Author's Accepted Manuscript

Identity, homophily and in-group bias

Sergio Currarini, Friederike Mengel

Cite this article as: Sergio Currarini and Friederike Mengel, Identity, homophily and in-group bias, *European Economic Review* http://dx.doi.org/10.1016/j.euroecorev.2016.02.015

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Identity, Homophily and In-Group Bias[*]

Sergio Currarini[†]   Friederike Mengel[‡,§]

March 9, 2016

Abstract

Many instance of social interaction display either or both of the following well documented phenomena. People tend to interact with similar others (*homophily*). They also tend to treat others of shared social identity more favorably (*in-group bias*). While both phenomena involve some degree of discrimination towards others, a systematic study of their relations and interplay is yet missing. In this paper we report the findings of an experiment designed to address this issue. Participants are exogenously and randomly assigned to one of two groups. Subsequently they play a sequence of eight games with either an in-group or an out-group member. In treatment EXO in- and out-group matches are formed exogenously, while in ENDO participants can choose between in- and out-group matches. We find strong evidence of in-group bias in EXO, and strong evidence of homophily in ENDO. In-group biases, however, either decrease or disappear altogether under endogenous matching. We show that self selection of homophilous agents into in-group matches cannot explain this fact. We also find that homophily is strongly correlated with risk aversion, and we build on this evidence to derive a rationale for both the existence of homophily and the disappearance of in-group biases under endogenous matching.

Keywords: In-Group Bias, Homophily, Endogenous Matching, Experiments, Game Theory, Segregation, Discrimination.

JEL-classification: D03, D01, C91, C92, C7.

[†]University of Leicester. Email: sc526@le.ac.uk

[‡]Department of Economics, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom, e-mail: fr.mengel@gmail.com

[§]Department of Economics (AE1), Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands

1 Introduction

Socio-economic discrimination refers to the widespread influence of various dimensions of people's identity on their choices and socio-economic opportunities. Depending on the context, discrimination may affect either the formation of social contacts, or the way in which people treat their counterparts in economic transactions, or both. *Homophily* - the tendency of people to interact with similar others - is a well documented feature of most social networks, and is present along many dimensions of similarity (such as ethnicity, gender, religious views,...) and typologies of social ties (from the intimate relations of friendship and marriage to business collaborations and everyday interactions). *In-group bias* - the tendency to treat others more favorably if they are perceived to belong to the same group - has been well documented in social identity research both in Social Psychology and more recently in Economics. Both phenomena have important welfare consequences, and the range of related policy issues include the discussion about "parallel societies", (sex-) segregated education, the costs and benefits of cultural diversity, the management of ethnic conflicts and the design of fair and efficient matching institutions among many others.

Despite the fact that both homophily and in-group bias reflect some degree of discrimination towards others, a systematic joint analysis of these phenomena and of their interplay is yet missing. The segregation of social contacts implied by homophily may both affect and be affected by discrimination in economic transactions. Homophily might for example originate from a rational reaction to the expectation of preferential treatment from the opponent, i.e. to the anticipation of in-group biases. Hence homophily might be a consequence of in-group biases. On the other hand, by affecting the patterns of interaction in favour of homogeneous contacts, homophily may also affect agents' behaviour and in-group bias.

In this paper we study homophily and in-group bias in a controlled laboratory experiment. Participants are randomly assigned to one of two groups, called the RED group and the BLUE group, with no further identity enhancing activity. We adopt, hence, what is called the "near-minimal" group design paradigm (see Tajfel and Turner (1979) and the literature cited below). Choosing such arbitrary assignment to minimal groups reduces the possibility that stereotypes or prejudices participants may have about specific identities, such as gender, ethnicity or religion are triggered in the experiment. Everything participants know about the participant(s) they interact with in the experiment is whether s/he is from the RED or BLUE group.

Most of our results come from the analysis of two treatments, that differ in the adopted matching institution: EXO, in which participants are matched according to a color-blind uniform random process, and ENDO, in which participants are allowed to affect their probability of being matched with RED and BLUE types (see below for details). After matching has occurred, in both EXO and ENDO participants play a series of 8 games designed to elicit their degree of

altruism, positive reciprocity and negative reciprocity.[1]

In treatment ENDO homophily is measured by the willingness of participants to be matched with opponents from the same RED/BLUE group. We elicit two measures of homophily. First we ask for a (non-incentivized) expression of preference for a RED or BLUE match. Second, we elicit participants' willingness to pay (wtp) for an in-group (or out-group) match.[2] Participants are then matched according to their wtp, such that agents with a higher wtp for an in-group (out-group) match are more likely to be matched with in-group (out-group) opponents. Afterwards we let participants play the 8 games.

We find evidence of pervasive *homophily* even in our minimal setting. About 45 percent of participants in treatment ENDO indicate a strictly positive wtp for an in-group match and 70 percent indicate a (weak) preference for an in-group match. We also record significant *in-group bias* in the EXO treatment. Participants are about 34% *more* likely to reward an in-group match, but 39% *less* likely to punish an in-group match compared to out-group matches.

However, while we find substantial in-group biases in EXO, in-group biases either decrease or vanish altogether in the ENDO treatment. More precisely, while in EXO participants are 34% more likely to act positively reciprocal and 39% less likely to act negatively reciprocal in in-group matches compared to out-group matches, there is no statistically significant difference between the two in ENDO. This evidence leads to a first key insight: to the extent that participants' expectations are correct, homophily cannot stem from strategic anticipation of in-group bias. This is also confirmed by a comparison of levels of homophily in ENDO with two benchmark treatments, one with no scope for in-group bias and a second with substantial scope for in-group bias. We find no significant difference with the first, and significantly lower levels of homophily in ENDO compared to the second.

The substantial reduction of in-group bias in the ENDO treatment provides a second insight: the nature of the matching institution affects the degree of discrimination in economic transactions. In particular, when homophily is allowed to shape the patterns of social interaction, less in-group bias (on average) obtains as a result. We show that self-selection of homophilous agents into in-group matches cannot, by itself, explain the observed decrease from EXO to ENDO. Some shift in behavior must have therefore occurred in response to the change in the matching institution. Somewhat surprisingly, all this suggests that while homophily does not result from expected in-group bias, the amount of realised in-group bias depends on whether homophily has a "playing field".

But if not the expectation of preferential treatment, what causes homophily in our experiment? One possible clue might lie in the evidence we gathered in our post-play questionnaire on risk attitudes, cognitive abilities, gender, age and nationality. In particular we found that ho-

---

[1] We use variants of some of the games described in Charness and Rabin (2002).

[2] In the experiment we never use the terms in-group or out-group, but only RED and BLUE group.

mophily is positively correlated with a (non-incentivized) measure of risk aversion, but not with any of the other measures elicited. In an online questionnaire posted on Amazon Mechanical Turk we then reproduced this finding. Respondents who are less willing to take risks tend to be more homophilous. These findings echo recent sociological theories, which interpret homophily as a way to reduce subjective uncertainty (Hogg, 2000).[3]

There is a final, somewhat provoking, message coming out of our exercise. Even though more homophily would probably lead to more discrimination (in-group bias) "for any given matching institution", letting agents decide about their match (moving from exogenous to endogenous matching) decreases aggregate discrimination. In particular, letting agents be in control of their own economic relations has two opposite effects on social structure and economic outcomes: the degree of segregation (measured by the share of in-group interactions) will increase due to homophily but, at the same time, social discrimination (measured by in-group bias) may decrease due to a combination of self selection and changed individual behavior. Hence self-selected segregation need not necessarily be detrimental to the level of pro-social behaviour in a society and need not increase discrimination. While one should be clear that there is a big gap between the minimal design in this experiment and the complex life outside the lab, the minimal design has allowed us to uncover a mechanism, which could potentially be of crucial relevance for the assessment of policies that affect social and economic segregation.

Previous literature has documented homophily in several field studies. See e.g. Currarini et al. (2009), Currarini et al. (2010), Centola (2011) or McPherson et al. (2001) for surveys of the sociological literature. As already mentioned above this literature is mostly focused on documenting homophily as a biased matching pattern, but has found it difficult to identify the sources of homophily. To our knowledge there are no systematic experimental studies on homophily yet.[4]

There has been much more experimental work on in-group bias. A number of papers in psychological research have found evidence in support of in-group favoritism under a "minimal paradigm" design, where the assignment of agents to groups is made with no reference to previous interaction, correlation of preferences or pre-formed identities (see the pioneering work by Tajfel and Turner (1979)). This has been confirmed in a recent study in experimental economics by Chen and Li (2009), who have found in-group biases when agents are sorted in ad-hoc manner into two groups labeled with colours (maize and blue). Our treatment EXO is closely related to the "RandomBetween treatment" in the study by Chen and Li (2009) and our results from this treatment are in line with their findings from their "RandomBetween treatment". Other work in experimental economics has not found biases in a truly minimal design (see Charness et al.

---

[3]Shifting the focus on risk attitudes as the primitive source of homophily requires some rationale for agents to perceive in-group interaction as less uncertain. We discuss possible explanations in detail below.

[4]Experimental studies with endogenous group formation that are not directly measuring homophily are discussed below.

(2007) in the context of prisoner's dilemma and battle of sexes, Eckel and Grossman (2005) in the context of team production and Chen and Chen (2011) in minimum effort games.). Strong in-group biases have been found by Chen and Chen (2011), Charness et al. (2007) or Chakravarty and Fonseca (2014) in designs that were not truly minimal. Ioannou et al. (2015) find in-group biases only in settings where group identity is reinforced by displaying group payoffs. There is also strong evidence of the role of priming in reinforcing identity-based behavior and preferences (Benjamin et al. (2010), Van Bavel et al. (2008)).

Attempts to study in-group bias when the pattern of interaction is to some extent endogenous include Foddy et al. (2009) who find that agents prefer to receive donations from in-group members rather than from out-group members. They also show that this can be imputed to the expectation of better treatment inside the group. Our experimental evidence only partially confirms and points to a perceived reduction in uncertainty as a main motive for homophily (see below). A number of papers have considered social dilemma games with an endogenous group structure (Coricelli et al. (2004), Keser and van Winden (2000), Grimm and Mengel (2009) and Ahn et al. (2009) among others). Since selection and exclusion - that are indeed found to affect behaviour - are based on behavior rather than group membership or identity in these studies, they are somewhat less related to our work. One study where selection and exclusion are based on both behaviour and (possibly identity enhancing) team building tasks is Charness et al. (2014). They allow for endogenous group formation in three stages involving expulsion, voluntary exit and reformation in the context of a public good game. They find that endowments are more important than the team-building task in determining segregation.

Our paper is organized as follows. Section 2 describes our experimental design in full detail. Section 3 and 4 present the evidence homophily and in-group bias. Section 5 discusses the possible determinants for these phenomena, and for the decrease in in-group bias under endogenous matching. Section 6 concludes.

2 Design

Our experiment was conducted at the BEE-Lab at Maastricht University between March-May 2011. 258 participants participated in one of our 5 main treatments. Our basic treatments are treatments ENDO, EXO and CONTROL, which we describe next.

Treatment ENDO In treatment ENDO participants were first randomly and exogenously allocated to the BLUE and RED group and informed about which group they belong to. Subsequently the experiment developed as follows: First participants were asked whether they preferred to be matched with a member of the RED or BLUE group. Subsequently (on a new screen) their willingness to pay (*wtp*) for their choice was elicited. (Details on the elicitation

mechanism can be found below). They were then matched with a member of the RED or BLUE group (with probabilities depending on their *wtp*) and informed about the group of their match (RED or BLUE). Subsequently the two players in a given match were randomly allocated the role of Player A and B (with equal probabilities) and played 8 games with their match in fixed sequence.

	First Mover (A) chooses	Second Mover (B) chooses	
G1	no choice	(400,400) or (750,375)	Altruism
G2	no choice	(100,300) or (400,200)	Altruism
G3	(250,250) or let B choose	(100,100) or (500,100)	Negative Reciprocity
G4	(50,650) or let B choose	(0,100) or (100,100)	Negative Reciprocity
G5	(500,0) or let B choose	(300,300) or (600,275)	Positive Reciprocity (Inequ. increasing)
G6	(250,0) or let B choose	(100,100) or (250,50)	Positive Reciprocity (Inequ. increasing)
G7	(350,100) or let B choose	(300,300) or (100,350)	Positive Reciprocity (Inequ. decreasing)
G8	(400,0) or let B choose	(200,200) or (0,400)	Positive Reciprocity (Inequ. decreasing)

Table 1: The 8 games. Payoffs are in format (π_A, π_B) where π_i is the payoff of player i.

Table 1 describes the 8 games. They are variants of some of the games described in Charness and Rabin (2002). In Games 1 and 2 there is no choice for Player A and Player B chooses between two allocations. These games indicate how altruistically Player B behaves. In all other games Player A moves first and either ends the game by picking an allocation or lets Player B choose, who then chooses between two allocations. Games 3 and 4 indicate how negatively reciprocal Player B behaves and Games 5-8 indicate how positively reciprocal Player B behaves. We will sometimes distinguish between the cases where positive reciprocity is inequality decreasing (Games 7 and 8) or increasing (Games 5 and 6). We can also distinguish whether altruism is inequality decreasing (Game 1) or not (Game 2) and whether negative reciprocity is inequality decreasing (Game 3) or increasing (Game 4).

Participants in the role of Player B were asked to make a conditional choice indicating what they would do if player A decided to let them choose (strategy method). This means we observed the choice of *each* participant in the role of Player B irrespective of what Player A did actually choose. Hence even if there is little variation in player A's behaviour or if player A's behaviour is very different across treatments, we get a full set of observations from player B. Participants did not receive feedback about each other's choices until all eight games had been played. This implies that each participant can be treated as an independent observation in all the games.

Treatment EXO Our second treatment (EXO) coincides with ENDO except for the fact that participants were randomly and exogenously matched with either someone from the RED or BLUE group to play the 8 games. Comparing behaviour in EXO and ENDO enables us to understand the connection between homophily (which can manifest itself only under endogenous matching) and in-group biases at the population level.

6

CONTROL In our CONTROL condition participants were randomly matched to play the 8 games, but there were no RED or BLUE groups in this treatment.[5] The control condition allows us to see how behaviour is affected by the introduction of groups. Understanding how the creation of these artificial differences among our participants affects behaviour has relevant implications regarding the role of heterogeneity within a society.

Those are our main treatments that we will use to understand homophily and in-group bias. Let us briefly define these two key notions, that we will discuss in more detail later.

Homophily As we mentioned before, by homophily we will refer to a *preference* for interacting with agents from the same group. Most of the time, we will measure homophily by the willingness to pay for an in-group match (and heterophily as the *wtp* for an out-group match). It is important to note, though, that participants faced two subsequent screens. On the first, they express a non-incentivized preference for RED or BLUE and only on the second screen they are asked for their *wtp*. We will use the second measure as a robustness check for some results. Given our design, more homophily (in the sense defined above) will lead to a higher expected number of in-group matches in the ENDO treatment.

In-group Bias By in-group bias we will refer to differences in choices in the eight games across in- and out-group matches, which we will measure by the *behavior of player B*.

Additional Treatments We conducted two additional treatments to understand the reasons behind homophily. Both LOWB and COORD coincide with ENDO, but the 8 games were different in each case. In LOWB games were such that there is no scope for in-group biases. In other words in LOWB there are no strategic reasons to be homophilous and hence we consider the amount of homophily observed in this treatment a lower bound. In COORD, however, games were such that being from the same group could potentially help to resolve coordination problems. Hence we expected there to be more homophily in COORD compared to ENDO compared to LOWB. Understanding where ENDO lies in the range between LOWB and COORD can help us understand to which extent homophily is strategic. Sample Games from these treatments can be found in Appendix G. Table 2 summarizes the treatment structure together with the number of (independent) observations per treatment.[6]

---

[5]Sample Instructions for treatments ENDO, EXO and CONTROL can be found in the Appendix.

[6]54 additional participants participated in two more treatments LABEL and FIXED. The design and some results of treatment LABEL can be found in the Appendix. The results of treatment FIXED will be used for another study and are available upon request. Other than the treatments reported we did not run any additional sessions or treatments, nor did we conduct pilot studies.

Label	Matching	Minimal groups	Games	Participants	Role B
T1 ENDO	endogenous	yes	SocialPreferences	67	40
T2 EXO	exogenous	yes	SocialPreferences	73	40
T3 CONTROL	exogenous	no	SocialPreferences	44	25
T4 LOWB	endogenous	yes	No Scope for Bias	38	-
T5 COORD	endogenous	yes	Coordination	36	-

Table 2: Main Treatments with number of participants.

Mechanism to elicit *wtp*   We use the following mechanism to elicit the *wtp* of participants in treatments ENDO, LOWB and COORD. Participants are endowed with 500 ECU at the beginning of the experiment. This endowment was *not* given to them in connection with the elicitation of their *wtp*. To elicit their *wtp* they are asked to state a number between 0 and 100 that indicates how much they would be willing to pay to be matched with their preferred group. We then draw a random number between 0 and 100 from a uniform distribution. If the randomly drawn number exceeded the number stated by the participant they were matched randomly. Otherwise they were matched with their preferred group and an amount corresponding to their number was deducted from their endowment.[7] Since we had a finite number of participants in the experiment, there was a small chance that this mechanism is infeasible. In this case (which didn't happen) we would have matched participants randomly and not deducted anything from their endowment. Participants were fully informed about all these details (see Instructions in Appendix D).

Our mechanism is essentially a first-price version of the well known Becker-deGroot-Marschak (BDM) mechanism. The latter has recently been criticized for being too complex to understand for participants and hence possibly leading to distorted decisions (Cason and Plott (2014)).[8] Since in the first-price version the price that participants state is also the price they pay, the first-price version is closer to everyday experience of most participants. We, hence, hoped that the first-price version might be easier for participants. In addition to the *wtp* elicited in this manner we also have the non-incentivized binary preference for in- and out-group matches (see paragraph "Homophily").

Minimal Group Design   The design we used to induce in- and out- groups is called the near-minimal group paradigm in social psychology (See e.g. Tajfel and Turner (1979)). According to the near-minimal group design non-overlapping groups are created using trivial and sometimes

---

[7]Participants under this mechanism should maximize $\frac{wtp}{100} * (u_{IN} - u_{OUT} - wtp) + (1 - \frac{wtp}{100})(\frac{u_{IN} - u_{OUT}}{2})$, where u_{IN} is the utility derived from an in-group match and u_{OUT} the utility they derive from an out-group match. Assuming that $u_{IN} > u_{OUT}$, the interior optimum is to bid $wtp = \frac{u_{IN} - u_{OUT}}{4}$. If $u_{IN} = u_{OUT}$, the optimum is to bid zero and if $u_{IN} < u_{OUT}$, the optimum is to bid $wtp = \frac{u_{OUT} - u_{IN}}{4}$ for an out-group match.

[8]See also Plott and Zeiler (2005) and the failed replication by Fehr et al. (2015). Earlier research has questioned the incentive compatibility of the BDM mechanism pointing out its dependence on the independence axiom (Karni and Safra, 1987), its failure under Non-EU preferences (Horowitz, 2006) and circular validation issues when relying on such mechanisms more generally (Guala, 2000).

meaningless tasks. Group membership has to be anonymous and no social interaction (face-to-face or computer-mediated chat) should take place between subjects. Summarizing 15 years of sociological research Tajfel and Turner (1986) conclude that "the trivial, ad-hoc intergroup categorization leads to in-group favoritism and discrimination against the out-group". Chen and Li (2009) have compared ad-hoc categorization according to two colours (maize and blue) with categorization according to expressed preferences for paintings by different artists in an economic experiment. They found that both procedures lead to significant in-group biases, while there are no significant effects of the procedure on either size or direction of in-group biases. In our study we used colours RED and BLUE as group labels to avoid hierarchical labels (such as group 1 and 2 or group A and B) and to avoid that labels are correlated with things we can't control for (such as stereotypes regarding gender, race etc.).

Many studies have shown that expectations about positive and negative reciprocity vary greatly between different cultures and across genders and there are interaction effects between the two as well. See e.g. Gaechter and Herrmann (2009) or Bohnet et al. (2010). Hence to be sure that the *wtp* for in-group matches captures homophily and not e.g. differing expectations across dimensions such as gender, culture etc.., it is important to use neutral groups. If we used non-neutral groups (such as e.g. gender), then we couldn't be sure that what we call homophily does not simply reflect a gender stereotype. Of course in reality gender- and other stereotypes might well create homophily. In this study, however, we want to focus on deeply rooted culturally learned preferences for in-group matches, which are activated even for meaningless labels such as RED and BLUE. In a sense we will measure the component of homophily that is common to all dimensions (race, gender, social status etc.) in which homophily will manifest itself outside the lab. In the instructions we also alternated between "RED and BLUE" and "BLUE and RED" to avoid creating a hierarchy between the groups. We also test whether the color affects behavior *per se* and find that generally it doesn't (see Appendix B.3).

The abstract minimal group environment is certainly less rich than the natural environment in the field. However, as we have outlined in Section 1, most existing studies in this literature suffer from the problem, that their environment is "too rich" to disentangle different sources for homophily and in-group biases. We will see below that our very stylized setting will help us to identify clear links between these phenomena.

Other Details We used the experimental software z-tree by Fischbacher (2007) and the recruitment system Orsee by Greiner (2004). Participants were paid the sum of payoffs obtained in all games (with an exchange rate 1 Euro=500 ECU), a show up fee of 2 Euros as well as (in treatments ENDO, LOWB and COORD) whatever remained from their initial endowment of 500 ECU. In some of the sessions a participant in the role of Player A was matched to two participants in the role of player B. This was done because of low show-up. In this case participants
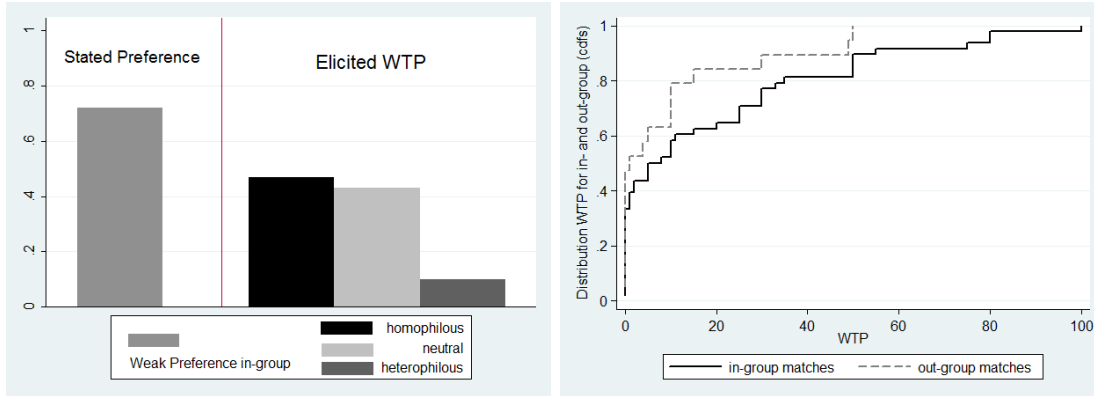
Figure 1: 72 percent of participants prefer an in-group match according to stated preferences. 47 percent are strictly homophilous, 43 percent neutral and 10 percent strictly heterophilous according to the elicited *wtp*. The distribution of wtp for in-group matches first-order stochastically dominates that for out-group matches (Kolmogorov-Smirnov test, $p < 0.0001$).

in the role of Player A only received feedback and payments from one of their matches which was randomly selected. At the end of the experiment participants filled in a questionnaire where we elicited their risk aversion and a measure of cognitive ability and asked for their gender, age and nationality.[9] Details on these descriptive statistics in our sample can be found in Appendix A, where also randomization checks based on these observables can be found. Each session lasted between 30min (CONTROL) - 70min (ENDO) and participants earned on average 13,40 with a minimum of 5,70 and a maximum of 24,60 Euros.

3 Homophily

We start by assessing how prevalent homophily is in our experiment. To these ends, we define three types of agents, based on their declared willingness to pay and on the type of preferred match. *Strictly homophilous* agents are those with a strictly positive willingness to pay for an in-group match; *Strictly heterophilous* agents have a strictly positive willingness to pay for an out-group match; and *Neutral* agents have a willingness to pay of zero.

Figure 1 shows the share of these types in the population. The left-most bar in figure 1(a) measures the share of agents that declared to prefer an in-group match, including those who afterwards declared a willingness to pay of zero. The three right bars in Figure 1(b) show the percentages of strictly homophilous, neutral and strictly heterophilous participants in ENDO.

There are about 45% of strictly homophilous agents, about the same proportion of neutral

---

[9]We also posted an independent questionnaire on Amazon Mechanical Turk to get additional evidence on the correlation between risk aversion and homophily. More details about this questionnaire are found in Sections 5 and C.3.

agents, and only few strictly heterophilous agents that make up for about 10% of the population. Figure 1(b) shows the distribution (cdfs) of *wtp* for in-group and out-group matches, respectively, conditioning on those agents with a weak preference for the respective type of match. It can be seen that the distribution of *wtp* for in-group matches dominates that for out-group matches in the sense of first-order stochastic dominance.

The average *wtp* computed as the average of all agents who stated to prefer an in-group or out-group match (including neutral agents) is 13.86 for in-group matches and 4.23 for out-group matches. Among strictly homophilous agents the average *wtp* for an in-group match is 29.03 and among strictly heterophilous agents it is 23.06.[10] In Appendix B we also compute the ex post optimal bid for an agent who perfectly anticipates in-group biases (discussed extensively in Section 4). We find that this bid is ≈ 23. Hence, while many participants put in zero bids, resulting in underbidding on average, those who display homophily overbid by about 6 units. This could possibly be because they anticipate stronger in-group biases than there are or because the bidding we observe is largely non-strategic.

Having assessed the presence of substantial homophily in our experimental setting, our aim is to use our controlled environment to better understand what the sources of homophily in social behaviour are. One natural conjecture is that agents prefer to match with similar others because they strategically anticipate potential favourable in-group biases in the interaction to follow, i.e. higher payoffs in in-group matches compared to out-group matches. This conjecture views homophily as driven by in-group bias, through agents' anticipation of potential gains. To test this conjecture, we compare our main treatment ENDO to treatments LOWB and COORD. In LOWB, strategic interaction is such that players cannot favour or discriminate others on grounds of group membership. In other words, in LOWB there is no scope for in-group bias. In COORD, instead, there are clear incentives for in-group matching (sample games from treatments LOWB and COORD can be found in Appendix G). Where ENDO stands with respect to these treatments helps us understand to what extent homophily in ENDO is driven by expected game payoffs. Comparing ENDO and LOWB we record almost no difference in average *wtp*, which is even slightly higher in LOWB, and exactly the same percentage of strictly homophilous agents (see Figure 2). In contrast, behaviour in treatment COORD significantly differs from that in both ENDO and LOWB, both in terms of the average *wtp* and in terms of the percentage of agents with a strict preference for an in-group match (two-sided ranksum test, $p < 0.0019$ and $p < 0.0581$, respectively). This evidence suggests that while agents do respond for when they face explicit incentives to bid for in-group matches (as in COORD), the homophily we observe in ENDO is not strategic in that it is not driven by the anticipation

---

[10]The entire distribution of the *wtp* in treatment ENDO as a bar chart can be found in Figure 8 in Appendix B. Note that the distribution of the *wtp* for an in-group match is not truncated at zero, since agents can pay up to 100 for an out-group match, i.e. to avoid an in-group match. A t-test of the one-sided hypothesis that this (non-truncated) *wtp* > 0 returns a p-value of 0.0035.
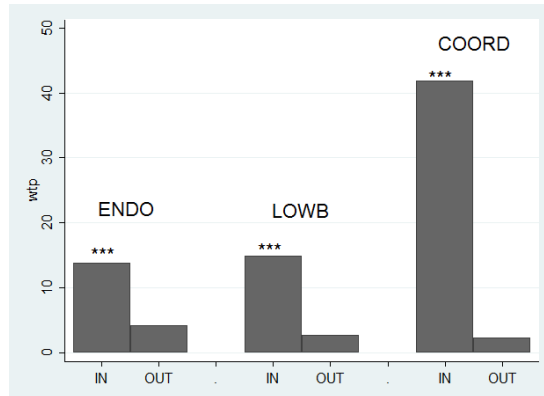
Figure 2: Comparison of *wtp* across treatments ENDO, LOWB and COORD. In all treatments bids are significantly different from zero in in-group, but not out-group matches. Bids for an in-group match are statistically indistinguishable in ENDO and LOWB (two-sided ranksum test, $p > 0.4$), and significantly higher in COORD compared to either (two-sided ranksum test, $p < 0.0019$).

of more favorable treatment in in-groups (or in-group bias). As we will see in Section 4 any expectation of substantial in-group bias would be misguided, since, as it turns out, in-group biases are statistically no different from zero in ENDO.

Result 1: Homophily in a minimal design

- *There are about 45% of strictly homophilous agents, 45% neutral agents, and 10% strictly heterophilous agents in the ENDO treatment.*

- *Homophily is not a matter of opportunities and is not driven by the anticipation of in-group biases in the social preference games.*

We will discuss possible alternative explanations for homophily in ENDO in Section 5. Some of our discussion there will be based on the relation between in-group bias and homophily, and on how in-group bias and discrimination are affected by the matching institution. These topics are covered in detail in the next section.

4   In-Group Bias and the Role of the Matching Institution

We start by giving a descriptive overview of behaviour in the 8 social preferences games, then define our measure of in-group bias, and present our evidence for the EXO and ENDO treatments. Descriptive Statistics on all main variables used in the following can be found in Table

	EXO A	ENDO A	EXO B	ENDO B
G1	–	–	0.30	0.29
	–	–	(0.37, 0.23)	(0.31, 0.27)
G2	–	–	0.15	0.28
	–	–	(0.23, 0.06)	(0.26, 0.30)
G3	0.49	0.26	0.37	0.35
	(0.53, 0.44)	(0.24, 0.29)	(0.18, 0.61)	(0.43, 0.26)
G4	0.61	0.61	0.35	0.18
	(0.70, 0.50)	(0.59, 0.69)	(0.18, 0.55)	(0.23, 0.15)
G5	0.34	0.34	0.32	0.38
	(0.36, 0.32)	(0.62, 0.08)	(0.50, 0.12)	(0.44, 0.30)
G6	0.25	0.37	0.23	0.20
	(0.30, 0.19)	(0.62, 0.15)	(0.37, 0.08)	(0.29, 0.18)
G7	0.16	0.30	0.65	0.65
	(0.24, 0.07)	(0.54, 0.08)	(0.69, 0.62)	(0.59, 0.70)
G8	0.05	0.15	0.82	0.72
	(0.08, 0.00)	(0.24, 0.08)	(0.91, 0.72)	(0.85, 0.67)

Table 3: Overview of behaviour in treatments EXO and ENDO. Percentages in A columns indicate the frequency with which player A moves "In", i.e. lets Player B choose. Percentages in B columns indicate the mean frequency with which Player B chooses the altruistic (G1-G2), negatively reciprocal (G3-G4) or positively reciprocal (G5-G8) options, respectively. Numbers in brackets indicate the mean frequency in in- and out-group matches.

11 in Appendix A (questionnaire variables) and Table 12 in Appendix B (main experimental variables).

A first look at Table 3 suggests that there are not many differences in average behaviour of Player B between ENDO and EXO. However, there seem to be consistently higher differences between in- and out-group behaviour in treatment EXO compared to treatment ENDO. For player A we find differences also in means: Player A's seem to be somewhat more willing to let Player B choose in ENDO (compared to EXO) in positive reciprocity games and less willing in negative reciprocity games. Below we will try to uncover some of the mechanisms behind these numbers.

We define in-group bias as the difference in behaviour of player B between in-group and out-group matches. We focus on Player B in measuring in-group bias since in-/out-group differences in the behaviour of Player A could be due not only to a genuine tendency to discriminate between groups (which the notion of in-group bias aims to capture), but also to the (strategic) anticipation of differential behaviour by Player B in in/out-group matches. Player B, by contrast, moves last and has therefore no such strategic reasons, and any behaviour differences for player B are clean evidence of in-group biases.

Since our design does not rely on the strategy method, all participants play the games *either* in an in-group situation *or* in an out-group situation. One advantage is that our design is not leading, since it avoids conditional choices of the type "if I was matched within the "RED" group

I would do this, otherwise I would do that". Such contingent choices might be seen as suggestive, i.e. participants might expect that they should choose differently just because a difference is made by the experimenter.[11] The downside of our design is that for any given sample size we will have fewer observations in each situation and that we can't directly observe in-group bias at the individual level.[12] So, in measuring in-group bias we will compare the behaviour of different participants in in-group and in out-group matches.

4.1 In-Group Bias with Exogenous Matching

We start with treatment EXO, where agents have no way of affecting their matching probabilities. Table 4 shows the results of simple logit regressions where a binary variable, indicating whether a participant in the role of player B displayed altruistic, negatively or positively reciprocal behavior is regressed on another binary variable that indicates whether behaviour took place in an in-group match (variable "in-group"). We clustered standard errors by individual, since we observed each participant in two games for each category. Remember the categories are altruism (games 1,2), negative reciprocity (3,4), positive reciprocity (inequality increasing case 5,6) and positive reciprocity (inequality decreasing 7,8).[13]

We find that there is significantly *less negative reciprocity* in in-group compared to out-group matches (column (2) in Table 4) and *more positive reciprocity* in Games 5-6 (where positive reciprocity is inequality increasing) (column (3) in Table 4). In particular there is a 34 percent increase in positive reciprocity and a 40 percent decrease in negative reciprocity in in-group matches according to the marginal effects of the logit regression below. (See also figure 3). Those results are consistent with Chen and Li (2009) who, with a within subject design, find a 19 percent increase in positive reciprocity and a 13 percent decrease in negative reciprocity in in-group compared to out-group matches. If anything, the in-group biases we find are slightly stronger. This is consistent with evidence from e.g. political science, where it has been found that people tend to suppress discriminatory tendencies when directly confronted with multiple options ((Kuklinski et al., 1997)).

Table 5 shows the same regression except that we included additional variables from the

---

[11]Of course the opposite might also occur and participants might suppress discriminatory tendencies when directly confronted with conditional choices (Kuklinski et al., 1997).

[12]Between subject designs also always rely on the assumption that there is no selection (e.g. on social preferences) into the different treatments. Randomization checks in Appendix A show that there was no selection on age, gender, nationality or risk aversion.

[13]Since we will pool the two games in each category in the following let us have a brief look at within-pair consistency. The conditional probability to choose the altruistic option in G2 conditional on having done so in G1 is 0.84 in ENDO and 0.80 in EXO. The conditional probability to choose negatively reciprocal in G4 conditional on having done so in G3 is 0.98 in ENDO and 0.92 in EXO. The probability to choose positively reciprocal in G6 conditional on having done so in G5 is 0.78 in ENDO and 0.83 in EXO. And, finally, the probability to choose positively reciprocal in G8 conditional on having done so in G7 is 0.88 in ENDO and 0.97 in EXO. Individual behaviour is hence reasonably stable across the two games in each category.

	(1) G1-G2 Altruism	(2) G3-G4 Neg. Recp	(3) G5-G6 Pos. Recp (i)	(4) G7-G8 Pos. Recp (d)
in-group	0.956	-1.841***	2.123**	0.665
	(0.651)	(0.562)	(0.839)	(0.589)
Constant	-1.825***	0.336	-2.398***	0.693*
	(0.557)	(0.441)	(0.781)	(0.400)
Groups	40	40	40	40
Observations	80	80	80	80

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: In-group Bias in EXO. Logit Regression. 40 individuals (groups) in the role of Player B, 80 observations. Standard errors clustered by id. 53 percent of all matches were in-group matches. ***1%,**5%,*10% significance.

EXO	G1-G2 ("Altruism")	G3-G4 ("Neg. Recp")	G5-G6 ("Pos.Recp.(I)")	G7-G8 ("Pos.Recp.(d)")
constant	-8.137^{**}	1.995	-9.746^{**}	8.944^{*}
	(3.741)	(4.498)	(4.395)	(5.029)
in-group	1.816^{**}	-1.572^{***}	2.141^{***}	1.008
	(0.785)	(0.605)	(0.807)	(0.748)
age	0.069	-0.117	0.288^{*}	-0.315
	(0.164)	(0.187)	(0.173)	(0.212)
gender	-1.100	0.500	-0.019	-1.706^{*}
	(0.697)	(0.655)	(0.694)	(0.904)
RED	2.063^{*}	-0.590	0.208	-0.465
	(0.917)	(0.625)	(0.625)	(0.672)
risk aversion	0.591^{**}	0.221	-0.038	0.265
	(0.240)	(0.273)	(0.252)	(0.246)
cognitive reflection	0.902^{**}	0.054	0.593^{*}	-0.527
	(0.350)	(0.287)	(0.320)	(0.395)
Groups	39	39	39	39
Observations	78	78	78	78

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5: In-group Bias in EXO. Logit Regression with additional variables (Standard errors clustered by id). 39 individuals (groups) in the role of Player B, 78 observations. One individual dropped who preferred not to answer gender/age question in questionnaire. 53 percent of all matches were in-group matches. ***1%,**5%,*10% significance.

questionnaire as well as a dummy "RED" that takes the value 1 for participants from the RED group to see whether colour matters per se. The questionnaire variables are age, a gender dummy (1=female), a measure of risk aversion and a measure of cognitive reflection. All the variables are described in detail in Appendix C.

The main message is that in-group biases identified in Games 3-4 and 5-6 remain significant when these variables are introduced and the coefficients are of about the same size. In fact statistical significance is even higher for positive reciprocity when controlling for these variables (G5-G6). Interestingly a new in-group bias appears for the category Altruism (Games 1-2) when controlling for these variables. Participants are more altruistic towards in-group members. In addition, it seems that more risk averse, more cognitively reflected people and people from the RED group are more altruistic. We have no convincing explanation for why more cognitively reflected participants might be more altruistic.[14] However none of these variables can robustly explain behavior in any of the other game categories. The single most important variable is the in-group dummy.

4.2 In-Group Bias with Endogenous Matching

We next measure in-group bias in treatment ENDO with endogenous matching. Our main result here is that when agents choose who to match with, the aggregate in-group bias *either diminishes or totally vanishes* in statistical terms. This is illustrated in Figure 3, depicting in-group bias in both treatments EXO and ENDO, for the different pairs of games.

One important issue in comparing the EXO and ENDO treatments is that in the latter, whether or not a person ends up in an in-group match is not exogenous to behavior. In particular, the same characteristic that leads an agent to show differential behavior across in-group and out-group matches may affect that agent's preference over who to be matched with.[15] In Appendix B we reproduce the equivalent of Tables 4 and 5 for treatment ENDO, but because of the endogeneity issue discussed these tables should be read with care. In addition, to control for the willingness to pay, and therefore isolate the effect of a match's type on behavior, we have included the variable "*wtp × in-group*" in some of the logit regressions reported in Table 13 in Appendix B. The right column for each of the pairs of games shows coefficients where the *wtp* variable is omitted. We find that, irrespective of whether we control for *wtp × in-group* or not,

---

[14] Dual process theory sometimes maintains that looking for self-interest is an automated and primed response, while understanding one's ethical obligations to others is a more conscious and reflective process (Moore and Loewenstein (2004)). This could explain why more cognitively reflected participants are more altruistic. Indeed, there is some experimental evidence that delayed/slower decisions imply less negative reciprocity (Grimm and Mengel (2011); Oechssler et al. (2015)) and more altruism (Piovesan and Wengstroem (2009)) and sometimes also cooperative behaviour (Rand et al. (2012)), even though the latter has not been put in doubt by Tinghoeg et al. (2013) or Recalde et al. (2014). Since we only find the effect for altruism and since the link between cognitive reflection and response times seems less than clear (Grossman et al. (2014); Recalde et al. (2014)), we do not want to push this interpretation too much.

[15] As we will show below, this is indeed the case. "Homophilous" agents do behave differently from others.
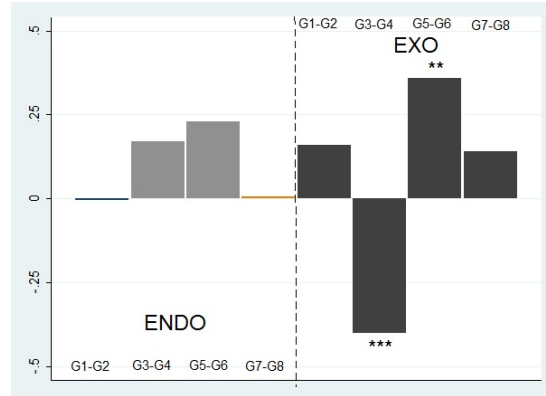
16

Figure 3: In-Group Biases. The graph shows the difference in the percentages of B-players displaying "altruism" (Games 1-2), "negative reciprocity" (Games 3-4), "positive reciprocity" (inequality increasing (Games 5-6)) or "positive reciprocity" (inequality decreasing (Games 7-8)) between in-group matches and out-group matches. On the left hand side is ENDO and on the right hand side is EXO. *** 1%,** 5%,* 10% significantly different from zero (minimum significance level across Tables 4, 5, 13 , 14 or 6).

the coefficients on in-group are not significant. The exception is a marginally significant positive bias for games 5 and 6 (positive reciprocity), which disappears once we control for *wtp*.

Table 14 in Appendix B shows the results of the same regression as the base regressions in Table 13, but again we have included some variables from the questionnaire. In this regression as well we do not find any significant in-group biases for treatment ENDO. The marginally significant coefficient on in-group in Games 5-6 disappears here and instead in-group appears as marginally significant in Games 3-4. Taken together there is no evidence for robust in-group biases in ENDO.

Table 6 shows the correlation coefficients from a Spearman correlation test, where we correlate behaviour with the binary variable "in-group". This table illustrates that biases (correlation coefficients) are uniformly larger in EXO compared to ENDO (in terms of their absolute value) and are only statistically significant in EXO, with the exception of a marginally significant coefficient for Games 5 and 6 (positive reciprocity) in ENDO. Again, the evidence from the table confirms previous analysis: there is little evidence of in-group bias in ENDO.

	ENDO	EXO
Altruism (Games 1-2)	−0.0314	0.2728*
Neg. Reciprocity (Games 3-4)	0.2363	−0.4551***
Pos. Reciprocity (I) (Games 5-6)	0.3103*	0.5392***
Pos. Reciprocity (d) (Games 7-8)	0.0205	0.1870

Table 6: Spearman correlation coefficients. Correlation between the frequency of altruistic, negatively reciprocal etc.. behavior and in-group match dummy. 40 individuals in the role of Player B, 80 observations. *** 1%,** 5%,* 10%.

17

4.3 Selection

We need to be careful in interpreting the reduction of in-group bias from EXO to ENDO in terms of a change in individual behaviour. Because of self-selection, homophilous (heterophilous) agents tend to interact in in-group matches more (less) frequently in ENDO than in EXO. But can this self-selection of types alone explain the observed change in in-group bias? Or can we conclude that some shift in individual behaviour has happened as a result of the change in the matching institution?

Let us first take a step back and note that while self selection may potentially drive the result, the direction in which it operates is not obvious. If homophilous agents, for example, displayed more reciprocity in *both* in- and out-group matches compared to non-homophilous agents, then we would necessarily have an increase in in-group biases as more homophilous agents assemble in in-groups in ENDO compared to EXO. If, however, homophilous agents were less reciprocal in all matches compared to others, we would have a decrease in the bias from EXO to ENDO etc. Hence, it is not sufficient to observe the direction of the bias to rule out self-selection as the unique cause of the decline in in-group bias.
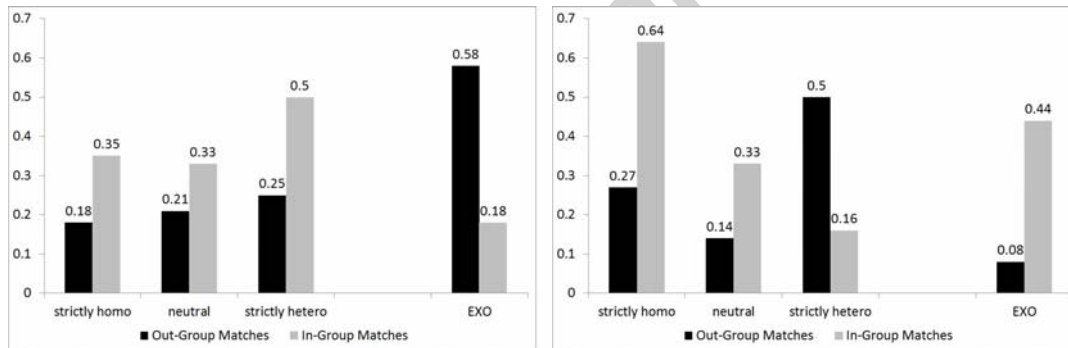

Figure 4: Levels of reciprocity in ENDO (by type of matching preference) and in EXO. Remember that 45 percent of agents are classified as homophilous and neutral, respectively and 10 percent as heterophilous.

To do so we have to answer two questions: (i) are there possible realizations of the matching process that can produce the observed bias in EXO when holding contingent behavior in ENDO for all three types constant? and (ii) is contingent behavior in in-group and out-group matches significantly different between ENDO and EXO? We answer these questions separately for the case of negative and positive reciprocity. Note that by doing so we are testing a weaker hypothesis, since it is the same realization of the matching process that has to produce *both* the bias in negative and positive reciprocity. We start with the case of negative reciprocity illustrated in Figure 4(a).

Negative Reciprocity   For all types, negative reciprocity is weakly higher in in-group matches compared to out-group matches in ENDO. At the same time, in the EXO treatment we observed significantly *less* negative reciprocity at the aggregate level in in-group compared to out-group matches . Furthermore, in ENDO all types are *more* negatively reciprocal in in-group matches compared to average behavior in in-group matches in EXO (one-sided ranksum test $p < 0.07$). And all types are *less* negatively reciprocal in out-group matches in ENDO compared to average behavior in out-group matches in EXO (one-sided ranksum test $p < 0.01$). Hence, even allowing for any distribution of types, there is no realization of the matching process that could have produced this outcome without a shift in behavior.

Positive Reciprocity   The case of positive reciprocity is illustrated in Figure 4(b). Here we see that all types are more positively reciprocal in out-group matches in ENDO compared to EXO (one-sided ranksum test $p < 0.06$).[16] This implies that again there is no realization of the matching process in EXO that may have generated the observed average level of reciprocity, clearly demonstrating that self-selection *alone* cannot explain the changes in aggregate behaviour we observe by moving from exogenous to endogenous matching.

We can now summarise the main insights we have obtained in this section about the role of the matching institution for in-group biases.

Result 2: In-group bias and the matching institution

- *Aggregate in-group biases vanish or decrease in the transition from exogenous to endogenous matching.*

- *Self selection alone cannot explain the change in in-group bias from exogenous to endogenous matching, and a shift in behaviour due to the change in the matching institution must have occurred.*

4.4   Homophily, Discrimination and the Matching Institution

A somewhat provoking implication of the above results is that providing social actors with greater control on their interaction patterns would decrease discrimination, as behavioural differences between in-group and out-group matches would, on average, vanish. So, while on the one hand endogenizing matches has the effect of increasing the degree of *segregation* of social contacts, on the other hand it may mitigate the degree of *discrimination* in social interaction through the joint effect of self selection and of the shift in behaviour describe above. Two observations,

---

[16]We can also conduct this test separately for all types and find significance levels of $p = 0.0929$ (strictly homophilous) and $p = 0.0656$ (neutral) for the one-sided test, but no significance for heterophilous agents due to their small number.

and *caveats*, are in order here. First, due to the design of our experiment, the decrease in discrimination does not refer to a change in the behaviour that the same agent adopts in different types of matches, but rather to the change in the behaviour that is observed, in aggregate, in society in different types of matches. Second, the fall in discrimination mainly comes from an increase in the punishment adopted in in-group matches and not, as one may expect or even wish, from an increase in positive reciprocity or even altruism towards out-group opponents.

A related but conceptually different issue is whether increases in homophily would affect the degree of discrimination for a given fixed matching institution and, if so, in what direction. Figure 5 shows in-group biases separately for the three different types of matching preferences in negative and positive reciprocity games (3-4 and 5-6, respectively). In the positive reciprocity games, we record a significant and positive in-group bias only for homophilous agents, while no bias is present on average for neutral and heterophilous agents. No significant bias is present in negative reciprocity games. The type of bias we find in the positive reciprocity games is of the type one would naturally expect: agents that prefer to match with similar agents, are also prone to reward similar others in return for kind actions; by the same logic agents that prefer to match with dissimilar others would be expected to reward similar agents less. We find evidence for both intuitions, although only in the case of homophilous agents the effect is statistically significant.
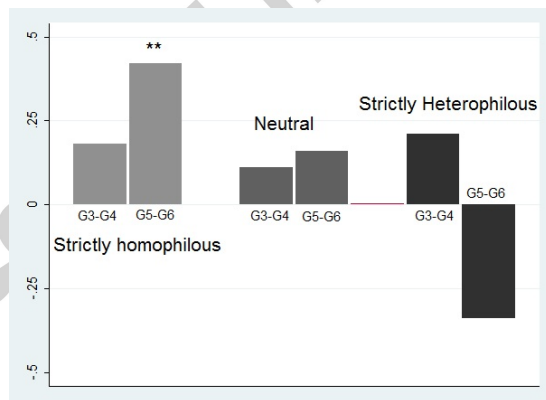

Figure 5: In treatment ENDO in-group biases are only found when restricting to homophilous agents. Stars indicate significance level according to two-sided ranksum test comparing in- and out-group matches *** 1%,** 5%,* 10%. effect sizes are as follows: strictly homophilous: 0.17 in G3-G4, 0.42** (G5-G6); neutral: 0.11 (G3-G4), 0.19 (G5-G6); heterophilous: 0.25 (G3-G4), −0.31 (G5-G6). As a robustness check we also do this analysis for stated preferences only and find the following effect sizes: weakly homophilous 0.18* (G3-G4), 0.48*** (G5-G6); weakly heterophilous 0.07 (G3-G4), −0.114 (G5-G6).

Result 3 *Under endogenous matching, in-group biases can be found only among homophilous agents, who are more positively reciprocal towards the in-group than towards the out-group.*

20

5 Discussion

5.1 On the Sources of Homophily

Within our minimal design, homophily is certainly not a matter of opportunities or stereotypes. Agents meet according to a matching process which is in all respects anonymous except for group membership (RED or BLUE) and the way in which agents affect probabilities according to their declared *wtp*; likewise, the way in which agents are assigned to the BLUE and RED group is independent of any individual characteristic of participants. In Section 3 we have also seen that in our minimal design homophily is not driven by the strategic reaction to the anticipation of in-group biases. Hence we can rule out all of these as possible channels behind homophily in our experiment. In this section we develop and discuss some (evidence-based) conjectures on what the roots of homophily might be, and relate them to the results of the previous section.

We start by reporting the results from the post experimental questionnaire in which we elicited risk aversion of participants. As a measure of risk aversion we use a variable that counts in how many Holt-Laury style lottery choice questions in the questionnaire the participants preferred a sure outcome to a non-degenerate lottery.[17] The questions can be found in Appendix C.1. The variable ranges from 0 (least risk averse) to 7 (most risk averse). It should be noted that this is a non-incentivized measure. Table 7 shows the distributions of the variable risk aversion in treatments ENDO and EXO, which are remarkably similar (see also Figure 7 in Appendix A).

	ENDO	EXO
0 (least risk averse)	0.10	0.10
1	0.07	0.08
2	0.29	0.21
3	0.31	0.34
4	0.18	0.17
5	0.02	0.05
6	0	0
7 (most risk averse)	0	0.01

Table 7: Distribution of the variable Risk aversion.

Participants with a higher willingness to pay for an in-group match are more risk averse (Spearman test $\rho = 0.3586^{***}$). Figure 6 (left panel) illustrates the predicted values of a linear regression, where we explain the *wtp* for an in-group match via the risk aversion variable, as well as data points in risk aversion - *wtp* space. The figure omits strictly heterophilous agents and displays fitted lines from two regressions - one without and one with a square term. Clearly, the *wtp* for an in-group match seems to increase with an agent's risk aversion.

---

[17]Since all the lotteries were presented to participants on the same page and ordered according to the implied CRRA range, we observe very few instances (less than 1%) of multiple switching.
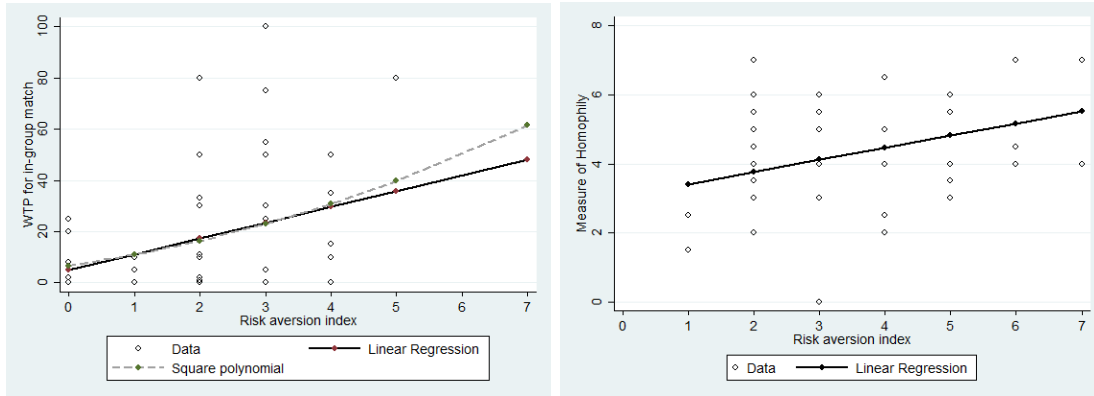
Figure 6: Data points and fitted values from a linear OLS regression of the *wtp* for an in-group match on risk-aversion in treatment ENDO (left panel) and in the AMT study (right panel).

We also conducted a cognitive reflection test in the questionnaire and found no significant correlation between cognitive reflection and a binary measure of homophily nor between cognitive reflection and willingness to pay. Table 8 shows OLS regressions, where we regress the *wtp* for homophilous agents on all our questionnaire variables. It can be seen that the only variable that is significant is the degree of risk aversion of the participant. We run three different regressions where we control for different sets of questionnaire variables. The estimated coefficient on our risk aversion variable is pretty stable across all these.

	(1)	(2)	(3)
constant	9.766	0.4308	8.0820
	(42.376)	(11.828)	(9.517)
age	0.121		
	(1.925)		
gender	−9.022		
	(11.450)		
risk aversion	7.863*	9.6014**	8.8207**
	(4.382)	(3.612)	(3.549)
cognitive reflection	2.918	4.4167	
	(4.688)	(4.076)	
observations(groups)	32	32	32

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 8: Homophily (wtp) . OLS Regression of *wtp* on questionnaire variables for 32 homophilous participants in ENDO. ***1%,**5%,*10% significance.

Result 4 *Homophily is positively correlated with risk aversion.*

To get some independent evidence of this correlation we posted a questionnaire on Amazon

22

Mechanical Turk (AMT) in January 2014.[18] We asked participants a simple question about their risk preferences, namely "How willing are you in general to take risks"? This question has been found to explain risky behavior, such as holdings stocks, smoking and occupational choices better than other standard measures of risk aversion (Dohmen et al. (2011)). We use the answer to this question (on a scale from 0,...,7) to explain respondents' homophily, which we measure with two questions: (i) How strongly do you prefer to "mix with your own kind"? and (ii) How strongly do you prefer to interact with people similar to you? also ranging from 0,...,7. The spearman correlation coefficient between these measures is $\rho = 0.3119^{**}$, amazingly close to the one found in our experiment. Figure 6 (right panel) illustrates the distribution of data points and the linear regression found in column (2) of Table 9. Column (1) also controls for gender and age of our AMT respondents.

	(1)	(2)
constant	3.538***	3.057***
	(1.045)	(0.601)
age	−0.024	
	(0.027)	
gender	0.645	
	(0.645)	
risk aversion	0.355**	0.352**
	(0.159)	(0.152)
observations(groups)	42	42

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 9: OLS Regression of homophily indicator on risk aversion, gender and age in our AMT questionnaire with 42 respondents. ***1%,**5%,*10% significance.

This evidence points to a possible psychological source of homophily. Suppose agents perceived the behavior of similar agents to be more predictable (for reasons to be discussed below). Those who are more risk averse would end up investing more in relationships with similar others (by declaring a higher *wtp*), in which they expect ex-ante less strategic uncertainty. Interestingly, an interpretation based on risk aversion as the primitive force behind homophily is consistent with the recent sociological theory that explains group identification as a reflection of agents' desire to decrease their perceived uncertainty over the outcome of social interaction (Hogg and Abrams (1993), Hogg (2000)). Perceptions and beliefs also play a role in recent theories in Economics on the emergence of homophily (see e.g. Kets and Sandroni (2014)).[19]

But why should agents believe to better predict behavior in in-group matches? One possible

---

[18]The full questionnaire and details about our respondents and Amazon Mechanical Turk can be found in Appendix C.3.

[19]Compare with the search based theory of Currarini and Vega-Redondo (2011) and the explanation based on social preferences by Chen and Chen (2011).

explanation is that agents expect in-group interaction to be regulated by behaviors that stabilize outcomes on some implicit norm. Such behaviors may take the form of harsh punishments of bad behavior from in-group opponents, possibly grounded in the psychological distress of having invested resources (wtp) in vain, or in a reaction to unexpected ill treatment from other agents from the same group. It is not uncommon in reality to see in-group members being punished more harshly than out-group members for a given behaviour. Through self selection, risk averse agents tend to cluster in in-group matches, forming disjoint close-knit communities within cultural groups. Such communities tend to be characterized by strong social norms, which are enforced by means of selective punishments and rewards. Phenomena such as the punishing of *acting white* behavior which refers to a person's perceived betrayal of identity can be viewed as instances of such effects.[20],[21]

5.2 The Effect of the Introduction of Groups

Before we conclude we have a brief look at our CONTROL condition where agents played the same 8 games as in EXO and ENDO, but where no mention was made of different groups. Understanding how behaviour in EXO differs from CONTROL can help understand how the introduction of artificial group differences or identities affects behavior.

	EXO in-group	EXO out-group	CONTROL	p-value EXO vs CONTROL
G1-G2	0.30	0.13	0.22	$p = 0.6511$
G3-G4	0.18	0.58	0.26	$p = 0.3321$
G5-G6	0.43	0.08	0.30	$p = 0.8316$
G7-G8	0.79	0.66	0.74	$p = 0.9817$

Table 10: Comparison of Behavior EXO and CONTROL (Means). p-values from two-sided ranksum tests.

Table 10 shows the result of this comparison. In all four categories of games behavior in CONTROL lies in between average behaviour in in- and out-groups in EXO. This shows that - at least in our setting - the introduction of different identities does not lead to substantially different behaviour on average. An exception is the case of negative reciprocity (games 3-4), where behaviour in the CONTROL condition is much closer to in-group behavior in EXO and significantly different from out-group behavior in EXO ($p < 0.05$). Hence the introduction of artificial identities leads to a loss in welfare (i) due to an increase in negative reciprocity as well

---

[20]It is reasonable to assume, of course, that such heuristics are learned in real groups outside the laboratory and triggered in the unfamiliar situation encountered in the laboratory.

[21]Figure 3 shows that, indeed, there might be more negative reciprocity in in-group matches compared to out-group matches in ENDO, even though the effect is not significant. Consistently with more negative reciprocity, player A opts out, i.e. does *not* let player B choose, more often in in-group matches (56 percent of the time) compared to out-group matches (46 percent) in ENDO (see Appendix B.2). Again, this difference is not significant, however.

as (ii) due to discrimination (different behaviour in in- and out-group matches) in our case. This is true only, as our previous results have shown, under exogenous matching conditions.

6   Conclusions

We proposed a minimal design to study homophily in the laboratory and found evidence for homophily in such a design even in the absence of stereotypes, differing opportunities or the strategic anticipation of preferential treatment by others. We further studied the connection between homophily and in-group bias. We found that they are tightly linked: giving homophily a playing field (by allowing for endogenous matching) significantly reduces in-group biases. Furthermore in-group biases are only found among agents that are also homophilous. We also found that risk aversion is strongly correlated with homophily in two independent samples.

These results should be of interest to any social scientist studying discrimination, segregation and the like. They also have implications for a number of important policy dimensions, especially for situations where matching is an issue. Those include matching workers into teams, children to schools or social workers or field agents to different neighborhoods. Our results show that allowing some degree of choice in who to work with, can reduce in-group biases and hence discrimination at the work place. Our results have also shown, though, that whether or not discrimination will be reduced depends on the incentives (whether or not there is scope for positive, negative reciprocity etc.). In addition, of course, other factor that were blended out in this experiment, need to be considered as well, such as information asymmetries between different groups or complementarities in skills etc. The interaction of such effects with the parameters considered in this experiment should be understood in future research.

Future research, both in the lab and in the field, is also needed to understand the sources of homophily and the reasons why in-group biases decrease with homophily at work through endogenous matching. One direction for this research could lie in the theory brought forward in this article which is based on two assumptions: In-group interactions among homophilous agents obey behavioral norms and deviations from these norms are harshly punished (hence the reduction in in-group bias). Secondly, these norms (seem to) make behavior in in-group matches more predictable, hence providing a motive for risk averse agents to be homophilous.

References

Ahn, T., R. Isaac, and T. C. Salmon (2009). Coming and going: Experiments on endogenous group sizes for excludable public goods. *Journal of Public Economics 93(1-2)*, 336–351.

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

Bohnet, I., B. Herrmann, and R. Zeckhauser (2010). Trust and the reference point for trustworthiness in gulf and western countries. *Quarterly Journal of Economics 125(2)*, 811–828.

Buhrmester, M., T. Kwang, and S. Gosling (2011). Amazon's mechanical turk: A new source of inexpensive, yet high-quality data? *Perspectives on Psychological Science 6*, 3–5.

Cason, T. and C. R. Plott (2014). Misconceptions and game form recognition: Challenges to theories of revealed preference and framing. *Journal of Political Economy 122*, 1235–1270.

Centola, D. (2011). An experimental study of homophily in the adoption of health behavior. *Science 334*(6060), 1269–1272.

Chakravarty, S. and M. A. Fonseca (2014). The effect of social fragmentation on public good provision: an experimental study. *Journal of Behavioral and Experimental Economics 53*, 1–9.

Charness, G., R. Cobo-Reyes, and N. Jimenez (2014). Identities, selection, and ccontribution in a public good game. *Games and Economic Behavior 87*, 322–338.

Charness, G. and M. Rabin (2002). Understanding social preferences with simple tests. *American Economic Review 117(3)*, 817–869.

Charness, G., L. Rigotti, and A. Rustichini (2007). Individual behavior and group membership. *American Economic Review 97*, 1340–1352.

Chen, R. and Y. Chen (2011). The potential of social identity for equilibrium selection. *American Economic Review 101*, 2562–2589.

Chen, Y. and S. X. Li (2009). Group identity and social preferences. *American Economic Review 99:1*, 431–457.

Coricelli, G., D. Fehr, and G. Fellner (2004). Partner selection in public goods experiments. *Journal of Conflict Resolution 48(3)*, 356–378.

Currarini, S., M. Jackson, and P. Pin (2009). An economic model of friendship: Homophily, minorities and segregation. *Econometrica 77(4)*, 1003–1045.

Currarini, S., M. Jackson, and P. Pin (2010). Identifying the roles of choice and chance in network formation: Racial biases in high school friendships. *PNAS 107*, 4857–4861.

Currarini, S. and F. Vega-Redondo (2011). A simple model of homophily in social networks. mimeo.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. Wagner (2011). Individual risk attitudes: Measurement, determinants and bebehavior consequences. *Journal of the European Economic Association 9(3)*, 522–550.

Eckel, C. and P. Grossman (2005). Managing diversity by creating team identity. *Journal of Economic Behavior and Organization 3*, 371–392.

Fehr, D., R. Hakimov, and D. Kuebler (2015). The willingness to pay - willingness to accept gap: A failed replication of plott and zeiler. mimeo.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics 10(2)*, 171–178.

Foddy, M., M. Platow, and T. Yamagishi (2009). Group-based trust in strangers. *Psychological Science 20(4)*, 419–422.

Gaechter, S. and B. Herrmann (2009). Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment. *Philosophical Transactions of the Royal Society B - Biological Sciences 364*, 791–806.

Greiner, B. (2004). An online recruitment system for economic experiments. In K. Kremer and V. Macho (Eds.), *Forshcung und wissenschaftliches Rechnen*. GDWG Bericht 63.

Grimm, V. and F. Mengel (2009). Cooperation in viscous populations - experimental evidence. *Games and Economic Behavior 66(1)*, 202–220.

Grimm, V. and F. Mengel (2011). Let me sleep on it: Delay reduces rejection rates in ultimatum games. *Economics Letters 111(2)*, 113–115.

Grossman, Z., J. van der Weele, and A. Andrijevik (2014). A test of dual-process reasoning in charitable giving. SSRN working paper.

Guala, F. (2000). Artegoods in experimental economics: preference reversals and the becker-degroot-marschak mechanism. *Economics and Philosophy 0(1)*, 47–75.

Hogg, M. (2000). Subjective uncertainty reduction through self-categorization: a motivational theory of social identity processes and group phenomena. *European Review of Social Psychology 11*, 22355.

Horowitz, J. (2006). The becker-degroot-marschak mechanism is not necessarily incentive compatible, even for non-random goods. *Economics Letters 93(1)*, 6–11.

Horton, J., D. Rand, and R. Zeckhauser (2011). The online laboratory: conducting experiments in a real labor market. *Experimental Economics 14*, 399–425.

Ioannou, C., A. Rustichini, and S. Qi (2015). Group payoffs as public signals. *Journal of Economic Psychology in press.*

Karni, E. and Z. Safra (1987). Preference reversal and the observability of preferences by experimental methods. *Econometrica 55(3)*, 675–685.

Keser, C. and F. van Winden (2000). Conditional cooperation and voluntary contribution to public goods. *Scandinavian Journal of Economics 102(1)*, 23–29.

Kets, W. and A. Sandroni (2014). A belief based theory of homophily. mimeo.

Kuklinski, J. H., P. M. Sniderman, K. Knight, T. Piazza, P. Tetlock, G. Lawrence, and B. Mellers. (1997). The politics of affirmative action. *American Journal of Political Science 41*, 402–419.

McPherson, M., L. Smith-Lovin, and J. Cook (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology 27*, 415–444.

Moore, D. and G. Loewenstein (2004). Self-interest, automaticity, and the psychology of conflict of interest. *Social Justice Research 17(2)*.

Oechssler, J., A. Roider, and P. Schmitz (2015). Cooling off in negotiations - does it work? *Journal of Institutional and Theoretical Economics*.

Piovesan, M. and E. Wengstroem (2009). Fast or fair? a study of response times. *Economics Letters 105 (2)*, 193–196.

Plott, C. and K. Zeiler (2005). The willingness to paywillingness to accept gap, the endowment effect, subject misconceptions, and experimental procedures for eliciting valuations. *The American Economic Review 95(3)*, 530–546.

Rand, D. (2012). The promise of mechanical turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology 299*, 172–179.

Rand, D., J. Greene, and M. A. Nowak (2012). Spontaneous giving and calculated greed. *Nature 489*, 427–430.

Recalde, M., A. Riedl, and L. Vesterlund (2014). Error prone inference from response time: the case of intuitive generosity. *CESIFO working paper 4987*.

Suri, S. and D. Watts (2011). Cooperation and contagion in web-based, networked public goods experiments. *PLoS ONE 6*, e16836.

Tajfel, H. and J. Turner (1986). *The social identity theory of inter-group behavior.* Chicago: Nelson Hall.

Tajfel, H. and J. C. Turner (1979). An integrative theory of intergroup conflict. In *The social psychology of integroup relations.* ed. W.G. Austin and S. Worchel Monterey, CA: Brooks/Cole.

Tinghoeg, G., D. Andersson, C. Bonn, H. Boettiger, C. Josephson, G. Lundgren, D. Vast-fjall, M. Kirchler, and M. Johannesson (2013). Intuition and cooperation revisited. *Nature 498*(7452), E1–E2.

Van Bavel, J., D. Packer, and W. Cunningham (2008). The neural substrates of in-group bias. *Psychological Science 19*, (11).

SUPPLEMENTARY MATERIALS: FOR ONLINE PUBLICATION ONLY

Contents

A   Appendix: Sample Characteristics and Randomization Checks

In this section we describe some of the characteristics of our sample and conduct report randomization checks based on five observables: age, gender, nationality, risk attitude and a measure of cognitive reflection. The variable gender takes the value 1 for female and zero for male. The variable nationality was elicited in 9 categories (1=german/dutch; 2= western europe (non-german, non-dutch); 3= eastern europe; 4=north america; 5= south america; 6= east asia; 7=south asia; 8=africa; 0=others). The variables risk aversion and cognitive reflection are described in detail in Appendix C.
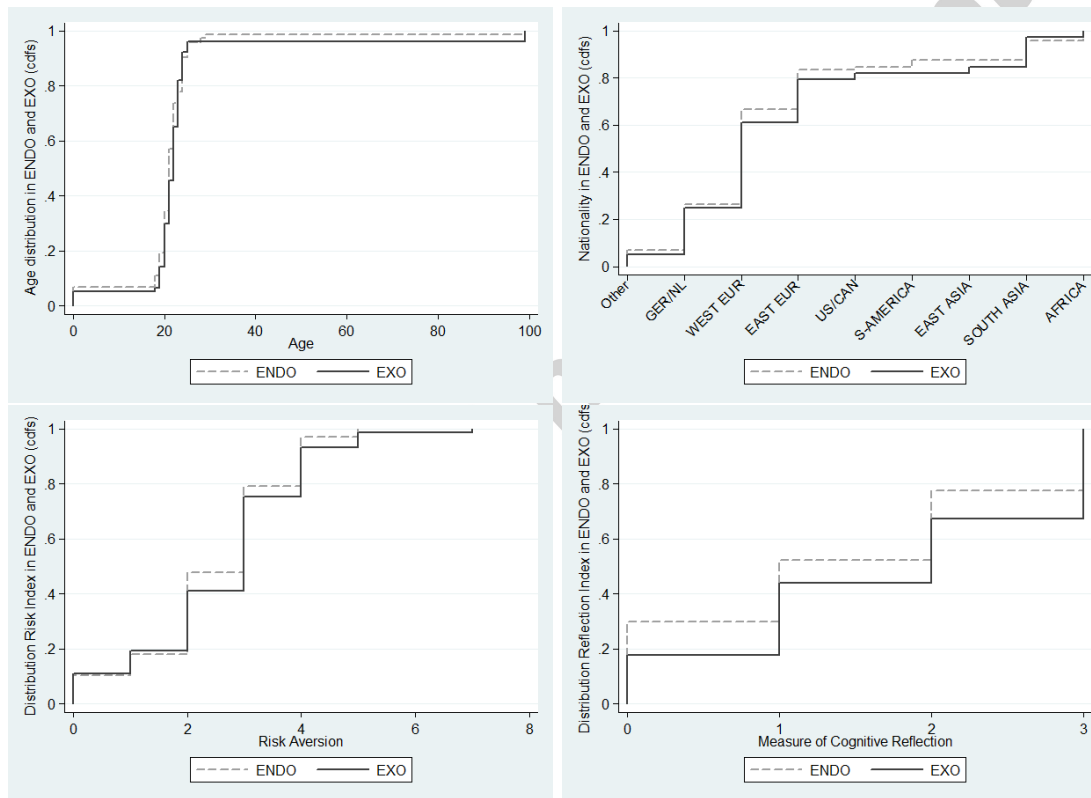
Figure 7: CDF's of age, nationality, risk aversion index and cognitive reflection index for the two samples in ENDO and EXO.

Figure 7 shows the cdf's of four of these (gender is binary) for both treatments ENDO and EXO. It can be seen that the distribution of age, nationality and our risk aversion index are roughly similar across the two treatments EXO and ENDO. There seem to be somewhat more very little reflected participants (score 0) in ENDO as opposed to EXO.

Table 11 shows the mean and standard deviation of these variables in ENDO and EXO as well as randomization checks using OLS and two-sided ranksum tests. The table shows that the differences in means tend to be small and the distributions not significantly different according

31

	ENDO		EXO				two-sided
	mean	SD	mean	SD	$\Delta_{ENDO-EXO}$	OLS	ranksum test
age	21.71	9.69	21.66	15.53	0.05	$p = 0.351$	$p = 0.3868$
gender	0.671	0.50	0.575	0.52	0.094	$p = 0.271$	$p = 0.2597$
nationality	2.805	1.99	3	2.10	- 0.195	$p = 0.577$	$p = 0.6269$
risk aversion	2.477	1.25	2.630	1.42	- 0.153	$p = 0.506$	$p = 0.5131$
cognitive reflection	1.402	1.142	1.712	1.111	- 0.310	$p = 0.107$	$p = 0.1065$

Table 11: Observable Sample Characteristics and Randomization Checks.

to either test. The cognitive reflection measure stands out a bit, with participants in ENDO being somewhat less cognitively reflected. The difference in distributions is not statistically significant, but the p-values quite close to the 10 percent level. We have seen in Sections 4 and 5, though, that cognitive reflection does not predict wtp nor behaviour in games 3-8. Hence, even if there are differences between the two samples in terms of cognitive reflection, they do not seem to explain our experimental results (with respect to wtp (homophily) and behaviour in games 3-8 (in-group bias)).

B Appendix: Additional Tables and Results

B.1 Additional Tables and Results Sections 3-5

This section collects additional results and tables. Table 12 shows descriptive statistics (mean and SD) for our main experimental variables.

| | CONTROL | EXO | | ENDO | |
		IN	OUT	IN	OUT
Altruism	0.22 (0.38)	0.29 (0.33)	0.13 (0.28)	0.26 (0.38)	0.27 (0.34)
Neg Recp	0.26 (0.35)	0.18 (0.24)	0.58 (0.46)	0.38 (0.41)	0.20 (0.34)
Pos Recp (I)	0.30 (0.25)	0.43 (0.35)	0.08 (0.25)	0.46 (0.43)	0.21 (0.31)
Pos Recp (d)	0.74 (0.35)	0.79 (0.33)	0.66 (0.38)	0.69 (0.38)	0.68 (0.34)
wtp (IN)	-	-	-	13.86 (23.56)	
wtp (OUT)	-	-	-	4.23 (11.53)	

Table 12: Descriptive Statistics (Mean (SD)) of experimental variables across treatments EXO, ENDO, in- and out-group matches and the control condition.

Figure 8 shows the distribution of the variable *wtp*. It peaks at zero, reflecting the relatively high percentage of neutral agents. There are few heterophilous agents, whose *wtp* ranges from 0 to 50 and many more homophilous agents whose *wtp* ranges from 0 to 100. We next ask how much of homophily observed can be rationalized by the anticipation of in-group biases.
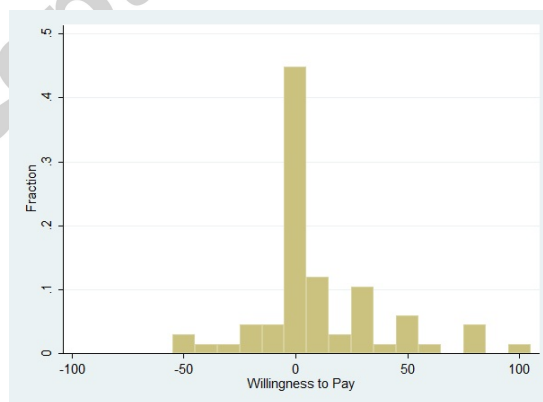

Figure 8: The distribution of willingness to pay in treatment ENDO. Negative Numbers indicate *wtp* for an out-group match and positive numbers *wtp* for an in-group match. Bin width is 10.

Ex post empirical optimality of bids How much should one be willing to pay for an in-group match if one could anticipate the choices of the population perfectly? We start with player

type A and we assume standard preferences. In treatment ENDO player type A does not face significantly different treatment in G1-G2 by player B, in G3-G4 anticipating player B's choices would make player A opt for "OUT" in in-group matches resulting in a payoff of 250 and for "In" in out-group matches resulting in an expected payoff of 272, an expected loss of 22. In games G5-G6 player A would stay out in both in- and out-group matches, implying no payoff difference in these games. For player type B again there would be no differences in G1-G4, since in none of these games player A behaviour differs across in- and out-group matches (see Section B.2 right below). In G5-G6 the expected gain is 104 and in G7-G8 the expected gain is 101. Given that participants do not know yet when they indicate their *wtp* whether they will be type A or type B (and both are equally likely), the expected gain from playing in an in-group match is $\frac{1}{2} * (-22) + \frac{1}{2} * 205 \approx 92$. Given our mechanism to elicit *wtp* players should maximize

$$\max \frac{wtp}{100} * (92 - wtp) + \left(1 - \frac{wtp}{100}\right) * \frac{92}{2} \tag{1}$$

which implies a choice of $wtp = 23$.

Table 13 reproduces the regression in Table 4 for treatment ENDO. It should be noted, however, that in ENDO this regression has endogenity issues due to precisely the endogenous selection into in-group matches. Table 14 reproduces Table 5 from the main text for treatment ENDO with the same caveat.

	G1-G2	G1-G2	G3-G4	G3-G4
constant	−0.9555***	−1.3332***	−1.3633***	−1.3633**
in-group	−0.2615	−0.0430	0.9167	0.8933
wtp x in-group	0.0094		−0.0011	
groups	40	40	40	40
observations	80	80	80	80
	G5-G6	G5-G6	G7-G8	G7-G8
constant	−1.3633***	−1.3633***	0.7777**	0.7777**
in-group	0.9375	1.2091*	0.6442	0.0332
wtp x in-group	0.0134**		−0.0239	
groups	40	40	40	40
observations	80	80	80	80

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 13: Logit Regression with standard errors clustered by id. In-group Bias in ENDO. 40 individuals (groups) in the role of Player B, 80 observations. 62 percent of all matches were in-group matches. *** 1%,** 5%,* 10%.

Table 15 compares how much more positively or negatively reciprocal participants of different types (homophilous, neutral or heterophilous) are in in-group matches compared to out-group

ENDO	G1-G2 ("Altruism")	G3-G4 ("Neg. Recp")	G5-G6 ("Pos.Recp.(I)")	G7-G8 ("Pos.Recp.(d)")
constant	-2.364^{**}	-1.174^{**}	-0.538	-0.151
	(0.746)	(0.383)	(3.149)	(2.685)
in-group	-0.190	1.429^{*}	0.663	0.307
	(0.746)	(0.838)	(0.700)	(0.734)
age	0.140	0.044	-0.015	0.034
	(0.151)	(0.148)	(0.137)	(0.114)
gender	-0.913	0.483	-0.123	0.123
	(0.779)	(0.629)	(0.641)	(0.636)
RED	0.285	-0.794	0.927	-0.505
	(0.748)	(0.929)	(0.778)	(0.671)
risk aversion	-0.313	-0.206	-0.149	0.171
	(0.289)	(0.308)	(0.318)	(0.311)
cognitive reflection	-0.305	-0.106	-0.242	-0.087
	(0.245)	(0.273)	(0.324)	(0.239)
groups	40	40	40	40
observations	80	80	80	80

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 14: In-group Bias in ENDO. Logit Regression with additional variables (Standard errors clustered by id). 40 individuals (groups) in the role of Player B, 80 observations. 62 percent of all matches were in-group matches. ***1%,**5%,*10% significance.

matches.

	Homophilous	WTP=0	Heterophilous
Neg. Reciprocity (Games 3-4)	0.3158	0.0558	0.3043
Pos. Reciprocity (Games 5-6)	0.5190^{**}	0.1484	-0.6667
Number of Observations (Share)	18(45%)	17(43%)	5(12%)

Table 15: Spearman correlation coefficients. Correlation between the frequency of altruistic, negatively reciprocal etc.. behavior and in-group match dummy. 40 individuals in the role of Player B, 80 observations. *** 1%,** 5%,* 10%

B.2 Appendix: Behavior Player Role A

This section is devoted to understanding the behaviour of participants in the role of player A. Player A's decision to let Player B decide or not can be motivated by (i) concerns on the distribution of payoffs (such as efficiency, equality etc.) and (ii) by anticipation of Player B's behaviour. If a participant in the role of Player A lets Player B choose in Game 3, then we can reasonably interpret this to mean that s/he does not expect B to be negatively reciprocal. If s/he lets Player B choose in Games 7-8, then we can interpret this to mean that s/he does expect Player B to be positively reciprocal. In Games 4,5 or 6 a case could be made that Player A may let B choose because this will be inequality reducing. This makes player A's behaviour more difficult to interpret in terms of in-group bias, which is why we have focused on Player B's decision for the main part of the paper.

Games 3 and 4 - Negative Reciprocity There are no significant differences between in- and out-group matches in terms of the behavior of Player A in ENDO, where 56 percent and 46 percent of them, respectively, decide *not* to let player B make a choice (the difference is marginally significant for game 3 alone). There is also no significant difference between in- and out-group matches in EXO, where 38 percent and 53 percent respectively decide *not* to let player B make a choice. The difference between ENDO and EXO is marginally significant for in-group matches and insignificant for out-group matches.

Games 5 and 6 - Positive Reciprocity In ENDO there are strong and significant differences between in-group and out-group matches. In out-group matches only about 10 percent of agents let Player B choose, while in in-group matches 62 percent let Player B choose ($p = 0.0010$). In treatment EXO there are no significant differences between in- and out-group matches with 33 percent and 25 percent respectively letting Player B choose.

Games 7 and 8 - Positive Reciprocity In ENDO only 7 percent of agents let Player B choose in out-group matches, while 38 percent let Player B choose in in-group matches. The difference is significant ($p = 0.0273$). In treatment EXO 3 percent let Player B choose in out-group matches, while 14 percent let Player B choose in in-group matches. This difference is not significant.

B.3   Appendix: Effect of Group Labels

This subsection presents some results regarding the effect of group labels. We first ask whether people have a higher willingness to pay to play with a RED rather than the BLUE group. The average willingness to pay to play with the RED group is 20.41 in ENDO and the average w.t.p to play with BLUE is 19.77. The two are not significantly different according to a Mann-Whitney test. We then ask whether behavior is affected by being in a RED or BLUE group per se.

	red	blue
Altruism	0.23	0.15
Neg. Rec. (Games 3-4)	0.32	0.32
Pos. Rec. (Games 5-6)	0.35	0.16
Pos. Rec. (Games 7-8)	0.67	0.76

Table 16: Colours don't matter per se. ENDO

	red	blue
Altruism	0.23	0.12
Neg. Rec. (Games 3-4)	0.31	0.43
Pos. Rec. (Games 5-6)	0.29	0.28
Pos. Rec. (Games 7-8)	0.71	0.73

Table 17: Colours don't matter per se. EXO

There are no significant differences between blue and red players in either of the treatments and for any of the games with the exception of Altruism in treatment EXO. The colour of the group does not seem to matter per se.

B.4   Appendix: Sample Games LOWB and COORD

Tables 18 and 19 show sample games from treatments LOWB and COORD.

	X	Y
X	0,0	200,200
Y	200,200	0,0

Table 18: Game 1 in treatment LOWB

In treatment LOWB, players played a series of pure (anti-) coordination games where players have to choose different actions to reach the good payoffs. There is no possibility for a player to favor an in-group opponent here by choosing either X or Y.

In treatment COORD players played the same games as in LOWB. However, actions here are labeled by group labels RED or BLUE, in such a way that - if players choose the action that corresponds to their group label - players of the same group will manage to coordinate while players of different groups won't.

	BLUE	RED
RED	0,0	200,200
BLUE	200,200	0,0

Table 19: Game 1 in treatment COORD

B.5   Appendix: Additional Treatment LABEL

This section collects some additional results from our treatment LABEL. This treatment coincides with ENDO, but we labeled the actions in the 8 games by either RED or BLUE to see whether such labels can induce different norms of behaviour in the two groups and whether such norms could possibly crowd out existing norms of behaviour. Our evidence shows that this was not the case.

	red	blue
prefer in-group	0.65	0.65
wtp in-group	14.4	12.41
wtp out-group	0.75	6.58

Table 20: Percentage of red and blue participants preferring an in-group match, as well as average willingness to pay for a in-group or out-group match in Period 1.

None of the differences between red and blue participants are significant (Mann-Whitney, $p > 0.7267$). However there is a higher willingness to pay for a in-group than for a out-group match across all participants (Mann-Whitney, all periods, $p = 0.0092$).

	red	blue
Altruism	0.20	0.30
Neg. Recip.	0.25	0.30
Pos. Recip.	0.30	0.37
Ant. PosR	0.15	0.18
Ant. PosR	0.15	0.18

Table 21: Are labels followed? Altruistic and Reciprocal Behavior by Red and Blue participants as well as Anticipation of Positive and Negative Reciprocity (Player A) if the match is blue/red.

The table shows that labels are not followed. Red participants are not more altruistic or reciprocal than blue participants. (None of the differences are significant).

C Appendix: Questionnaire

We first report the questions asked in the questionnaire and then describe the variables derived from it.

C.1 Appendix: Questions

- General Questions with multiple options to tick.

  - What is your nationality?
  - What is your gender?
  - What is your age?
  - What is your field of study?

- Cognitive Reflection Test

  - A bat and a ball cost 1.10 Euros in total. The bat costs 1 Euro more than the ball. How much does the ball cost?
  - If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?
  - In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

- Risk Aversion

  - If you had the choice between throwing a coin and receiving 100 Euros if heads comes up and 0 Euros if tails come up or a deal where you get 10 Euros for sure which would you prefer?
  - If you had the choice between throwing a coin and receiving 100 Euros if heads comes up and 0 Euros if tails come up or a deal where you get 20 Euros for sure which would you prefer?
  - If you had the choice between throwing a coin and receiving 100 Euros if heads comes up and 0 Euros if tails come up or a deal where you get 30 Euros for sure which would you prefer?
  - If you had the choice between throwing a coin and receiving 100 Euros if heads comes up and 0 Euros if tails come up or a deal where you get 40 Euros for sure which would you prefer?

39

- – If you had the choice between throwing a coin and receiving 100 Euros if heads comes up and 0 Euros if tails come up or a deal where you get 50 Euros for sure which would you prefer?

- – If you had the choice between throwing a coin and receiving 100 Euros if heads comes up and 0 Euros if tails come up or a deal where you get 60 Euros for sure which would you prefer?

- – If you had the choice between throwing a coin and receiving 100 Euros if heads comes up and 0 Euros if tails come up or a deal where you get 70 Euros for sure which would you prefer?

C.2 Appendix: Overview Variables

Table 22 shows some summary statistics of the questionnaire variables used in the regressions in the main text.

	interpretation	range EXO	mean EXO	range ENDO	mean ENDO
risk aversion	7 most risk averse, 0 least	[0,7]	2.63	[0,5]	2.48
gender	1= female , 0=male	[0,1]	0.55	[0,1]	0.65
age		[18,25]	21.71	[18,29]	21.66
cognitive reflection	number of correct answers	[0,3]	1.71	[0,3]	1.40

Table 22: Questionnaire Variables used in Regressions in the main text.

C.3 Appendix: AMT questionnaire

AMT is an online labor market in which employers can employ workers to complete short tasks (generally less than 10 minutes) for relatively small amounts of money (generally less than 1 US dollar). One major advantage of AMT is it allows experimenters to easily expand beyond the college student convenience samples typical of most economic game experiments. Among American subjects, AMT subjects have been shown to be significantly more nationally representative than college student samples (Buhrmester et al. (2011)). Running experiments online necessarily involves some loss of control, since the workers cannot be directly monitored as in the traditional lab; hence, experimenters cannot be certain that each observation is the result of a single person (as opposed to multiple people making joint decisions at the same computer). Moreover, although the sample of subjects in AMT experiments is more diverse than samples using college undergraduates, we are obviously restricted to people that participate in online labor markets. To address these potential concerns, recent studies have explored the validity of data gathered using AMT (for an overview, see Rand (2012)). Suri and Watts (2011), e.g., find quantitative agreement in contribution behavior in a repeated public goods game between experiments conducted in the physical lab and those conducted using AMT with approximately

10-fold lower stakes (see also Horton et al. (2011)). It has also been shown that AMT subjects display a level of test-retest reliability similar to what is seen in the traditional lab on measures of political beliefs, self-esteem, Social Dominance Orientation, and Big-Five personality traits (Buhrmester et al. (2011)) as well as belief in God, age, gender, education level and income (Rand (2012)). Hence we are confident that our measures obtained on AMT are at least somewhat representative. The following questions were asked in our AMT questionnaire.

- What is your age?

- What is your gender?

- What is your highest level of education?

- What is your income?

- How willing are you to take risk in general?

- How strongly do you prefer to "mix with your own kind?

- How strongly do you prefer to interact with people similar to you?

In addition we also asked them the same lottery choices as we did in the questionnaire of our experiment. The variables used in Section 5 are summarized in the Table below.

	interpretation	range	mean
risk aversion	7 most risk averse, 0 least	[0,6]	3.38
homophily indicator	7 most homophilous, 0 least	[0,7]	4.33
gender	1= female , 0=male	[0,1]	0.57
age		[20,67]	35.64

Table 23: Questionnaire Variables AMT Questionnaire used in Section 5.

D   Appendix: Sample Instructions

D.1   ENDO

Welcome and thanks for participating in this experiment. Please read these instructions carefully. They are identical for all the participants with whom you will interact during this experiment. If you have any questions please raise your hand. One of the experimenters will come to you and answer your questions. From now on communication with other participants is not allowed. If you do not conform to these rules we are sorry to have to exclude you from the experiment. Please do also switch off your mobile phone at this moment.

For your participation you will receive 2 Euros. During the experiment you can earn more. How much depends on your behavior and the behavior of the other participants. During the experiment we will use ECU (Experimental Currency Units) and at the end we will pay you in Euros according to the exchange rate 1 Euro =500 ECU. All your decisions will be treated confidentially.

These instructions are exactly the same for all participants in the room.

THE EXPERIMENT

Phase 1: The RED and BLUE groups

At the beginning of the experiment each participant receives an endowment of 500 ECU.

All participants in this room are randomly assigned to one of two groups, the BLUE group and the RED group. Once the software starts you will be informed about which group you belong to.

Phase 2: Your choice of who to interact with

In Phase 3 of the Experiment that we describe below you will interact in a number of games. Before that - in Phase 2 - you can state a preference for whom to interact with. In particular you can first state a preference for whether you would like to interact with someone from the RED or BLUE group.

Afterwards you will see a new screen. If you have previously stated that you would like to interact with someone from a particular group (either BLUE or RED), then you will have the possibility to increase the chances that you will actually be matched with someone from that group on the third screen.

You can increase your chances to be matched with someone from your desired group as follows:

First you will indicate a number between 0 and 100 that indicates how much you would be willing to pay to be matched with your preferred group that you indicated on Screen 2. We will draw a random number between 0 and 100 (where all numbers have the same probability). If our

number is smaller than your number, you will be matched for sure with your preferred group. In this case an amount corresponding to your number will be deducted from your endowment. If our number is larger than your number you will be matched randomly with a member of either group. Example: Here is an example of how the procedure works: Assume you indicate the number 21 and we draw randomly the number 46. Then since $46 > 21$ you will be matched randomly with someone from either the RED or the BLUE group. However if e.g. you indicate 83 and we draw randomly the number 16, then since $83 > 16$ you will be matched with the group you prefer. However, in this case your endowment will be reduced by 83 (the number that you stated).

Hence the higher the number you state, the more likely it is that you will be matched with your desired group. On the other hand your endowment is also reduced by more if you state a higher number. There is a small probability that it is not feasible to match everyone with their desired group (e.g. if everyone states 100). If this is the case you will be randomly matched and your endowment will not be reduced.

Phase 3: The Games

In Phase 3 you will play the following 8 games together with an interaction partner. Whether your interaction partner will be from the BLUE or RED group will depend on your choices in Phase 2. Before Phase 3 begins you will be informed about which group your interaction partner belongs to. Similarly your interaction partner will be informed about which group you belong to.

Here is a list of all 8 games you will play. You will play each game exactly once. Before the games begin either you or your interaction partner will be randomly determined to be Player A. If you are not Player A then you will be Player B in all 8 games. Whether you or your interaction partner will be Player A is fully independent of your or your interaction partner's group.

1. Game 1: Player B chooses between two outcomes. (400,400) meaning that each player gets 400 or (750,375) meaning that Player B gets 375 and Player A 750.

2. Game 2: Player B chooses between two outcomes. (100,300) meaning that player A gets 100 and Player B 300 or (400,200) meaning that Player A gets 400 and Player B 200.

3. Game 3: Player A chooses between two options. (250,250) meaning that each player gets 250 or s/he lets player B choose. If s/he lets B choose then B can choose between (100,100) meaning that each player gets 100 and (500,100) meaning that player A gets 500 and player B gets 100.

4. Game 4: Player A chooses between two options. (50,650) meaning that player A gets 50 and player B 650 or s/he lets player B choose. If s/he lets B choose then B can choose

between (0,100) meaning that A gets nothing and B gets 100 and (100,100) meaning that each player gets 100.

5. Game 5: Player A chooses between two options. (500,0) meaning that player A gets 500 and player B gets 0 or s/he lets player B choose. If s/he lets B choose then B can choose between (300,300) meaning that each player gets 300 and (600,275) meaning that player A gets 600 and player B gets 275.

6. Game 6: Player A chooses between two options. (250,0) meaning that player A gets 250 and player B gets 0 or s/he lets player B choose. If s/he lets B choose then B can choose between (100,100) meaning that each player gets 100 and (250,50) meaning that player A gets 250 and player B gets 50.

7. Game 7: Player A chooses between two options. (350,100) meaning that player A gets 350 and player B gets 100 or s/he lets player B choose. If s/he lets B choose then B can choose between (300,300) meaning that each player gets 300 and (100,350) meaning that player A gets 100 and player B gets 350.

8. Game 8: Player A chooses between two options. (400,0) meaning that player A gets 400 and player B gets 0 or s/he lets player B choose. If s/he lets B choose then B can choose between (200,200) meaning that each player gets 200 and (0,400) meaning that player A gets 0 and player B gets 400.

It is important to note that you won't get any feedback about what your interaction partner chose until the last game has been played.

Phase 4: A post-experimental Questionnaire

At the end of the experiment there will be a short questionnaire to be filled in which will take about 5 minutes.

Summary

1. First you will be informed about which group (RED or BLUE) you belong to.

2. You can state a preference for being matched with someone of the BLUE or RED group and you can increase your chances of being matched with someone of that group via the procedure explained above.

3. You will play the 8 games described above with your interaction partner. Whether your interaction partner is from the RED or BLUE group will depend on your choices in Step 2.

44

4. There will be a short questionnaire at the end.

Enjoy the Experiment!

D.2 EXO

Welcome and thanks for participating in this experiment. Please read these instructions carefully. They are identical for all the participants with whom you will interact during this experiment. If you have any questions please raise your hand. One of the experimenters will come to you and answer your questions. From now on communication with other participants is not allowed. If you do not conform to these rules we are sorry to have to exclude you from the experiment. Please do also switch off your mobile phone at this moment.

For your participation you will receive 2 Euros. During the experiment you can earn more. How much depends on your behavior and the behavior of the other participants. During the experiment we will use ECU (Experimental Currency Units) and at the end we will pay you in Euros according to the exchange rate 1 Euro =500 ECU. All your decisions will be treated confidentially.

These instructions are exactly the same for all participants in the room.

THE EXPERIMENT

Phase 1: The RED and BLUE groups

All participants in this room are randomly assigned to one of two groups, the BLUE group and the RED group. Once the software starts you will be informed about which group you belong to.

Phase 2: Your match

In Phase 2 of the Experiment you will be randomly matched with an interaction partner from either the RED or BLUE group to play the games we will describe below.

Phase 3: The Games

Before the start of Phase 3 you will be informed about whether your interaction partner is from the BLUE or RED group. Similarly your interaction partner will be informed about which group you belong to.

Here is a list of all 8 games you will play. You will play each game exactly once. Before the games begin either you or your interaction partner will be randomly determined to be Player A. If you are not Player A then you will be Player B in all 8 games. Whether you or your interaction partner will be Player A is fully independent of your or your interaction partner's group.

45

1. Game 1: Player B chooses between two outcomes. (400,400) meaning that each player gets 400 or (750,375) meaning that Player B gets 375 and Player A 750.

2. Game 2: Player B chooses between two outcomes. (100,300) meaning that player A gets 100 and Player B 300 or (400,200) meaning that Player A gets 400 and Player B 200.

3. Game 3: Player A chooses between two options. (250,250) meaning that each player gets 250 or s/he lets player B choose. If s/he lets B choose then B can choose between (100,100) meaning that each player gets 100 and (500,100) meaning that player A gets 500 and player B gets 100.

4. Game 4: Player A chooses between two options. (50,650) meaning that player A gets 50 and player B 650 or s/he lets player B choose. If s/he lets B choose then B can choose between (0,100) meaning that A gets nothing and B gets 100 and (100,100) meaning that each player gets 100.

5. Game 5: Player A chooses between two options. (500,0) meaning that player A gets 500 and player B gets 0 or s/he lets player B choose. If s/he lets B choose then B can choose between (300,300) meaning that each player gets 300 and (600,275) meaning that player A gets 600 and player B gets 275.

6. Game 6: Player A chooses between two options. (250,0) meaning that player A gets 250 and player B gets 0 or s/he lets player B choose. If s/he lets B choose then B can choose between (100,100) meaning that each player gets 100 and (250,50) meaning that player A gets 250 and player B gets 50.

7. Game 7: Player A chooses between two options. (350,100) meaning that player A gets 350 and player B gets 100 or s/he lets player B choose. If s/he lets B choose then B can choose between (300,300) meaning that each player gets 300 and (100,350) meaning that player A gets 100 and player B gets 350.

8. Game 8: Player A chooses between two options. (400,0) meaning that player A gets 400 and player B gets 0 or s/he lets player B choose. If s/he lets B choose then B can choose between (200,200) meaning that each player gets 200 and (0,400) meaning that player A gets 0 and player B gets 400.

It is important to note that you won't get any feedback about what your interaction partner chose until the last game has been played.

Phase 4: A post-experimental Questionnaire

At the end of the experiment there will be a short questionnaire to be filled in which will take about 5 minutes.

Summary

1. First you will be informed about which group (RED or BLUE) you belong to.

2. You will be matched with someone of the BLUE or RED group.

3. You will play the 8 games described above with your interaction partner. Whether your interaction partner is from the RED or BLUE group will be determined in Step 2.

4. There will be a short questionnaire at the end.

Enjoy the Experiment!

D.3   CONTROL

Welcome and thanks for participating in this experiment. Please read these instructions carefully. They are identical for all the participants with whom you will interact during this experiment. If you have any questions please raise your hand. One of the experimenters will come to you and answer your questions. From now on communication with other participants is not allowed. If you do not conform to these rules we are sorry to have to exclude you from the experiment. Please do also switch off your mobile phone at this moment.

For your participation you will receive 2 Euros. During the experiment you can earn more. How much depends on your behavior and the behavior of the other participants. During the experiment we will use ECU (Experimental Currency Units) and at the end we will pay you in Euros according to the exchange rate 1 Euro =500 ECU. All your decisions will be treated confidentially.

These instructions are exactly the same for all participants in the room.

THE EXPERIMENT

In the experiment you will play 8 games with another randomly selected participant. Here is a list of all 8 games you will play. You will play each game exactly once. Before the games begin either you or your interaction partner will be randomly determined to be Player A. If you are not Player A then you will be Player B in all 8 games. Whether you or your interaction partner will be Player A is fully independent of your or your interaction partner's group.

1. Game 1: Player B chooses between two outcomes. (400,400) meaning that each player gets 400 or (750,375) meaning that Player B gets 375 and Player A 750.

2. Game 2: Player B chooses between two outcomes. (100,300) meaning that player A gets 100 and Player B 300 or (400,200) meaning that Player A gets 400 and Player B 200.

47

3. Game 3: Player A chooses between two options. (250,250) meaning that each player gets 250 or s/he lets player B choose. If s/he lets B choose then B can choose between (100,100) meaning that each player gets 100 and (500,100) meaning that player A gets 500 and player B gets 100.

4. Game 4: Player A chooses between two options. (50,650) meaning that player A gets 50 and player B 650 or s/he lets player B choose. If s/he lets B choose then B can choose between (0,100) meaning that A gets nothing and B gets 100 and (100,100) meaning that each player gets 100.

5. Game 5: Player A chooses between two options. (500,0) meaning that player A gets 500 and player B gets 0 or s/he lets player B choose. If s/he lets B choose then B can choose between (300,300) meaning that each player gets 300 and (600,275) meaning that player A gets 600 and player B gets 275.

6. Game 6: Player A chooses between two options. (250,0) meaning that player A gets 250 and player B gets 0 or s/he lets player B choose. If s/he lets B choose then B can choose between (100,100) meaning that each player gets 100 and (250,50) meaning that player A gets 250 and player B gets 50.

7. Game 7: Player A chooses between two options. (350,100) meaning that player A gets 350 and player B gets 100 or s/he lets player B choose. If s/he lets B choose then B can choose between (300,300) meaning that each player gets 300 and (100,350) meaning that player A gets 100 and player B gets 350.

8. Game 8: Player A chooses between two options. (400,0) meaning that player A gets 400 and player B gets 0 or s/he lets player B choose. If s/he lets B choose then B can choose between (200,200) meaning that each player gets 200 and (0,400) meaning that player A gets 0 and player B gets 400.

It is important to note that you won't get any feedback about what your interaction partner chose until the last game has been played.

A post-experimental Questionnaire

At the end of the experiment there will be a short questionnaire to be filled in which will take about 5 minutes.

Enjoy the Experiment!

48