Identities, Selection, and Contributions in a Public-goods Game

Gary Charness[1], *Ramón Cobo-Reyes*[2,3], *Natalia Jiménez*[2,4]

(1)University of California at Santa Barbara
(2) University of Granada, Globe
(3) University of Exeter, Business School
(4) Middlesex University, Business School

November 26, 2013

Abstract: The notions of one's social identity and group membership have recently become topics for economic theory and experiments, and recent research has shown the importance of identity in a wide array of economic environments. But predictions are unclear when there is some trade-off between one's identity (e.g., race, gender, handedness) and potential monetary considerations. We conduct a public-goods experiment in which we permit endogenous group-formation. In a 2x2 design, we vary whether people participate in a team-building exercise and whether some people receive an endowment twice as much as others receive. We find that when both identity and financial considerations are present, high-endowment participants are strongly attracted to each other, with one's word-task-group affiliation eclipsed by the opportunity to earn more. Nevertheless, the team-building exercise greatly increases the level of contribution whether or not one is linked to people from one's team-building exercise.

JEL Classification: A13, C91, C92, D03, J41

Keywords: Experiment, identity, team building, public goods contributions

Contact: Gary Charness, UCSB, 2127 North Hall, Santa Barbara, CA, 93106-9210, charness@econ.ucsb.edu. Ramón Cobo-Reyes, University of Granada, Campus de la Cartuja s/n, E-18071, Granada, Spain, ramonc@ugr.es. Natalia Jiménez, University of Granada, Campus de la Cartuja s/n, E-18071, Granada, Spain, natjj@ugr.es.

1. Introduction

Identity and group membership have become prominent issues in the economics literature in the past decade. Previously, economic theory has typically presumed that behavior is determined solely at the individual level. However, the seminal paper by Akerlof and Kranton (2000) highlights the notion that one's *identity*, or sense of self, can affect behavior and economic outcomes in a variety of environments. In their model, identity is considered in relation to social categories or groups, leading to identity-based utility; how a person perceives her own affiliations with these groups affects her choices in many areas such as labor disputes, discrimination, politics, and even criminal behavior. Furthermore, the interaction between or amongst people may well be affected by their perception of shared values or shared outcomes.

Also related to identity but in a social-network framework, Currarini, Jackson, and Pin (2009) develop an economic model of friendship networks, where people derive type-dependent utility from group (network) membership. The central idea is *homophily*, which means that similar types of people like to associate with each other.[1] This principle can lead to segregation, either due to preferences or the nature of the meeting process. The authors match data from friendship networks in an American high school to their model, finding considerable support. Typically, people wish to associate with their friends.

Yet this finding may not prevail in many economic environments. A critical issue is to understand which sense of identity becomes salient in which circumstances, as one typically "wears many hats". The question is especially interesting when identity and economic motivations pull in opposite directions. For example, a person may be female, a Republican, and Hispanic; which of these elements comes to the fore regarding the debate about immigration

[1] The folk wisdom is that "birds of a feather flock together", which goes back to at least the 17th century (see footnote 5 in Currarini, Jackson, and Pin, 2009).

policy in the U.S.? There has been little work done in this area.[2] In this paper, we explore two forms of identity in an experimental public-goods game with endogenous group formation, one involving a sense of group identity from participating in a group task and the other involving different levels of endowment in the game. According to which criterion will the segregation resulting from endogenous group formation occur? To the best of our knowledge, there is no study in experimental economics that analyzes this issue.

Revisiting the notion of homophily, we could find that people like to associate with other people with a similar type, but that they share certain features with some people and different features with other people. For example, consider the case of Catalonia, a region in Spain. On one hand, there is the identity conflict between nationalists and non-nationalists, while on the other hand there are different political ideologies. Will people vote for the party that matches their identity or they will vote for the party that matches their political ideology?[3] For another example, consider the case of India. On one hand, there is the religious conflict between Muslims and Hindus, and on the other hand, there are rich and poor people. Do people prefer to associate, one who shares the same religion or shares the same economic status?[4] A final example is a work place in which workers are not completely homogeneous but instead differ across several features such as ability, motivation and productivity (Hamilton et al., 2003), as well as with respect to personal features such as hobbies or political views. To put this into an academic environment, does a recruiting committee choose to hire someone with whom it would

[2] There is one contemporaneous study in experimental economics (Chen, Li, Liu, and Smith, 2010) that considers the issue of multi-dimensional identities. They find that depending on the natural identity they prime in their experiment, the effect on cooperation might be either negative (ethnicity in a minimum effort game) or positive (school identity in a prisoner's dilemma game).

[3] Due to this conflict of interests, there are four political parties in Catalonia: a non-nationalist/conservative party, a non-nationalist/liberal party, a nationalist/conservative party and a nationalist/liberal party.

[4] A reviewer points out that there may very well be correlations between different dimensions of identity, as with wealth and religion views or between political views and degrees of nationalistic fervor. Nevertheless, to the extent that this correlation is imperfect, there is scope for conflicting identities to play a role.

be great to go to lunch or someone who will publish well but has no social graces? For proximity, we are going to focus our attention on this academic example to motivate our design. In the line of the homophily hypotheses, we aim to explore whether academics prefer to associate with more skilful co-authors or with researchers with whom they have a good relationship. Which one of these two identities will be more salient? An academic department may be considered as a team work in which the research ideas, publications, research projects of one of its members may have a positive effect for the other members. Thus, a Public Goods Game with endogenous formation may be appropriate to capture this idea. The first dimension of the identity is the ability. Although there are several forms to capture the concept of research ability, one of the simplest one in a PGG is to allocate a higher endowment to the more skilful individuals. In this way, the skilful individuals have the potential to contribute more in the department (for instance in a research article collaboration) but they can decide not to do so. The second dimension of the identity is the fact of having a relationship. One way to proceed will be to recruit groups of friends. Nevertheless, this procedure will entail several issues [5] which will diminish the control of the experimenter. We opted instead for a endogenous minimal relationship (subjects perform a task prior to the PGG so they may build a certain relationship among them) being aware that we were scarifying significance in favor of control and transparency in the interpretation of the results.

This paper contributes to bringing economic theory into alignment with other social sciences (e.g., anthropology, political science, psychology, and sociology), where identity is considered a central concept, by considering the effects of identity in a strategic environment. One insight from the theory of social identity is that one derives self-esteem or ego utility

[5] Among them: problems to conduct a session if an insufficient number of friends shows up, lack of anonymity, friends may agree to share experimental payoffs which will modify the strategic behavior of PGG.

(Koszegi, 2006) from group membership. People tend to conform to the stereotypes that are salient with respect to the relevant group (see, e.g., Benjamin, Choi, and Strickland, 2010). An issue from the beginning of the literature is the difference in how ingroup and outgroup members are treated. In the Akerlof and Kranton (2000) neo-classical model of identity, deviations from what one believes are the relevant group norms leads to a loss of utility. They apply their general model to a variety of economic environments, including poverty, social exclusion and discrimination, and even contract theory.

We take the previous examples of homophily as a starting point for our experimental design. In our experiments, eight participants play a public-goods game with possible endogenous changes to the initial setting consisting of two groups of four people; contributions have more value with increasing group size (up to a group size of four). As is standard in such games, one earns less individually by contributing to the public good, but the social optimum occurs when everyone contributes.

We vary two factors in a 2x2 design. One factor is whether there is a team-building exercise prior to the game. In this exercise, people are assigned to 4-person groups; in sessions without this team-building exercise, people performed this task individually. In each case, there is a monetary reward for forming a sufficient number of words from a series of letters; we set the threshold low enough so that the reward was received in every case. The second factor concerns the endowments assigned to the players. In one treatment everyone received the same endowment, while in another treatment half of the participants received twice this amount. Thus, we pit these two aspects against each other and observe the resulting behavior.

Our results are intriguing. Despite the fairly minimal team-building activity, we find a striking benefit from having word-task groups, as the contribution rate is greatly increased.

While endogenous group-formation *per se* is effective at sustaining a fairly high contribution rate, this rate is greatly enhanced (up to well over 90 percent of the endowment) by first having a group word-task. Thus, team building offers promise for enhancing the provision of public goods.

We would like to point out the contrast between this result and those in other experiments involving identity/membership. Those studies find support for favoritism towards their own group members (except results of Pan and Houser (2013) with the cooperative production system), whereas we find a positive effect in contributions towards "non-members". We do not find that the team-building exercise leads to group formation on the basis of one's word-task-group identity. Instead, it seems that the endowment type is a much more central consideration in the group-formation process. While there is some evidence in the data of sorting by word-task group when everyone receives the same endowment, there is much more sorting on the basis of endowment types when these are present. However, we find that when there is heterogeneity, the word-identity facilitates segregation although in an indirect manner, that is, it facilitates endowment-type segregation.

Since it is something of a puzzle that the group-word-task exercise increases contribution rates without people sorting on the basis of membership in the same word-task group, we conduct two extra treatments to see if it is the knowledge that everyone (not just one's own word-task group) participated in a team-building exercise that is driving the results. In this design, only half of the participants perform the group word task, while the other half do the word task individually. Our results indeed suggest that this awareness affected behavior. Contributions were much lower than when all people were in word-task groups, with the difference driven by lower contributions for those in the word-task group. Thus, it appears that

in this environment it is not sufficient for one to have the good feeling from having been engaged in a team-building task; it is also important to know that all of the participants have been in these groups. This indicates that there may be a positive spillover from arranging team-building tasks for people, when these can include the preponderance of the parties in an organization or society.

In the remainder of the paper, we present a review of the relevant literature in section 2, and describe our experimental design and implementation in section 3. Our experimental results are given in section 4, and discussed in section 5. We conclude in section 6.

2. Related previous work

In the past decade economists have begun to do research on the topics of identity, group membership, and homophily. Nevertheless, the social-psychology literature on social identity is quite extensive, and the notion typically refers to the sense of self that is derived in part from the groups with which one is affiliated. Tajfel, Billing, Bundy, and Flament (1971) established that there are strong effects of induced identity, even with "minimal groups" formed only on the basis of participants' preferences for one of two painters. This weak form of identity led to substantial behavioral effects, where people favored their own group members in the distribution of real rewards, with own material payoffs unaffected by the distribution. Maximizing the profit for one's own group and maximizing the difference between the ingroup and outgroup payoffs substantially affected the distribution of rewards.

New studies in experimental economics also confirm that group membership indeed can have a strong effect on behavior. The first paper on the effects of group membership on behavior is that of Charness, Rigotti, and Rustichini (2007).[6] Previous work on the ingroup-outgroup

[6] Brown-Kruse and Hummels (1993) and Cadsby and Maynes (1998) introduce a previous questionnaire where the information is shared with the rest of the subjects so as to create a sense of group membership among individuals. Kirkwood and Solow (2002) conduct an experiment with groups comprised of members of an already-existing

almost exclusively had considered only pure allocation tasks, rather than strategic ones. Thus, an open question has been whether considerations of identity affect behavior in games. It was found that having minimal groups (i.e., groups being formed by pure random assignment) is not sufficient to affect behavior in two games (Battle of the Sexes and Prisoner's Dilemma). However, when group membership is made more salient by having shared payoffs and having members of one's own group observing one's choice, there are strong effects. Players become more aggressive in their strategies when they are under observation by their peers, so that social identity clearly plays a role in their choice behavior.

Chen and Li (2009) study the effect of induced group identity on social preferences in allocation tasks and simple games borrowed from Charness and Rabin (2002). In addition, they provide an excellent literature review on the topic of social identity. They employ the minimal-group paradigm used by Tajfel, Billing, Bundy, and Flament (1971), where people are divided into two groups based on their preference for either Klee or Kandinsky. Participants offer more to ingroup members than to outgroup members whether the allocator has a higher material payoff or a lower material payoff than the recipient; in general, one is more likely to display social-welfare-maximizing preferences (Charness and Rabin, 2002) when paired with another person from one's group. In addition, in simple binary-response games, people are more likely to reward ingroup members for good behavior, and are also less likely to punish ingroup members

group, the Iowa Marching Band. They find that the stronger the social identity, the higher the contributions in a public-goods game. Croson and Marks (2003) use members of fraternities and sororities as experimental participants. They show that creating a sense of group identity among women increases efficiency and equity in public goods games, whereas this group identity among men decreases these measures. See also Charness and Rustichini (2011), where group membership has divergent effects on males and females in the Prisoner's Dilemma. Chen, Li, Liu and Shih (2013) find that depending on the natural identity they make salient in their experiment, the effect on cooperation might be either negative (ethnicity in a minimum effort game) or positive (school identity in a prisoner's dilemma game).

for misbehavior.[7]

Pan and Houser (2013) set up an experiment in which there was endogenous formation of cultural groups and in which they use different production settings in order to diminish the membership effect. They find that a cooperative production environment prevents favoritism towards in-group members while an independent production task foments discrimination towards out-group members.

Currarini, Jackson, and Pin (2009) implicitly extend the notion of group membership to friendship networks, where people derive utility from membership in the network; all else equal, people prefer to associate with similar types.[8] As with the Akerlof and Kranton (2000) model, this can lead to segregation and discrimination. Using the high-school data on social networks, they find relative homophily, which means that an individual in a larger network is more likely to form a friendship with (bring into the network) someone from their ingroup than an individual in a smaller network; indeed, people in larger groups form more friendships. An additional and interesting observation is that people in middle-sized groups are likely to "inbreed" (form more same-type friendships than the ratio in the population) than are people in very small or very large networks. To explain these patterns of behavior, preferences must depend on more than the number of one's friends. In particular, these preferences must be sensitive to types and inbreeding appears to be the result of a bias in the process that favors meeting one's own type.

There are a number of studies regarding endogenous group-formation (assortative matching) in the public-goods game.[9] The idea behind this approach is to enable would-be

[7] Hargreaves Heap and Zizzo (2009) also find experimental evidence of discrimination against outsiders in an investment ("trust") game.

[8] Of course, an invitation to join a network might signal that the people in the network are not similar to one's self, as is seen in the Groucho Marx statement that "I don't want to belong to any club that will accept people like me as a member." This is type-dependent utility in the Currarini, Jackson, and Pin (2009) model.

[9] For a detailed survey, see Chaudhuri (2011). Other related papers on endogenous group-formation include (but are not limited to) Riedl and Ule (2002), Cinyabuguma, Page, and Putterman (2005), Page, Putterman, and Unel (2005),

cooperators to avoid free riders and join with like-minded individuals. Ehrhart and Keser (1999) were the first to consider a public-goods game with endogenous group-formation. Nine participants were randomly placed into three initial groups and played the game. Each person was then told the sizes and average contributions for each group, and could unilaterally decide, at a fixed cost, to switch groups or to form a new (one-person) group. However, although there was considerable movement, this device had limited success. Without exclusion or an entry restriction, the pattern is one of more cooperative participants being on the run from free riders, who constantly chase them.

The study that is closest to the design of our public-goods game is Charness and Yang (2009). As in Erhard and Keser (1999), nine participants were randomly placed into three initial groups and played the game. Each person knew about the behavior of each group (and its size) and the individual members of her own group, and could unilaterally decide to exit, to exclude individuals from the group, or to merge with other groups or with singles. The value of a contribution was increasing in terms of the group size (although the per-capita return was decreasing in group size), so there was an incentive to form large groups and to contribute to avoid exclusion. This mechanism was very successful, with resulting contribution rates over 90 percent. Our mechanism is similar, except that a group could only add one person in a period.[10]

Brosig, Margreiter, and Weimann (2005), Croson, Fatás, and Neugebauer (2006), Gunnthorsdöttir, Houser, McCabe, (2007), Ahn, Isaac, and Salmon (2009), and Maier-Rigaud, Martinsson, and Staffiero (2010).

[10] Yang, Xu and Tang (2009) report an experiment that applies a simplified version of the current group-formation mechanism to an endogenous-size large group coordination problem. This design works without exclusion, and without allowing compound mergers in one period.

Team building, induced identity

There are two primary techniques that have been employed in research on social identity. The first is to make salient one's natural identity through subtle priming; the second is to induce artificial identities. Our approach involves the latter technique, using a team-building task. Team building has a long history in social psychology and organizational research, going back at least to Campbell (1958). Team building affects social identity to the extent that there is a sense of shared purpose or even solidarity.[11] Team building can consist of either sedentary joint problem-solving or activities such as ropes (or challenge) courses, where people must work together to achieve a physical goal. Buller (1986, p. 147) states: "Team building is one of the most popular interventions in organizational development."

One purpose of team building is to improve the effectiveness of work teams within organizations. For example, Eden (1986) reports the results of a field quasi-experiment with combat units in Israel. Seven experimental command teams underwent an intensive three-day workshop that included team-building activities such as conflict resolution, problem solving, role negotiation, and role definition. The results indicate that team building had significant positive effects on organizational concerns such as teamwork and conflict management. Carron and Spink (1993) examined whether cohesion could be enhanced in fitness classes through a psychological intervention program focusing on team-building concepts. The team-building strategies were implemented in classes in the experimental condition. The experimental (team-building) and control conditions could be differentiated on the basis of their cohesiveness and the team-building program significantly enhanced individual satisfaction. On the other hand, Buller and Bell (1986) conducted a field experiment with 53 hard-rock miners in an underground metal

[11] Kramer and Brewer (1984) and Brewer and Kramer (1986) find that ingroup cooperation is enhanced with salient group identity; Charness, Rigotti, and Rustichini (2007) also find this effect with ingroup-ingroup matching (as opposed to ingroup-outgroup matching) in the Prisoner's Dilemma.

mine examined the effects of two interventions, team building and goal setting, on miners' productivity and strategy development; the results were inconclusive.

To the best of our knowledge, the only research on team building in experimental economics is Eckel and Grossman (2005). They confront the issue of diversity in the workplace and test whether team building can ameliorate problems of group effectiveness, here measured by contributions in a team-production (public-goods game) task. The experiment consists of six treatments, including a control; the treatments varied the degree of group identification and the interaction amongst team members (in groups of five). In one treatment, teams received color tags, with one person from each team seated at a table and had no further interaction; a second treatment added quiz questions, with the people with the highest scores allocated to group one, etc. Their third treatment added an unpaid group task for which team members needed to exchange pieces to complete a puzzle; a fourth treatment added a nominal financial incentive. Finally, the fifth treatment introduced tournament competition between groups.[12]

The experimental results indicate a highly-significant difference of about 15 percentage points in contributions across (pooled) weak and strong identity (the latter three) treatments. Unlike the typical results in the public-goods game, contributions are rather stable over time. In all, it is found that "actions designed to enhance team identity contribute to higher levels of team cooperation." Working together on a team-building task enhanced cooperative tendencies even with little or no financial incentive for performing the task. Nevertheless, exogenous group formation and homogeneous endowments are considered in this study.

[12] See also Bornstein, Gneezy, and Nagel (2002) for the effects of intergroup tournaments in the minimum-effort game.

3. Experimental design

The experiment was conducted at the University of Granada. Participants in the experiment were students in economics, business administration, tourism, and market research who had not participated in previous experiments. We run six different treatments (which will be explained in more detail below) with a total of 22 sessions with 8 subjects in each session. There were 3 sessions in each treatment except that there were 5 sessions in each of Treatments 3 and 4. The average earnings were 22.93€. All instructions can be found in the online appendix.

The experiment was comprised of two different activities. The first activity involved creating words from a number of letters during 5 minutes. Once participants finished with this activity, they received the instructions for the second activity that consisted of a public-goods game (PGG hereafter) with endogenous group formation.[13] Treatments differed in two features: The word task was performed individually (NG) or in groups (G) and there were one (NT) or two types (T) of endowment in the PGG. We were attempting to induce salient group identity with an incentivized group word-task. The team-building task is designed to create a sense of *bonhomie* amongst the parties in the group involved with the task, a proxy for friendship. Having different endowments is a proxy for different levels of wealth or talent.

In Treatment 1 (hereafter the "*NoGrNoTy*" treatment), subjects participated individually in the word-building task. They would be paid 3€ if they could find at least six correct words and 0.5 extra Euros for each additional word beyond six.[14] People were then placed into two initial groups of four people and played a PGG. Each person was endowed with 25 tokens that could be allocated between a private and a public account. The social value of an allocation to the public account depended on the group size, as shown in Table 1.

[13] The mechanism we used is a simplified version of the one proposed by Charness and Yang (2009).
[14] We chose this calibration so that nearly everyone would succeed at receiving the bonus. In fact, only two people of 64 did not receive the bonus.

Table 1: Group returns and MPCR in all treatments, by group size

Group size	Group return	MPCR
1	1.000	1.000
2	1.250	0.625
3	1.500	0.500
4	1.750	0.438

After each period, individuals learned about the contribution of each individual (by identification number) in one's current group in that period, as well as the size and the average contribution in other groups.[15] Then, we proceeded with the group formation. We divided this group formation into three stages involving expulsion, voluntary exit, and re-formation.[16]

i) Stage 1: People voted to keep or exclude each and every other person in the group, if more than 50 percent of the voters chose to exclude an individual, that individual was expelled from the group.

ii) Stage 2: Each subject who was not expelled from the group could unilaterally decide whether to exit the group.

At this point there were a number of groups and possibly some number of single participants not in multi-person groups. We ranked the groups remaining by size, with random ordering for groups with the same size.

iii) Stage 3: Each person assigned a value from {0,1,2,3} to each of the single individuals, where 0 meant that the voter wished to not invite the single into the group and higher

[15] In the GrTy and NoGrTy treatments (two different endowments in the PGG), individuals had information also about the type of the other members of the group, but not about types in the other groups. We chose this approach in part to avoid overloading participants with information, as this often tends to decrease comprehension and performance. In any event, it was also fairly easy in practice to make inferences about the distribution of types in the other groups, given knowledge of the types in one's own group and average contributions for the other group or groups.

[16] Expulsion from a group is not uncommon. For example, consider social clubs, political parties, recreational sports and even co-authorship. Frequently an individual is excluded due to a perceived shift in beliefs or due to poor performance, and is often replaced with a new person.

values reflected more favorable opinions.[17] Singles were ranked according to the sum of the values received, with random ordering for singles with the same value.[18] The largest group was considered first; the single with the highest score received an offer to join this group, and chose whether to accept the invitation.[19] If she declined, then the second candidate in the ranking could receive an invitation. This process was repeated until there were no more candidates.

Once the first group had finished, the second group in order of size faced the same choices from the available candidates. This process continued until all groups had gone through this stage or until there were no more singles remaining in the society; note that all groups (including singles) could enter in this stage, so singles had the chance to join other singles. There were two restrictions for incorporating new subjects to a group. The first was that the maximum group size was 4, so groups of four members did not have the chance of incorporating a new single member; the second was that groups were allowed to incorporate at most one single in each period.

Once the third stage had finished, new groups would play again the PGG as explained above. This process would be repeated for two segments of 12 periods each. After the first 12 periods, subjects IDs were redrawn. We chose to have a re-start, as an individual might get locked into a situation that is difficult to escape during a segment, but instead would receive a fresh start in the second segment.

In Treatment 2 (*GrNoTy* hereafter) the structure of the game was the same as in the NoGrNoTy treatment. The only difference was that participants were randomly assigned into

[17] The people in the voting group observe the IDs, allocations to the group account in the last period, the type of each person who is single (just for GrTy and NoGrTy) and could distinguish whether the person was single at the beginning of the round, became single during stage 1, or became single in stage 2.

[18] Note that, in order to receive an invitation from a group, a single individual must receive more positive votes than zero votes from the group members.

[19] The single person who was ranked first would observe the size and the contribution per capita to the group account in the previous period for each of the groups and other singles. She would observe also the id #, the size and the per capita contribution of the group extending the invitation. Singles did not know the distribution of types nor the word-task members in the group. They could infer the number of highs and lows from the average contribution but they did not know the exact composition of the group.

two groups of four subjects upon arrival to the lab.[20] They participated in the word-building task in their group. The total earnings of the group were divided equally, independently of the number of words each subject had proposed. A group is paid 12€ if its members could find at least 12 correct words and 1 extra Euro for each additional word beyond 12.[21] Thus, in the GrNoTy Treatment, participants had the same information as participants in the NoGrNoTy Treatment plus some additional information about the word-task group. In this treatment, IDs numbers appeared in green and orange depending on each person's word-task group.

Treatment 3 (*NoGrTy*, hereafter) differed in only way from the NoGrNoTy. Half of participants in the PGG were endowed randomly with 25 tokens (the same as in the NoGrNoTy), while the other participants were endowed with 50 tokens. That is, we had a society with two types of participants: "high types", with an initial endowment of 50 tokens and "low types", with an initial endowment of 25 tokens (these labels were used neither in the instructions nor at any point during the experiment).[22] Participants were placed into two initial groups of four people, two "high types" and two "low types" in each group and then played a PGG. Participants had the same information as in the NoGrNoTy, as well as information about the types of all the other participants with whom they interacted. Note that one's type never changed. After the re-start (in period 13), ID numbers were randomly re-assigned but the type was the same. This was common information.[23]

[20] When participants arrived to the experiment, they were asked to pick one number from an opaque bag. Subjects who picked an even number were assigned to the "orange" group and subjects who picked an odd number were assigned to the "green" group.

[21] Once again, we chose this calibration so that nearly everyone would succeed at receiving the bonus. In fact, all groups received the bonus. We would like to point out that one's success on a word task *per se* is essentially the same across NG and G sets of treatments and nearly everyone achieves the threshold. The only difference is that people achieve success in the word-task in groups or individually.

[22] In the instructions we label "Type 1" and "Type 2" for endowments 50 and 25, respectively.

[23] We note that the total efficiency of contributions is independent of the distribution of types across groups, so that there is no efficiency reason per se to form same-type groups. However, there are practical deterrents that inhibit the

Treatment 4 (hereafter *GrTy*) included both the group word-task and two types of participants, with the concomitant information.

After observing the results in Treatments 1-4, we designed Treatments 5 and 6. In Treatment 5 (*MixNoTy*), half of the participants performed the word-task in groups and the other half individually. All subjects were endowed with 25 tokens in the PGG. Treatment 6 (*MixTy*) only differs from Treatment 5 in that half of the people receive endowments of 25 tokens and the other half receive endowments of 50.

4. Results

This section is structured as follows. We first analyze subjects' behavior regarding their contribution to the public-good and then study patterns and determinants of exclusion. Next, we consider exit behavior and we focus on the decision of whether or not to invite a single to join with one's group. Then, we study the density and segregation in the society. We also analyze the stability of the networks and the group size. Finally, we focus on the behavior in mixed groups.

4.1 Contributions to the Public Good

Table 2 shows the average (non-single) contribution for each type and for each treatment, both in absolute and relative terms. We also calculate the average earnings for each endowment type. The online appendix provides these data on the session level.

We observe that contributions are higher in every case when there are word-task groups than when there are not. When there are no types, people in the GrNoTy contributed 92.9 percent of their endowment, compared to 73.9 percent for those people in the NoGrNoTy. A Wilcoxon-Mann-Whitney ranksum test using each individual's (non-single) average contribution

contributions of high types in mixed groups. For example, a high type in a group with three low types will always lose money by contributing more than 41 tokens, even if the low types contributed the maximum possible.

as one observation gives $Z = 4.210$, $p = 0.000$, while a more conservative and less powerful test using session-level data gives $p = 0.100$ (one-tailed test).[24] [25]

Table 2: Average contributions and earnings, by treatment and type

Treatment/type	Contribution	Percentage	Period earnings
NoGrNoTy	18.47	0.739	32.68
GrNoTy	23.22	0.929	40.41
NoGrTy-high	29.28	0.586	51.21
NoGrTy-low	20.30	0.812	38.36
NoGrTy-overall	24.79	0.699	44.79
GrTy-high	43.96	0.879	70.17
GrTy-low	23.65	0.946	44.75
GrTy-overall	34.80	0.912	57.46

Contributions by singles are not included in calculation contributions. We calculate numbers for NoGrTy-overall and GrTy-overall by giving equal weight to the figures for high types and low types.

When there are types, low types in the NoGrTy contributed 81.2 percent of their endowment, compared to 94.6 percent in the GrTy; the ranksum test using each individual's (non-single) average contribution as one observation gives $Z = 4.031$, $p = 0.000$, while the session-level test gives $p = 0.016$ (one-tailed test).[26] High types in the NoGrTy contributed 58.6 percent of their endowment, compared to 87.9 percent in the GrTy; the ranksum test using each individual's (non-single) average contribution as one observation gives $Z = 5.195$, $p = 0.000$, while the session-level test gives $p = 0.004$ (one-tailed test).

[24] If one runs this test on Stata, the result is $Z = -1.528$ and $p = 0.063$, one-tailed test. However, a simple argument demonstrates otherwise. The sum of the ranks of the three GrNoTy sessions is seven. A sum this small can occur in only two ways, with ranks of either (1^{st}, 2^{nd}, 3^{rd}) or (1^{st}, 2^{nd}, 4^{th}). The number of possible orders is (6!)/(3!3!) = 20, so the one-tailed probability is 0.1. Siegel and Castellan (1988) of course also gives the latter value.

[25] With the aim of analyzing in a more transparent manner the effect of the word task in groups on first-period contributions (eliminating learning and other uncontrolled effects), we conduct the same test (NoGrNoTy vs GrNoTy) but only for the first period and we obtain $Z = -1.602$ and $p = 0.055$, one-tailed test. This result, although less significant, yields the same qualitative result as when considering an individual average of all the periods.

[26] If we conduct the same test (NoGrTy versus GrTy) using only data from the first period, we obtain $Z = -6.145$ and $p = 0.000$, one-tailed test.

Next, we explore the effect of endowment heterogeneity in cooperation. People in the NoGrNoTy contributed 73.9 percent of their endowment, compared to 69.9 percent in the NoGrTy; the ranksum test using each individual's (non-single) average contribution as one observation gives $Z = 0.569$, $p = 0.570$ (two-tailed test) and the session-level test gives $p = 0.882$ (two-tailed test). However, the difference approaches significance in the presence of the group word-task $Z = 1.412$, $p = 0.079$, one-tailed test with only one observation for each individual). Thus, we have:

Result 1: *The previous group word-task significantly increases contributions to the public good in both homogeneous and heterogeneous environments. Although endowment heterogeneity has no effect in cooperation without a group word task, this effect becomes slightly negative when the group word-task is present.*

We now conduct regression analysis to study the main determinants of contributions to the public-good. Table 3 shows that the average contribution by the other group members in the previous period has a positive and highly statistically-significant effect on the amount contributed to the public-good even though our design allows for changes in the composition of the groups. An intuition for this might be that, even if one subject changes her group from one period to another, the fact that in the previous period the group contributed a high amount may create some kind of positive tendency and this subject will increase her contribution in the next period even playing with a different group. The average contribution in a given period may also be interpreted as a reference point if subjects intend not to be excluded in the next round.

[Table 3 about here]

The Dummy t1t2 term captures the effect of the group word-task on contribution rates for the case in which there are no types (in specification (1), the value of the dummy is 1 when the

data are from Treatment 1 and 0 for data from Treatment 2). The Dummy t3t4 term captures the effect of the group word-task on contribution rates for the case in which there are types (in specifications (2) - (4), the value of the dummy is 1when the data are from Treatment 3 and 0 for data from Treatment 4). Thus, we see that contributions are higher when there has been a group word-task regardless of whether or not there are high and low types.

Group size also has a positive and significant effect on the level of contributions both when there are types and when there are no types; however, this is only true for high types when we consider percentage contributions. The explanation for the positive effect of the group size could be that, given the structure of the game, the return from the public-good is increasing with the size of the group, so subjects may have stronger incentives to contribute when they know that the return will be larger.

An individual who is a high type also contributes more when in a group where the majority of other people are high types; however, low types contribute a high amount even when in a group with high types. This result could be related to high types contributing less when they are in groups with low types; a high type is reluctant to contribute more than 25 in this case, except to try to merge with other high types. This reflects the notion that one's perception of fairness is self-serving and this often causes a tension between parties with differing interests.

As we have shown above, the fact of having the previous word task increases significantly the contribution levels. From Table 3, we may conclude that the increment in contribution is due to the word-task *per se*, not to the fact of being paired in the PGG with the partners from the former word-task group. The coefficient for being in the same word-task group as the majority of the contribution group is insignificant in all specifications.

4.2 Exclusion

In this section we analyze the main factors determining the probability that an individual is excluded from a group. On the whole, there is not a great deal of exclusion in any of our treatments. The exclusion rates were 12.47 percent, 2.75 percent, 10.69 percent, and 10.73 percent in Treatments 1-4, respectively; there is a relatively high rate of exclusion at the beginning of a 12-period segment (recall that new groups were formed after 12 periods) and this rate diminishes steadily over the course of the segment. Once again, we see very little movement in Treatment 2, where there is type homogeneity and previous successful word-task-group experience. The very high contribution rate of 93 percent, presumably engendered by the experience of a successful word-task group, actually makes it unnecessary to seek out others of one's word-task group (given the absence of multiple types). Nevertheless, the possibility of being excluded may rein in selfish behavior, as this rate depends greatly on the contribution the potential outcast made. The exclusion rate was 80.9 percent when this contribution was 10 or less, 26.7 percent when the contribution was between 10 and 20, and 9.5 percent when the contribution was at least 20. This plays a part in the result that people contribute at least 80 percent in 76.9 percent of the cases. There is a floor, however, at least in the treatments in which there are high and low types. In that case, low types who contributed all 25 units were excluded 4.69 percent of the time (however, not once in 367 cases was a high type who contributed at least 40 ever excluded).

Table 4 provides a regression analysis of the determinants of exclusion. We see that coefficient on the difference in contribution between the group and the player is positive and highly statistically significant. That means that the more an individual deviates (by contributing less) from the average contribution of the group, the more likely that she will be excluded. Also, a higher maximum average contribution in the other groups leads to a greater chance of

exclusion. Thus, it seems that participants voting on exclusion also take into account how other groups are performing. Specifications (1) and (2) also respectively show that there is significantly more exclusion in the NoGrNoTy treatment than in the GrNoTy treatment, and significantly more exclusion in the NoGrTy treatment than in the GrTy treatment. Thus, there is something about having a group word-task that induces a lower rate of exclusion, even accounting for the other factors mentioned above. Specification (3) shows that, unsurprisingly, high types are excluded significantly less frequently than low types; it also shows that one is less likely to be excluded if one is the same type as the majority of the other members of the group.

[Table 4 about here]

Result 2: *There is more exclusion when there are two endowment types, with high types less likely to be excluded; there is less exclusion when people have participated in a group word-task. The level of one's contribution, the disparity between this contribution and the group average, and the maximum contribution observed in other groups are important factors, all in the expected directions.*

4.3 Exit

We now investigate the main factors explaining the probability that one chooses to remain in one's group. In fact, exit is somewhat uncommon in our treatments. The exit rates were 4.80 percent, 1.01 percent, 7.65 percent, and 9.48 percent in Treatments 1-4, respectively, so that exit is more common when there are types. The rate for high types leaving their groups is higher than for low types, as the respective exit rates are 9.95 percent and 5.28 percent in Treatment 3, and 10.95 percent and 7.57 percent in Treatment 4; presumably high types are more likely to exit in order to seek out other high types.

[Table 5 about here]

We present an econometric analysis in Table 5. First, the higher one's own contribution in relation to the group average, the more likely one is to exit. Here the maximum contribution of the other groups is only significant in Treatments 3 and 4, and there is no difference in exit rates between Treatments 1 and 2, or between Treatments 3 and 4. Specification (3) confirms that high types are significantly more likely to exit; it also shows that high types (but not low types) are less likely to exit when the majority of the other group members are high types.

Result 3: *The greater the difference between own contribution and one's group's average contribution, the more likely one is to exit from the group. High types are more likely to exit than low types, although not when most of the people in the group are high types.*

4.4 Voting on singles

This section analyzes how groups vote in order to add singles. We attempt to identify the main variables that influence groups' decisions. Everyone not in a 4-person group voted on all of the single participants who are available. Singles almost invariably voted to join a group when invited (350 of the 407 times, a rate of 86 percent). For this analysis, we employ a logit model clustered by the single participants who are invited to join a group.[27] We present these regression results in Table 6.

Specifications (1) and (2) address the contribution of the single being voted upon in Treatments 1 and 2. An increase in contribution has a significantly positive effect on positive votes whether it is in absolute or relative terms. Moving to Treatments 3 and 4, specification (3) offers a very similar result in terms of the percentage contribution, and also includes a dummy that indicates whether the vote is by a majority-high-type group in relation to a candidate high-

[27] The reason why we do not use a random-effects logit model as in the previous subsections is as follows. The number of data is very reduced in each period due to relatively small number of changes in the groups (especially in the G-treatments). Therefore, the corresponding panel data is extremely unbalanced.

type single; the coefficient of this dummy is positive and highly significant, so high-type groups favored high-type singles. Specification (4) looks at the same issue for majority-low-type groups and these also prefer high-type candidates. Everyone is attracted by the possibility of large contributions, even when the actual contribution is taken into account.

In specification (5), we include a general term for high-type singles and a term that separately considers whether the single is the same type as the majority of the group of the voter. Both terms are statistically significant, but it matters much more if the candidate single is a high type than if she is the same type per se as that of the majority of one's group. Finally, specification (6) indicates that people in the GrNoTy treatments are significantly more likely to vote for a single from the same word-task group, but that this effect is not significant in the GrTy treatment (this is due to the fact that the variable "Single and majority in same word-task group" is insignificant when alone but it is positive and significant when interacted with the homogeneous endowment treatment dummy, that is, Dummyt2t4).[28] This suggests that there is some affinity to one's word-task group, but that this is entirely crowded out when there are different contribution types.

[Table 6 about here]

Result 4: *The contribution level of the available single and its relation to the group average are important determinants of whether the single is invited to join a group. High types are more likely to be invited to join. Singles who were in the same word-task group as the majority of the contribution group members are more likely to be invited when there is only one endowment type, but this effect vanishes when there are types.*

[28] We believe that the explanatory variable "Single and majority in same word-task group" (and analogously "Single's type same as majority of group") captures the favoritism of the voting group towards the same word-task group (same type) members. The intuition is the following. As we have controlled for other variables (single's contribution, single's type, single has been excluded, etc), if the single is the most popular candidate we infer that this is due to the support of the members of the voting group who were initially in the same word-task group.

4.5 Density and segregation

As the result of all the previous stages, groups change over time; we study whether groups segregate or not. It turns out that there is much more sorting on the basis of endowment types than on the basis of word-task groups. We find that 89 percent of the links in the final six periods (of 24) in the treatment with a group word-task and two endowment types are between participants with the same endowment type; this compares to 66 percent without the word-task groups but with different endowment types.

We first consider the case of groups with "high" and "low" types. To analyze the segregation in a group, we compute the *density*. *Density of high types* is defined (in the spirit of Jackson, 2008) as the number of links between "high types" over the maximum number of possible links between "high types".[29,30] In fact, the density is higher in the GrTy treatment (0.641 and 0.750 for all 24 periods and the last 12 periods, respectively) than in the NoGrTy treatment (0.447 and 0.469). Session-level Wilcoxon rank-sum tests show that this difference is significant at $p = 0.008$ (two-tailed test).

Next, we explore whether the group word-task has an effect on how people segregate on the basis of the endowment type. A simple but intuitive way to do this would be to compare across the G-treatments and NG-treatments the number of times in which there was no segregation (on the basis of endowment types) at all at the end of each segment. In the NoGrTy treatment this happened 50% (5 out of 10) of the times while only 20% (2 out of 10) of the time in the GrTy treatment.[31] A different way to account for this segregation is to use the number of periods the society remained segregated (all same-endowment groups) in each segment. On

[29] In our setting, all links are undirected.
[30] It is difficult to establish a perfect benchmark against which these rates of segregation should be compared, as one must know the distribution of group sizes to compute the random rate of segregation; here we use the efficient case, as Table 7 shows that around 75 percent of the networks in these treatments connected four people.
[31] $Z = 1.371$, $p = 0.085$ (one-tailed test).

average, such segregation lasted 3.8 periods in the NoGrTy treatment, compared to 4.75 periods in the GrTy treatment.[32]

Result 5: *The group word-task facilitated segregation on the basis of the endowment types. Specifically, when the group word-task is present, the segregation by high types is higher.*

The intuition behind this result may be the following. Table 2 shows that performing the group word-task increases contributions. If subjects are contributing close to their maximum, it is more salient that playing with a high type is more profitable than playing with a low type.

Finally, we analyze segregation from the point of view of the word-task group. That is, we study if subjects who participated together in the word-task choose to play together in the PGG. Here we find that the density is slightly higher in GrNoTy (0.316 and 0.324 for all 24 periods and the last 12 periods, respectively) than in the GrTy treatment (0.293 and 0.289); one might well expect this directional result, as the affiliation to one's word-task group tends to be less salient when the alternative identity from one's contribution type is added to the picture.

In fact, Wilcoxon rank-sum tests on session-level data show that this difference is significant for both all 24 periods and the last 12 periods ($Z = 1.650$, $p = 0.050$ and $Z = 2.306$, $p = 0.011$, respectively, one-tailed tests). So in the absence of the indirect financial incentive (at least for the high types) to group by types, word-task groups appear to have a small but significant effect on one's network in the contribution game; this is particularly the case in the final 12 periods, when everyone has had a full segment of experience.

4.6 Stability, group size, and earnings

One measure of the stability of a network is the frequency with which a network changes from one period to the next. It turns out that the GrNoTy sessions, with a no-change rate of 75.8

[32] $Z = -3.464$, $p = 0.000$ (one-tailed test) .

percent, are much more stable than those of the other treatments (27.3 percent, 28.2 percent, and 28.2 percent).[33] There is little difference across the other treatments. If we consider only the final six periods of each session, we see similar patterns; the no-change rate is 100 percent in the GrNoTy sessions, compared to 33.3 percent, 36.7 percent, and 43.3 percent for the NoGrNoTy, NoGrTy, and GrTy treatments, respectively, so there is slightly (but not significantly) more stability at the end of the sessions in the GrTy treatment than in the NoGrNoTy and NoGrTy treatments.

The difference between the stability in the GrNoTy and GrTy treatments is significant even using session-level data ($p = 0.036$, two–tailed test). This appears to reflect the idea that if people are making high contributions (presumably due to the goodwill/warm glow generated by the successful group word-task), there is no reason to rock the boat with changes if there are no different types. However, given the initial high contributions in GrTy, the high types try to join other high types, decreasing the stability of the network. When contributions are not so high, subjects have incentives to change their group for looking for larger profits and this makes difficult to find a stable structure.

Table 7 shows the distribution of group sizes in each treatment, as well as the proportion of the total population in each size group. It is notable that the group sizes tend to be larger when there has been a group word-task, as between 96 percent and 97 percent of participants are in groups of size three or four in both the GrNoTy and GrTy treatments; this compares to 82 percent and 86 percent for the NoGrNoTy and NoGrTy treatment, respectively.

[33] In fact, there was absolutely no change in network structure at any point in Session 2

Table 7: Number of groups, by group size and treatment

	Size of group				
Treatment	1	2	3	4	Avg. size
NoGrNoTy	54 (9.4%)	25 (8.7%)	48 (25.0%)	82 (56.9%)	2.76
GrNoTy	21 (3.7%)	0 (0.0%)	21 (10.9%)	123 (85.4%)	3.49
NoGrTy	60 (6.2%)	38 (7.9%)	60 (18.8%)	161 (67.1%)	3.01
GrTy	63 (6.6%)	41 (8.5%)	61 (31.8%)	158 (65.8%)	2.97

Note: The proportion of the population in each group size is shown in parentheses.

We can also compute a group-efficiency index for each treatment, where the maximum group-efficiency occurs when all eight people are in 4-person networks and so the effective multiplier is 1.75.[34] The respective group-efficiency levels for the NoGrNoTy, GrNoTy, NoGrTy, and GrTy treatments are 0.899, 0.969, 0.924, and 0.920. Having different types (and the concomitant motivation for high types to seek out other high types) lowers efficiency when there is a group word-task and so contributions are high from the start; however, having types may serve as a focus to increase efficiency in the absence of a group work-task.[35] The combination of contributions and group-efficiency yields earnings; these are 35.97, 41.16, 52.89, and 58.80 for the NoGrNoTy, GrNoTy, NoGrTy, and GrTy treatments, respectively.

Profits are higher when there has been a group word-task, even with the motivation to restructure due to the presence of two different endowment types. Using session-level data, while the small number of observations helps to limit the significance of the difference between earnings in the NoGrNoTy and GrNoTy treatments (p = 0.100, one-tailed test), the difference

[34] For example, we have (1*54 + 1.25*50 + 1.50*144 + 1.75*328)/(576*1.75) = 0.899, for the group-efficiency rating for the NGNT treatment.

[35] One possible explanation could be that when the word task was played individually, contributions are lower. When we introduce the types, at least there is one focus of stability (high types know where they want to be), leading to motivated contributions, larger groups, and more stability. With no types, there is no goal to achieve, so we have more movement and lower efficiency.

between earnings in the NoGrTy and GrTy treatments is highly significant ($p = 0.004$, one-tailed test).

Result 6: *Sessions of the GrNoTy treatment are significantly more stable than in any of the other treatments. Four-person groups are by far the most common in all treatments. And earnings are higher when there has been a team-building exercise.*

4.7 Behavior with mixed groups

While we did anticipate that having everyone participate in team-building exercises might increase contributions, we had thought that this would be driven by a sense of solidarity with one's fellow group members, leading to networks comprised primarily of people from these original word-task groups. However, we instead primarily find evidence that people sort much more on the basis of endowment types. Moreover, people still do not sort on the basis of word-task groups even when there is only one endowment type. To a certain extent, this latter result may reflect on the friction and inefficiency involved with re-forming groups. Nevertheless, contributions increase substantially when everyone participates in a team-building exercise, from 73.9 percent to 92.9 percent when there is only one endowment type, and from 69.9 percent to 91.2 percent when there are two endowment types. Why are contributions so much larger with the team-building exercise, even though people are often in groups with non-team members?

One possibility for the higher contribution rates could be that subjects participating in the previous word-task in groups might simply feel more cooperative and, hence, they would increase their contributions. On this view, contributions would be higher in the team-building treatments from the start. A second possibility is that over time people can see that contributions by people from the other word-task group are high; so that there is no reason to especially associate with one's word-task group members. In this learning story, there should be some period of discovery. However, contributions are already much higher from the very beginning,

casting doubt on this explanation. Another possibility is that one might feel that since those people who were in the other word-task group also participated in a team-building exercise, they will also have a positive degree of goodwill (or *bonhomie*). Then, subjects anticipate that other participants will contribute more to the PGG, which increases subjects' own contribution.[36]

To attempt to understand which of these notions apply, we conduct Treatments 5 and 6, in which four of the eight people in a session participate in a group word task, while the other four people perform the word task individually. If the first explanation holds, we should see similar contributions (relative to those in the treatments with everyone participating in team-building exercises) by the people in the group word-task and no increase in contributions (relative to those in the treatments with no group word-task) for people performing the task individually. If, on the other hand, it is the third explanation that drives the results, we should observe smaller contributions by subjects in the word-task group relative to the treatment in which all subjects performed the word task in groups.

In fact, we find considerable support for the third explanation. Table 8 presents the data from Treatments 5 and 6 (session-level data can be found in the online appendix). For convenience, we include some data from Table 2 for Treatments 1-4. The overall contribution rate in MixNoTy is significantly lower than in GrNoTy ($Z = 4.316$, $p = 0.000$) but not significantly different than the overall contribution rate in NoGrNoTy ($Z = -0.330$, $p = 0.741$). By the same token, the overall contribution rate MixTy is very similar to the rate in NoGrTy ($Z = 0.749$, $p = 0.454$), but quite different from the rate in GrTy ($Z = 2.933$, $p = 0.003$).[37]

[36] This increment in contribution is due to subjects' strategic behavior. They contribute more in order to keep the contribution rate high and in order to not be excluded from the group.

[37] All of these tests are at the individual level. Session-level tests give $Z = 0.218$, $p = 0.414$ for MixNT vs. NGNT, $Z = 1.528$, $p = 0.064$ for MixNT vs. GrNoTy, $Z = 0.745$, $p = 0.228$ for MixT vs. NoGrTy, and $Z = 1.640$, $p = 0.050$ for MixT vs. GrTy (all one-tailed tests).

Table 8: Results from the Mix Treatments

Treatment/type	Contribution	Percentage	Period earnings
MixNoTy-group	19.42	0.777	38.19
MixNoTy-indiv.	18.55	0.742	38.23
MixNoTy-overall	19.00	0.760	38.21
MixTy-group	26.70	0.715	55.01
MixTy-indiv.	25.49	0.676	56.04
MixTy-high	33.98	0.680	67.55
MixTy-low	17.82	0.713	42.89
MixTy -overall	26.10	0.696	55.52
NoGrNoTy – overall	18.47	0.739	32.68
GrNoTy - overall	23.22	0.929	40.41
NoGrTy- overall	24.79	0.699	44.79
GrTy - overall	34.80	0.912	57.46

Furthermore, we see that there is no difference in contribution rates in either MixNoTy or MixTy for people in the word-task group and people who performed the task individually (Z = 0.981, p = 0.326; Z = 0.462, p = 0.454, two-tailed tests). The contribution rate by high-endowment types in MixTy (0.680) is intermediate, but closer to that in NoGrTy (0.586) than in GrTy (0.879); the rates for low-endowment types in MixTy (0.713) and MixNoTy (0.760) are actually lower than in both NoGrTy (0.812) and GrTy (0.946).[38]

Result 7: *The enhancement in first period contributions when the word-task is performed in groups is due to participants anticipating that other participants will contribute more to the PGG.*

Regarding segregation, that for word-task groups in MixNoTy (.336 for all 24 periods and .281 for the last 12 periods) and MixTy (.299 for all 24 periods and .296 for the last 12

[38] It also turns out that the MixT treatment is approximately as stable as the GrTy and NoGrTy treatments, with no changes from one period to the next 27.3 percent of the time; the stability in the MixNT treatment is intermediate (at 58.7 percent) between the NGNT treatment and the GrNoTy treatment.

periods) is roughly the same as in GrNoTy and GrTy, respectively.[39] This suggests that there is not more allegiance to one's own word-task group in the Mix treatments than in the G treatments. The segregation for high types is intermediate between the individual and group treatments, at 0.517 and .597 for all 24 periods and the last 12 periods, respectively. Regarding average group size, this is not dramatically different for MixNoTy (3.31) and MixTy (2.97) than for GrNoTy and GrTy, respectively, while the group size is considerably higher in MixNoTy than for NoGrNoTy, but not different for MixTy and NoGrTy. The concomitant group-efficiency levels for the MixNoTy and MixTy treatments are 0.926 and 0.952, respectively; these levels are much closer to those observed in the GrNoTy and GrTy treatments than in the NoGrNoTy and NoGrTy treatments.

5. Discussion

We induce identity by having a team-building task and we also introduce heterogeneity in the initial endowment in a public-goods game. We find that people flock together on the basis of endowment types. We had thought that group-word-task affiliations would lead to more cooperation amongst these group members, which would presumably lead to sorting on the basis of this identity. Alternatively, we thought that if people instead sort on the basis of efficiency types (and perceived financial interest, even though the equilibrium strategy, with standard preferences, is to contribute nothing in all cases), it would mean that the group word task had no effect. Instead, we find a strong effect from the group word-task.

Even though the word-task identity has a weaker effect on segregation than the endowment type, the team-building activity greatly increases the contribution rate, from 73.9% to 92.9% without endowment types and from 69.9% to 91.2% with endowment types. We find

[39] Note that we only take into account the subjects who previously participated in the word task in groups.

no significant effect of endowment heterogeneity in contributions. This is in line with results found in Sadrieh and Verbon (2006) although the evidence in the experimental literature is inconclusive. Interestingly, heterogeneity has a negative effect when the word-task group is present.

The effect of the team-building activity is an unexpected effect, since previous evidence indicates that one does not tend to treat non-team members the same as one treats one's ingroup. Two additional treatments were conducted to examine the underpinnings of this effect. One explanation could be that people become more altruistic *per se* after participating in a team-building task. A second explanation could be that expectations about the contributions of non-team members differ according to whether the non-team members have participated in a team-building task in another team. Since behavior is quite different according to whether the non-team members engage in a team-building task, the latter explanation is supported by the data. We note that this effect does not appear to diminish over time, (see Figure A1 in the appendix) as both the contribution rate in each condition and the difference across these remains the same (except in the final period of the sessions, were we do see some unraveling).[40]

In a sense, it seems that there is some form of more generalized homophily (which here seems related to the notion of *warm glow*; Andreoni, 1990) present as a result of the group team-building task and the awareness that others have engaged in a team-building task. This leads to a higher initial level of contributions in the public-goods game, which is sustained through the game by the threat of possible exclusion and sharply reduced potential earnings.[41] When there

[40] These results are completely in line with those obtained in Charness and Yang (2009).
[41] Of course, this should unravel all the way to the beginning of each segment, since the game has a finite horizon. Nevertheless, there are reasonable beliefs for which cooperation is the best strategy for most of the game. See Charness and Yang (2009) for a discussion of this issue.

are no groups and no types, there is scope for neither identity nor endowment heterogeneity to affect behavior.

How does this feeling of community spirit transfer from one's word-task-group identity to those who were not in one's group but who were in a similar one? We can only speculate. It seems likely, however, that there is a change in beliefs about either what other people are going to do or what one 'should' do when one has had the experience of cooperating successfully in a small team project.[42] This result seems to be outside of existing theories of social identity and friendship networks, although it is intuitively appealing. Does one perceive that a different social norm is in effect for the entire set of people in the session? In fact we do find that those people who are in the word-task group in the Mix treatments make smaller contributions than in the Gr treatments, perhaps since they know the others were not in a word-task group. Those people who performed the word task individually contributed about the same as those in the NoGr treatments, perhaps since they are not aware of the positive feeling that emerges from being in a word-task group. Nevertheless, people in the word-task group in the Mix treatments do not sort on this basis any more than in the Gr treatments.[43]

Regarding segregation, we observe that the group word-task facilitates segregation on the basis of the endowment type, although in an indirect way. One plausible explanation might be that as the word-task significantly raises contributions, this makes more salient the fact that it is more profitable for high-endowment types to join with each other. In addition, another positive effect of the group word task when there is no endowment heterogeneity is that stability and efficiency in the network are enhanced.

---

[42] We felt it most effective to set the threshold for success low enough to ensure success, but it is true that we have not eliminated the possibility that simply *attempting* the group task (and failing) would create the same degree of success for the team-building task.

[43] $Z = 0.218$, $p = 0.827$; $Z = 0.150$, $p = 0.881$, two-tailed tests, for Types and all 24 periods at a session level, $Z = 0.696$, $p = 0.487$; $Z = 0.000$, $p = 1.000$, two-tailed tests, for No-Types and the last 12 periods at a session level.

In terms of the group-formation process, there is a modest rate of exclusion; people are more likely to be excluded when they make smaller absolute and relative contributions and generally less likely to be excluded when they have higher endowments; there is less exclusion after there has been a team-building exercise, even accounting for contribution levels. Exit is relatively rare, but is observed more for high types than low types; high types who wish to network with other high types must exit to do so, as it is necessary to make larger contributions in order to be invited to join a high-powered group and one who makes such contributions will not be excluded. Singles are invited to join groups based on their contributions, and are also more likely to be invited to join when they are the same endowment type as the majority of the voting group and, if there are no endowment types, when they were part of the same word-task group as the majority of the voting group. Having a team-building task is correlated with more group stability and larger groups.

6. Conclusion

One's social identity, as exemplified by group membership, has been found to have a considerable effect on behavior in economic contexts. We conduct public-goods experiments in which we vary whether there is a team-building exercise and whether everyone receives the same endowment or some people receive twice as much as others. It is an open question as to whether word-task identity or possible economic considerations is more salient; we are unaware of previous research on this topic. If social identity and group membership are important determinants of behavior, it seems vital to examine whether financial incentives are stronger than identity tin explaining subjects' behavior. Our study is a first step in this direction.

We start with two 4-person groups and permit endogenous group-formation via exclusion, exit, and being able to add an unattached person to one's group. We do see some

evidence of one's word-task-group identity affecting the endogenous networks when there is only one endowment type; however, when two endowment types are introduced, the high-endowment participants are strongly attracted to linking up with each other, crowding out the word-task-group identity. Thus, endowment ("money") is more important than word-task group ("friendship" in our setting). While we make no claim that this generalizes to all field settings, it does represent the first experimental data in this regard.

One surprising and interesting result is that the team-building exercise greatly increases the level of contribution without respect to whether one is linked to people from one's team-building exercise. However, this result mainly does not carry over when there is only one word-task group, with others performing the word task individually. This suggests that the positive feeling engendered by the group exercise spills over to other participants only when these other participants are known to have also participated in a team-building exercise. While in some sense this seems intuitive, this is neither the prediction in the current theoretical models nor the results found in most previous experiments on group membership, where mistrust across different groups seems to be the rule.

All in all, the simple and relatively inexpensive group word task leads to considerable benefits for individual earnings and, by extension, social welfare. It may be particularly effective to have an interactive common activity where there is a threshold level for group success and identical payments to the group members. Previous research on group membership and social identity has largely found negative social effects, but our study suggests that there may be ways to make social identity and group membership a positive force in our society. This research is at a very early stage and more study is surely needed.

Finally, our results seem to support the idea that the monetary identity is the most salient force driving segregation in a PGG with endogenous formation. Thus, one interesting and not evident question to explore would be: How segregation will evolve if we confront two different kind of monetary identities? For instance, in our setting, will it be a preference for individuals with a higher MPCR or a higher endowment?[44] A different issue will be to check whether the segregation results will remain if we contemplate a stronger degree of relationship than the one obtained after the word task. One possibility may be to consider an exogenous group already formed such as a sport team, a fraternity, a music band… Another option may be to change the task in the experiment so that they will share more time and/or (personal) information. All these remain as open questions.

[44] We thank an anonymous referee for this appealing suggestion.

References

Akerlof, G. and R. Kranton (2000), "Economics and Identity," *Quarterly Journal of Economics* CVX (3): 715-753.

Ahn, T., Isaac, R. M. and T. C. Salmon (2009). "Coming and Going: Experiments on Endogenous Group Sizes for Excludable Public Goods," *Journal of Public Economics* 93 (1-2): 336-351.

Andreoni, James (1990), "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving," *Economic Journal*, 100 (401): 464–477.

Benjamin, D., Choi, J. and J. Strickland (2010) "Social identity and preferences," *American Economic Review*, 100 (4): 1913-1928.

Brewer, M. and R. Kramer (1986), "Choice Behavior in Social Dilemmas: Effects of Social Identity, Group Size, and Decision Framing, *Journal of Personality and Social Psychology* 50, 543-549.

Brosig, J., M. Margreiter, and J. Weimann (2005), "Endogenous Group Formation and the Provision of Public Goods: The Role of Promises and Lies," mimeo.

Brown-Kruse, J. and D. Hummels (1993), "Gender Effects in Laboratory Public Goods Contribution: Do Individuals Put Their Money Where Their Mouth Is?," *Journal of Economic Behavior and Organization* 22(3), 255-67.

Buckley, E. and Croson, R. (2006). "Income and wealth heterogeneity in the voluntary provision of linear public goods," *Journal of Public Economics*, 90:935-955.

Buller, P. (1986), "The Team Building – Task Performance Relation: Some Conceptual and Methodological Refinements," *Group & Organization Studies* 11(3), 147-168.

Buller, P. and C. Bell(1986), "Effects of Team Building and Goal Setting on Productivity: A Field Experiment," *Academy of Management Journal*, 29(2), 305-328.

Cadsby, C. and E. Maynes (1998), "Gender and Free Riding in a Threshold Public Goods Game: Experimental Evidence," *Journal of Economic Behavior and Organization* 34(4), 603-20.

Campbell, D. (1958), "Common Fate, Similarity, and Other Indices of the Status of Aggregates of Persons as Social Entities," *Behavioral Science* 3, 14-25.

Carron, A. and K. Spink (1993), "Team building in an exercise setting," *The Sport Psychologist*, 7(1), 8-18.

Cinyabuguma, M., T. Page, and L. Putterman (2005), "Cooperation under the Threat of Expulsion in a Public Goods Experiment," *Journal of Public Economics*, 89: 1421-1435.

Chan, K. S., Mestelman, S., Moir, R., and Muller, R. A. (1996). "The voluntary provision of public goods under varying income distributions," *Canadian Journal of Economics*, 29:54-59.

Chan, K. S., Mestelman, S., Moir, R., and Muller, R. A. (1999). "Heterogeneity and the voluntary provision of public goods," *Experimental Economics*, 2:5-30.

Charness, G., Rigotti, L. and A. Rustichini (2007) "Individual Behavior and Group Membership," *American Economic Review*, 97: 1340-1352.

Charness, G. and M. Rabin (2002), "Understanding Social Preferences with Simple Tests," *Quarterly Journal of Economics*, 117: 817-869).

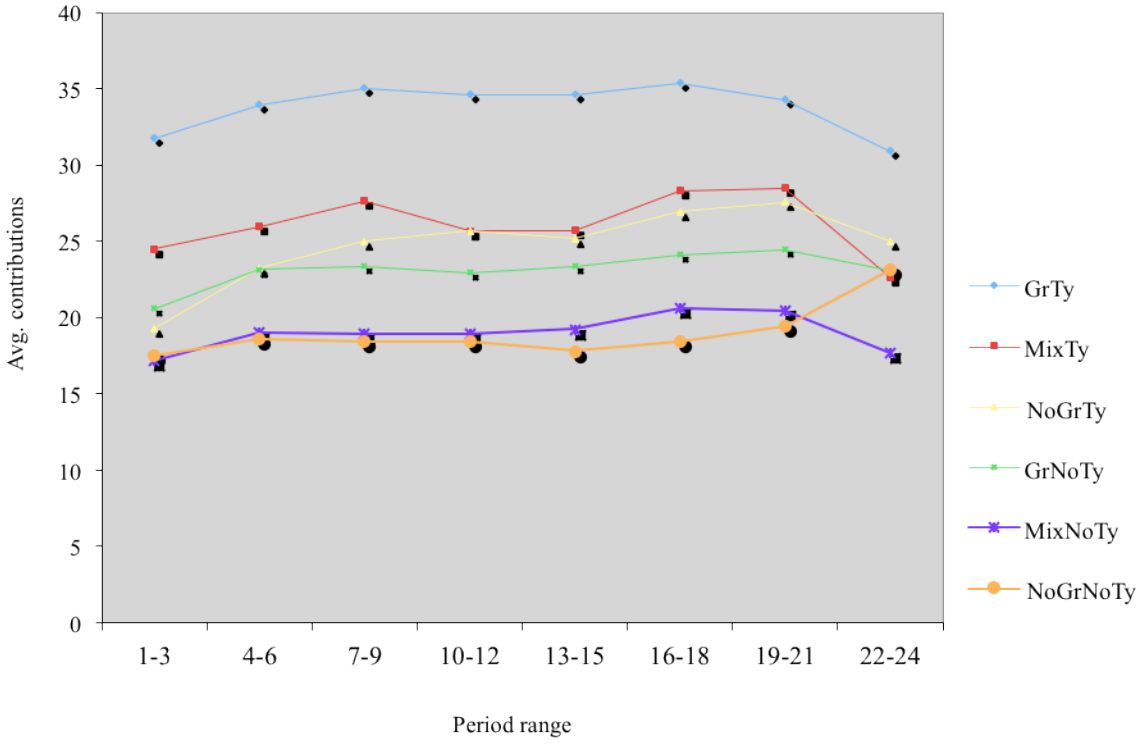
Charness, G. and A. Rustichini (2011), "Gender differences in cooperation with group membership," *Games and Economic Behavior*, 72: 77-85.

Charness G. and C.-L. Yang (2009), "Public Goods Provision with Voting for Exclusion, Exit, and Mergers: An Experiment," mimeo.
Chaudhuri, A. (2011), **"**Sustaining cooperation in laboratory public goods experiments: a Selective Survey of the Literature," *Experimental Economics*, forthcoming
Chen, Y. and S. Li (2009), "Group Identity and Social Preferences," *American Economic Review* 99 (1): 431-457.
Chen, Y., S. X Li, T. X Liu, M. Shih (2010), "Which Hat to Wear? Impact of Natural Identities on Coordination and Cooperation," University of Michigan working paper.
Cherry, T. L., Kroll, S., and Shogren, J. (2005). "The impact of endowment heterogeneity and origin on public good contributions: Evidence from the lab," *Journal of Economic Behavior & Organization*, 57:357-365.
Croson, R., E. Fatás, and T. Neugebauer (2006), "Excludability and Contribution: A Laboratory Study in Team Production," mimeo.
Croson, R. and M. Marks, (2003) "Groups Work for Women: Gender and Group Identity in the Provision of Public Goods. Mimeo.
Currarini, S., Jackson, M. and P. Pin (2009), "An Economic Model of Friendship: Homophily, Minorities and Segregation," *Econometrica,* 77 (4): 1003–1045.
Eckel, C. and P. Grossman (2005), "Managing Diversity by Creating Team Identity," *Journal of Economic Behavior and Organization* 58, 371-392.
Eden, D. (1986), "Team Development: Quasi-Experimental Confirmation Among Combat Companies," *Group & Organization Management* 11 (3), 133-146.
Ehrhart, K.-M. and C. Keser (1999), "Mobility and Cooperation: On the Run," mimeo
Gunnthorsdöttir, A., Houser, D. and K. McCabe (2007), "Disposition, History and Contributions in Public Goods Experiments," *Journal of Economic Behavior and Organization,* **62**: 304-315.
Hamilton, B. H., Nickerson, J. A., and Owan, H. (2003), "Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation," *Journal of Political Economy*, 111:465-497.
Hargreaves Heap, S.P. and D. Zizzo (2009), "The Value of Groups," *American Economic Review*, 99(1): 295–323.
Jackson, M. (2008), *Social and Economic Networks*, Princeton: Princeton University Press.
Kirkwood, N. and J. L. Solow (2002),"Group identity and gender in public goods experiments," *Journal of Economic Behavior and Organization* 48, 403-412.
Koszegi, B. (2006), "Ego Utility, Overconfidence, and Task Choice." *Journal of the European Economic Association*, 4(4): 673-707.
Kramer, R. and M. Brewer (1984), "Effects of Group Identity on Resource Use in a Simulated Commons Dilemma, *Journal of Personality and Social Psychology* 46, 1044-1057.
Maier-Rigaud, F., Martinsson, P. and G. Staffiero (2010), "Ostracism and the Provision of a Public Good: Experimental Evidence*," Journal of Economic Behavior and Organization*, 73: 387–395.
Ostrom, E., Gardner, R., and Walker, J. (1994), "Rules, Games, and Common-pool Resources," *University of Michigan Press*, Ann Arbor.
Page, T., L. Putterman, and B. Unel (2005), "Voluntary Association in Public Goods Experiments: Reciprocity, Mimicry, and Efficiency", Economic Journal, 115: 1032-1053.
Pan, X. S. and D. Hauser (2013), "Cooperation during cultural group formation promotes trust towards members of out-groups," *Proceedings of the Royal Society B, 280: 1762.*

Riedl, A. and A. Ule (2002), "Exclusion and Cooperation in Social Network Experiments," mimeo.
Sadrieh, A. and Verbon, H. (2006), "Inequality, cooperation, and growth: An experimental study," *European Economic Review*, 50:1197-1222.
Tajfel, H. and J.C. Turner (1979), "An Integrative Theory of Intergroup Conflict," In W. G. Austin & S. Worchel (Eds.), *The Social Psychology of Intergroup Relations*. Monterey, CA: Brooks-Cole.
Tajfel, H., Billig, M., Bundy, R. P. and C. Flament (1971), "Social categorization and intergroup behavior," *European Journal of Social Psychology, 2*: 149-178.
van Dijk, F., Sonnemans, J., and van Winden, F. (2002), "Social ties in a public good experiment," *Journal of Public Economics*, 85:275-299.
Visser, M. and Burns, J. (2006), "Bridging the great divide in south Africa: Inequality and punishment in the provision of public goods," *Working paper*, Göteborg University.
Yang, C.-L., Xu, M. and F. Tang (2009), "Efficient Large-size Coordination in Experiment via Endogenous Group Formation," mimeo.

Appendix: Contributions over time, by treatment

Figure A1: Contributions over time



Regression tables

Table 3: GLS Random Effects Regressions on Contribution

	(1) Contribution	(2) Contribution	(3) Contribution	(4) % Contribution
Average lagged group contribution	0.233*** (0.040)	0.225*** (0.026)	0.148** (0.026)	
Group size	1.631*** (0.323)		0.738*** (0.292)	0.006 (0.009)
Group size*high type				0.031** (0.013)
Dummy t1t2	-4.148*** (0.812)			
High type			13.071*** (0.631)	-0.257*** (0.055)
Number of highs in the group		1.306*** (0.278)		
Same type as the majority of the group		-0.594 (0.602)		
Same type as majority of the group*high type		6.184*** (0.986)		
Same word-task group as majority of contribution group	1.068 (0.750)	0.898 (0.595)	0.543 (0.600)	0.007 (0.014)
Dummy t3t4		-7.192** (1.026)	-8.171*** (1.532)	-0.208*** (0.028)
Average lagged % group contribution				0.072*** (0.026)
Constant	11.284*** (1.525)	21.765*** (1.143)	6.381*** (1.532)	1.195*** (0.086)
R-squared	0.309	0.529	0.554	0.298
N	1032	1726	1726	1717

Notes: ***, **,* denote significance at $p = 0.01$, 0.05, and 0.10, respectively (two-tailed tests). Robust standard errors are in parentheses. The variable "Dummy t1t2" takes the value 1 when the observation is from Treatment 1 (NoGrNoTy) and 0 when it is from Treatment 2 (GrNoTy). The variable "Dummy t3t4" takes the value 1 when the observation is from Treatment 3 (NoGrTy) and 0 when it belongs to Treatment 4 (GrTy). The number of observations in (1) represents period 2 to period 12 of the 2 segments in Treatments 1 and 2, because we have included the lagged contribution. For (2) and (3), these are observations from Treatments 3 and 4 for periods 2-12 of the 2 segments.

Table 4: Random-effects logit regressions on exclusion (Marginal effects)

	(1)	(2)	(3)	(4)	(5)
Avg. group contrib. - average contrib.	0.006*** (0.002)	0.007*** (0.0007)			
Average percent of own contribution				-0.352*** (0.052)	-0.148*** (0.034)
Max. avg. contrib. in other groups	0.001** (0.0006)	0.002** (0.0006)			
Dummy t1t2	0.041*** (0.015)				
High type			-0.059*** (0.018)		
Number of highs in the group			-0.003 (0.006)		
Same type as majority of group			-0.103*** (0.017)	-0.141*** (0.026)	-0.026** (0.011)
Dummy t3t4		0.032** (0.013)	-0.010 (0.017)	-0.129*** (0.030)	-0.020 (0.012)
LL	-196.755	-484.627	-555.018	-302.054	-167.418
N	1077	1797	1789	875	922

Notes: ***, **,* denote significance at p = 0.01, 0.05, and 0.10, respectively (two-tailed tests). Robust standard errors are in parentheses. Specification (1) includes observations from Treatments 1 and 2, while (2)-(5) include observations from Treatments 3 and 4. The variable "Dummy t1t2" takes the value 1 when the observation is from Treatment 1 (NoGrNoTy) and 0 when it is from Treatment 2 (GrNoTy). The variable "Dummy t3t4" takes the value 1 when the observation is from Treatment 3 (NoGrTy) and 0 when it belongs to Treatment 4 (GrTy).

Table 5: Random-effects logit regressions on exit (Marginal effects)

	(1)	(2)	(3)
Avg. group contribution – average contribution	-0.002* (0.0009)	-0.004*** (0.0005)	
Maximum avg. contribution in other groups	0.00009 (0.0002)	0.002** (0.0006)	
Dummyt1t2	0.006 (0.006)		
High type			0.103*** (0.021)
Same type as the majority of the group			0.020 (0.014)
Same type as the majority of the group*high type			-0.098*** (0.022)
Dummy t3t4		0.012 (0.013)	-0.012 (0.013)
Average percent of own contribution			0.039 (0.026)
LL	-87.818	-358.780	-381.205
N	978	1594	1772

Notes: ***, **,* denote significance at p = 0.01, 0.05, and 0.10, respectively (two-tailed tests). Robust standard errors are in parentheses. Specification (1) includes observations from Treatments 1 and 2, while (2)-(3) include observations from Treatments 3 and 4. The variable "Dummy t1t2" takes the value 1 when the observation is from Treatment 1 (NoGrNoTy) and 0 when it is from Treatment 2 (GrNoTy). The variable "Dummy t3t4" takes the value 1 when the observation is from Treatment 3 (NoGrTy) and 0 when it belongs to Treatment 4 (GrTy).

Table 6: Logit regressions on votes for singles (Marginal effects)

	(1)	(2)	(3)	(4)	(5)	(6)
Contribution of the single	0.037*** (0.008)					
Single has been excluded	-0.196 (0.101)	-0.153 (0.102)				
Avg. group contrib. – single contrib.		-0.054*** (0.011)				
Dummyt1t2	0.265** (0.114)	0.149 (0.126)				
Dummyt3t4			0.0001 (0.046)	0.053 (0.043)	0.032 (0.037)	
Single is high type					0.199*** (0.040)	
Single's type same as majority of group					0.080*** (0.022)	
Single and majority in same word-task group * Dummyt2t4						1.594** (0.731)
Single and majority in same word-task group						-0.027 (0.272)
Dummyt2t4						-0.210** (0.124)
% contribution of the single			0.250** (0.106)	0.409*** (0.077)	0.378*** (0.066)	0.119*** (0.021)
Type HH			0.223*** (0.036)			
Type LL				-0.121** (0.049)		
Dummies for half-segments	YES	YES	YES	YES	YES	YES
LL	-98.897	-88.854	-149.745	-93.265	-249.591	-151.779
N	192	192	340	306	646	387

Notes: ***, **,* denote significance at p = 0.01, 0.05, and 0.10, respectively (two-tailed tests). Robust standard errors are in parentheses. The variable "Type HH" is a dummy that takes value 1 when a group with majority of high types votes for a single who is a high type and 0 when a group with a majority of high types votes for a single who is a low type. The definition is analogous for the variable "Type LL". Dummies for

half-segments reflect periods 1-6, 7-12, 13-18, and 19-24, but with periods 12 and 24 excluded. Specification (1)-(2) includes observations from Treatments NoGrNoTy and GrNoTy, while (3)-(5) include observations from Treatments NoGrTy and GrTy. Specification (6) includes observations from only GrNoTy and GrTy. The variable "Dummy t1t2" takes the value 1 when the observation is from Treatment 1 (NoGrNoTy) and 0 when it is from Treatment 2 (GrNoTy). The variable "Dummy t3t4" takes the value 1 when the observation is from Treatment 3 (NoGrTy) and 0 when it belongs to Treatment 4 (GrTy). The variable "Dummy t2t4" takes the value 1 when the observation is from Treatment 2 (NoGrTy) and 0 when it belongs to Treatment 4 (GrTy).