation shares separately for the two different groups reveals that in all treatments (except $x = 1$) the majority of subjects in group A cooperates, while almost no group B–member does (see table 4).

[22] Whereas in the case of conditional cooperation the treatment dummies can be used to control for endogeneity, in the case of norm-enforcement this cannot be done, as norm-enforcement is highly correlated with treatment. In this case there are true feedback effects.

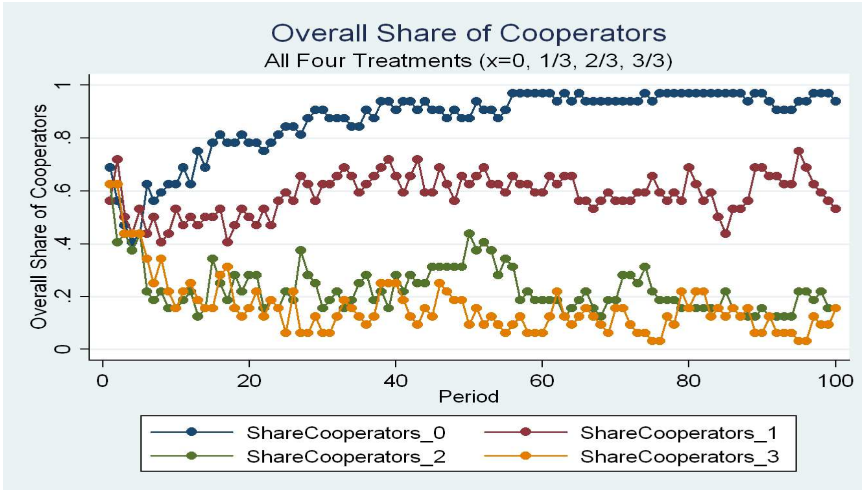

Figure 2: Shares of Cooperators.

Treatment	Group A	Group B	Overall
$x = 0$	.95	.14	.94
$x = \frac{1}{3}$	.90	.06	.86
$x = \frac{2}{3}$	.70	.02	.66
$x = 1$	.42	.08	.29

Table 6: Average Cooperation Rates

We obtain the following results which are again all significant at the 1% level:[23]

Result 2 (Cooperation)

(i) Subjects in group A cooperate significantly more than subjects in group B.

(ii) The more subjects are in group A the more subjects cooperate in group A (except treatment $x = 0$, where subjects almost always cooperate in group A).

(iii) "Norm Enforcers" are more likely to cooperate in all treatments.

(iv) "Conditional Cooperators" and "Altruists" cooperate significantly more than all other cooperator types in all treatments.

Again the interpretation of the results seems clear. (i) and (ii) simply reflect the incentive structure. Defection makes only sense in group B as in group A it is punished. The higher p_A the more likely an agent is matched with a group A member

[23] We ran random effect logit regressions for the variable *cooperate* (= 1 if an agent chooses to cooperate and zero otherwise). The independent variables were *cooperator*, *dpunish*, the treatment dummies and interaction terms.

and thus cooperation in group A is more likely to occur. This is true for all treatments except $x = 0$ where matching is independent of p_A. (iii) and (iv) show that behavior in the experiment is roughly consistent with attitudes expressed in the questionnaire. Norm Enforcers (who state that defection should be punished) are more likely to cooperate. Similarly those that claim to be altruists or conditional cooperators in the questionnaire did indeed cooperate more in the experiment.

5.4 Profits

The observed behavior (concerning group selection and cooperation) had clear consequences on profits. Recall that overall rates of cooperation were higher, the higher population viscosity. Consequently, payoffs were highest in treatment $x = 0$, lowest (and close to the payoffs from mutual defection) for $x = 1$ and in between for the remaining treatments with intermediate degrees of separation. Members of group A have a higher payoff than members of group B in treatments $x = 0$ and $x = \frac{1}{3}$ and vice versa in treatments $x = \frac{2}{3}$ and $x = 1$. We summarize our results in table 5.

Treatment	Group		Overall
	A	B	
$x = 0$	383	223	366
$x = \frac{1}{3}$	326	265	303
$x = \frac{2}{3}$	206	248	235
$x = 1$	202	245	229

Table 7: Profits.

We find the following regression results.[24]

Result 3 (Profits)

(i) Average profits in the population are highest in treatment $x = 0$ (1%), followed by $x = \frac{1}{3}$ (1%) and treatments $x = \frac{2}{3}$ and $x = 1$ (no significant difference between the latter).

(ii) The profit of a group A–member is higher than the profit of a group B–member in treatments $x = 0$ and $x = \frac{1}{3}$ and vice versa in treatments $x = \frac{2}{3}$ and $x = 1$ (all significant at 1%)

Agents that claim to be conditional cooperators in the questionnaire seem to make somewhat higher profits in all treatments but the effect is not significant.

[24] We ran panel data regressions for the variable $totalprofit$ (that reports overall profits in the experiment). The independent variables were *cooperator*, $defector$, the treatment dummies and interaction terms.

5.5 Rate of Convergence and Learning Dynamics

How often do the participants switch groups during the experiment? Table 6 reveals that participants switch most often in the treatments with intermediate degrees of viscosity, and least often if groups are perfectly separated. While in treatment $x = 0$ most of the "group switching" takes place during the first quarter (Q1, the first 25 rounds) of the experiment, in the treatments with intermediate degrees of viscosity there is still a substantial number of switches even in the last quarter (Q4). Consistently with theory (either the evolutionary or the reinforcement model) convergence to equilibrium is fastest in the $x = 0$ treatment and slowest for the intermediate treatments. The higher payoff differences between the two groups in the $x = 0$ and $x = 1$ treatments effectively seem to speed up learning, as the reinforcement model predicts. The observed behavior could also reflect a higher transparency of the economic incentives in these two treatments, though.

Treatment	Q1	Q2	Q3	Q4	Overall
$x = 0$	48	12	8	3	71
$x = \frac{1}{3}$	51	32	39	45	167
$x = \frac{2}{3}$	52	38	42	30	162
$x = 1$	45	25	18	14	102

Table 8: Switching Frequencies per Quarter and Overall, All Treatments.

Result 4 (Convergence) *The subjects' play converges fastest in treatment $x = 0$ and slowest in treatments $x = \frac{1}{3}$ and $x = \frac{2}{3}$.*

We feel that estimating and comparing different learning models is not of central interest in this paper. Instead let us take a closer look at what are the factors that induce agents to switch groups. The candidates we consider are the share of participants in either group, the experience of meeting a defector and the experience of being punished. We get the following results:[25]

Result 5 (Learning Dynamics)

(i) The share of subjects in group A triggers switches into group A in treatments $x = 0$ and $x = \frac{1}{3}$ and switches into group B in treatments $x = \frac{2}{3}$ and $x = 1$ (all 1%).

(ii) Meeting a defector triggers switches into group A in treatments $x = 0$ and $x = \frac{1}{3}$ and into group B in treatment $x = \frac{2}{3}$ (all 1%). There is no significant effect in treatment $x = 1$.

[25] We ran random effects logit regressions for the variable *switch* (= 1 if an agent had the opportunity to switch groups and did so). The independent variables were *shareA* (reporting the share of participants in group *A*) *pexperience* (= 1 if an agent has been punished at least once during the last 4 periods and zero otherwise) and *oppdefectexperience* (= 1 if an agent's match has defected at least once in the last 4 rounds).

(iii) The experience of punishment leads to both: switches into group B (since there it is cheaper to defect) and switches to cooperative behavior in group A (1%).

All of these findings are consistent with theory.

6 Social Norms in Experiments

Apart from analyzing the effect of population viscosity on cooperation, we also view this project as a first step to tackle the difficult question of experimentally investigating the dynamics of social norms and its relation to the matching structure. As should have come clear from the previous sections though, there are many difficulties in bringing social norms to the lab. In our experiment punishment in group A induces material incentives that resemble incentives of norm-guided agents. If one wants to truly study the dynamics of social norms in experiments though, a way has to be found to measure social norms in the laboratory. This is a very challenging question. Social norms are particularly interesting in this context because of the feedback effects between equilibrium and norm. These effects have been studied theoretically by Benabou and Tirole (2005), Lindbeck, Nyberg and Weibull (1999), Mengel (2006) and Traxler (2005).

We try to shed some light on these feedback effects through the use of our post experimental questionnaire, especially designed to gain insight into possible feedback effects between norm and equilibrium. Our results suggest that such feedback effects do indeed exist: participants of the $x = 0$ and $x = \frac{1}{3}$ treatments that play cooperative equilibria more often state (in the questionnaire) that they are in favor of norm-enforcement than participants of the $x = \frac{2}{3}$ and $x = 1$ treatments do (that have been playing deficient equilibria). This is only very preliminary evidence of such effects though. It would be desirable to be able to estimate "norms" directly from the experimental data instead of relying on questionnaire data only. In a follow up study we investigate this issue.

7 Conclusion

In this paper we have experimentally investigated the impact of population viscosity on cooperation in social dilemma situations. Participants in our experiment could repeatedly choose between two groups, where in one of them an institutionalized norm fosters cooperation. The degree of population viscosity was varied between treatments. Our results are largely in line with theory. In particular we found that the share of participants that choose into the normative group rises with the degree of population viscosity. Participants almost always cooperate in the normative group and almost never in the other group. Average profits for participants in the normative group are the higher the more separated groups are. Finally participants are able to quickly understand the economic incentives implied by different matching structures (degrees of population viscosity) and their behavior seems consistent with e.g. a simple reinforcement learning model.

Our experiment thus provides evidence on the importance of cultural group selection for sustaining sanctioning institutions that lead to cooperation. Using a post-experimental questionnaire we also find evidence for possible feedback effects. Participants of treatments characterized by high viscosity tend to be more in favor of norm enforcement. In short, population viscosity seems a powerful and important mechanism for sustaining cooperation. To understand the way population viscosity acts on economic incentives and social norms gives rich potential for further research, both theoretically and experimentally.

References

[1] Benabou, R. and J.Tirole (2005), "Incentives and Prosocial Behavior", *NBER working paper 11535*.

[2] Brandts, J. and A.Schram (2001), "Cooperation and noise in public goods experiments: applying the contribution function approach", *Journal of Public Economics 79*, 399-427.

[3] Boyd,R. and P.Richerson (2005), "The Origin and Evolution of Cultures (Evolution and Cognition)", University of Chicago Press..

[4] Coricelli, G., D. Fehr and G. Fellner (2004), " Partner Selection in Public Goods Experiments", *Journal of Conflict Resolution 48*, 356-378.

[5] Fischbacher, U. (1999), Z-tree Zurich Toolbox for Readymade Experiments, *IERE working paper No 21*, University of Zurich.

[6] Fischbacher, U., Gächter, S. and E.Fehr (2001), "Are people conditionally cooperative ? Evidence from a public goods experiment", *Economics Letters 71,* 397-404.

[7] Fischbacher, U. and S. Gächter (2006), "Heterogenous social preferences and the dynamics of free-riding in public goods", *CeDEx Discussion paper*.

[8] Goette, L., D. Huffman and S. Meier (2006), "The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups", *working paper 06-07* Federal Reserve Bank of Boston.

[9] Guererk, Ö., B. Irlenbusch, and B. Rockenbach (2006) "The Competitive Advantage of Sanctioning Institutions," *Science 312 (5770)*, 108-111.

[10] Kosfeld, M. (2003), "Network Experiments", *IERE working paper No 152*, University of Zurich.

[11] Kosfeld, M. and A. Riedl (2004), "The Design of Decentralized Punishment Institutions for Sustaining Cooperation", mimeo University of Zurich.

[12] Lindbeck,A., S.Nyberg and J.Weibull (1999), "Social Norms and Economic Incentives in the Welfare State", *Quarterly Journal of Economics 114,* 1-35.

[13] Mengel, F. (2006) "A Model of Immigration, Integration and Cultural Transmission of Social Norms", *IVIE working paper AD 2006-08*.

[14] Mitteldorf,J. and D.S. Wilson (2000), "Population Viscosity and the Evolution of Altruism", *Journal of Theoretical Biology 204*, 481-496.

[15] Myerson R.B., G.B.Pollock and J.M.Swinkels (1991), "Viscous Population Equilibria", *Games and Economic Behaviour 3,* 101-109.

[16] Page, T., L. Putterman and B.Unel (2005), "Voluntary Association in Public Goods Experiments: Reciprocity, Mimicry and Efficiency, *The Economic Journal 115*, 1032-1053.

[17] Richerson, P., R.Boyd and J.Henrich (2003), "Cultural Evolution of Human Cooperation" in: P.Hammerstein (ed.), *Genetic and Cultural Evolution of Cooperation,* MIT-Press.

[18] Roth, A.E. and I. Erev (1995), Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term, *Games and Economic Behavior 8*, 164-212.

[19] Traxler, C. (2005), "Social Norms, Voting and the Provision of Public Goods", mimeo, University of Munich.

[20] Vega-Redondo, F. (1996), "Evolution, Games and Economic Behaviour", Oxford University Press.

[21] Weibull, J. (1995), "Evolutionary Game Theory", Cambridge: MIT-Press.

[22] Wilson, D.S. and E.Sober (1994), "Re-Introducing Group Selection to the Human Behavioural Sciences", *Behavioral and Brain Science 17*, 585-654.

A Proof of Proposition 1

First note that whereas in the theory outlined in Section 2 there is a continuum of agents this is obviously not the case in the experiment. In the following we provide a proof of the proposition for the discrete case.[26] Denote thus the number of agents in group A (B) by n_A (n_B) and the total number of agents by n.

[26] The proof works analogously for the continuous case.

Furthermore note that agents in group B will always defect (it is a dominant strategy to do so in this group independently of the number of subjects in groups A and B). Assuming that all agents in group A cooperate, the payoff of an agent in group A from cooperating given that all agents in group A cooperate (denoted $\Pi_A(C|C)$) is given by

$$\Pi_A(C|C) = \left(1 - \frac{n_B}{n-1}x\right) a + \left(\frac{n_B}{n-1}x\right) b$$

and the payoff from defection is

$$\Pi_A(D|C) = \left(1 - \frac{n_B}{n-1}x\right) (c-\gamma) + \left(\frac{n_B}{n-1}x\right) d,$$

where $\frac{n_B}{n-1}x$ is the probability for an agent from group A to interact with an agent from group B.[27] An agent in group A has incentives to deviate from cooperation and to defect in group A whenever $\Pi_A(D|C) > \Pi_A(C|C)$ or, equivalently, whenever

$$\begin{aligned}\left(\frac{n-1-n_B}{n-1}\right) &\geq \frac{x(d-b)-(1-x)(a-(c-\gamma))}{x(d-(c-\gamma)-b+a)} \\ &= \frac{200x-50}{200x},\end{aligned} \tag{3}$$

where the last equality follows from substituting in the parameter values from our experiment. Only if $\left(\frac{n-1-n_B}{n-1}\right) = \frac{n_A-1}{n-1} \geq \frac{200x-50}{200x}$ an equilibrium where agents in group A cooperate can exist.

Analogously it can be shown that an equilibrium in which members of group A defect can exist only if

$$\frac{n_A-1}{n-1} < \frac{150-(1-x)200}{x}. \tag{4}$$

Now recall that the payoff of a group A member if all agents in group A cooperate is given by

$$\Pi_A(C|C) = \left(1 - \frac{n_B}{n-1}x\right) a + \left(\frac{n_B}{n-1}x\right) b.$$

If all group A members defect they receive

$$\Pi_A(D|D) = \left(1 - \frac{n_B}{n-1}x\right) (d-\gamma) + \left(\frac{n_B}{n-1}x\right) d.$$

Agents in group B always defect. If agents in group A cooperate they receive[28]

$$\Pi_B(D|C) = \left(\frac{n_A}{n-1}x\right) c + \left(1 - \frac{n_A}{n-1}x\right) d.$$

[27] Note that, for the matching probabilities, the number of *other* subjects in groups A and B matter (exclusive of the subject under consideration). Thus, in the discrete case, the equivalent to p_B in the matching probability is $\frac{n_B}{n-1}$.

[28] $\Pi_B(D|C)$ ($\Pi_B(D|D)$) denote the payoff of a group B member if all members of *group A* cooperate (defect).

If agents in group A defect their payoff is

$$\Pi_B(D|D) = d.$$

Now we are in the position to prove proposition 1:

Case (i) If $x < \frac{1}{4}$ agents in group A will always (independently of n_A) cooperate as can be read from (3) and (4). But note that if this is the case and $x < 1/4$, we have that $\Pi_A(C|C) > \Pi_B(D|C)$ and the dynamic equation implies that $\dot{p}_A > 0 \; \forall p_A \in [0, 1]$. All agents will thus end up in group A ($n_A^* = n$).

Case (ii) Now consider the interval $x \in [\frac{1}{4}, \frac{4}{7}]$. Note that in this case an equilibrium where agents in group A cooperate exists only if (3) holds (if there are sufficiently many agents in group A), whereas an equilibrium where agents in group A defect exists if and only if (4) holds. Furthermore, $\Pi_A(C|C) > \Pi_B(D|C)$ for high n_A and $\Pi_A(D|D) < \Pi_B(D|D) \; \forall n_A \neq 0$. Consequently in this parameter range both equilibria $n_A^* = n$ and $n_A^* = 0$ coexist.

Case (iii) Finally consider the interval $x > \frac{4}{7}$. Note that in this case $\Pi_A(C|C) < \Pi_B(D|C) \; \forall n_A$ and $\Pi_A(D|D) < \Pi_B(D|D) \; \forall n_A \neq 0$. Consequently, independently of whether agents in group A cooperate or defect (and independently of how many they are) agents in group B will always receive a higher payoff than agents in group A and thus, $\dot{p}_A < 0 \; \forall p_A \in [0, 1]$. Group B will proliferate. The unique equilibrium will have $n_A^* = 0$.

B Instructions Treatment $x = 0$

Welcome and Thanks for participating at this experiment. Please read these instructions carefully. They are identical for all the participants with whom you will interact during this experiment.

If you have any questions please raise your hand. One of the experimenters will come to you and answer your questions. From now on communication with other participants is forbidden. If you do not conform to these rules we are sorry to have to exclude you from the experiment. Please do also switch off your mobile phone at this moment.

For your participation you will receive 2,50 Euro. During the experiment you can earn more. How much depends on your behavior and the behavior of the other participants. During the experiment we will use ECU (Experimental Currency Units) and at the end we will pay you in Euros according to the exchange rate 1 Euro = 2500 ECU. All your decisions will be treated confidentially.

The Experiment

At the beginning of the experiment we will split you and the other participants equally into two groups - **group A and group B**. In each round of the experiment you play a game against a "representative member" either from group A or group B that we will call in the following your **interaction partner**.

To begin you will play at least 4 rounds as a member of the group that you have been assigned to at the beginning of the experiment. In each of these 4 rounds you play a game, that we will describe in the next section.

Starting with round 5 each round has two phases:

- **phase 1**: Each round some participants can decide whether to change groups or not. You can make this decision for the first time between round 5 and 8 and from then on every 4 rounds.

- **phase 2**: You play the game that we will describe in the next section.

The Experiment consists of 100 rounds.

The Game and the payments

Independently from which group (A or B) you are in, you play during the first 4 rounds and in the second phase of every following round the following game with a randomly selected interaction partner:

In each round you and your interaction partner can choose between two alternative, C and D. How much you **earn** in each round depends on **what you and your interaction partner have chosen** and in **which group you are**.

In the payment table below **your** actions and the payments are given in **red** and those of your interaction partner in blue. The table reads as follows:

- if both choose D, each gets 200 ECU (down right)

- if you choose D and your interaction partner C, you get 550 ECU and your interaction partner 50 ECU (down left)

- if you choose C and your interaction partner D, you get 50 ECU and your interaction partner 550 ECU (up right)

- if both choose C, each gets 400 ECU (up left)

		your interact.partner chooses	interact. partner chooses
		C	D
you choose	C	400 ECU 400 ECU	50 ECU 550 ECU
you choose	D	550 ECU 50 ECU	200 ECU 200 ECU

How does my group-membership impact my payment ?

Group membership impacts payments as follows:

- if you are in **group B**, you always get the payments described in the table above.

- if you are in **groupA**, you also get the payments described in the table above. In the case though that you chose **action D** and **your interaction partner in this round is also from group A**, 200 ECU will be deducted from your payments.

In addition your group membership impacts with which other participants you interact in the experiment. How exactly we will explain in the following section.

Who do I play with and how does this depend on my group-membership ?

In each round your interaction partner will be chosen randomly. You will interact exclusively with participants of the experiment, that in the round in question are in the same group (A or B) as you are. In other words, you always only interact with members of your own group. Obviously the composition of the groups can change during the course of the experiment as both you and the other members can change their group-membership in fixed intervals.

In each round, before the second phase, we will give you information about your group-membership and about which share of participants is currently in group A and B.

Your interaction partner

Your interaction partner in each round is not another participant of the experiment, but a "representative member" of the group in which you are at the moment. He chooses the actions C and D with probabilities that correspond to the shares with which the other members of your group have chosen C and D.

If you are the only member of your group, the behavior of your interaction partner will be simulated by the computer (**but only in this case**). In all other cases the behavior of your interaction partner depends **exclusively** on the **behavior of the other members** of your group.

These rules obviously are the same for all other participants of the experiment.

Example: You are in group A and consequently your interaction partner will also be from group A.

- if among the other members in group A 70% chose action C and 30% chose action D, your interaction partner will choose with probability 70% action C and with probability 30% action D.

-if all other members of group A have chosen action C, your interaction partner will choose action C with probability 100%.

Information you receive

In each round you get the following information

At the beginning of the second phase you are informed

- in which group you are
- what share of agents is in group A and B respectively.

After the second phase you are informed about

- which action you and your random interaction partner have chosen
- your payment.