Efficiency, Team building, and Spillover in a Public-goods Game

May 4, 2012

Abstract: The notions of one's social identity, group membership, and homophily have recently become topics for economic theory and experiments. Yet, since people are members of many groups (e.g., race, gender, handedness) what determines which identity or identities are the most salient in different environments? Further, how do these factors trade off against one's financial interest? We conduct public-goods experiments in which we permit endogenous group-formation and vary whether there is a team-building exercise and whether some people receive an endowment twice as much as others receive. We do see evidence that team identity affects endogenous networks when there is only one endowment type; however, when both identities are present, high-endowment participants are strongly attracted to linking up with each other. One 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. Apparently the positive feeling engendered by the group exercise spills over to participants who were in another team; however, this is not the case when one group has been in a 4-person team and the other four participants have not.

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

Keywords: Experiment, identity, team building, homophily

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.

More recently, 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.

These papers bring economic theory into alignment with other social sciences (e.g., anthropology, political science, psychology, and sociology), where identity is considered a central concept. In social identity theory (Tajfel and Turner, 1979), a person who identifies with a group receives benefits from group membership and tries to behave in accordance with the perceived group prescription. And yet, there are many aspects that have not yet been explored.

Tajfel, Billing, Bundy, and Flament (1971) established 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. Charness, Rigotti, and Rustichini (2007) extend the research on group membership to strategic settings and observe strong effects when group membership is salient, but not with minimal groups, while Chen and Li (2009) find that people are more likely to reward an ingroup member for good behavior, but less likely to punish

[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).

an ingroup member for misbehavior.

However, it appears that a critical element has not yet been explored in any of this research. In reality, an individual has many affiliations at the same time, so that identity is multi-faceted. The process by which it is determined which of these affiliations is salient in a particular environment is unclear. Individuals may not only choose their aggregate identity, but may also (consciously or unconsciously) choose which dimension of identity applies at any time and circumstance. This choice may be a key determinant of behavior, particularly when different aspects of one's identity pull in different directions.

We take this 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, each of which can be seen as reflecting a form of identity. 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 almost 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. While this factor is less explicit with respect to group membership; there is potentially a strong financial motivation for better-endowed people to join together. Thus, we pit two aspects of one's experimental identity 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. However, 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 homophily 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.

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. 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-group task. 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 all 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.

1. Related previous work

As was discussed briefly in the previous section, 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. One insight from the theory of social identity is that one derives self-esteem or ego utility (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.

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.[2] 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

[2] 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.

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.

In economics, one experimental paper on the effects of group membership on behavior is that of Charness, Rigotti, and Rustichini (2007).[3] Previous work on the ingroup-outgroup 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 for misbehavior.[4]

There are a number of studies regarding endogenous group-formation (assortative matching) in the public-goods game.[5] The idea behind this approach is to enable would-be cooperators to avoid free

---

[3] 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 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.

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

[5] 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

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 were 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 only one person could be added to a group in a period.[6]

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.[7] 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-

---

Unel (2005), 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).

[6] 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.

[7] 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.

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 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 help to 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.[8]

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, only one dimension of identity is considered in this study.

2. 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)

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

with a total of 22 sessions with 8 subjects in each session. 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. 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.

In Treatment 1 (hereafter the "*NGNT*" 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.[9] 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.

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.[10] Then, we proceeded with the group formation. We divided this group formation into three stages.

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.

[9] 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.

[10] In the GT and NGT 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.

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 values reflected more favorable opinions.[11] Singles were ranked according to the sum of the values received.[12] 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.[13] 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 ID 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.

[11] 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 GT and NGT) 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.

[12] 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.

[13] 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.

In Treatment 2 (*GNT* hereafter) the structure of the game was the same as in the NGNT treatment. The only difference was that participants were randomly assigned into two groups of four subjects upon arrival to the lab.[14] 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.[15] Thus, in the GNT Treatment, participants had the same information as participants in the NGNT 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 (*NGT*, hereafter) differed in only way from the NGNT. Half of participants in the PGG were endowed randomly with 25 tokens (the same as in the NGNT), 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 or 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).[16]

Participants had the same information as in the NGNT, 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.

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

3. 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 analyze the stability of the networks and the group size. We also study the density and segregation in the society. Finally, we focus on the behavior in mixed groups.

[14] 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.

[15] 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.

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

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.

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

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

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

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 GNT contributed 92.9 percent of their endowment, compared to 73.9 percent for those people in the NGNT. A Wilcoxon-Mann-Whitney ranksum test using each individual's (non-single) average contribution 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).[17] When there are types, low types in the NGT contributed 81.2 percent of their endowment, compared to 94.6 percent in the GT; 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). High types in the NGT contributed 58.6 percent of their endowment, compared to 87.9 percent in the GT; 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). Thus, we have:

Result 1: *The previous word-task significantly increases contributions to the public good.*

[17] 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 GNT 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.

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 NGNT treatment than in the GNT treatment, and significantly more exclusion in the NGT treatment than in the GT 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 who was 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). We present 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-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, specifications (6) and (7) indicate that people in the GNT treatment are significantly more likely to vote for a single from the same word-task group, but that there is no significant difference in the GT treatment. 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.*

4.5 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 GNT sessions, with a no-change rate of 75.8 percent, are much more stable than those of the other treatments (27.3 percent, 28.2 percent, and 28.2 percent).[18] 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 GNT sessions, compared to 33.3 percent, 36.7 percent, and 43.3 percent for the NGNT, NGT, and GT treatments, respectively, so there is slightly (but not significantly) more stability at the end of the sessions in the GT treatment than in the NGNT and NGT treatments.

The difference between the stability in the GNT and GT 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 GT, the high types try to join other high types, decreasing the stability of the

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

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: Number of groups, by group size and treatment

	Size of group				
Treatment	1	2	3	4	Avg. size
NGNT	54 (9.4%)	25 (8.7%)	48 (25.0%)	82 (56.9%)	2.76
GNT	21 (3.7%)	0 (0.0%)	21 (10.9%)	123 (85.4%)	3.49
NGT	60 (6.2%)	38 (7.9%)	60 (18.8%)	161 (67.1%)	3.01
GT	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.

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 GNT and GT treatments; this compares to 82 percent and 86 percent for the NGNT and NGT treatment, respectively.

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.[19] The respective group-efficiency levels for the NGNT, GNT, NGT, and GT 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.[20] The combination of contributions and group-efficiency yields earnings; these are 35.97, 41.16, 52.89, and 58.80 for the NGNT, GNT, NGT, and GT treatments, respectively.

Profits are higher when there has been a group word-task, even when there is the imperative 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 NGNT and GNT treatments ($p = 0.100$, one-tailed test), the difference between earnings in the NGT and GT treatments is highly significant ($p = 0.004$, one-tailed test).

[19] 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.

[20] 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.

Result 5: *Sessions of the GNT treatment are far 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.*

4.6 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".[21] In our setting, the maximum level of segregation in four-person (efficient) groups would be to find one group composed of four "high types" and other group composed of four "low types".[22] In fact, the density is higher in the GT treatment (0.641 and 0.750 for all 24 periods and the last 12 periods, respectively) than in the NGT 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).

Result 6: When *the word-task is present, segregation of high types is higher.*

The intuition behind this result may be the following. Table 2 shows that performing the word-task in groups 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.

Next, 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 GNT (0.316 and 0.324 for all 24 periods and the last 12 periods, respectively) than in the GT 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

[21] In our setting, all links are undirected.
[22] 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.

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.

Result 7: *Overall, there is slightly but significantly less same-word-task grouping when there are two contribution types.*

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.[23]

[23] 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.

To attempt to understand which of these notions apply, we conduct sessions 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. Accordingly, we add two treatments. In Treatment 5 (MixNT), 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 (MixT) only differs from Treatment 5 in that half of the people receive endowments of 25 tokens and the other half receive endowments of 50.

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 MixNT is significantly lower than in GNT ($Z = 4.316$, $p = 0.000$) but not significantly different than the overall contribution rate in NGNT ($Z = -0.330$, $p = 0.741$). By the same token, the overall contribution rate MixT is very similar to the rate in NGT ($Z = 0.749$, $p = 0.454$), but quite different from the rate in GT ($Z = 2.933$, $p = 0.003$).[24] Furthermore, we see that there is no difference in contribution rates in either MixNT or MixT 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 MixT (0.680) is intermediate, but closer to that in NGT (0.586) than in GT (0.879); the rates for low-endowment types in MixT (0.713) and MixNT (0.760) are actually lower than in both NGT (0.812) and GT (0.946).[25]

[24] 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. GNT, $Z = 0.745$, $p = 0.228$ for MixT vs. NGT, and $Z = 1.640$, $p = 0.050$ for MixT vs. GT (all one-tailed tests).

[25] It also turns out that the MixT treatment is approximately as stable as the GT and NGT 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 GNT treatment.

Table 8: Results from the Mix Treatments

Treatment/type	Contribution	Percentage	Period earnings
MixNT-group	19.42	0.777	38.19
MixNT-indiv.	18.55	0.742	38.23
MixNT-overall	19.00	0.760	38.21
MixT-group	26.70	0.715	55.01
MixT-indiv.	25.49	0.676	56.04
MixT-high	33.98	0.680	67.55
MixT-low	17.82	0.713	42.89
MixT -overall	26.10	0.696	55.52
NGNT - overall	18.47	0.739	32.68
GNT - overall	23.22	0.929	40.41
NGT- overall	24.79	0.699	44.79
GT - overall	34.80	0.912	57.46

Regarding segregation, that for word-task groups in MixNT (.336 for all 24 periods and .281 for the last 12 periods) and MixT (.299 for all 24 periods and .296 for the last 12 periods) is roughly the same as in GNT and GT, respectively.[26] 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 MixNT (3.31) and MixT (2.97) than for GNT and GT, respectively, while the group size is considerably higher in MixNT than for NGNT, but not different for MixT and NGT. The concomitant group-efficiency levels for the MixNT and MixT treatments are 0.926 and 0.952, respectively; these levels are much closer to those observed in the GNT and GT treatments than in the NGNT and NGT treatments.

4. Discussion

In accordance with recent work on social identity and group membership, we find strong behavioral effects from a successful group team-building task. This effect is not solely due to the feeling of success, as people in other treatments also succeed in the corresponding individual task. Thus, there is something about the group nature of the team-building task that induces the effects that we observe. We note that this effect does not appear to diminish over time, (see Table A1 in the appendix) as both the

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

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).

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 on contributions from the team-building activity, whether or not different types of people in the contribution game, and sorting primarily on the basis of contribution types. There are some modest effects on non-contribution choice behavior on the basis of one's word-task group, but this is essentially crowded-out when there is more than one endowment type, and high types try to sort into their own complete 4-person network.

We are not aware of any previous study that considers spillover effects from a team-building task to non-team members. In principle, one might not expect that there would be such spillover effects, since there is no reason to treat non-team members the same as one's ingroup. Yet as we shall see, people do not really differentiate across word-task group membership. One might feel that this is due to financial concerns, since a high-endowment type may well believe that her earnings; however, we also observe this effect when there is only one endowment type, where there is no such consideration.

In a sense, it seems that there is some form of more generalized homophily (literally, "love of man") present as a result of the group team-building task. This task 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.[27] When there are no groups and no types, there is no scope for either form of identity to affect behavior. Here the forces of endogenous group-formation and efficiency alone lead to a contribution rate of 73.9 percent. Adding team building by itself dramatically increases this rate to 92.9 percent.

When we add endowment types instead of team building, the rate for the low type is increased to 81.2 percent, perhaps as a means of competing to be in a network with high types and their larger endowments. The contribution rate for high types in not very impressive, however, at 58.6 percent. There seems to be less fear of exclusion, since even this rate leads to more value than a low type who contributes everything. But when both team building and different endowment types, rates for both low and high types increase sharply, to 94.6 percent and a dramatically higher 87.9 percent. In this case, the high types seem to have a stronger sense of cooperation, apparently stemming from their experience of

[27] 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.

having cooperated in a successful group enterprise. The fact that the different types need time to sort themselves out is in part due to our one-at-a-time design, and in any case seems likely to be relevant with a longer horizon.

How does this feeling of community spirit transfer from one's word-task-group identity to one's endowment-type identity? 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 very small team project.[28] 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 G 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 NG 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 G treatments.[29]

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 are from 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.

5. 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

[28] We felt it most effective to set the threshold for success low enough to ensure near-universal 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.

[29] $Z = 0.218$, $p = 0.827$; $Z = 0.150$, $p = 0.881$, 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 last 12 periods at a session level.

some people receive twice as much as others. These two factors can be seen as different forms of identities and it is an open question as to which of these is more salient; we are unaware of previous research on this topic. 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.

Perhaps the most 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. We also find that 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 also participated in a team-building exercise. While in some sense this is intuitive (one acts more positively in a positive environment), 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.

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.

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.

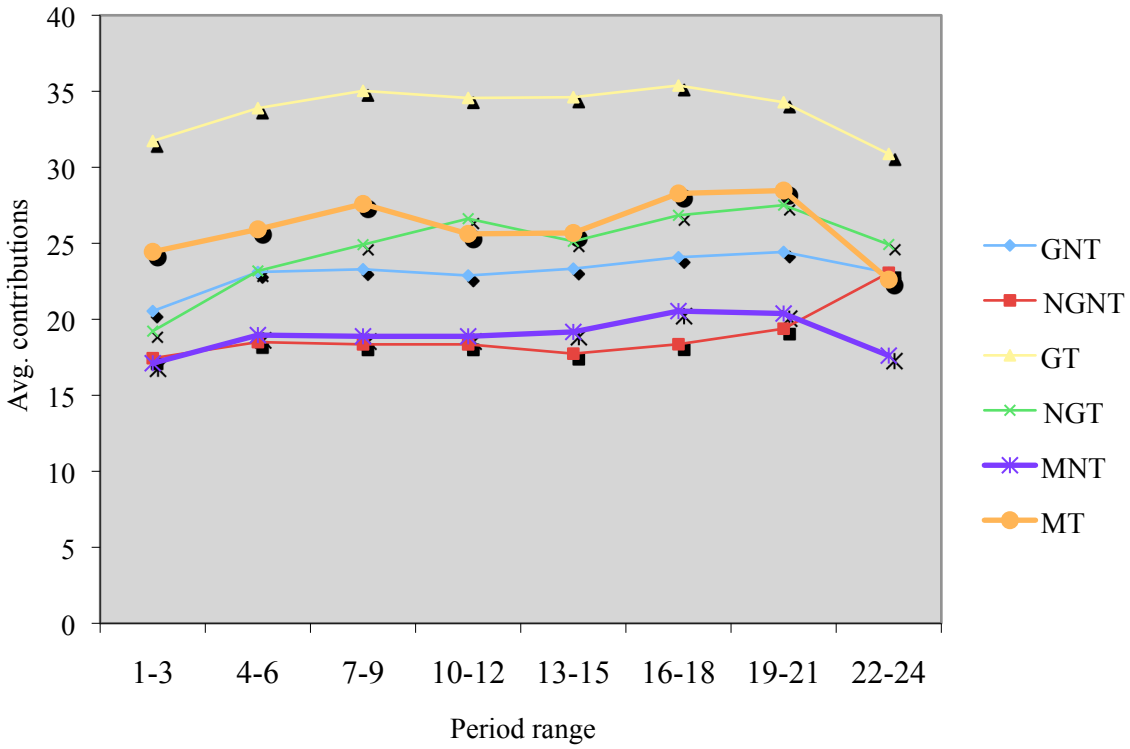
Brown-Kruse, J. and D. Hummels (1993), "Gender Effects in Laboratory Public Goods Contribution: Do Individuals Put Their Money Where Their Mouth Is?," *Journal of Economic Behavior and Organization* 22(3), 255-67.

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.
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.
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.
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.
Page, T., L. Putterman, and B. Unel (2005), "Voluntary Association in Public Goods Experiments: Reciprocity, Mimicry, and Efficiency", Economic Journal, 115: 1032-1053.
Riedl, A. and A. Ule (2002), "Exclusion and Cooperation in Social Network Experiments," mimeo.

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.
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	0.296*** (0.037)	0.072*** (0.023)	0.038* (0.023)	
Group size	1.139*** (0.317)		0.862*** (0.293)	0.006 (0.009)
Group size*high type				0.031** (0.013)
Dummy t1t2	-3.448*** (0.774)			
High type			14.243*** (0.828)	-0.256*** (0.055)
Number of highs in the group		1.267*** (0.325)		
Same type as the majority of the group		0.260 (0.771)		
Same type as majority of the group*high type		4.070*** (1.266)		
Same word-task group as majority of contribution group	0.389 (0.679)	0.514 (0.594)	0.276 (0.586)	0.002 (0.014)
Dummy t3t4		-8.413*** (1.026)	-9.060*** (0.938)	-0.204*** (0.028)
Average lagged % group contribution				0.073*** (0.026)
Constant	11.960*** (1.416)	27.271*** (1.269)	8.129*** (1.797)	1.198*** (0.086)
R-squared	0.345	0.473	0.554	0.298
N	972	1611	1611	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 (NGNT) and 0 when it is from Treatment 2 (GNT). The variable "dummy t3t4" takes the value 1 when the observation is from Treatment 3 (NGT) and 0 when it belongs to Treatment 4 (GT).
The number of observations in (1) represents period 1 to period 12 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 1-12. Specification (5) includes all observations in Treatments 3 and 4.

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.

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.

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
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						0.448** (0.179)	-0.001 (0.031)
% contribution of the single			0.250** (0.106)	0.409*** (0.077)	0.378*** (0.066)	0.959*** (0.327)	0.268*** (0.071)
Type HH			0.223*** (0.036)				
Type LL				-0.121** (0.049)			
Dummies for half-segments	YES	YES	YES	YES	YES	YES	YES
LL	-98.897	-88.854	-149.745	-93.265	-249.591	-21.616	-139.944
N	192	192	340	306	646	40	343

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. Specifications (6) and (7) include observations from only GNT and GT, respectively.