American Economic Review 2009, 99:1, 295–323
http://www.aeaweb.org/articles.php?doi=10.1257/aer.99.1.295

The Value of Groups

By Shaun P. Hargreaves Heap and Daniel John Zizzo*

We present the results of an experiment that attempts to measure the social value of groups. In the experiment, group membership is induced artificially: subjects interact with insiders and outsiders in trust games and periodically enter markets where they can trade group membership. We find that trust falls with groups because of negative discrimination against outsiders. Against this, however, there is evidence that group membership provides a psychological benefit, albeit one that may induce social inertia. Overall, the welfare effects of groups are at best neutral and could be negative. (JEL: D17, Z13*)*

People belong to groups. Groups vary enormously. Some are religious, others are ethnic. Some arise from family or kinship ties, others from work, shared interests, or a political commitment. A person's attachment to their group can be strong or weak, and the sense in which one can be said to belong to a group varies greatly. Notwithstanding these nuances and complexities, group membership, broadly understood, is a ubiquitous feature of economic and social life. Perhaps somewhat surprisingly in this context, economists have not been, at least until recently, especially interested in how belonging to a group affects an individual.

There are notable exceptions to this inattention. It has long been recognized that groups might form around a collective action problem, and so deliver, often sectional, benefits for their members (see Mancur Olson 1965). More recently, groups have been cast as a form of social capital because they allow members to trust each other in ways that nonmembers do not. As a result, exchanges between group members are usually thought to incur smaller transaction costs (i.e., waste fewer resources) than would otherwise be the case (see, e.g., Francis Fukuyama 1995; Economic Journal 2002).

These benefits from group membership are material in character. By contrast, group membership might also be a source of a separate and distinctive kind of psychological benefit. George A. Akerlof and Rachel E. Kranton (2000), for instance, suggest that simply being able to identify with a group is itself an important source of individual well-being. Adam Smith (1759/1976) argued in a similar fashion that people enjoy the "special pleasure of mutual sympathy" that comes from belonging to a group. The recent empirical work on happiness and well-being lends support to this idea. Belonging to a rich set of social networks seems to be one of the determinants of reported happiness in these studies (e.g., Richard Layard 2005); and there is some neurobiological evidence that being a member of a group produces an endorphin rush (see Robin Dunbar 2006).

* Hargreaves Heap: School of Economics, University of East Anglia, Norwich NR4 7TJ, UK (e-mail: s.hargreavesheap@uea.ac.uk); Zizzo: School of Economics, University of East Anglia, Norwich NR4 7TJ, UK (e-mail: d.zizzo@uea.ac.uk). We wish to thank Graham Loomes, Bob Sugden, Lyndsey Stonebridge, Caroline Lowton, three anonymous referees, and participants to presentations in Aberdeen, Athens, Dublin, East Anglia, Nottingham, Nottingham Trent, and Rome for useful advice, Kei Tsutsui and Lindsay Steenberg for research assistance, and the Nuffield Foundation and the University of East Anglia for financial support. The usual disclaimer applies. The experimental instructions, together with the other electronic appendices, are available online at: http://www.aeaweb.org/articles.php?doi=10.1257/aer.99.1.295.

Finally, there is some evidence from experimental economics that membership in a group matters for individuals in the sense that it can affect their behavior in prisoner's dilemma and battle of the sexes games (see Gary Charness, Luca Rigotti, and Aldo Rustichini 2007; Lorenz Goette, David Huffman, and Stephan Meier 2006), in the public goods game (e.g., Jonathan H. W. Tan and Friedel Bolle 2007) and in bargaining settings (Hargreaves Heap and Yanis Varoufakis 2002; Zizzo 2003).

In this paper, we investigate experimentally the potential significance of the social capital and the psychological benefits of group membership on individual welfare in a trust game. Both effects are intuitively plausible and they could have important policy implications (see John F. Helliwell 2006). For example, insofar as groups have these effects on individual welfare, the cost-benefit evaluations of policies should (but typically do not at present) pay attention to how any policy influences the constellation of groups in society. Likewise, firms may benefit from the development of a corporate culture that brings a sense that employees belong to a group; and, if this is the case, it will be important for employers, in realizing these potential gains, to be able to disentangle the social capital effect from the psychological one because the latter is, in effect, a nonmaterial benefit experienced by employees and should therefore affect the design of the employment contract.

We have chosen an experimental investigation because, while there is evidence from survey data linking groups with trust and well-being, there are doubts as to whether the answers to questions in these surveys about the extent to which people trust others or experience happiness are real in the sense that they are correlated with differences in actual behavior. For example, Edward L. Glaeser et al. (2000) found that the reported answers to survey questions on trust were often not associated with how subjects actually trusted one another in an experimental trust game. The advantage of the experimental method is precisely that it allows us to examine in a controlled fashion whether belonging to a group really affects individuals' behavior.

Both the social capital and the psychological benefits of group membership potentially arise through two conceptually distinct routes. There could be a "pure" effect that comes from belonging to a group per se, and an idiosyncratic influence which owes its character to the particular constitutive norms and other aspects of the group in question. In natural groups the two effects combine and are difficult to disentangle. For this reason, we induce group membership artificially within the experiment. This avoids importing any existing stereotypes or expectations that come with natural groups, and the results are therefore more likely to distinguish what, if any, the "pure" effects of groups on behavior and welfare are. This approach has a further advantage. Once the "pure" effects of group membership have been identified, they can be used in the future to disentangle the two types of influence in natural groups. In other words, our results form a potential baseline for future studies that attempt to identify the particular contribution that comes from an actual group's constitutive norms.

Thus we test for the potential "pure" group social capital effect by considering whether the existence of "artificial" groups increases trust in a trust game experiment where subjects can belong to one of two groups. The use of a trust game provides a direct connection to the work on social capital. However, our experiment differs in one key respect from related work on public goods games where there is some evidence that induced feelings of group identity have a positive effect on contributions (e.g., Richard Cookson 2000). In those studies, group feelings were encouraged in a context where there was only *one* group. We believe, like Charness, Rigotti, and Rustichini (2007), however, that group feelings more commonly arise where there is more than one group and our experimental design reflects this.

An additional virtue of the two-group frame is that it allows us to explore an important question about whether any difference in trust between insiders and outsiders comes from positive discrimination in favor of insiders or negative discrimination against outsiders. Both forms of

discrimination produce a difference between interactions with fellow members as compared with those involving nonmembers, but while the former would make group creation a genuine type of social capital formation in the sense that the existence of groups improved welfare, the latter would make groups a form of negative social capital because their existence actually lowers welfare. In our experiment, there is a difference in trust between insiders and outsiders, but we find that it arises because of negative discrimination against outsiders. Thus, contrary to what seems to be the presumption in most of the social capital literature (though not all of it: see Robert Putnam 2000), the existence of groups in our experiment tangibly reduces trust in the aggregate, and thus is welfare reducing.

We examine the possible additional psychological benefits of group membership by introducing an experimental market into the play of the trust game. In this market subjects can trade group membership. This trading opportunity provides an incentive-compatible mechanism for assessing the value that individuals place on their own group membership. This can then be used to generate estimates of the extent to which individuals value group membership for psychological or nonmaterial reasons (that is, for reasons beyond the material effects that arise from the influence of groups on the level of trust). We find that people do attach positive value to membership of their own group, beyond what would be expected from the material effects of groups. Interpreting this finding is complicated by a well-known wedge in experiments between the willingness-to-pay (WTP) and the willingness-to-accept (WTA) compensation that can arise through, for example, the influence of reference dependence effects. If the result is taken at its face value, by combining this positive psychological effect with the negative social capital result, we estimate that the net social value of groups is at best roughly neutral. Alternatively, if the psychological benefit revealed in the market phase is discounted for reasons of reference dependence or the like, then the net social value is negative.

Therefore, while the general impression from the literature on social capital and on well-being is that groups are a good thing because they boost trust and additionally improve their members' reported subjective happiness, the evidence from our experiment is more mixed. We do *not* find that membership of groups increases trust; indeed the presence of groups seems to lower trust. But there is some, albeit significantly qualified, evidence that group membership yields a distinct psychological benefit. Section I sets out the experimental design. Sections II and III present the results, Section IV discusses the results, and Section V concludes.

I. Experimental Design

A. *Outline and Stage 1*

The experiment was conducted between March 2006 and July 2007 at our university.[1] Apart from the experimental instructions and a control questionnaire, the experiment was fully computerized. Almost all subjects were university students, from a wide variety of subject backgrounds. A total of 308 subjects participated in the 26 sessions: we scheduled 12 subjects per session, but one session was run with 8 subjects due to no-shows. We had an international mix of subjects, with around 40 percent of the subjects (129) non-British; the second most represented nationality was Chinese (25 subjects), and an overall 45 subjects were East Asian (online Appendix B reports more details). Subjects were randomly seated in the laboratory. Computer terminals were partitioned to avoid communication by facial or verbal means. Subjects read the experimental instructions and answered a control questionnaire, to check understanding of the instructions

[1] The experimental instructions are provided in online Appendix A.

Table 1—Experimental Sequence

Experimental sequence	Task	Number of rounds
Stage 1	Trust games	3
Stage 2	Market 1 or Waiting Period 1, trust games	6
Stage 3	Market 2 or Waiting Period 2, trust games	6
Stage 4	Market 3 or Waiting Period 3, trust games	6

Notes: At the start of each of stages 2, 3, and 4, the experiment had a waiting period in the B treatment, and markets for groups in the C, SG, and SI treatments (technically, there were two markets, one to pay for membership of each of the two groups); in the SM treatment, there was a waiting period at the start of stage 2 and there were markets at the start of stages 3 and 4.

before proceeding with the tasks. Experimental supervisors individually advised subjects with incorrect answers in the questionnaires. The experimental instructions had a neutral frame (e.g., did not refer to "trust," "trusters," or "trustees"). The experiment used "experimental points" as currency, each worth 4 UK pence (0.04 pounds).

There were six experimental treatments: baseline (B), color group assignment (C), group segregation (SG), and three variants on the SG treatment: group segregation with reduced framing (SF); group segregation with reduced markets (SR); and group segregation and incentives (SI). Each session was divided into four stages. There were five sessions in the B treatment and four planned sessions in each of the others, but, to compensate for the fact that one session with eight subjects was run in the SG treatment, we ran a fifth session in this treatment as well.

Stage 1 had three rounds and was common to all treatments. Each round was a standard Joyce Berg, John W. Dickaut, and Kevin A. McCabe (1995) basic trust game. The truster (the "first mover") received 24 experimental points and had to decide how many points (if any) to give to the other person and how many (if any) to keep. All the points given were multiplied by a *conversion rate* equal to 3 before they were received by the trustee (the "second mover"). The trustee then decided how much (if any) to keep and how much (if any) to return to the truster. Subjects were matched randomly and anonymously each round, with the constraint that they would hold the role of truster and that of trustee at least once.[2] The only information they received was about their round coplayer's decision and about their own round earnings; for example, in treatments with groups (C, SG, SF, SR, and SI), they had no information about the color group of coplayers. The key purpose of stage 1 was to provide subjects practice and experience with trust games. We now move on to the specifics of each treatment. For the sake of clarity, the flow of the experiment is represented in Table 1.

B. *The Baseline (B) Treatment*

In the B treatment, stages 2, 3, and 4 were very similar to stage 1. Each stage had six rounds rather than three. As in stage 1, each round consisted of the basic trust game, but at the start of each round 48 points rather than 24 were given to trusters. To mirror the information provided in stages 2, 3, and 4 of the other treatments (as described below), the computer screen displayed information on average giving rate and average return rate, with a summary table on average giving and return rates from stage 2 onward being provided at the end of each stage. Each stage was otherwise identical to stage 1.

[2] They were asked to make decisions within one and a half minutes, and a small clock on the computer display informed them how much time they had. In practice, however, they could take more, though they rarely did.

Between stages there was a two-minute waiting period, at the start of which subjects were paid an additional 48 points. Again, this was meant to mirror the other treatments, both by providing the same money amounts and by creating a temporal wedge between trust game tasks.

C. *The Color Group Assignment (C) Treatment*

At the start of the experiment, subjects were randomly assigned to either the blue group or the red group; six participants were assigned to each group.[3] Stages 2 through 4 were divided into two phases.

Trust Games Phase.—In stages 2, 3, and 4 subjects played six trust games as in stage 1, but with the following differences. Each round, trusters were allocated 48 points rather than the 24 of stage 1. They were randomly matched with coplayers within their group for three rounds out of six, and with coplayers from the other group for the remaining three rounds; they were told in the experimental instructions that this would be the case.[4] In each round they were informed whether the coplayer belonged to the blue group or to the red group, though they were not told their identity.[5] They were assigned at least once the role of trusters and at least once that of trustees with respect to both insiders and outsiders. They were provided, on a round-by-round basis, with a table containing information on average giving rates and average return rates by members of each group with respect to insiders and outsiders (see Figure 1): that is, from blue group to blue group members, from blue to red, from red to blue, and from red to red.

In addition, they received a summary table with average giving and return rates for each stage from the second onward by members of each group with respect to insiders and outsiders.

Market for Groups Phase.—Before stages 2, 3, and 4 of the trust games were played, subjects had an opportunity to change color groups, provided there was a trader belonging to the other group willing to swap places at a mutually acceptable price. We introduced this phase because, in principle, market mechanisms supply an incentive-compatible mechanism for eliciting individual valuations and in practice there is evidence that when they are repeated some of the well-known experimental decision anomalies notably diminish (see James C. Cox and David M. Grether 1996; Jason F. Shogren et al. 2001).[6]

In particular, subjects were given an endowment of 48 points and first asked to state whether, if they could choose and both options were free, they would rather stay in their group or switch to the other. If they stated they would rather switch, then they became potential buyers for the membership of the other group and they were asked how much they were willing to pay to swap places with a member of the other group. They could state any value between 0 and 48 points, the

[3] A similar minimal group manipulation has been used by Hargreaves Heap and Varoufakis (2002).

[4] It is possible that, in addition to an effect associated with being a member of the blue or red group, people also experience a sense of belonging to the group formed by being part of the same experiment. This effect should *reduce* any discrimination between blue and red (sub)groups (since they would be moderated by the all coplayers belonging to the same experiment) and thus make any discrimination that we actually find the more convincing (see Section II for further discussion).

[5] Although anonymity should make each interaction a one-shot game, it is possible that people falsely believe they are engaging in a repeated game where they can influence a coplayer's future behavior through their own behavior now. We attempted to avoid this by ensuring that, in stages 2–4, the average likelihood of playing with the same person again is approximately the same with C treatment insiders (1 out of 11 for any given round, in sessions with 12 subjects), C treatment outsiders (1 out of 12), and B treatment coplayers (again, 1 out of 12). Over 15 rounds, this means that subjects are likely to be matched on average just a little more than once. We note that insofar as this aspect of the design fails to avoid the impression of playing a repeated game, this effect would likely induce positive rather than negative discrimination.

[6] Thanks to Graham Loomes who suggested this mechanism to us.

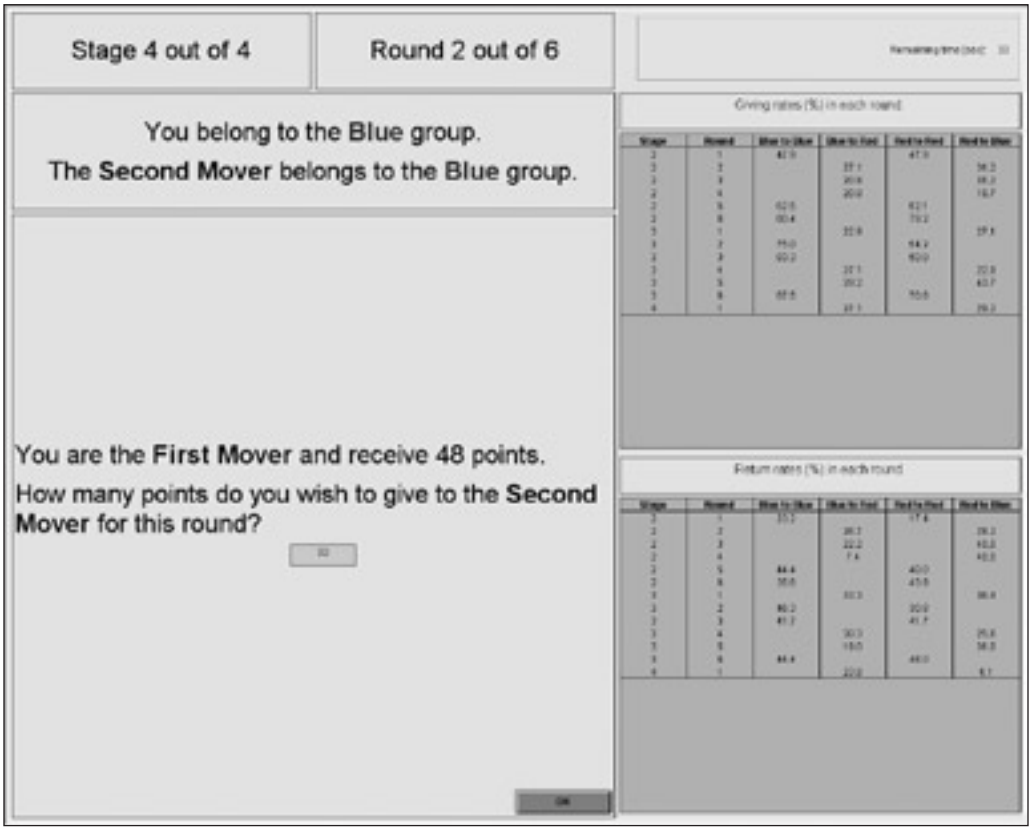

FIGURE 1. SAMPLE COMPUTER DISPLAY

Notes: The sample computer display is from the C treatment. After each round a new line was added to the giving rate and the return rate tables. In each stage 2–4, round subjects were either all matched with insiders or all matched with outsiders.

value of their endowment. Using this method we measured the WTP of agents, with a common upper limit of 48 points chosen to avoid bankruptcy problems or the dependence of the WTP range on previously earned money.

Similarly, if subjects stated they would rather stay, they became a potential seller of group membership and were asked to state how much they would need to be paid by a member of the other group in order to swap places, again with a an upper limit of 48 points. Subjects were also given the option to state that they were not willing to switch groups at any price within the allowed range (0 to 48 points). Using this method, we obtained information on the WTA of agents.

The market then operated as a Walrasian clearinghouse, where the price was set so that the number of sellers was equal to the number of buyers of membership of the other group. Whenever there was a range of possible market-clearing prices, the lowest market-clearing price was chosen. Crucially, the mechanism operated only by swapping players between groups, so that each group remained with six subjects throughout the experiment.[7]

[7] Or the group had four players, in relation to the one session with eight subjects. Subjects were told that they should make their market decisions within four minutes.

D. *The Group Segregation (SG) Treatment*

This treatment was exactly as the C treatment, but with one difference designed to pick up on the way that members of groups often interact more frequently with each other than with outsiders. In evolutionary game theory this is referred to as "associative" matching and has been used to explain how cooperation occurs within groups but not between them (see Theodore C. Bergstrom 2002). Specifically, in the SG treatment they played twice as frequently with insiders as with outsiders. In each stage they were matched four times with insiders (twice as trusters, twice as trustees) and only two times with outsiders (once as trusters and once as trustees). Subjects were informed about this in the initial experimental instructions.[8]

E. *The Group Segregation with Reduced Framing (SF) Treatment*

This treatment was the same as the SG treatment except for the framing. Instead of having a red and blue group, we used "blue" and "not blue" as labels for the group. This could reduce the feeling of group identity in the not blue group and so help disentangle whether our measure of psychological value is picking up a genuine psychological benefit or simply reflects the well-known WTP/WTA bias due, say, to reference dependence effects, which would apply equally to blue and not blue.

F. *The Group Segregation with Reduced Markets (SM) Treatment*

This SM treatment variant differed from the other segregated ones by substituting a waiting period for the market phase at the beginning stage 2. This is to test for the possible influence that the first market phase has on the salience of group membership in stage 2.

G. *The Group Segregation and Incentives (SI) Treatment*

This treatment was the same as the SG treatment except that we introduced an element of material competition between the groups. Again this variation was designed to introduce a feature that is sometimes present in intergroup relations, and we implemented it through a variable multiplication factor for gifts. If blue trustees had been given more in a certain stage, all points given to them were multiplied by four and those given to red trustees were multiplied by two. If red trustees had been given more, all points given to them were multiplied by four and those given to blue trustees were multiplied by two.[9]

This incentive structure was a trust game adaptation of the marginal incentive scheme present in the public good literature on team competition (Jonathan H. Tan and Friedel Bolle 2007). For comparability with the literature on team competition,[10] which provides the natural benchmark, we chose incentives to be a function of round (rather than, say, stage or session overall)

[8] Another possible source of difference between behavior in C and SG arises when anonymity does not have the effect of people treating interactions as a one-shot game (see footnote 5). In these circumstances, we might expect that the greater frequency of interaction with insiders than with outsiders should encourage positive discrimination (though not negative discrimination) and so increase the gap between how insiders and outsiders are treated in SG as compared with SI.

[9] Subjects were told that, if blue and red trustees received the same, the group that got their points multiplied by four rather than by two was chosen at random. In practice, though, a tie never occurred in the experiment.

[10] See, for example, Haig R. Nalbantian and Andrew Schotter (1997), Gary Bornstein, Uri Gneezy, and Rosmarie Nagel (2002), and Tan and Bolle (2007).

performance. The incentive structure could induce large disparities in winnings between groups, which, in principle, could have then been picked up by markets for groups in later stages.

H. *Payments*

Each session lasted a little over one hour. The average earning was 12.50 UK pounds per subject (approximately 25 US dollars). Payment was based on the earnings from each of the markets (or of the waiting periods for the B treatment) plus those from a randomly chosen round from each of the four stages.[11] Subjects were privately paid and left the laboratory one at a time in an order designed to minimize the likelihood of seeing each other.

II. Experimental Results on Behavior in Trust Games

Let the *giving rate* be the fraction of the endowment given by trusters to trustees, and let the *return rate* be the fraction of the amount received by trusters which is returned by the trustees to the trusters (where the amount received by trustees is three times what was given in all treatments except SI, where it is either two or four times what was given). Figure 2 and Table 2 show the average giving and return rates in each experimental treatment.

In all the group treatments (C, SG, SF, SM, and SI) we observe discrimination between insiders (i.e., members of the same group) and outsiders (i.e., members of the other group) (see Figure 2). In all 21 sessions, the mean giving rate to a fellow insider was higher than to an outsider (Wilcoxon $p < 0.001$). In 17 sessions out of 21, the mean return rate was higher when interacting with a fellow insider than with an outsider (Wilcoxon $p < 0.01$), though possibly with less across-treatment robustness.[12]

This leaves open whether this discrimination is positive or negative in origin. Stage 1 is, of course, the practice stage common to all treatments, whereas stages 2 through 4 are differentiated across treatments, and we find in the aggregate that introducing groups did not raise trusting behavior. In fact, the opposite happened. Although giving rates were roughly the same in all treatments in stage 1, their mean value in stages 2–4 with respect to outsiders was statistically significantly lower in the 21 sessions with color groups (C, SF, SG, SM, and SI treatments) than in the 5 B sessions (Mann-Whitney $p = 0.003$).[13] Likewise, the mean return rates toward outsiders in the B sessions were statistically significantly higher than those in the color sessions (Mann-Whitney $p = 0.041$). In marked comparison, we cannot reject the hypothesis that giving and return rates toward insiders are the same in the color treatments as in the baseline (Mann-Whitney $p = 0.380$). This is the preliminary evidence for negative discrimination.

To build on this, we run regressions controlling for session-specific effects with the stages 2–4 mean giving rate by subject to insiders or outsiders as dependent variables (regressions 1–4 in Table 3).[14] Differentiating between insiders and outsiders in these regressions allows us to test in a precise way whether there is positive or negative discrimination relative to the B treatment baseline.

[11] Since stage 1 had 3 rounds and 24 points given to trusters, and stages 2 through 4 had 6 rounds but double the number of points given to trusters, the marginal incentives were the same for each of the four stages.

[12] Two of the four exceptions are among the five SG sessions; there is also one exception each from the SF and SI treatments.

[13] Statistical significance is estimated by treating session averages as the unit of observation, in order to control for possible nonindependence of choices within each session. In addition, in this paper all reported p values are two tailed, except where otherwise specified.

[14] Distributions of mean giving and return rates by subject to insiders and outsiders can be found in online Appendix C. In relation to the B treatment, we simply use the overall average mean giving and return rates for each subject as the dependent variable, as there is no differentiation between insiders and outsiders.

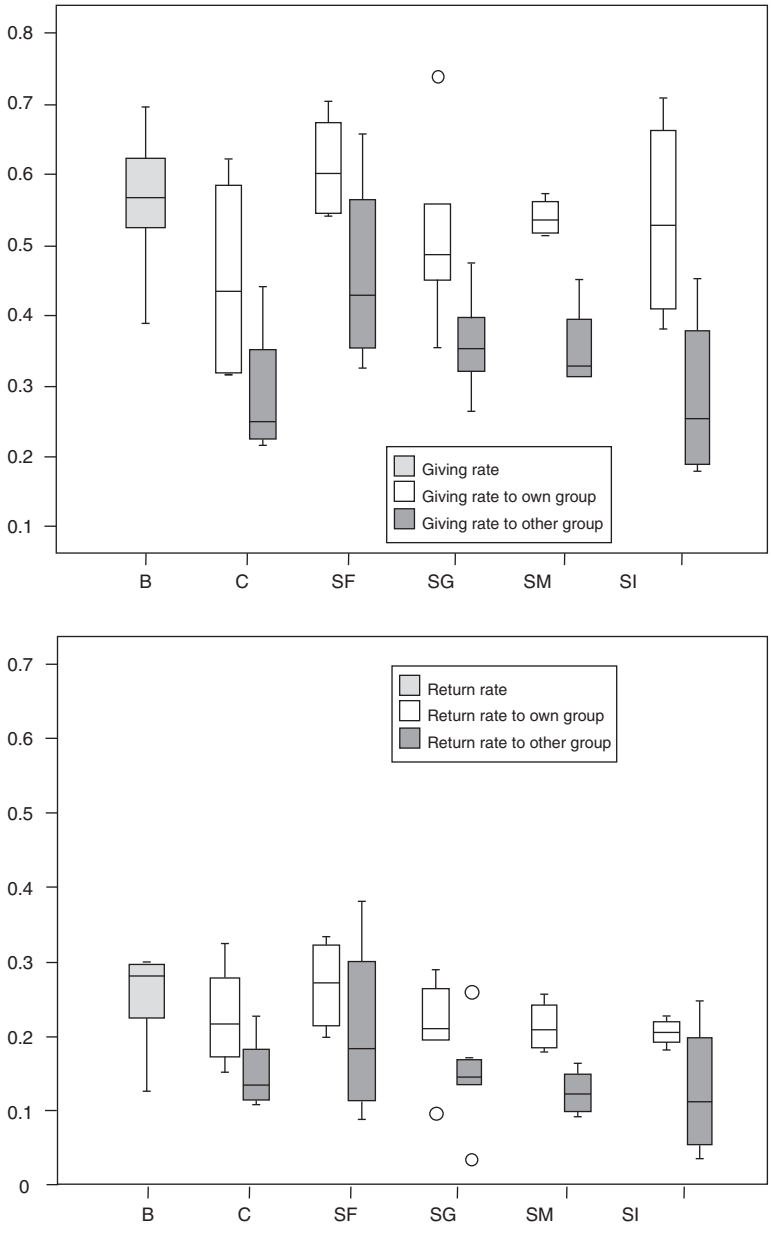

FIGURE 2. GIVING AND RETURN RATES IN STAGES 2–4

Notes: Giving and return rates by session in each treatment, as employed in the statistical analysis in the text ($n = 4$ in the C, SF, SM, and SI treatments; $n = 5$ in the B and SG treatments). The median value is the middle bar, the edges of the box represent the 25th and 75th percentile, whiskers include observations within 1.5 of the box length, and circles represent any other observation.

We perform the same analysis for the return rate, distinguishing again between whether it is to insiders or outsiders (regressions 5–8), except there is one difference. One problem with interpreting the difference in return rates between the baseline and group treatments is that subjects may simply return proportionally less because they have been given less. This might occur

TABLE 2—GIVING AND RETURN RATES

	B	C	SG	SF	SM	SI
Stage 1 giving rate	0.562	0.492	0.516	0.534	0.511	0.498
Stages 2–4 giving rate	0.558	0.369	0.453	0.559	0.476	0.498
to insiders		0.449	0.500	0.609	0.538	0.534
to outsiders		0.288	0.359	0.457	0.354	0.284
Stage 1 return rate	0.318	0.202	0.267	0.298	0.326	0.296
Stages 2–4 return rate	0.252	0.186	0.193	0.253	0.187	0.194
to insiders		0.229	0.210	0.269	0.213	0.205
to outsiders		0.149	0.158	0.217	0.125	0.137

for a number of psychological motives that had been documented in other experiments, such as inequality aversion (Ernst Fehr and Klaus M. Schmidt 1999), reciprocity (Armin Falk and Urs Fischbacher 2001), or trust responsiveness (Gerardo Guerra and Zizzo 2004). The return rates regressions have the Giving Rate received by the Second Mover as an independent variable, that is, the mean stages 2–4 giving rate the subject has received when playing as a trustee. This allows us to control for the positive relationship which we might expect between giving rate and return rate.

In the regressions, error clustering is used to take into account the possible nonindependence of observations by different subjects in the same session.[15] We have a dummy variable, session 3, equal to one for the one session (session 3) which had 8 subjects rather than 12 to control for possible group size effects. There are also a number of dummy variables that capture individual-specific heterogeneity: stage 1 giving and return rate, age, and dummies for gender (= 1 for women), economics or management educational background (= 1 if applicable), nationality (UK = 1 for UK subjects and China = 1 for Chinese subjects) and religious affiliation (Christian = 1 for Christian subjects and Agnostic Atheist = 1 for agnostic or atheist subjects, the two largest affiliations). There are also key dummy variables for the experimental treatment, using as a baseline the B treatment with no groups. In regressions 1, 3, 5, and 7 we employ dummy variables for each treatment with groups (*C*, *SF*, *SG*, *SM,* and *SI* = 1 for sessions in the C, SF, SG, SM, and SI treatment, respectively). In regressions 2, 4, 6, and 8 we use, instead, a single groups dummy variable (= 1 for all treatments with groups).

There are no significant coefficients on any of the group dummies in regressions 1 and 2 on giving to insiders.[16] This apparent absence of positive discrimination is reinforced by the fact that virtually the same proportion of subjects gave and returned 0 in the B treatment (9.5 and 31.4 percent of the subjects, respectively) as in the color groups with respect to insiders (10.6 and 31.6 percent, respectively), whereas in the group treatments over twice as many gave 0 and around 50 percent more returned 0 to outsiders (23.3 and 47.1 percent, respectively).

RESULT 1: *The creation of groups did not affect mean giving rates to insiders, so there is no evidence of positive discrimination.*

[15] Online Appendix D reports broadly the same picture when standard or Tobit random effects regressions are used instead.

[16] There are actually signs that insiders may be returning less than baseline subjects (notably, the *F*-test restricting all treatment dummies coefficients to zero in regression 1 is rejected at $p = 0.048$), but the corresponding coefficients are always smaller in size than with respect to outsiders and the global groups dummy in regression 2 remains statistically insignificant ($p = 0.548$).

TABLE 3—REGRESSIONS ON MEAN GIVING AND RETURN RATE

Regressions on stage 2–4 mean giving rate	Regression 1			Regression 2		
	β	t	p	β	t	p
To own group						
Stage 1 giving rate	0.026	0.46	0.652	0.018	0.32	0.755
Stage 1 return rate	0.54	9.23	0	0.533	9.57	0
Groups				−0.027	−0.61	0.548
C	−0.112	−1.53	0.137			
SF	−0.016	−0.36	0.725			
SG	−0.055	−1.17	0.255			
SM	0.052	0.88	0.389			
SI	−0.054	−0.7	0.493			
Session 3	−0.117	−1.89	0.07	−0.044	−1.25	0.222
Gender	−0.042	−1.38	0.18	−0.05	−1.68	0.106
Christian	0.03	0.76	0.454	0.028	0.68	0.506
Agnostic Atheist	0.051	1.08	0.292	0.052	1.08	0.292
EcMgt	0.061	1.66	0.11	0.063	1.71	0.099
UK	0.012	0.41	0.686	0.006	0.18	0.857
China	−0.119	−2.11	0.045	−0.132	−2.32	0.029
Age	0.004	1.74	0.093	0.004	1.72	0.098
Constant	0.196	2.32	0.029	0.186	1.99	0.057
R^2	0.34			0.325		
	Reg 1 groups coeffs. = 0: $F = 1.24$ ($p = 0.318$)					
	Regression 3			Regression 4		
To other group						
Stage 1 giving rate	−0.123	−2.04	0.052	−0.127	−2.18	0.039
Stage 1 return rate	0.432	7.2	0	0.429	7.13	0
Groups				−0.214	−4.91	0
C	−0.294	−6.12	0			
SF	−0.171	−3.04	0.006			
SG	−0.205	−4.35	0			
SM	−0.131	−1.8	0.084			
SI	−0.319	−5.08	0			
Session 3	−0.132	−2.16	0.04	−0.053	−1.27	0.216
Gender	−0.044	−1.28	0.214	−0.046	−1.39	0.175
Christian	0.025	0.69	0.497	0.033	0.82	0.421
Agnostic Atheist	0.034	0.73	0.473	0.038	0.79	0.435
EcMgt	−0.022	−0.66	0.513	−0.004	−0.12	0.902
UK	−0.042	−1.45	0.159	−0.049	−1.56	0.132
China	−0.133	−3.13	0.004	−0.15	−3.15	0.004
Age	0.005	2.99	0.006	0.005	2.86	0.008
Constant	0.362	5.58	0	0.338	4.65	0
R^2	0.334			0.299		
	Reg 3 groups coeffs. = 0: $F = 15.85$ ($p < 0.001$)					

TABLE 3—REGRESSIONS ON MEAN GIVING AND RETURN RATE (*continued*)

Regressions on stage 2–4 mean giving rate	Regression 5			Regression 6		
	β	t	p	β	t	p
To own group						
Trust rate as 2nd mover	0.231	6.2	0	0.226	5.58	0
Stage 1 giving rate	−0.046	−1.52	0.142	−0.047	−1.55	0.133
Stage 1 return rate	0.194	5.13	0	0.193	5.39	0
Groups				−0.029	−1.51	0.143
C	−0.015	−0.64	0.525			
SF	−0.026	−1.17	0.252			
SG	−0.033	−1.28	0.211			
SM	−0.016	−0.51	0.616			
SI	−0.064	−2.97	0.006			
Session 3	−0.065	−2.25	0.033	−0.053	−3.6	0.001
Gender	−0.038	−2.04	0.052	−0.038	−2.02	0.055
Christian	−0.001	−0.02	0.982	0.001	0.05	0.962
Agnostic Atheist	−0.001	−0.04	0.966	0	−0.01	0.992
EcMgt	0.006	0.26	0.794	0.006	0.25	0.804
UK	−0.029	−1.7	0.102	−0.03	−1.77	0.089
China	−0.028	−0.84	0.408	−0.023	−0.71	0.482
Age	0.004	2.81	0.01	0.005	2.98	0.006
Constant	0.005	0.09	0.925	−0.001	−0.03	0.979
R^2	0.267			0.258		
	Reg 5 groups coeffs. = 0: $F = 2.64$ ($p = 0.048$)					
	Regression 7			Regression 8		
To other group						
Trust rate as 2nd mover	0.152	3.64	0.001	0.172	3.35	0.003
Stage 1 giving rate	−0.115	−2.6	0.015	−0.112	−2.7	0.012
Stage 1 return rate	0.173	4.82	0	0.176	4.92	0
Groups				−0.069	−2.71	0.012
C	−0.104	−5.17	0			
SF	−0.05	−1.15	0.262			
SG	−0.061	−2.19	0.038			
SM	−0.072	−1.6	0.121			
SI	−0.087	−2.53	0.018			
Session 3	−0.072	−1.95	0.063	−0.068	−3.35	0.003
Gender	−0.021	−0.95	0.351	−0.02	−0.94	0.355
Christian	−0.012	−0.41	0.689	−0.009	−0.3	0.77
Agnostic Atheist	−0.025	−0.87	0.392	−0.023	−0.78	0.445
EcMgt	−0.025	−0.95	0.35	−0.022	−0.89	0.381
UK	0	−0.02	0.986	−0.002	−0.08	0.933
China	0.058	0.99	0.333	0.05	0.86	0.396
Age	0.007	3.08	0.005	0.007	2.89	0.008
Constant	0.02	0.25	0.804	0.007	0.08	0.938
R^2	0.246			0.238		
	Reg 7 groups coeffs. = 0: $F = 5.64$ ($p = 0.001$)					

Notes: Sample size: $n = 308$ (regressions 1–6) and 302 (regressions 7–8). The lower number of observations in regressions 7–8 is due to zero giving from outsiders. We employ error clustering to control for session level effects. The p-values provided are two tailed. The F-tests test whether the restriction that the coefficients on *C*, *SF*, *SG*, *SM*, *SI* be jointly equal to zero is accepted.

The comparison with the regression on behavior toward outsiders is marked: all the group dummies in the giving to outsiders dummies (regressions 3 and 4) are significantly negative. Only the *SM* dummy is at the borderline of statistical significance ($p = 0.084$), all other dummies being significant at $p < 0.001$. A *F*-test restricting all the treatment dummies to zero is clearly rejected ($p < 0.001$), and the aggregate groups dummy of regression 4 is again significant ($p < 0.001$). Overall, controlling for all the other variables, outsiders are given 21 percent less (regression 4) and returned 7 percent less (regression 8) as the result of negative discrimination relative to the no groups baseline.

RESULT 2: *The creation of groups reduced mean giving rates to outsiders. The minimal group inducement of the C treatment was sufficient to produce this negative discrimination result.*[17]

Turning to the regressions on the return rate, the giving rate received when second mover is statistically significant ($p < 0.005$) in these equations: subjects who receive more return proportionally more. This fits with previous findings on trust games. The negative discrimination effect is less unequivocal for return rates: the coefficients on *SF* and *SM* in regression 7 are not individually significant; however, an *F*-test restricting all the treatment dummies to zero is rejected ($p = 0.001$), and the aggregate Groups dummy of regression 8 is statistically significant ($p = 0.012$).

RESULT 3: *The creation of groups reduced mean return rates to outsiders. The effect is stronger in three of the five groups treatments (C, SG, and SI) and holds even after controlling for the possible relationship between the giving rate received and the return rate.*

Some individual variables are significant: older subjects appeared to give and return more; Chinese subjects gave less to both insiders and outsiders; stage 1 return rates were positively related to the giving and return rates to both insiders and outsiders, although subjects who were more generous in giving in stage 1 were nastier toward outsiders when groups were formed. There is some evidence of lower return rates in the session with eight subjects, with respect to both insiders and outsiders (the evidence is less clear for giving). There was no evidence that coming from the UK or being Christian or agnostic/atheist mattered, and the effect of an economics/management background seemed also limited (with only regressions 1 and 2 at the borderline of significance).

Looking at how giving and returning evolved in stages 2 through 4, the giving rate with respect to insiders increased with time if statistically insignificantly so (Spearman $\rho = 0.118$, *n.s.*), while it clearly decreased with respect to outsiders ($\rho = -0.492$, $p < 0.001$); the return rate decreased with respect to insiders ($\rho = -0.31$, $p = 0.001$) but more so with respect to outsiders ($\rho = -0.505$, $p < 0.001$). Overall, discrimination increased with time in both the giving rate ($\rho = 0.406$, $p = 0.001$) and the return rate ($\rho = 0.316$, $p = 0.012$), possibly because of negative reciprocation feedback.

Figure 2 and Table 2 suggest, therefore, and our statistical analysis confirms, that the difference in trust between insiders and outsiders arises because there is negative discrimination against outsiders, while the balance of the evidence is that insiders are treated no worse, and

[17] There is always a general problem that any experiment designed to test the effect of "something" must make that "something" salient in some way, especially in experiments with within-group manipulations (as subjects find themselves comparing and contrasting different tasks). As a result, the experimenter may actually encourage precisely a type of behavior that experiment is designed to test for. The same problem of interpretation arises here. However, while the salience given to groups in this experiment might explain why individual decisions turn on group information, it cannot explain why we systematically observe negative as opposed to positive discrimination.

certainly no better, than in the absence of groups.[18] This result is, however, undoubtedly sensitive to what one takes to be the no-group benchmark. In our experiment, the mean giving rate in the baseline treatment where there are no groups was 55.8 percent in stages 2–4 (Table 2). While this is high compared with some trust game experiments, it is consistent with the range of results reported for these games in Colin F. Camerer (2003). In addition, whereas we have interpreted this baseline as a case where there are no groups, it might be argued that the subjects here actually belong to one big group (the one formed by all subjects in the laboratory) and so the difference captured in results 1 and 2 would really be best described as what happens when the constellation of groups changes to be less inclusive. In defense of the interpretation offered in results 1 and 2, we followed the practice of most trust game experiments in the baseline (i.e., we did nothing to make the "big" group salient) and the results in those trust games are not usually interpreted as in part a consequence of some "big" group effect (even though this, too, is always a possibility).

III. Experimental Results from Market Phase

A. *Estimation of the Psychological Value*

We now consider what, if any, additional psychological value subjects placed on own group membership. We measure this by the extent to which subjects place a value on own group membership in excess of its material value. We discuss later whether this is a genuine psychological benefit rather than the artifact of well known mechanisms, like reference dependence (Section IVC). The markets at the start of stages 2, 3, and 4 in the C, SG, SF, and SI treatments (and stages 3 and 4 in SM) provided an incentive-compatible mechanism for the revelation of individual preferences for staying in one's own group (the WTA value) or for switching groups (the WTP value). As there were repeated markets, subjects also had the opportunity to gain experience both about the nature of the social commodity being traded (i.e., membership of a given group) and about the market mechanism itself.

WTA(own) is the positive price a subject needed to be paid to be willing to switch, and WTP(own) is equivalent to a *negative* price on own group membership, and so they reveal the value that an individual places on membership of his or her group.[19] Figure 3 provides information on WTA and WTP values observed in the experiments.

We define the measure M1 as equal for each subject and market to WTA(own) or −WTP(own), whichever is the case.[20] Suppose, further, that individuals hold a common expectation that members of group k enjoy a material advantage of x over members of the other group. It follows that if subjects enjoyed no nonmaterial, psychological value from belonging to their group, then the individual M1 for each of the six members of k should be equal to x and $-x$ for the six members of the other group, with the result that the average M1 is zero. In these circumstances, with one qualification noted below, the extent to which *average* M1 deviates from zero is a measure of the additional average psychological value placed on own group membership. In other words, people

[18] As shown in online Appendix C, the discrimination result is fairly widespread across subjects, as opposed to being mostly driven by a small minority of subjects.

[19] A price of zero was also a possibility, allowed by the experimental program in relation both to WTA and WTP valuations. Therefore, technically we always had a *nonnegative* price (WTA(own)) or a *nonpositive* price (WTP(own)), with one further qualification to be mentioned shortly.

[20] In each market, subjects *either* choose a WTA *or* choose a WTP; they do not do both. Hence, a strategy by which a subject places both a WTA and a WTP is not feasible. In addition, a strategy of putting a high WTA just in case there is an opportunity of making money, rather than because of a desire to stick with one's own group, is not incentive compatible: if I do not care about which group to belong to, even if I am offered a single unit I should rationally oblige and accept, and therefore I should place a WTA = 1.

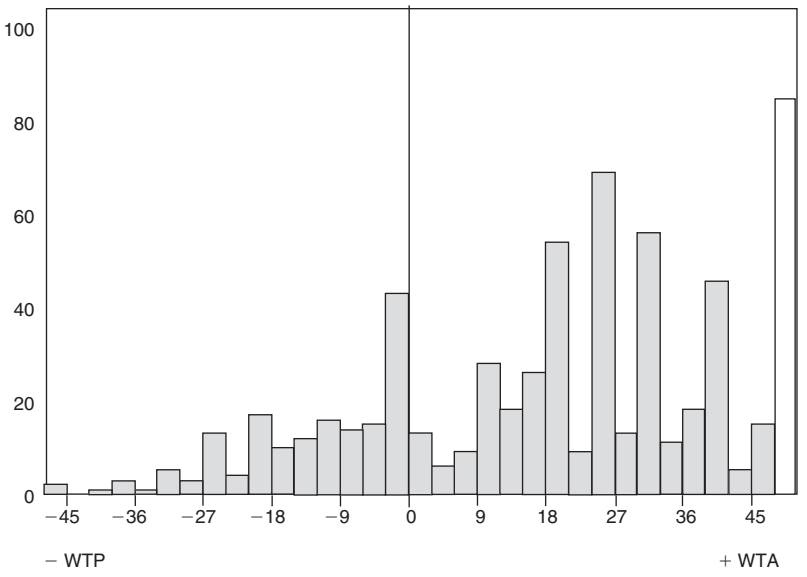

FIGURE 3. HISTOGRAM OF WILLINGNESS TO PAY (WTP) AND WILLINGNESS TO ACCEPT (WTA) IN RELATION TO THE OTHER GROUP

Notes: The histograms are built from individual choices; $n = 161$ for −WTP observations (mean = 12.416, standard deviation (s.d.) = 11.335) and $n = 461$ (mean = 28.84, s.d. = 13.516) for WTA observations. The white column stands for choices of not being willing to accept any price up to 48 to switch group.

on average are placing more (or less) value on own group membership than can be associated with its material effects and so reveal an additional psychological benefit (or cost) from belonging to their own group.[21]

One problem with the average M1 measure of the mean psychological benefit from own group membership is that subjects could state that they were not willing to lose their membership at any allowed price between 0 and 48 points. Nine percent of the choices were of this "definite stay" kind (between 8 and 11 percent in different treatments). We opted for two routes to deal with this problem. M1 contains all observations, but conservatively introduces a valuation of 49 for these definite stays:[22] since the lower bound for valuations was −48 (the budget), if anything,

[21] This last point is worth drawing out. The psychological value that is revealed is related to remaining in one's own group rather than joining another and so, to the extent that subjects anticipate the psychological value of belonging to the other group, this assessment might understate the psychological value that a subject places on belonging to a group as compared with not belonging to a group at all. The SF treatment may be useful to evaluate this possibility because, in this treatment, the blues value membership in the blue group as compared with that of the not blue group, and the latter can be interpreted as being closer, in frame, to not being in a group. As discussed more fully in Section IV, the valuation of the blues is not different from the valuation placed on group membership in the other treatments where there are clearly two groups. On the basis of this evidence, the understatement seems to be insignificant and this would not be surprising if subjects were imperfectly able to introspect on how it would feel to be a member of a different group.

[22] We chose the closest integer value to 48 in keeping with the experimental procedure, where for simplicity subjects could provide only integer valuations: therefore, 49 is the lowest value in keeping with this constraint. In terms of upward bias of psychological value estimates, the "worst case scenario" for this modeling choice would be if all six agents had a true value of 48.001 and preferred not to round their valuation to 48; even in this scenario, the implied

for $x > 49$, this introduces a downward bias. M1b simply omits "definite stay" observations and also introduces a downward bias in average M1 estimates (larger than for M1a). Either way, in the light of possible downward biases, our average M1 estimates should be interpreted as conservative estimates of the revealed psychological values of own group membership.

The virtue of the *average* M1 measure of the psychological value placed on own group membership is that it does not depend on any precise method for forming expectations regarding the material value of being in one group rather than another. All that matters is that individual expectations are homogenous on this matter (or, if they are heterogeneous, all that matters is that the individual expectations in each group vary randomly around a common value for both groups: see the Appendix). Nevertheless, in practice, even this might be a strong assumption, unless one accepts that those who share the same information set should draw the same inferences, since subjects might well wish to take into account who else might be swapping, and conjectures could plausibly lead to divergent expectations. The effect of this uncertainty is, however, constrained when group size is held constant; and this is a virtue of the experimental design. Whatever is the number swapping between groups, the ensuing uncertainty over membership attaches in equal measure to a subject's initial group composition and the one they might swap to. Hence if subjects are risk averse, this fact should not lead them to value differently their initial group as compared with the other.[23]

It is worth noting that the *individual* revealed psychological values *are* sensitive to the precise expectations regarding the material consequences of belonging to different groups, and since we wish to run some regressions using individual psychological values, we make two possible assumptions here. For individuals, M2 is equal to M1 minus the expectation of the material gain from a switch when these expectations are formed adaptively (i.e., the expected material gain/loss is the same as that in the previous stage except for the market at the start of stage 2 when it is set equal to zero, as there is no past information on relative group trustworthiness at the time); and M3 is equal to M1 minus the expectation of the material gain from a switch when these expectations are formed rationally (i.e., the expected relative material gain/loss is the same as actually occurs). One illustrative piece of evidence that subjects did take expectations of material gain into account is that, in the SI treatment, more successful groups seem to have revealed comparatively greater valuations in the following stage: the fraction of wins from being blue in the previous stage has a Spearman correlation $\rho = 0.62$ ($p = 0.101$) with the blue group price (including definite no stayers, valued at 49); whereas $\rho = -0.845$ ($p = 0.008$) with the corresponding red group price.

B. *Results on Psychological Values*

Table 4 depicts mean psychological values for each measure, treatment, and stage, and Figure 4 provides illustrative histograms of the distribution of mean psychological values.

In all sessions, and using any of the measures, we find that mean psychological values are above zero (Wilcoxon $p < 0.001$). Figure 4 exemplifies the scale of the effect: only 25 out of 248 subjects had mean M1 values lower than zero, and only one had a mean value below -20. As subjects learn more about the task and the nature of the commodity, it is unsurprising that

upward bias would be only 0.5. In practice, none of our key results would change if we were to choose a value, say, of 48.001 for "definite stay" cases.

[23] The lack of knowledge regarding group membership might also be thought to weaken any group identity effects. In this context, our finding that subjects are willing to pay more than can be accounted for by the material benefits of belonging to a group becomes more striking.

TABLE 4—MEAN PSYCHOLOGIAL VALUES

Treatment stage		M1, M2, M3	M1b	M2b	M3b
C	2	21.042	19.826	19.826	19.571
	3	24.646	23.022	23.121	22.937
	4	21.063	17.814	17.654	17.766
	Total	22.250	20.254	20.235	20.122
SG	2	23.393	22.444	22.444	22.147
	3	26.054	23.804	24.435	24.488
	4	18.054	15.020	15.985	15.727
	Total	22.500	20.462	20.983	20.813
SF	2	22.167	21	21	20.757
	3	16.542	12.767	12.789	12.774
	4	17.167	14.272	14.51	14.375
	Total	18.625	16.113	16.198	16.065
SM	3	25.271	22.512	22.39	22.506
	4	18.792	16.045	17.001	17.252
	Total	22.031	19.241	19.665	19.849
SI	2	19.958	18.022	18.022	17.556
	3	18.813	17.5	17.697	17.792
	4	18.271	13.024	13.799	14.838
	Total	19.014	16.288	16.597	16.794

Notes: M1, M2, M3, M1b, M2b, M3b are different measures of psychological value, as described in the main text. The (M1, M2, M3) column values also apply to any other psychological value measure that assigns a constant value (such as 49) to definite stayer cases and that is based on homogenous or symmetric expectations (see the Appendix). There was no market at the beginning of stage 2 in the SM treatment.

psychological value measures may become smaller,[24] but in stage 4 they are still at least 17 points according to M1 (or equivalent measures) and at least 13 according to the measures that omit definite stayer observations.[25] Projected over the three stages, a psychological value of 13 or 18 implies a valuation of about 12.8 percent or 15.6 percent of mean stages 2–4 experimental winnings for the sake of own group membership.[26] The mean actual valuation in the experiment depends on the treatment and the measurement used, but ranges between 14.6 percent (SF treatment, M3b measure) and 22.4 percent (C treatment, M1 and equivalent measures) of experimental winnings. [27]

[24] The Spearman correlation between M1 (or equivalent measures) and stage number is $\rho = -0.285$ ($p < 0.05$). The correlation is larger when definite stayer observations are removed (ρ (M1b, stage) $= -0.368$, $p < 0.01$; ρ (M2b, stage) $= -0.353$, $p < 0.01$; (ρ (M3b, stage) $= -0.332$, $p = 0.01$), but this is likely to be an artifact of the fact that later stages have a larger number of definite stayers (6 percent in stage 2, 9.6 percent in stage 3, and 13.5 percent in stage 4), and so of high psychological values being removed, hence exacerbating the downward bias problem in measured mean psychological values. The larger extent to which there is capping of psychological values at 49 for definite stayers in later stages may also be artificially increasing the correlation for the M1 or equivalent measures correlation.

[25] The latter are likely to overemphasize the decrease in psychological value (see previous footnote). Nevertheless, there is clearly no evidence that subjects appear to suffer special emotional discomfort from joining subjects they have discriminated against since, in the light of the observed discrimination, this would mean that the psychological values should become larger with time, which they do not. Spearman correlations between amount of discrimination and later psychological value are also no larger than those between psychological value and later discrimination, again suggesting no evidence of such a special discomfort.

[26] Specifically, the lowest cost of 13.024 is for stage 4, SI treatment, M1b measurement: the corresponding percentage (projected over all three stages 2–4) can be found as $(13.024 \times 3 \times 0.04)/12.185$ (mean gains in SI treatment) $=$ 12.8 percent. The lowest total M1 cost of 17.167 is for stage 4 and the SF treatment: it can be found as $(17.167 \times 3 \times 0.04)/13.236$ (mean gains in SG treatment) $=$ 15.6 percent.

[27] Average WTA by subjects willing to sell own group membership (at 48 or less) is 28.454 (30.736, 29.211, and 25.269 in stages 2, 3, and 4, respectively); average WTP by subjects willing to buy own group membership is 11.784 (7.882, 12.651, and 15.378 in stages 2, 3, and 4 respectively). These values overstate convergence as they fail to take into

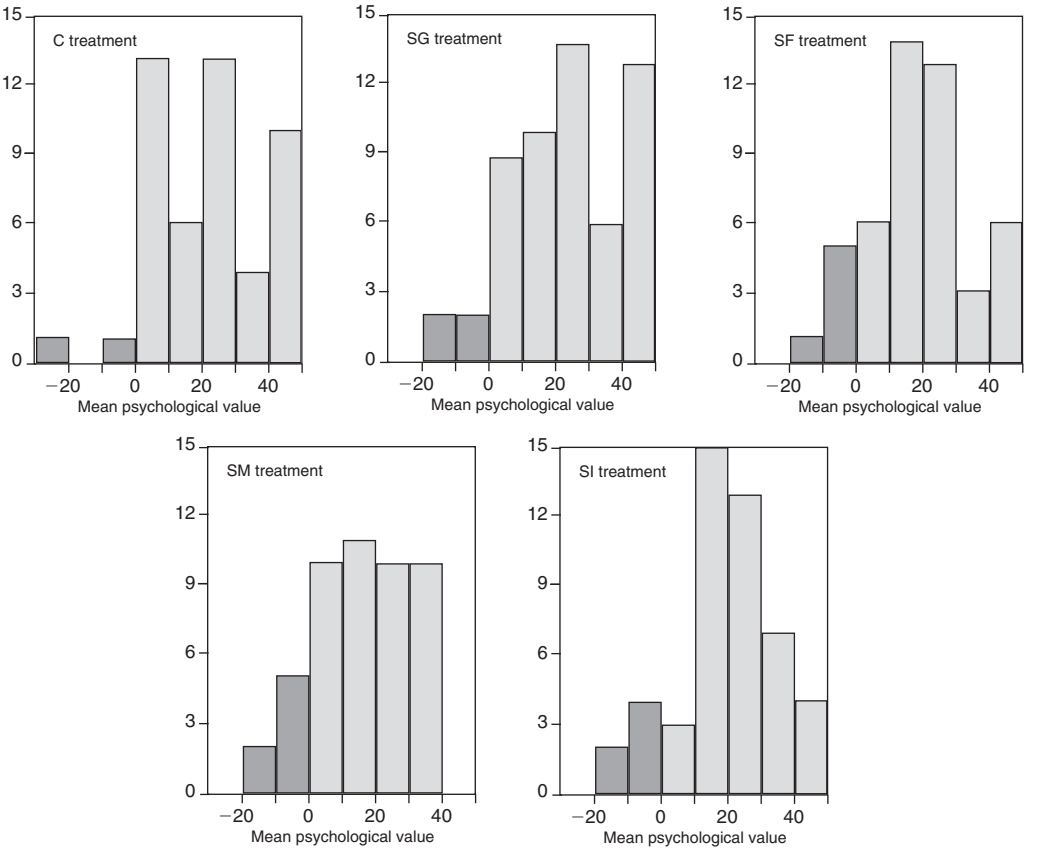

FIGURE 4. HISTOGRAMS OF PSYCHOLOGICAL VALUES BY SUBJECT

Notes: The histograms depict the distribution of M1 mean psychological values associated with each subject (n = 48, 56, 48, 48, and 48 for the C, SG, SF, SM, and SI treatments, respectively). Histograms using other measures are similar.

Between-treatment differences in mean psychological values are comparatively small. They are not statistically significant for any stage in Kruskal-Wallis χ^2 tests. In stage 3 there is suggestive evidence of lower values in the SF treatment ($p < 0.1$ in Mann-Whitney tests), but not otherwise.

RESULT 4: *Almost all subjects revealed a positive psychological value for own group membership. While decreasing slightly, this remained high even with experienced subjects. The positive psychological valuation was estimated between 14.6 percent and 22.4 percent of experimental winnings, depending on the experimental treatment.*

Our next result comes from examining whether these revealed psychological benefits can be used as a predictor of discrimination taking into account the same set of control variables that

account that the proportion of buying decisions is actually lower in stages 3 and 4 (43 and 45, respectively) than in stage 2 (51), and the proportion of definite stay decisions more than doubles with time (see footnote 23).

we considered in Table 3. Let DAvgGivingRate (DAvgReturnRate) be equal to mean giving (return) rate by a subject to insiders minus mean giving (return) rate by the same subject to outsiders. We ran regressions controlling for session-specific effects on DAvgGivingRate and DAvgReturnRate. The results of these regressions, which control for possible session-level specific effects by using error clustering, are given in Table 5.[28]

These regressions can only use data from stages 2–4 in the groups treatments, since elsewhere either there were no groups (B treatment) or subjects were not provided information on coplayers' groups (stage 1 of the treatments with groups). A similar set of control variables is used, as for the regressions in Table 3, but in the DAvgReturnRate regressions (13 through 16) we now use "Diff. in Trust Rate as a Second Mover," equal to mean giving rate received as second mover from insiders *minus* mean giving rate received as second mover from outsiders. We use this because it measures the degree of discrimination trustees have experienced as second movers, and this may indirectly produce discrimination against outsiders due to reciprocity or some similar motivation. There was more discrimination in the session with eight subjects only (session 3); there is also some evidence of lower discrimination in giving in the SM and possibly SG treatments. Control variables such as religion, gender, age, and (mostly) educational background appear not to matter, while there is only some borderline evidence of a slightly lower discrimination in return rates on the part of UK and Chinese subjects than those from the rest of the world.

By multiplying psychological value measures (*PV* in the regression tables) by the relevant treatment dummy (*C*, *SG,* or *SI*, equal to one in the respective treatments, otherwise zero), we can estimate their explanatory power on discrimination in each treatment. Regressions 9, 10, 13, and 14 use M1 as a measure of psychological value, while regressions 11, 12, 15, and 16 use M1b. We also have a control interaction dummy variable $PV \times$ Session 3, to check that any predictive power is not being influenced by the one session with eight subjects. We find a mixed picture with results that are sensitive to the treatment. The psychological value measures have no predictive power in the weakest group manipulations (C and SF); they have predictive power on discrimination in mean giving rates and return rates in the SG and SM treatments ($p \leq 0.02$) and, less conclusively, in the SI treatment in relation to mean giving rates (in one-tailed tests of the hypothesis that PV predicts discrimination, $p < 0.05$). Given this picture, perhaps the strongest tests of some connection between measures and behavior are the *F*-tests restricting all $PV \times$ treatment interaction coefficients to be equal to 0 (as all of them reject the restriction, $p < 0.01$); and the statistical significance of the global *PV* dummies used in regressions 10, 12, 14, and 16 ($p \leq 0.05$).

Overall, the results suggest that an extra point in psychological value, as measured by M1 and M1b, increases discrimination in giving rates by 4 percent and in return rates by 2 percent. While M1 and M1b are more successful predictors, a similar picture is obtained if M2 and M2b, but not M3 and M3b, are used.[29]

In the SI treatment, discrimination may be driven more by the material incentives, to the detriment of the predictive power of our measures. This notwithstanding, that their predictive power is lowest in the weakest group manipulations (C and SF), while their average value remains high (Table 4), suggests that our measures of psychological value are only partially capturing psychological value itself, and may be partially proxying for other psychological effects such as reference dependence (see Section IVC). The noisiness of the results, with fairly low R^2, certainly

[28] Again, broadly the same picture is obtained if standard or Tobit random effects regressions are used instead (see online Appendix D).

[29] The global coefficient on M2 or M2b is equal to 0.030 ($p < 0.01$) in the difference in giving regression and equal to 0.013 ($p < 0.05$) in the difference in return regression. These results are robust to using random effects or Tobit random effects regressions. M3 and M3b are, instead, mostly unsuccessful in predicting discrimination.

TABLE 5—REGRESSIONS ON DIFFERENCE IN MEAN GIVING AND RETURN RATE TO INSIDERS AND OUTSIDERS

Regressions on stage 2–4 difference in mean giving rate	Regression 9			Regression 10		
	β	t	p	β	t	p
PV measures used						
SF	0.017	0.24	0.811	0.005	0.07	0.945
SG	−0.198	−2.7	0.014	−0.06	−0.75	0.462
SM	−0.268	−2.91	0.009	−0.159	−1.95	0.066
SI	0.053	0.74	0.467	0.094	1.5	0.149
Session 3	0.211	4.5	0	0.213	3.54	0.002
PV				0.004	4.06	0.001
PV × C	0.001	1.11	0.282			
PV × SF	0	0.28	0.779			
PV × SG	0.008	7.76	0			
PV × SM	0.008	2.79	0.011			
PV × SI	0.003	1.78	0.09			
PV × Session 3	0	0.13	0.898			
Gender	−0.021	−0.61	0.549	−0.018	−0.5	0.621
Christian	−0.023	−0.43	0.674	−0.015	−0.28	0.781
Agnostic Atheist	0.01	0.2	0.842	0.018	0.41	0.686
EcMgt	0.087	1.81	0.085	0.077	1.57	0.132
UK	0.053	1.57	0.132	0.053	1.45	0.163
China	0.017	0.29	0.774	0.01	0.16	0.872
Age	−0.003	−1	0.329	−0.003	−0.83	0.417
Constant	0.169	1.48	0.154	0.095	0.82	0.423
R^2	0.151			0.121		
	Reg 9 PV coeffs. = 0: $F = 58.32$ ($p < 0.001$)					
	Regression 11			Regression 12		
PV b measures used						
SF	0.022	0.29	0.772	0.008	0.11	0.913
SG	−0.173	−2.33	0.03	−0.057	−0.7	0.493
SM	−0.276	−3.52	0.002	−0.161	−1.96	0.063
SI	0.032	0.41	0.683	0.1	1.53	0.142
Session 3	0.245	4.99	0	0.215	3.57	0.002
PV				0.004	3.69	0.001
PV × C	0.001	1.15	0.263			
PV × SF	−0.001	−0.38	0.706			
PV × SG	0.007	9.92	0			
PV × SM	0.006	2.53	0.02			
PV × SI	0.004	1.79	0.089			
PV × Session 3	−0.001	−1.25	0.227			
Gender	−0.022	−0.66	0.516	−0.016	−0.44	0.663
Christian	−0.02	−0.35	0.728	−0.021	−0.39	0.701
Agnostic Atheist	0.005	0.11	0.915	0.012	0.26	0.8
EcMgt	0.076	1.56	0.134	0.073	1.46	0.16
UK	0.056	1.58	0.13	0.051	1.35	0.192
China	0.023	0.39	0.702	0.003	0.05	0.959
Age	−0.002	−0.68	0.503	−0.002	−0.58	0.571
Constant	0.161	1.34	0.194	0.098	0.81	0.425
R^2	0.137			0.112		
	Reg 11 PV coeffs. = 0: $F = 32.05$ ($p < 0.001$)					

TABLE 5—REGRESSIONS ON DIFFERENCE IN MEAN GIVING AND RETURN RATE TO INSIDERS AND OUTSIDERS *(continued)*

Regressions on stage 2–4 difference in mean return rate	Regression 13			Regression 14		
	β	t	p	β	t	p
PV measures used						
D in trust rate as 2nd mover	0.199	4.11	0.001	0.2	4.41	0
SF	−0.039	−0.47	0.647	−0.032	−0.74	0.471
SG	−0.124	−1.77	0.092	−0.059	−1.46	0.159
SM	−0.152	−2.02	0.057	−0.066	−1.51	0.147
SI	−0.098	−1.22	0.238	−0.06	−1.69	0.106
Session 3	0.111	3.13	0.005	0.07	1.88	0.074
PV				0.002	2.07	0.052
PV × C	0	0.17	0.865			
PV × SF	0.001	0.37	0.717			
PV × SG	0.003	4.81	0			
PV × SM	0.003	3.12	0.005			
PV × SI	0.002	1.37	0.185			
PV × Session 3	−0.002	−1.9	0.072			
Gender	−0.038	−1.47	0.157	−0.036	−1.41	0.175
Christian	0.005	0.13	0.901	0.008	0.2	0.844
Agnostic Atheist	0.024	0.56	0.581	0.028	0.69	0.499
EcMgt	0.038	1.24	0.229	0.035	1.16	0.258
UK	−0.044	−1.68	0.109	−0.047	−1.79	0.088
China	−0.1	−1.7	0.106	−0.107	−1.79	0.088
Age	−0.004	−0.9	0.377	−0.004	−0.91	0.373
Constant	0.177	1.34	0.196	0.161	1.29	0.211
R^2	0.179			0.121		
	Reg 13 PV coeffs. = 0: $F = 6.11$ ($p < 0.001$)					

	Regression 15			Regression 16		
PV b measures used						
D in trust rate as 2nd mover	0.2	4.1	0.001	0.201	4.42	0
SF	−0.033	−0.44	0.663	−0.028	−0.57	0.574
SG	−0.103	−1.63	0.118	−0.057	−1.39	0.179
SM	−0.114	−1.66	0.113	−0.062	−1.4	0.176
SI	−0.076	−1.05	0.305	−0.059	−1.6	0.124
Session 3	0.077	2.31	0.031	0.071	1.95	0.065
PV				0.002	2.63	0.016
PV × C	0.001	0.43	0.669			
PV × SF	0.001	0.83	0.414			
PV × SG	0.003	3.76	0.001			
PV × SM	0.004	3.25	0.004			
PV × SI	0.002	1.23	0.235			
PV × Session 3	0	−0.23	0.821			
Gender	−0.039	−1.55	0.136	−0.037	−1.5	0.149
Christian	0.007	0.18	0.858	0.007	0.19	0.853
Agnostic Atheist	0.028	0.64	0.53	0.031	0.75	0.462
EcMgt	0.038	1.21	0.239	0.034	1.14	0.266
UK	−0.048	−1.87	0.076	−0.05	−1.85	0.079
China	−0.109	−1.87	0.077	−0.112	−1.85	0.079
Age	−0.004	−0.86	0.398	−0.003	−0.85	0.404
Constant	0.161	1.29	0.211	0.132	1.2	0.244
R^2	0.191			0.121		
	Reg 15 PV coeffs. = 0: $F = 3.93$ ($p = 0.009$)					

Notes: Sample size: $n = 248$ (regressions 9–10), 245 (reg. 11–14), 242 (reg. 13–14), 239 (reg. 15–16). PV stands for psychological value measures, and so for the M1 measure (or M2, or M3, or anyone else satisfying the conditions of the Appendix: see Section IIIA); PV b (psychological value b) is measured by the M1b value (see Section IIIA). Some observations are lost in regression subsets due to zero giving, undefined PV b values, or both. We employ error clustering to control for session-level effects. The *p*-values provided are two tailed. The dependent variable is DAvgGivingRate = mean giving rate to insiders *minus* mean giving rate to outsiders for reg. 9–12. It is DAvgReturnRate = mean return rate to insiders *minus* mean return rate to outsiders for reg. 13−16. D in trust rate as second mover stands for mean giving rate gifted as second mover from insiders *minus* mean trust rate gifted as second mover from outsiders. The *p*-values provided are two tailed. The *F*-tests test whether the restriction that the PV terms coefficients be jointly equal to zero is accepted.

should caution against placing too much of a hedonic interpretation on our measures, a point that has welfare implications to which we shall return.

RESULT 5: *M1 and M1b measures of psychological value have some predictive power on discrimination, especially in giving and in relation to the SG, SM, and SI treatments.*

Last, we analyze a possible effect of high revealed psychological values associated with group membership that would arise independently of whether these revealed values capture a genuine psychological benefit rather than the operation of something like a reference dependent effect: namely, a kind of social inertia. To be precise, suppose that the expected material gain from switching groups is x (which can be positive, zero, or negative) and that subjects' evaluations have some small unsystematic random noise around x; assume furthermore that there is no positive psychological value. The standard prediction in these circumstances, as is normally made in WTP–WTA market experiments, is that we should observe 50 percent of all the physically possible trades being implemented on average (e.g., Daniel Kahneman, Jack L. Knetsch, and Richard H. Thaler 1990).[30] This is because, on average, the half that undervalue being in the group that enjoys the material advantage will trade with the half in the other group which overvalues this material advantage. What we label social inertia occurs when trade is less than this prediction.

Table 6 shows that trade was between 6.7 percent and 16.7 percent depending on the treatment, which is much less than this standard prediction. In short, there was significant social inertia. There was not a single session, or a single stage, where mean trade was as much as 50 percent of possible trade (Wilcoxon $p < 0.001$). Stage mean values hover between 0 and 20.8 percent, and differences among treatments are not statistically significant. This is true even though the higher percentages in stages 2 and 4 of the SI treatment implied a proportionally greater fraction of markets having at least one group change, and so a well-defined market price. Just one successful deal was required in order for group changes to occur, and so the fact that in most markets (including slightly over 50 percent of SI markets) there was no group change is itself indicative of social inertia in our experiment.

Mean market prices by session are not quite statistically different from one another across treatments (Kruskal-Wallis $\chi^2 = 6.68$, $p = 0.154$). Nevertheless, while the SG, SM, and SI mean prices cluster together, and the SF mean price is a little lower, they all appear twice or thrice as large as the C mean price, and a Mann-Whitney test shows that the C mean price is significantly lower than in the other treatments ($p = 0.015$). Given the small number of trades for each treatment, not much should be read into these differences in mean market prices.

RESULT 6: *There is a form of social inertia associated with group membership: people traded group membership less than would be expected in the absence of psychological benefits. This was true across all treatments.*

IV. Discussion of Results

Should policymakers take account of how any policy affects individuals' membership of groups? The literatures on social capital and well-being would seem to encourage an affirmative

[30] The fact that 9–10 percent of the subjects made definite stay choices, implying (in the absence of positive psychological values) $x > 48$, may be a problem for this prediction if, in fact, we observed an equivalent choice of $x = -48$ by another rough 9–10 percent of the subjects. This would imply that in practice the $[-48 \ldots 48]$ valuation bounds may have prevented trade that would have otherwise occurred. However, there were only four valuation choices below -40 in the whole experiment, so this cannot be a plausible reason for observing less than 50 percent trade.

Table 6—Group Changes and Market Prices

Treatment	Stage	Number of group changes as a fraction of all possible deals	Number of markets with group changes	Average market price when deal is done
C	2	0.083	2	6
	3	0.125	3	9.333
	4	0.083	2	1
	Total	0.097	7 (out of 24)	6
SG	2	0	0	
	3	0.067	2	15.5
	4	0.133	3	20.667
	Total	0.067	5 (out of 30)	18.6
SF	2	0.083	2	17
	3	0.167	4	16.167
	4	0.167	4	16.5
	Total	0.139	10 (out of 24)	15.6
SM	3	0.083	2	15
	4	0.125	3	25.333
	Total	0.104	5 (out of 16)	21.2
SI	2	0.208	5	16.2
	3	0.083	2	25.5
	4	0.208	5	20.8
	Total	0.167	12 (out of 24)	19.667

Notes: In each market stage there are two markets (one for paying for the membership of each group), and so there are six markets per session (four in the SM treatment). The market price is defined over the markets in relation to which group changes deals are made.

answer, at least in principle. In practice, matters are rather more difficult because the evidence connecting group membership with trust and well-being largely derives from survey data that report on trusting attitudes and perceptions of happiness, and we know neither whether such group effects are significant in the sense that individual behavior is actually affected through membership of groups, nor how any such group effect is decomposed between the influence of groups *per se* and the contribution that comes from the specific character of the actual groups in question. Our experiment is potentially important because it addresses both these areas of ignorance.

With respect to the first issue, we focus on welfare effects of artificial groups and, in doing this, we create the basis from which one can begin to answer the second. The point is that when studying natural groups we will be able to discern the *specific* influence of an actual group only if we know something about how the existence of groups per se affects behavior. In this way, our experiment acts as a baseline control for those future studies of natural groups.

The contribution of our study to the first of these questions turns on our specific findings, which we discuss below. They apply to all the group treatments, as there were few interesting differences between the various types of groups that we induced experimentally. This lack of difference is perhaps surprising given, for example, the role that segregation can play in promoting within-group cooperation in evolutionary game theory (see Bergstrom 2002).[31] It also differs

[31] Although it provides some evidence that anonymity did have the effect of making each interaction a one-shot game for subjects (see footnotes 5 and 8).

from the recent experimental evidence in Charness, Rigotti, and Rustichini (2007), where the influence of group membership depended on the salience of the group, and in particular where they found that the kind of minimal groups we induced in the C treatment had no effect on behavior.

A. *The "Pure" Group Effect on Trust is Negative*

This result runs counter to the conventional wisdom in the literature where groups are frequently cast as a form of social capital. Insofar as the contrary supposition in the literature is based on the experience with natural groups, this result suggests that the impact on welfare of actual groups may depend rather more on the character of the constitutive norms of those actual groups than the fact that they are groups per se.

B. *"Pure" Groups Induce Negative Discrimination against Outsiders*

The existence of negative discrimination is, of course, the key to the negative social capital influence discussed above.[32] It is also noteworthy because, although this possibility is recognized in the literature (see Steven N. Durlauf 1999; Putnam 2000) and there is some support from another study of artificial groups (Zizzo 2003) and from some traditional social psychological experiments (e.g., Muzafer Sherif 1966), some evidence from natural groups seems to point in a different direction. Chaim Fershtman and Gneezy's (2001) trust game experiment on natural groups found no in-group bias effect but, in comparison to the results from other experiments they performed, found some gender-specific evidence of generalized adverse beliefs of the trustworthiness of the low-status group.[33] Their experiments, unlike ours, were one-shot and had a classroom and nonanonymity flavor that may have worked against negative outgroup discrimination, but may be consistent with psychological evidence that, with natural groups of different status, the subordinate group may form beliefs consistent with dominant group favoritism (see Henri Tajfel and John Turner 2001, for a discussion). There is also the claim in the Goette, Huffman, and Meier (2006) prisoner's dilemma experiment with weak natural groups that there is no evidence of negative discrimination against outsiders, although this is only indirectly inferred rather than directly tested in the manner of our experiment. Taking these papers at their face value, our specific result of negative discrimination in artificial groups tends to reinforce the earlier conclusion with respect to the potential importance of norms and beliefs of natural groups in influencing behavior. This, in turn, accords with some experimental findings in social psychology where attitudes toward the "outgroup," for example in Sven Waldzus and Amelie Mummendey (2004) and Waldzus, Mummendey, and Michael Wenzel (2005), depend on the character of the "superordinate category" used for evaluating the "ingroup" and "outgroup."

C. *A Positive Psychological "Pure" Group Effect on Welfare*

We have found that people, on average, place a value on their own group membership which exceeds the material advantages of belonging to that group. If we assume that preferences are revealed in behavior, then we can infer that there are genuine additional psychological benefits that people enjoy from belonging to their groups, and this evidence is consistent with arguments

[32] It is also inconsistent with the conjecture that group effects are driven by an expectation of more frequent repeated play with insiders relative to play with generic coplayers in the B treatment, since this should lead to positive rather than negative discrimination (see footnote 5).

[33] Only men discriminated, and did so in relation to men only.

like those of Adam Smith (1759/1976) and Akerlof and Kranton (2000) around the role of groups. While it is usual to assume that behavior reveals a genuine preference in this sense, there are reasons in this instance to pause.

First, the Akerlof and Kranton or Smith observations typically make sense or have a ring of truth in relation to natural groups where people interact with each other in a variety of rich ways. There is none of this richness in our experiment and it is, as a result, rather more surprising to find a psychological value when all that ties our subjects to their group is little more than a red or blue color identifier. In this context, it is perhaps more plausible to construe the results in terms of the way that the experimental identification of minimal groups triggers, in the subjects' minds, a sense of value that is actually either born out of their social experience of groups or is hardwired into them through some evolutionary process (as suggested by the neurobiological evidence on how group membership is correlated with endorphin levels, referred to earlier; see Dunbar 2006).

Second, there is an alternative explanation of why our subjects might place a monetary value on own group membership. By construction, our measure of this psychological value is revealed in the gap between people's WTP to join a group and their WTA compensation to leave a group, and it is well known from the experimental literature on, for example, the valuation of environmental goods that there is often a wedge between such WTP and WTA assessment (see Ian J. Bateman and Kenneth G. Willis 1999). This wedge is a puzzle in those settings because there is rarely an interpretation that is analogous to the one we have advanced here so far around the influence that group membership has, say, on one's sense of identity. Instead, it is often assimilated to the well-known class of effects that can arise when people have reference dependent preferences (perhaps due to forms of inexperience with the questions being asked: see Graham Loomes, Chris Starmer, and Robert Sugden 2003) and this creates the possibility that it should be discounted from a welfare perspective. In other words, it is still a psychological phenomenon, but it is a consequence of the particular psychological way that individuals think about decision problems and should not necessarily be taken into account when judging the welfare effects of a change.

There are, however, some reasons for believing that at least *a proportion* of what we have labeled as "psychological value" reflects genuine psychological benefits and not the influence of reference dependence effects (or at least for acknowledging the role of the genuine hedonic interpretation; see also Thomas C. Brown 2005). The SF treatment presented a reduced frame group in the form of the "not blue" group. If there is genuine psychological benefit, then we would expect that it would be lower for this group than for subjects in the blue group of the SF treatment or those in the color treatments. Whereas if the apparent psychological valuation of own group membership in this experiment arises from reference dependence, it would have equal force for not blues as for blues and (in relation to the other color group treatments) for reds.

There are two qualifications to this prediction. First, since we are considering the mean psychological value measure of a single group in each given SF treatment session, we cannot rely on the equivalence result between psychological value measures, nor, as a result, on the simple WTP–WTA difference as expressed by M1/M1b: expectations about the material gains of staying in a group or switching groups will matter. We can, however, rely on M2/M2b and M3/M3b as proxies for how such expectations are formed. Second, this prediction points only to a possible lower bound to the actual size of the genuine psychological value, since, if M2/M2b and M3/M3b are still positive in relation to the reduced-frame not blue group, this *could* be due at least partially to residual group identity feelings, although they could equally, and perhaps more plausibly, be due to reference dependence or other factors that contribute to a WTP–WTA gap.

The mean M2, M2b, M3, and M3b values were respectively 16.356, 14.287, 15.518, and 13.240 for the not blue group, against respective values of 20.893, 18.157, 21.731, and 18.932 for the SF

treatment blue group, and, as shown by Table 4, mean values around 20–22 for all other treatments except SI. A Mann-Whitney test that the not blue group has lower measures than the mean values in the other treatments receives some statistical support ($p = 0.06, 0.05, 0.05, 0.04$, respectively, in relation to M2, M2b, M3, and M3b).

There is a further possible interpretation of the WTP–WTA wedge. It could reflect a general aversion to change, which attaches in this instance to moving from a group. This is something akin to a halfway house between the two interpretations discussed so far. The benefit attached to being a member of one's group would still be real in the sense that this was how an individual avoided the distaste for change. However, one might plausibly suspect that, once a change had occurred, the experience of a psychological cost from leaving one's original group would recede with time and one would come to attach a similar benefit to staying with one's "new" group, as this would become the way in which one avoided change in the future.[34]

To summarize, we are making no claims that our measures *purely* reflect genuine psychological value. Caution is needed both because the correlation between our measures and discrimination behavior could be stronger and because the not blue group tests with the SF treatment discussed above suggest that as much as 75 percent of the mean revealed "psychological value" *could* be due to something other than genuine psychological value. A realistic assessment is that *a fraction* of our measures reflect genuine psychological value, and that noise in these measures produced by other sources like reference dependence reduces the correlation between them and discrimination behavior.

D. *A Negligible Net Benefit of "Pure" Groups?*

We have investigated experimentally two mechanisms that link the presence of groups positively with welfare in the literature on social capital and happiness/well-being, namely the degree of trust and an identity-related psychological benefit. Under the usual revealed preference interpretation that the excess valuation of group membership betokens a genuine psychological benefit, our results are consistent with one of these mechanisms but not the other (i.e., that groups raise trust). The question that arises is whether our experimental results nevertheless support the general message that groups have a beneficial effect on welfare. In other words, it is natural to ask which of the two conflicting effects of groups on welfare predominates in our experiment. The answer is given in Table 7.

There is a small net benefit in each of the group treatments, but it is not statistically significantly different from zero. Thus, it would appear that the existence of pure groups has a negligible impact on welfare. There are two important qualifications to this conclusion worth mentioning.

The first turns on the interpretation of the WTP–WTA wedge discussed above. It will be clear that, to the extent that either of the alternative interpretations is compelling, the value placed in the psychological column as a genuine benefit will decline and this will tip the calculation in the negative direction.[35]

Second, even if the valuation of group membership does reflect a genuine psychological benefit, this could, in other circumstances, be the source of material welfare losses. In our experiment, there is no aggregate material gain from the actual swapping of people between groups because the number of people in each group is held constant. In many contexts, however, the

[34] There are only very few subjects who switched more than once: 5 out of 248 in the groups treatments.

[35] Against this, it might be argued that the psychological value revealed here for own group membership is a net figure that comes from comparing the psychological benefits associated with each group. In this way, it understates the total value of the psychological benefits that arise from the existence of groups. We discuss this possibility, and largely discount it, in footnote 21.

Table 7—Net Welfare Effects of Groups

Treatment	Material gain	Net social value (MI, M2, M3)	Net social value (M1b)	Net social value (M2b)	Net social value (M3b)
C	−18.205	4.045	1.943	1.927	1.809
SG	−10.107	12.393	10.32	10.862	10.686
SF	0.031	18.656	16.077	16.161	16.029
SM	−7.885	14.146	11.074	11.536	11.693
SI	−10.371	8.643	5.917	6.226	6.423

Notes: In relation to each treatment, net social value is computed by subtracting mean psychological values, as reported in Table 4, according to a number of measures (M1, M2, M3, M1b, M2b, M3b), from the mean material gain estimate. Values are expressed as experimental points, each of which was worth 0.04 UK pounds.

numbers in a group are not fixed. If there were no psychological benefits in such cases and there were differences in the material benefits from belonging to each group, people would switch to the one with highest material rewards and there would be an increase in aggregate material benefits. The presence of psychological benefits could, if sufficiently large, though, cause a kind of social inertia by providing a counter to the material benefit that would come from switching. To the extent that this happens, there would be additional material losses which would push the net benefit test in the negative direction.

Granted these possible qualifications, our experiment does not provide any evidence that the possible positive psychological benefits to own group membership outweigh the negative welfare effects associated with lower trust when there are groups. This is interesting because, while our experiment casts doubt over one of the putative mechanisms linking the presence of groups positively with welfare in the social capital and happiness/well-being literatures, it might still be the case that groups overall are welfare enhancing. So the general conclusion concerning groups from the social capital and happiness/well-being literatures might be supported by our experiment, even if some of the particulars are not. This is not the case and this reinforces the earlier observation that there would appear to be interesting differences between artificial and natural groups which need to be explored. Toward this end, our paper is potentially significant because it supplies a baseline insight into the effects of groups per se, from which the study of natural groups can then extract the particular influence of that group's norms and other distinguishing features.

V. Conclusion

This paper supplies experimental evidence on how membership in groups affects behavior. Like Charness, Rigotti, and Rustichini (2007), we find that membership does affect behavior. Unlike that study, however, we find that the minimal creation of groups is sufficient to produce this effect. More particularly, we address the question of whether, as is sometimes supposed, the existence of groups boosts trust and well-being more generally. This is important because, if groups have these effects, then policy formation should take account of how any change affects the constellation of groups in society.

There are, of course, important differences between natural groups and the artificial ones in the experiment which make generalization of our results to real world settings difficult. For instance, groups are bound together by shared beliefs that are encoded in norms of behavior; they vary in size, and membership can be overlapping. Our experiment has none of those features, although it does have some distinguishing aspects of actual groups like segregation and intergroup competition. Furthermore, our experiment allows subjects to trade membership in their

group and, while some natural groups have a "price" in this sense and a correspondingly fluid membership, many real groups do not.

Nevertheless, our experiment on artificial groups gives some insights into the ways that the simple presence of groups can be both positive and negative. We find that the presence of groups lowers trust in our experiment and that there is some, albeit significantly qualified, evidence that being a member of a group yields a distinct additional psychological benefit. Overall, and in marked contrast with the social capital and well-being literatures, our experiment suggests that the presence of groups is at best neutral and may be negative in terms of welfare.

Appendix: Relationship between Means of Psychological Value Measures

PROPOSITION 1: *Assume that: (a) expectations are homogenous across subjects within each stage, and (b) all observations are included. Then, for any expectation about material gains, the mean psychological value by stage or by session is the same and is equal to the mean M1 (WTA or – WTP) by stage or by session.*

PROOF:

Let y_i be the M1 valuation by each subject i and let there be n subjects in a session. Then the mean M1 in each stage is equal to $\sum y_i / n$. Let x be the common expectation about material gains from belonging to group k. Then one-half of the subjects have an expected own group material value of $+x$ and the other half of $-x$. In each stage, half of the subjects (those belonging to k) will have an own group value of $+x$ and half of the subjects (those not belonging to k) of $-x$. Then, given that all observations are included, the expectation-dependent mean psychological value is equal to

$$\frac{\sum y_i - (n/2)\,x + (n/2)\,x}{n} = \frac{\sum y_i}{n},$$

which is to say that the mean psychological value is not a function of the expectation value x, and is the same as mean M1. Since the mean psychological value by session is the average of the mean psychological values in each of stage 2, 3, and 4, the same identities apply at the level of mean psychological values by session.

Now suppose that the individuals in each group have heterogeneous beliefs about the material gains or losses in the following sense. Let $x_i = x + e_i$, where e_i is a random variable for subjects within each group with mean zero. It follows that the same expression for mean psychological value as the above is derived.

REFERENCES

Akerlof, George A., and Rachel E. Kranton. 2000. "Economics and Identity." *Quarterly Journal of Economics*, 115(3): 715–53.

Bateman, Ian J., and Kenneth G. Willis. 1999. *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU and Developing Countries.* Oxford: Oxford University Press.

Berg, Joyce, John W. Dickhaut, and Kevin A. McCabe. 1995. "Trust, Reciprocity, and Social History." *Games and Economic Behavior*, 10(1): 122–42.

Bergstrom, Theodore C. 2002. "Evolution of Social Behavior: Individual and Group Selection." *Journal of Economic Perspectives*, 16(2): 67–88.

Bornstein, Gary, Uri Gneezy, and Rosmarie Nagel. 2002. "The Effect of Intergroup Competition on Group Coordination: An Experimental Study." *Games and Economic Behavior*, 41(1): 1–25.

Brown, Thomas C. 2005. "Loss Aversion without the Endowment Effect, and Other Explanations for the WTA–WTP Disparity." *Journal of Economic Behavior and Organization*, 57(3): 367–79.

Camerer, Colin F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton: Princeton University Press.

Charness, Gary, Luca Rigotti, and Aldo Rustichini. 2007. "Individual Behavior and Group Membership." *American Economic Review*, 97(4): 1340–52.

Cookson, Richard. 2000. "Framing Effects in Public Goods Experiments." *Experimental Economics*, 3(1): 55–79.

Cox, James C., and David M. Grether. 1996. "The Preference Reversal Phenomenon: Response Mode, Markets and Incentives." *Economic Theory*, 7(3): 381–405.

Dunbar, Robin. 2006. "We Believe." *New Scientist*, 2536(4): 30–33.

Durlauf, Steven N. 1999. "The Case 'Against' Social Capital." *Focus*, 20(3): 1–4.

Economic Journal. 2002. "Social Capital Symposium." *Economic Journal*, 112(483): F417–79.

Falk, Armin, and Urs Fischbacher. 2001. "Distributional Consequences and Intentions in a Model of Reciprocity." *Annales d'Economie et de Statistique,* (63–64): 111–29.

Fehr, Ernst, and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114(3): 817–68.

Fershtman, Chaim, and Uri Gneezy. 2001. "Discrimination in a Segmented Society: An Experimental Approach." *Quarterly Journal of Economics*, 116(1): 351–77.

Fukuyama, Francis. 1995. *Trust*. London: Penguin Books.

Glaeser, Edward L., David I. Laibson, Jose A. Scheinkman, and Christine L. Soutter. 2000. "Measuring Trust." *Quarterly Journal of Economics*, 115(3): 811–46.

Goette, Lorenz, David Huffman, and Stephan Meier. 2006. "The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups." *American Economic Review*, 96(2): 212–16.

Guerra, Gerardo, and Daniel John Zizzo. 2004. "Trust Responsiveness and Beliefs." *Journal of Economic Behavior and Organization*, 55(1): 25–30.

Hargreaves-Heap, Shaun, and Yanis Varoufakis. 2002. "Some Experimental Evidence on the Evolution of Discrimination, Co-Operation and Perceptions of Fairness." *Economic Journal*, 112(481): 679–703.

Helliwell, John F. 2006. "Well-Being, Social Capital and Public Policy: What's New?" *Economic Journal*, 116(510): C34–45.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. 1990. "Experimental Tests of the Endowment Effect and the Coase Theorem." *Journal of Political Economy*, 98(6): 1325–48.

Layard, Richard. 2005. *Happiness: Lessons from a New Science*. Penguin: London.

Loomes, Graham, Chris Starmer, and Robert Sugden. 2003. "Do Anomalies Disappear in Repeated Markets?" *Economic Journal*, 113(486): C153–66.

Nalbantian, Haig R., and Andrew Schotter. 1997. "Productivity under Group Incentives: An Experimental Study." *American Economic Review*, 87(3): 314–41.

Olson, Mancur. 1965. *The Logic of Collective Action*. Cambridge, MA: Harvard University Press.

Putnam, Robert. 2000. *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon & Schuster.

Sherif, Muzafer. 1966. *Group Conflict and Cooperation*. London: Routledge and Kegan Paul.

Shogren, Jason F., Sungwon Cho, Cannon Koo, John List, Changwon Park, Pablo Polo, and Robert Wilhelmi. 2001. "Auction Mechanisms and the Measurement of WTP and WTA." *Resource and Energy Economics*, 23(2): 97–109.

Smith, Adam. 1759. *The Theory of Moral Sentiments.* Oxford: Oxford University Press, 1976.

Tajfel, Henri, and John Turner. 2001. "An Integrative Theory of Intergroup Conflict." In *Intergroup Relations*, ed. Michael A. Hogg and Dominic Abrams, 94–109. Hove: Psychology Press.

Tan, Jonathan H. W., and Friedel Bolle. 2007. "Team Competition and the Public Goods Game." *Economics Letters*, 96(1): 133–39.

Waldzus, Sven, and Amelie Mummendey. 2004. "Inclusion in Superordinate Category, In-group Prototypicality, and Attitudes towards Out-group." *Journal of Experimental Social Psychology,* 40(4): 466–77.

Waldzus, Sven, Amelie Mummendey, and Michael Wenzel. 2005. "When 'Different' Means 'Worse': Ingroup Prototypicality in Changing Intergroup Contexts." *Journal of Experimental Social Psychology,* 41(1): 76–83.

Zizzo, Daniel J. 2003. "You Are Not in My Boat: Common Fate and Similarity Attractors in Bargaining Settings." Oxford University Department of Economics Discussion Paper 167.