Running Head = HETEROGENEITY, HOMOPHILY, AND COOPERATION

Effects of Heterogeneity and Homophily on Cooperation

Ozan Aksoy

Nuffield College Centre for Experimental Social Sciences and Department of Sociology, University of Oxford

Corresponding Author:
Ozan Aksoy, Nuffield College, New Road OX1 1NF, Oxford, UK

Email: ozan.aksoy@sociology.ox.ac.uk

Abstract

The paper provides a micro-behavioral model and an experimental design to understand the effect of heterogeneity in social identities on cooperation while accounting for endogenous sorting. Social identity is induced exogenously using the minimal group paradigm. The experiment manipulates sorting with three treatments: having subjects interact with both in- and outgroup members, giving them the choice to interact either with ingroup or outgroup members, and isolating the groups from the outset. Cooperation is measured by the Prisoner's Dilemma Games at the dyadic-level and by Public Goods Games at the tetradic-level. The results show that heterogeneity hampers between-group cooperation at the dyadic-level. In addition, endogenous sorting mitigates this negative effect of heterogeneity on cooperation. Heterogeneity hampers cooperation at the tetradic-level most substantially if there is a commonly known negative history between groups.

In the near future, the developed regions of the world will become much more heterogeneous ethnically and culturally, not only due to factors such as wars and conflicts that force emigration from developing countries but also due to the need in the developed world to attract more foreign labor (UK Office for National Statistics 2014). This increase in heterogeneity with respect to social identities may bring many challenges, including but not restricted to urban integration, interethnic conflict, and threats to social cohesion and social capital (Putnam 2001; Putnam 2007). Understanding the social consequences of heterogeneity is, thus, one of the most important questions in sociology.

In this paper, we focus on cooperative behavior in social dilemmas and investigate how heterogeneity in social identity affects cooperation. Social dilemmas pose a conflict between collective and individual interests (Dawes 1980; Kollock 1998). Problems of collective action (Olson 1965), trust (Dasgupta 1988; Gambetta 1988), and cooperation (Taylor 1987) are classical examples of social dilemmas. Studying cooperation in social dilemmas is important because the capacity of a society to peacefully solve the conflict between individual and collective interests for the benefit of the latter characterizes the level of social cohesion in the society (Coleman 1990; Roca and Helbing 2011).

Prior experimental work has shown that heterogeneity in social identity generally hampers cooperation in social dilemmas (e.g., Yamagishi and Kiyonari 2000; Fershtman and Gneezy 2001; Bouckaert and Dhaene 2004; Eckel and Grossman 2005; Simpson 2006; Habyarimana et al. 2007; Chen and Li 2009). The results of these experiments resonate well with the observational studies which also show, albeit not in all occasions, a negative association between heterogeneity and social cohesion (see Dinesen and Sønderskoy 2015 and for an overview Van der Meer and Tolsma 2014).

In most prior experimental and observational studies, however, a crucial factor, namely endogenous sorting of heterogeneous actors has been largely left out. It is well documented that when a social entity is composed of heterogeneous actors, these actors tend to self-select into homogeneous subgroups (Schelling 1971; Marsden 1987; McPherson, Smith-Lovin, and Cook 2001). Due to this self-selection, the society may be globally heterogeneous but locally homogeneous, making heterogeneity a "moving target" (Uslaner 2010). Exposure to outgroups is thought to be a key mediator of the heterogeneity-cooperation link (Dinesen and Sønderskoy 2015). To what extent exposure to outgroups increases with heterogeneity is very difficult to

assess due to endogenous sorting of actors. In the aforementioned experimental studies, the experimenter typically forces subjects of different social identities to interact and compares the level of cooperation under this condition with an alternative condition in which actors of the same social identity are forced to interact. This typical design leaves no room for sorting. In the observational studies of the heterogeneity-cohesion link, the researcher can only observe social cohesion in an environment *ex-post,* that is, only after actors have sorted into the environment.

In this paper, we allow endogenous sorting in a controlled experiment by giving subjects the opportunity to choose the social group of their interaction partner before the interaction takes place. We compare this treatment with other treatments in which partner selection is not allowed, similar to the treatments in previous experimental studies. We show, as predicted by our theory, that allowing endogenous sorting mitigates the negative effects of heterogeneity on cooperation. We thus illustrate, both theoretically and empirically, that ignoring endogenous sorting processes yields misleading conclusions on the heterogeneity-cooperation link.

A second important aspect of the current work is that we study the heterogeneity-cooperation link at two levels: local (*dyadic*) and aggregate (*tetradic*). Actors can sort into homogenous environments locally, but at the aggregate-level heterogeneity remains unchanged. For example, sorting into different neighborhoods in a city does not influence heterogeneity at the city-level. Yet, some collective goods are produced at the local-level and some at the aggregate-level. The level at which the heterogeneity-cooperation link is studied is important. Dinesen and Sønderskoy (2015), for example, show that it is heterogeneity at the immediate micro-context rather than heterogeneity at larger geographical units that hampers social cohesion. The level at which Dinesen and Sønderskoy (2015) measure social cohesion, however, still remains unclear. In this paper we aim to show that depending on how interactions at the local-level unfold, hererogeneity

may or may not substantially influence cooperation at the aggregate-level. In other words, we show that the heterogeneity-cooperation link at the aggregate-level is not independent from local-level interactions.

A further distinctive feature of the current work is that we provide an explicit micro-behavioral model that defines and formalizes the micro mechanisms that link heterogeneity, self-selection into homogenous groups, and cooperation. Any theoretical model that seeks to explain a macro relationship should, in principle, include some sort of individual-level mechanism (Coleman 1990; Hedström 2005; Raub, Buskens, and van Assen 2011). This micro mechanism is missing in most previous theoretical work. For example, Putnam (2007) argues that heterogeneity hampers all forms of cooperation and trust (i.e., between and within groups). But the mechanisms through which heterogeneity hampers these different forms of cooperation remains unclear in Putnam's original work (Van der Meer and Tolsma 2014).

THEORY

We present our theoretical model in two steps. In the first step, we start with the case in which actors do not have an option to choose their interaction partner. This resembles the situation in almost all intergroup experiments (e.g., Yamagishi and Kiyonari 2000; Eckel and Grossman 2005; Simpson 2006; Chen and Li 2009). In the second step, we allow actors to choose the group of their interaction partners before the interaction takes place. This allows actors to potentially sort into homogenous subgroups. Perhaps due to theoretical and practical difficulties of providing actors with an opportunity to choose their partners, past experimental research has mostly ignored endogenous sorting (for an exception, see Slonim and Garbarino 2008).[1]

Social identity theory (Tajfel 1970; Tajfel et al. 1971) predicts that actors classify self and others in a few categories to cope with cognitive complexity. This is called the categorization

process. The second important element of social identity theory is self-enhancement: actors are predicted to attribute positive characteristics to the ingroup and negative characteristics to the outgroup. In addition to categorization and self-enhancement, a further element of social identity theory is about actors' perceptions of others' interests: ingroup members' interests are perceived as interchangeable, and actors assign a greater weight to collective ingroup outcomes than the outcomes for the self (Kramer and Brewer 1984; Turner 1985). We capture these effects of social identity on perceptions of others' interests in the following way.

In our model, in their "utility functions" actors assign non-zero weights to the outcomes of others. A utility function is a function that maps a distribution of tangible objective outcomes, such as money, to a subjective evaluation of this distribution. In principle, outcomes can be negative or positive. In our application, all outcomes are positive. An actor has two weights in their utility function, one for ingroup members' outcomes and one for outgroup members' outcomes. These weights are random variables and actors may have different weights. For example, some actors may assign higher weights to others' outcomes. Others may assign zero weights and thus are selfish. Still others may assign negative weights to others' outcomes. This latter type of actors can be classified as competitors. Utilities can be negative while outcomes are positive. For example, a competitor actor who assigns a negative weight to others' outcomes may derive disutility from a distribution in which his/her own outcomes are lower than others' outcomes.

Furthermore, we assume that actors vary with respect to ingroup bias; some assign high weights to ingroup members and low weights to outgroup members, others are less biased and assign similar weights to ingroup and outgroup members. Some "multicultural" actors may even assign higher weights to outgroup members' outcomes than to ingroup members' outcomes. Yet, in line with social identity theory we expect that, on average, the weight assigned to the outcomes

6

of ingroup members is higher than that of outgroup members' outcomes. Formally, the utility of

actor i from an outcome allocation for the self (x) and for N_j others ($\mathbf{y}$) is:

$$U(x, \mathbf{y}) = x + \sum \theta_i^g \times y_j \ \text{ with } \theta_i^g = \begin{cases} \theta_i^I \text{if } j \in I \\ \theta_i^O \text{if } j \in O \end{cases} \text{ and } E(\theta_i^I) > E(\theta_i^O) \qquad (1)$$

where I indicates the set of ingroup members and O indicates the set of outgroup members, and θ_i^I

is the weight i assigns to the outcomes of ingroup others and θ_i^O to the outcomes of outgroup

others. The utility is thus simply the outcome for the self (x) plus the sum of the outcomes of others

weighted by θ_i^G. This is an extension of the social orientation model (Schulz and May 1989;

Van Lange 2000; Simpson 2004; Chen and Li 2009; Aksoy and Weesie 2012; Murphy,

Ackermann, and Handgraaf 2011; Aksoy and Weesie 2014). Given this utility function, actors

transform the outcomes in the PD and PG (Table 1) into utilities. Previous research has shown that

this model is successful in explaining cooperative behavior in social dilemmas (Aksoy and Weesie

2009; Aksoy and Weesie 2013).

The predictions summarized below are derived from the following two basic assumptions: (1)

the average weight for ingroup members' outcomes is higher than the average weight for outgroup

members' outcomes and (2) actors maximize utility given the function in equation 1 (the rationality

assumption).

Predictions for the PD with and without Partner Choice

Consider the PD in Table 1 A. Without assuming any social preferences, the predicted behavior

(which happens to be the Nash equilibrium) is mutual non-cooperation. This is because whatever

the other player does, not cooperating gives a higher individual payoff. In game-theoretic

language, not cooperating is a dominant strategy. Both actors would be better off by mutual

cooperation which is unsustainable due to individual incentives for non-cooperation.

The solution of the game is different once one assumes social preferences. If we assume that subjects transform the outcomes in the PD into utilities using equation (1), then those with sufficiently large θ^g will cooperate. For example, if $\theta^g = .5$ then, Player 1 prefers to cooperate if he/she thinks Player 2 will cooperate (because $30 + .5 \times 30 > 35 + .5 \times 15$) *and* even if he/she thinks Player 2 will not cooperate (because $15 + .5 \times 35 > 20 + .5 \times 20$). In fact, the critical value of θ^g above which cooperation becomes more desirable is 1/3. Because we assume that on average ingroup θ is higher than outgroup θ, there will be more actors with a θ exceeding 1/3 when two ingroup members play this game than when two outgroup members play.[2] Hence, we expect that *without partner choice, within-group cooperation will be higher than between-group cooperation (H1)*.

Let's now assume that there are two groups, A and B, and before playing the game members of A and B can choose to be matched with either an ingroup or an outgroup member. In this case the formal analysis is more complex. The reason is as follows. An actor may choose to interact in an ingroup environment (i.e., play the PD with an ingroup member) or in an intergroup environment (i.e., play the PD with an outgroup member). The composition of these environments with respect to θ is endogenous. For example, members of Group A who choose to interact with members of Group B have a different θ distribution than members of Group A who choose to interact within their group. One may think that those who have only very large θ^Os will choose to interact in intergroup environments. If that was the case, then the cooperation rates in intergroup environments would be very high, which would be exploited by others who do not have high θ^O but choose to play in the intergroup environment to not cooperate and earn a high payoff. Of

course, each actors' partner choice decision contributes to the composition of θ across environments.

While an analytical solution is complex, a numerical solution can easily be found with computer simulations. Table 2 includes some examples. The table presents a number of θ distributions and the resulting predicted equilibrium cooperation rates with and without partner choice.[3]

Extensive simulations show that *when they have the choice, the majority of actors choose to interact with ingroup members (H2a)* (see column 6 in Table 2). Furthermore, *even with partner choice, the level of between-group cooperation is lower than the level of within-group cooperation (H2b)* (column 4 versus column 5 in Table 2). But *the difference between within- and between-group cooperation will be lower when actors have partner choice than when they don't have partner choice (H2c)* (see the difference in differences between columns 4 and 5 and columns 8 and 9 in Table 2).

[Table-2-about-here]

These theoretical results point to a paradoxical effect of sorting. On the one hand, when actors are free to choose their interaction partners, they mostly interact with ingroup members which increases segregation. On the other hand, this sorting keeps the level of cooperation in between-group interactions higher compared to the situation in which actors are forced to interact with outgroup members. This latter phenomenon occurs because actors who would behave non-cooperatively toward outgroup others, had they been forced to interact with the outgroup, are more likely to play with ingroup members when they can choose. Thus, most of the actors who would behave non-cooperatively toward outgroup members endogenously sort into within-group

9

environments. This sorting keeps the level of cooperation in between-group environments relatively high.

In addition to linking heterogeneity, sorting, and cooperation, our theoretical scheme provides further interesting implications. For example, research has shown that minority groups display more ingroup bias, are less cooperative, and have lower levels of generalized trust than majority groups (Hewstone, Rubin, and Willis 2002; Schlueter and Scheepers 2010; De Vroome, Hoege, and Marien 2013). In addition, a larger minority size is associated with stronger anti-immigrant attitudes among the majority group. These findings are often explained by psychological mechanisms such as perceived threat (Hewstone et al. 2002; Schlueter and Scheepers 2010).

Our theoretical model provides an alternative, a more structural explanation. Minority groups are more constrained in their partner selection choices. That is, they have to interact with outgroup members more frequently than the majorities interact with minority groups. Majority group members, however, mostly lack the opportunity to interact with outgroup members. Translating this to our setup, the minorities interact more often in conditions that resemble the 9th column of Table 2 whereas the majorities interact more often in conditions that resemble the 8th or the 4th column of Table 2 because the majorities are often isolated or can choose to isolate from the minorities. Our theory predicts that cooperation will be lower among the minorities than the majorities because of these differences in the structure of interaction opportunities, not necessarily because of a perceived threat or any other psychological process.

Predictions for PG

Public Goods Game is the N-person version of the PD (Table 1B). Here, we will restrict the formal and empirical analysis of PG to the case in which actors cannot choose their interaction partners. Actors can sort into homogenous groups locally (in our case in the PD), but at the most aggregate

level (in our case PG) heterogeneity remains unchanged. For example, sorting into different neighborhoods does not influence heterogeneity at the city level. Yet, some collective goods are produced at the local-level and some still at the aggregate-level. Including PG alongside the PD, we aim to understand the interplay between the local- and aggregate-level dynamics.

Consider the PG in Table 1B. As in the PD case, without social preferences universal non-cooperation is the predicted outcome of PG because not cooperating always gives a higher payoff irrespective of how many others cooperate. When we assume that actors transform the outcomes of the game into utilities using the equation (1) we have a different prediction: those with sufficiently high θ^I and θ^O will cooperate and the rest will not cooperate.[4] The exact probability that an actor cooperates in PG depends on the actor's ingroup and outgroup θs, weighted by the number of ingroup and outgorup members involved in PG.

One can test several implications of this formal result. For example, the higher the number of ingroup members relative to the number of outgroup members in the game, the higher the likelihood of cooperation. To be clear, we are not testing this prediction which is effectively the same as H1 formulated above. We will be testing H1 using the PD. In our experimental design, we keep the level of heterogeneity in PG constant: in all experimental conditions our subjects play PGs with one ingroup and two outgroup members. This is because we are interested in a further factor that may influence behavior in PG, namely prior negative between-group contact, which is discussed next.

Negative between-group contact and cooperation. Contact between members of different groups is a double-edged sword. Under certain conditions, contact is predicted to reduce intergroup conflict (Allport 1954). Although this has been shown to be true in various situations (e.g., Brown and Hewstone 2005) in our case, the conditions for contact to be beneficial are not

present. Social dilemmas are situations in which there is a risk for cooperative actors to be exploited by non-cooperative others. The formal analysis presented above predicts that interactions between actors of different social identities in social dilemmas do not reduce conflict; on the contrary, in such situations cooperation is predicted to be low (also see Rutherford et al. 2014).

Recent social psychological studies on "negative contact" show that negative experiences between groups are more strongly related to increased conflict than positive experiences are with reducing conflict (Paolini, Harwood, and Rubin 2010; Barlow et al. 2012). This is consistent with the game-theoretic literature on reciprocity which shows that people adjust their social preferences based on the previous behaviors of their interaction partners. People respond to the negative intentions of others with less pro-social preferences (negative reciprocity), and the positive intentions of others with more pro-social preferences (positive reciprocity) (Gautschi 2000; Falk and Fischbacher 2006). Moreover, negative reciprocity is shown to be a stronger motivation than positive reciprocity (Aksoy and Weesie 2014).

The formal analysis of the PD predicts that the subjects will experience the lowest levels of cooperation in between-group interactions when actors cannot choose the group of their interaction partners. When actors have partner choice, cooperation will still be lower in between-group interactions than in within-group interactions, but this difference will not be as high as the difference in the no partner choice case. Hence, we expect that during and after intergroup interactions unfold, subjects will lower their θ^O (the weight assigned to outgroup others' outcomes), particularly when they could not choose the group of their interaction partners. This yields the prediction that *keeping the level of heterogeneity constant, the more prior negative between-group contact there is, the lower will be cooperation in heterogeneous PGs (H3).*

Note that negative contact can have an effect on cooperation if actors receive information about others' decisions. When no feedback about others' decisions is provided, actors will not be exposed to non-cooperative behaviors of others (ingroup or outgroup), hence there will be no reason for reciprocity. We thus test H3 rigorously by experimentally concealing and revealing the information subjects receive about the level of negative contact between groups.

METHODS

Participants

186 subjects were recruited using the online recruitment system ORSEE (Greiner 2004). To assess the external validity of the experiment, subjects from heterogeneous backgrounds were invited. Of the subjects, 42 percent were male, and 50 percent had UK citizenship. The remaining non-UK subjects were from 33 different countries. 75 percent of the subjects were students, mainly at the University of Oxford, and the remaining 25 percent were non-student citizens of the city of Oxford. The average age of the subjects was 29.8 years (S.D.=13.9).

Design and Procedure

The experiment comprised 10 sessions. In each session, subjects were seated randomly in a cubicle in the Nuffield Centre for Experimental Social Science (CESS) lab at Nuffield College, so that they could not see each other. The entire procedure took place on computers with the software z-tree (Fischbacher 2007).[5] Throughout the experiment, subjects earned tokens that were converted to cash at the end of the experiment at an exchange rate of 20 tokens = £1 (~1.6 USD). Subjects earned £14 (~22 USD) on average (min £11, max £17), including a £4 show-up fee.

The experiment included four between-subjects treatments, *mixed*, *choice*, *isolated*, and *control,* and two within-subjects treatments, *no-feedback* and *feedback*. The overall design of the experiment is shown in Table 3 and explained in detail below.[6]

[Table-3-about-here]

Inducing heterogeneity. Our theoretical model does not rely on the nature of the groups per se. The theory applies more generally to any case in which actors, on average, feel more altruistic toward an ingroup than toward an outgroup. Nonetheless, a key design issue is how to manipulate heterogeneity. One can use natural identities, such as ethnicity and gender or it can be induced in the lab. Both approaches have been implemented in the literature. Gil-White (2003), Bouckaert and Dhaene (2004), Buchan, Johnson, and Croson (2006), Goette, Huffman, and Meier (2006), Habyarimana et al. (2007), Kuwabara et al. (2007), Bornhorst et al. (2010) are examples of recent studies that use natural identities. Tajfel et al. (1971), Yamagishi and Kiyonari (2000), Simpson (2006), and Chen and Li (2009) are studies that induced identity in the lab using the minimal group paradigm, in which subjects are divided into groups based on a trivial criterion. Experiments with minimal groups show that such a categorization per se, no matter how trivial, is sufficient to induce a fairly strong group identity.

Natural identities are multi-dimensional; many factors in addition to social identity vary with natural identities (Tajfel et al. 1971; Yamagishi and Kiyonari 2000; Chen and Li 2009). Moreover, it is generally not known beforehand how natural identities will respond to experimental manipulations. It is also difficult to assess whether a subject identifies with her "natural" group at all or if she does so whether *others* categorize her by her social identity Consequently, the results

14

of studies that use natural identities are more ambiguous because they can be attributed to many causes. Inducing identities in the lab with the minimal group paradigm, in contrast, gives the experimenter greater control over the subject's prominent identity and greater confidence that results are due to the experimental manipulation. Based on this discussion, in the current experiment, social identity was induced in the lab using minimal groups.

Subjects received five pairs of paintings by Paul Klee and Wassily Kandinsky. Depending on their relative preferences, half of the subjects were classified as Klees and the other half as Kandinskys.[7] A subject's identity as a Klee or a Kandinsky remained fixed for the rest of the experiment. The design included two further procedures to enhance group identity. Because these procedures are quite standard in the minimal group literature, they are not discussed here but given in full detail in the Online Appendix B.

Once social identities were induced (or not induced, in the control condition) the primary experimental procedure took place. Below, we summarize the details of the *mixed*, *choice*, *isolated*, *control* and the *no-feedback* and *feedback* treatments.

Mixed condition. **Stage 2:** In stage 2, subjects played 10 rounds of PD each time with a randomly matched partner. The particular outcome parameters in the PD were chosen based on previous experimental studies to keep cooperation levels close to .5 which would provide the maximum statistical power to detect differences between experimental conditions (Isaac and Walker 1988; Aksoy and Weesie 2013). Also, this version of the PD kept the fear and greed components the same (Simpson 2003; Kuwabara 2005).

In the mixed condition, the group of the interaction partner in each of the 10 PDs was always alternated. In the odd-numbered (even-numbered) rounds, the interaction partner was always a randomly selected ingroup (outgroup) member. Participants were paid for two randomly selected

rounds, one ingroup and one outgroup round. Information on the rounds that were selected and how much subjects earned in those rounds was made available only after the subjects completed the experiment.[8]

Stage 3: Subjects played 5 rounds of binary choice Public Goods Games (Table 1B).[9] Because we conceptualize PG as a higher level collective good from which one cannot exclude actors of different social identities, in each of these five rounds two randomly selected Klees and two randomly selected Kandinskys were matched. Participants were paid for a randomly selected round. Information on the round that was selected and how much the subjects earned in this round was made available at the end of the experiment.

During stages 2 and 3, no feedback about other subjects' behavior was provided. Hence, this first block is called the no-feedback condition.

Stage 4: Subjects played another 10 rounds of PD, half of the rounds with ingroup and half with outgroup as in stage 2. Different from stage 2, before each round the subjects received aggregate-level feedback about: (1) the percentage of cooperative decisions made by Klees who were matched with other Klees, (2) the percentage of cooperative decisions made by Klees who were matched with Kandinskys, (3) the percentage of cooperative decisions made by Kandinskys who were matched with other Kandinskys, and (4) the percentage of cooperative decisions made by Kandinskys who were matched with Klees, averaged up to that round. The payment procedure was the same as in stage 2.

Aggregate feedback (rather than individual feedback) was provided for the following reasons. If individual feedback was provided, each subject would experience a unique experimental history. This would undermine the experimental control over the subjects' experiences and would increase unsystematic variance in the data, complicating statistical analysis. Aggregate-level

feedback ensured not only that each subject received the same aggregate information but also that each subject knew that each subject received the same information.

Stage 5: Participants played a final 5 rounds of PG in heterogeneous groups of two Klees and two Kandinskys with feedback. Before each round, the subjects were informed about the percentage of cooperative decisions made by Klees and Kandinskys, averaged up to that round.

Once stage 5 was completed, the randomly selected payment rounds and the amount the subjects earned throughout the experiment were revealed.

Choice condition. The procedure was identical to the mixed condition with one exception. Instead of alternating the group of the interaction partners exogenously, in stages 2 and 4, each subject submitted whether she wanted to be matched with a Klee or a Kandinsky before each PD round. Then, the subject was matched with another subject from the target group who wanted to be matched with someone from the subject's group.[10]

In stage 4 in which feedback was provided the subjects were informed before partner choice and the PD play about the same percentages of cooperative behaviors of each group against each group as in the mixed condition.

Stages 3 and 5 were exactly the same as in the mixed condition: subjects played 5 PGs in a heterogeneous group (two Klees two Kandinskys) formed randomly in every round, without and with feedback. In other words, in PGs, subjects could not choose the groups of their interaction partners.

Isolated condition. The procedure was identical to the other two conditions with the following exception. In stages 2 and 4, where subjects played 10 PDs without and with feedback, respectively, the subjects were matched with people only from their own group. Stages 3 and 5 in this condition were exactly the same as the other two conditions. To be clear, we do not expect any

difference between this condition and the ingroup cooperation in the mixed condition. Different from the two conditions above, however, in this condition there is no (negative) contact between the two groups which enables us to test H3.

Control condition. Control condition did not induce group identity. In stages 2 and 4 (3 and 5), subjects played 10 PDs (5 PGs) each round with randomly selected others without and with feedback, respectively. In stages 4 and 5, before each round, the subjects were informed about the overall percentage of cooperative decisions, averaged up to that round. To be clear, we do not have hypotheses about the control condition. This condition would reveal if the inducement of group identity influences cooperation in an unexpected way.

Note that the *mixed*, *choice*, *isolated*, *control* treatments were between-subjects and the six stages, hence the *no-feedback* and *feedback* treatments were within subjects. The feedback treatment always followed the no-feedback treatment. We chose not to include a further condition in which the order of the no-feedback and the feedback treatments was reversed (i.e., no-feedback followed feedback). This was because in this alternative order, even if no-feedback was provided in a subsequent block, at least some subjects would remember the feedback provided in the first block.

Post-Experiment Survey

The post-experiment survey included the following four items taken from Yamagishi and Kiyonari (2000) to measure the degree of identification with the ingroup and outgroup: "How strongly did you feel belongingness to the (Klee or Kandinsky) group?" (belongingness); "How much commonality did you think you shared with the members of the (Klee or Kandinsky) group?" (commonality); "How close did you feel toward the members of the (Klee or Kandinsky) group?" (closeness); and "How favorably did you feel toward the members of the (Klee or Kandinsky)

group?" (liking). The answer categories were from 1 (not at all) to 7 (very much). In the control condition which had no grouping, these questions were omitted.

Basic demographics (age, gender, study field, occupation, nationality), political party preference measured from 1 (extreme left) to 11 (extreme right), number of siblings, and level of experience in similar experiments were other variables measured in the post-experiment survey.[11]

RESULTS

Manipulation Check

Participants in the minimal group condition felt more closeness (difference $= 1.48$, $t(121) = 9.50$, $p < .05$), more belongingness (difference $= 1.91$, $t(121) = 10.59$, $p < .05$), and commonality (difference $= 1.34$, $t(121) = 7.89$, $p < .05$) toward the ingroup and liked their ingroup significantly more (difference $= 1.30$ $t(121) = 7.10$, $p < .05$) compared to the outgroup.[12] All of these differences remain statistically significant when the three experimental conditions are considered separately.[13] Because all differences are substantial (i.e., larger than 1) and statistically significant, we conclude that the inducement of heterogeneity was successful.

Cooperation

Figure 1 displays the average cooperation in the PD and PG by treatment. Before interpreting the results, we should note that there is no significant main effect of a preference for the paintings of Klee (versus Kandinsky) on cooperation, nor is there a significant interaction of a preference for the paintings of Klee (versus Kandinsky) with the experimental conditions for any of the four panels in Figure 1. Hence, we collapse Klees' and Kandinskys' responses in all analyses. Furthermore, there is no significant effect of being a student on cooperation, and there is no

significant interaction between being a student and the experimental condition. Similarly, being a British citizen has no significant effect on cooperation in the PD or in PG.

[Figure-1-about-here]

The significance tests reported below are based on multilevel logistic regression models with random effects for subjects (Table 4). The subject-level random effects capture roughly the θ parameter in the theoretical model. When included in the model as a second level, variance across sessions is not statistically significant. The regression models also include the control variables sex, political party preference, and age, the only subject-level covariates that influence cooperation significantly. The effect of feedback on the effects of treatment is captured by interaction terms in the regression models.

[Table-4-about-here]

Note that the formal tests of our hypotheses are performed via a number of linear combinations of the coefficients in Table 4. Table 5 presents how these linear combinations are calculated, together with the associated hypotheses. Table 5 also includes the standard errors and the z-values of the linear combinations which are used to test the hypotheses formally.[14]

[Table-5-about-here]

PD Behavior

As apparent from Figure 1, there is a substantial difference between within-group and between-group cooperation in the mixed condition (H1). This is in line with theoretical predictions, but the effect size is striking: the odds of cooperating are approximately 4.6 times higher (see Table 5) with feedback and 3.2 times without feedback when ingroup members interact compared to the situation in which outgroup members interact.

As predicted, when they have the choice, the majority of the subjects (more than 70 percent) play with an ingroup member (H2a). A separate multilevel logistic regression model with random effects for subjects predicting partner selection shows that the percentage of subjects who choose to interact with an outgroup member is significantly different from 50 percent ($z = -5.04$, $p < .05$). The predicted difference between within-group and between-group cooperation rates in the choice condition becomes statistically significant only when feedback is provided (H2b) ($\Delta b = .63$, $z = 1.86$, $p < .05$, see Table 5). This difference is substantially significant, too: when feedback is provided the odds of cooperating are about 1.9 times higher in the choice condition when ingroup members interact compared to the situation in which outgroup members interact.

We also find strong support for H2c: the difference between within-group and between-group cooperation is lower in the choice condition compared to the mixed condition (without feedback, difference in $\Delta b = -1.46$, $z = -3.21$, $p < .05$; with feedback, difference in $\Delta b = -.89$, $z = -2.03$, $p < .05$, see Table 5). Figure 1 is useful to interpret the substantial significance of this result. Without feedback, within-group and between-group average cooperation rates are virtually identical in the choice condition but substantially different (i.e., larger than 0.2) in the mixed condition. With feedback, the average between-group cooperation becomes lower than the average within-group

cooperation in the choice condition, too. But the difference remains much higher in the mixed condition.

As an aside, our results also show that when the mixed and the choice conditions are combined in a separate regression model, the difference between within-group and between-group cooperation is higher under the feedback condition than under the no-feedback condition ($\Delta(\Delta b)=$ .61, $z = 2.11$, $p < .05$). Thus, feedback moderates the negative effect of heterogeneity on cooperation.

PG Behavior

In the no-feedback condition, cooperation rates in PG across treatments are very similar. In fact, the differences between treatments are insignificant when subjects do not receive feedback about others' decisions. When feedback is provided, the odds of cooperating in the isolated condition are approximately 12 times higher than in the mixed condition, and the difference is statistically highly significant ($\Delta b = 2.46$, $z = 2.70$, $p < .05$). The difference between the isolated and the choice condition is also statistically significant with feedback ($\Delta b = 1.72$, $z = 1.89$, $p < .05$). This corresponds to an odds ratio between the isolated and the choice condition of about 5.6. These results support H3.

DISCUSSION

In this paper we provide a micro-behavioral model and an experimental design to understand the effects of heterogeneity in social identities on cooperation. The theoretical scheme is a typical example of a micro-macro model (Coleman 1990; Hedström 2005; Raub et al. 2011). The micro-behavioral model links the macro constructs of heterogeneity and cooperation via the social preferences of actors while accounting for partner selection decisions (and thus endogenous

sorting). We show that given the same distribution of micro-level social preferences, depending on the nature of initial macro conditions quite different macro outcomes may emerge. For instance, the freedom actors have in choosing the social group of their interaction partners has a very substantial effect on the link between heterogeneity and cooperation.

The experimental design induces heterogeneity independent of endogenous sorting and cooperation, thus provides a clear assessment of the causal effect of heterogeneity on cooperation. Moreover, the design allows us to naturally analyze cooperation at two levels, dyadic and tetradic.

Our results show that when dyadic interactions are considered, heterogeneity decreases cooperation substantially. Most importantly, however, this negative relationship is moderated by endogenous sorting. When the subjects have the option to choose their interaction partners, a somewhat paradoxical effect—which is precisely predicted by our theoretical model—occurs. On the one hand, most of the subjects (more than 70 percent) sort and choose to interact with other ingroup members. On the other hand, this sorting *reduces* ingroup bias with respect to cooperative behavior. This is because when the subjects have a partner choice, then those who would act non-cooperatively toward outgroup others are more likely to play with ingroup others. Consequently, the difference between within-group and between-group cooperation is lower in the choice condition compared to the case in which subjects are forced to interact with outgroup members.

This particular result has a number of methodological implications. Firstly, it shows that it is very difficult to measure heterogeneity independent of endogenous sorting in observational studies because in most real life situations actors have a partner choice. As a consequence, most observational studies might have *underestimated* the "true" causal effect of heterogeneity on social cohesion. This is because in those studies the researcher can measure heterogeneity and cohesion

only after participants have sorted into their environments. Conversely, most previous intergroup experiments might have *overestimated* the level of ingroup bias that occurs naturally. This is because in those experiments subjects typically do not have a partner choice but are forced by the experimenter to interact either with ingroup or outgroup members.

Taken together, our theoretical and experimental results help us understand under which conditions heterogeneity is expected to reduce cooperation and social cohesion. At the local-level heterogeneity clearly decreases cooperation. But the size of this negative effect depends on how much freedom actors have in selecting their social environments. In rigid environments in which sorting in or out is relatively difficult, such as residential areas or schools, heterogeneity will likely reduce cooperation. In more fluid environments in which actors can voluntarily choose their interaction partners, heterogeneity is expected to influence cooperation much less, if at all. For example, heterogeneity in social clubs or in one's personal network should not have strong effects on social cohesion.

In addition to endogenous sorting, how much heterogeneity reduces cooperation also depends on the history of intergroup relations and how much actors are primed about this history. Our experiment shows, as predicted by our theory, that information about intergroup relationships is another key moderator of the heterogeneity-cooperation link. Cooperation diminishes in heterogeneous environments quite substantially when actors are informed about a negative history between groups. If there is prior negative contact but the subjects are unaware of it or if there is no prior contact between groups, cooperation remains relatively high in heterogeneous environments. This also hints that the negative effect of heterogeneity on cohesion is partly a self-fulfilling prophecy. Our theory predicts that making intergroup conflict more salient by, for example, a sustained emphasis from the media or politicians will reduce social cohesion in heterogeneous

environments. Our results also show that the effect of information about negative contact spills over to more macro-level contexts. For instance, in our experiment negative contact happens initially in the PD context. Yet, this negative intergroup history in the PD reduces cooperation in a more macro level, namely in heterogeneous PG. This implies that making intergroup conflict in a micro-context more salient would reduce cooperation not only at the micro-context but also at more macro levels.

We must add that in real life group interactions are more complex than we theoretically and experimentally analyzed: many other factors co-vary with social identity. For example, some ethnic groups are often worse-off than others and there is often a status difference between groups. Precisely because of these confounding factors, it is difficult to isolate the causal effect of heterogeneity in observational studies. We should note, however, that many of these factors will likely reinforce the results presented in this paper. If, for example, social identity is coupled with inequality, this would likely strengthen the differences between groups because actors would be different not only with respect to social identities but also with respect to resources. Nevertheless, future work should extend the experimental design presented in this paper with additional elements. The most obvious next step is inducing social identity and inequality together.

SUPPLEMENTAL MATERIAL

Additional supporting information may be found at [insert link].

Oxford, the Social Sciences Seminar Series at Koc University, the Thurgau Experimental Economics Meeting in Stein am Rhein, the Seventh Analytical Sociology Conference in Mannheim, the International Conference on Experimental Social Science on Social Dilemmas in Utrecht, four anonymous SPQ reviewers, and the SPQ editors Richard T. Serpe and Jan E. Stets for their helpful comments. The experiment reported in this paper was conducted at Nuffield Centre for Experimental Social Sciences. Financial support from the Netherlands Organization for Scientific Research (NWO) under grant 446-13-004 is gratefully acknowledged.

NOTES

1. There are a number of experimental studies which investigate cooperation in dynamic networks (e.g., Rand, Arbesman, and Christakis 2011; Wang, Suri, and Watts 2012; Gallo and Yan 2015). These studies also allow partner selection, similar to our design. In these dynamic network experiments, however, a subject can form a tie (or severe an existing tie) with a specific and anonymous individual, rather than a randomly selected member of a social group. Hence, these experiments do not study the effects of heterogeneity in social identity on endogenous sorting and cooperation. A further difference between those dynamic network experiments and the current design is that in those experiments a subject can be matched with several other subjects simultaneously. Nevertheless, the results of those dynamic network

experiments unequivocally show that endogenous sorting increases cooperation which is in line with the core argument of our paper.

2. The average level of ingroup and intergroup cooperation depends on the distribution of θ^g, (i.e., they will be 1- $F^I(\frac{1}{3})$ and 1- $F^O(\frac{1}{3})$, respectively where F^I and F^O are the cumulative distributions of θ^I and θ^O). While $E(\theta^I) > E(\theta^O)$ does not ensure that $F^I(\frac{1}{3}) < F^O(\frac{1}{3})$, in most reasonable distributions ingroup cooperation will be higher than intergroup cooperation.

3. In these examples, (θ^I, θ^O) is assumed to have a multivariate normal distribution with specific parameter values assessed as reasonable based on previous research that estimates θ without group manipulation (Aksoy and Weesie 2012; Aksoy and Weesie 2014). Online Appendix A includes a detailed description of the simulation used to produce the results in Table 2.

4. Written formally what sufficiently high means, an actor i who has no partner-choice and plays the game with N^I_{-i} ingroup and N^O outgroup others is predicted to cooperate with probability: $\Pr_i(\text{Cooperation}) = Pr(N^I_{-i}\theta^I_i + N^O\theta^O_i > \frac{1}{3})$.

5. One of the 10 sessions had to be aborted prematurely due to a power failure and a subsequent software crash. Data recorded before the power failure were used in the analyses whenever possible. Removing or including this session from the analyses does not change the results.

6. All instructions and the screens subjects saw during the experiment are available from the authors upon request.

7. Splitting the session into to equally sized groups maximizes the Herfindahl index, a standard measure of heterogeneity.

8. After the PDs and PGs, beliefs about others' behavior were measured using the quadratic discount rule which is a "proper" scoring rule (i.e., truth telling is a dominant strategy) (Palfrey and Wang 2009). In this paper, we do not discuss beliefs.

9. The particular outcomes of PG and the PD in Table 1 kept the "Marginal Per Capita Return" the same between the PD and PG, viz., .75.

10. In some cases, the number of people in a focal group who wanted to be matched with a target group could be different than the number of people in the target group who wanted to be matched with the focal group. In such cases, the computer performed a one-to-many matching. If there was nobody in the target group, the computer randomly selected a decision from previous rounds from the target group that wanted to match with the focal group. If there was no one in any of the previous rounds, the computer randomly selected a decision. Subjects were fully informed about this matching procedure.

11. Generalized trust was also measured in the post-experiment survey. We do not discuss the results on trust in detail for brevity. The levels of trust across the four experimental conditions are almost a copy of the levels of cooperation in PG with feedback. That is, trust in the mixed condition is significantly lower than trust in all other three conditions. This is inline with past research which shows a deep connection between cooperation and trust (Kuwabara et al. 2014; Irwin, Mulder, and Simpson 2014).

12. We report one-tailed p-values when we test explicit hypotheses (e.g., Table 5). All other p-values (manipulation checks and those in Table 4) are two-tailed.

13. See Table C1 in the online appendix for the means, the standard deviations, and the standard errors of the means of these items calculated separately for the ingroup and the outgroup.

14. The standard errors of the linear combinations are calculated using the Stata command "lincom".

REFERENCES

Aksoy, Ozan and Jeroen Weesie. 2009. "Inequality and Procedural Justice in Social Dilemmas." *Journal of Mathematical Sociology* 33:303–22.

Aksoy, Ozan and Jeroen Weesie. 2012. "Beliefs about the Social Orientations of Others: a Parametric Test of the Triangle, False Consensus, and Cone Hypotheses." *Journal of Experimental Social Psychology* 48:45–54.

Aksoy, Ozan and Jeroen Weesie. 2013. "Social Motives and Expectations in One-Shot Asymmetric Prisoner's Dilemmas." *Journal of Mathematical Sociology* 37:24–58.

Aksoy, Ozan and Jeroen Weesie. 2014. "Hierarchical Bayesian Analysis of Outcome- and Process-Based Social Preferences and Beliefs in Dictator Games and Sequential Prisoner's Dilemmas." *Social Science Research* 45:98–116.

Allport, Gordon W. 1954. *The Nature of Prejudice*. Cambridge, MA: Perseus Books.

Barlow, Fiona Kate, Stefania Paolini, Anne Pedersen, Matthew J. Hornsey, Helena R. M. Radke, Jake Harwood, Mark Rubin, and Chris G. Sibley. 2012. "The Contact Caveat: Negative Contact Predicts Increased Prejudice more than Positive Contact Predicts Reduced Prejudice." *Personality and Social Psychology Bulletin* 38:1629–43.

Bornhorst, Fabian, Andrea Inchino, Oliver Kirchkamp, Karl H. Schlag, and Eyal Winter. 2010. "Similarities and Differences when Building Trust: the Role of Cultures." *Experimental Economics* 13:260–83.

Bouckaert, Jan and Geert Dhaene. 2004. "Inter-Ethnic Trust and Reciprocity: Results of an Experiment with Small Businessmen." *European Journal of Political Economy* 20:869–86.

Brown, Rupert and Miles Hewstone. 2005. "An Integrative Theory of Intergroup Contact." Pp. 255–343 in *Advances in Experimental Social Psychology*, volume 37, edited by M. Zana. San Diego, CA: Acedemic Press.

Buchan, Nancy R., Eric J. Johnson, and Rachel T. A. Croson. 2006. "Let's Get Personal: an International Examination of the Influence of Communication, Culture and Social Distance on Other Regarding Preferences." *Journal of Economic Behavior and Organization* 60:373–98.

Chen, Yan and Sherry Xin Li. 2009. "Group Identity and Social Preferences." *American Economic Review* 431–57.

Coleman, James J. 1990. *Foundations of Social Theory*. Cambridge, MA: Belknap Press.

Dasgupta, Parta. 1998. "Trust as a Commodity." Pp. 49–72 in *Trust: Making and Breaking Cooperative Relations*, edited by D. Gambetta. Oxford: Blackwell.

Dawes, Robyn M. 1980. "Social Dilemmas." *Annual Review of Psychology* 31:169–93.

De Vroome, Thomas, Marc Hooghe, and Sofie Marien. 2013. "The Origins of Generalized and Political Trust among Immigrant Minorities and the Majority Population in the Netherlands." *European Sociological Review* 29(6):1336–50.

Dinesen, Peter Thisted, and Kim Mannemar Sønderskov. 2015. "Ethnic Diversity and Social Trust: Evidence from the Micro-Context." *American Sociological Review*: 0003122415577989.

Eckel, Catherine C. and Philip J. Grossman. 2005. "Managing Diversity by Creating Team Identity." *Journal of Economic Behavior & Organization* 58:371–92.

Falk, Armin and Urs Fischbacher. 2006. "A Theory of Reciprocity." *Games and Economic Behavior* 54:293–315.

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

Fischbacher, Urs. 2007. "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics* 10:171–78.

Gallo, Edoardo and Chang Yan. 2015. "The Effects of Reputational and Social Knowledge on Cooperation." *Proceedings of the National Academy of Sciences* 112(12):3647–52.

Gambetta, Diego. 1998. "Can We Trust." Pp. 213–37 in *Trust: Making and Breaking Cooperative Relations*. edited by D. Gambetta. Oxford: Blackwell.

Gautschi, Thomas. 2000. "History Effects in Social Dilemma Situations." *Rationality and Society* 12:131–62.

Gil-White, Francisco J. 2003. "Ultimatum Game with an Ethnicity Manipulation. Results from Khovdiin Bulgan Sum, Mongolia." Pp. 260–304 in *Foundations of Human Sociality: Ethnography and Experiments in 15 Small-Scale Societies*, edited by J. Henrich, R. Boyd, S. Bowles, and H. Gintis. Oxford: Oxford University Press.

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

Greiner, Ben. 2004. "The Online Recruitment System Orsee 2.0 – a Guide for the Organization of Experiments in Economics." *Working Paper Series in Economics* 10:1–15. Mimeo, University of Cologne.

Habyarimana, James, Macartan Humphreys, Daniel Posner, and Jeremy M. Weinstein. 2007. "Why Does Ethnic Diversity Undermine Public Goods Provision?" *American Political Science Review* 101:709–25.

Hedström, Peter. 2005. *Dissecting the Social: on The Principles of Analytical Sociology*. Cambridge: Cambridge University Press.

Hewstone, Miles, Mark Rubin, and Hazel Willis. 2002. "Intergroup Bias." *Annual Review of Psychology* 53:575–604.

Irwin, Kyle, Laetitia Mulder, and Brent Simpson. 2014. "The Detrimental Effects of Sanctions on Intragroup Trust Comparing Punishments and Rewards." *Social Psychology Quarterly* 77(3):253–72.

Isaac, R. Mark and James M. Walker. 1988. "Group Size Effects in Public Goods Provision: the Voluntary Contributions Mechanism." *Quarterly Journal of Economics* 103:179–99.

Kollock, Peter. 1998. "Social Dilemmas: the Anatomy of Cooperation." *Annual Review of Sociology* 24:183–214.

Kramer, Roderick M. and Marilynn B. Brewer. 1984. "Effects of Group Identity on Resource Use in a Simulated Commons Dilemma." *Journal of Personality and Social Psychology* 46:1044–57.

Kuwabara, Ko. 2005. "Nothing to Fear but Fear Itself: Fear of Fear, Fear of Greed and Gender Effects in Two-Person Asymmetric Social Dilemmas." *Social Forces* 84(2):1257–72.

Kuwabara, Ko, Robb Willer, Michael W. Macy, Rie Mashima, Shigeru Terai, and Toshio Yamagishi. 2007. "Culture Identity, and Structure in Social Exchange: a Web Based Trust Experiment in the United States and Japan." *Social Psychology Quarterly* 70:461–79.

Kuwabara, Ko, Sonja Vogt, Motoki Watabe, and Asuka Komiya. 2014. "Trust, Cohesion, and Cooperation after Early Versus Late Trust Violations in Two-Person Exchange: the Role of Generalized Trust in the United States and Japan." *Social Psychology Quarterly,* 77(4):344–60.

Marsden, Peter V. 1987. "Core Discussion Networks of Americans." *American Sociological Review* 52:122–31.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook. 2001. "Birds of a Feather:
Homophily in Social Networks." *Annual Review of Sociology* 27:415–44.

Murphy, Ryan. O., Kurt. A. Ackermann, and Michel. J. J. Handgraaf 2011. "Measuring Social
Value Orientation (SVO)." *Judgment and Decision Making* 6:771–81.

Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups.*
Cambridge, MA: Harvard University Press.

UK Office for National Statistics. 2014. "Migration Statistics Quarterly Report of UK Office for
National Statistics, February 2014." Retrieved September 22, 2015
(http://www.ons.gov.uk/ons/dcp171778_352080.pdf).

Palfrey, Thomas R. and Stephanie W. Wang. 2009. "On Eliciting Beliefs in Strategic Games."
Journal of Economic Behavior and Organization 71:98–109.

Paolini, Stefania, Jake Harwood, and Mark Rubin. 2010. "Negative Intergroup Contact Makes
Group Memberships Salient: Explaining why Intergroup Conflict Endures." *Personality and
Social Psychology Bulletin* 36:1723–38.

Putnam, Robert D. 2001. *Bowling Alone: the Collapse and Revival of American Community*.
New York: Simon and Schuster.

Putnam, Robert D. 2007. "E Pluribus Unum: Diversity and Community in the Twenty-First
Century the 2006 Johan Skytte Prize Lecture." *Scandinavian Political Studies* 30:137–74.

Rand, David G., Samuel Arbesman, and Nicholas A. Christakis. 2011. "Dynamic Social
Networks Promote Cooperation in Experiments with Humans." *Proceedings of the National
Academy of Sciences* 108(48):19193-98.

Raub, Werner, Vincent Buskens, and Marcel ALM van Assen. 2011. "Micro-Macro Links and
Microfoundations in Sociology." *Journal of Mathematical Sociology* 35:1–25.

Roca, Carlos P and Dirk Helbing. 2011. "Emergence of Social Cohesion in a Model Society of Greedy, Mobile Individuals." *Proceedings of the National Academy of Sciences* 108:11370–74.

Rutherford, Alex, Dion Harmon, Justin Werfel, Alexander S. Gard-Murray, Shlomiya Bar-Yam, Andreas Gros, Ramon Xulvi-Brunet, and Yaneer Bar-Yam. 2014. "Good Fences: the Importance of Setting Boundaries for Peaceful Coexistence."*PloS One* 9(5):e95660.

Schelling, Thomas C. 1971. "Dynamic Models of Segregation." *Journal of Mathematical Sociology* 1:143–86.

Schlueter, Elmar and Peer Scheepers. 2010. "The Relationship Between Outgroup Size and Anti-Outgroup Attitudes: a Theoretical Synthesis and Empirical Test of Group Threat and Intergroup Contact Theory." *Social Science Research* 39:285–95.

Schulz, U. and T. May. 1989. "The Recording of Social Orientations with Ranking and Pair Comparison Procedures." *European Journal of Social Psychology* 19:41–59.

Simpson, Brent. 2003. "Sex, Fear, and Greed: a Social Dilemma Analysis of Gender and Cooperation." *Social Forces* 82(1):35–52.

Simpson, Brent. 2004. "Social Values, Subjective Transformations, and Cooperation in Social Dilemmas." *Social Psychology Quarterly* 67(4):385–95.

Simpson, Brent. 2006. "Social Identity and Cooperation in Social Dilemmas." *Rationality and Society* 18:443–70.

Slonim, Robert, and Ellen Garbarino. 2008. "Increases in Trust and Altruism from Partner Selection: Experimental Evidence." *Experimental Economics* 11(2):134–53.

Tajfel, Henri. 1970. "Experiments in Intergroup Discrimination." *Scientific American* 223:96–102.

Tajfel, Henri, Michael G. Billig, Robert P. Bundy, and Claude Flament. 1971. "Social Categorization and Intergroup Behaviour." *European Journal of Social Psychology* 1:149–78.

Taylor, Michael. 1987. *The Possibility of Cooperation*. Cambridge: Cambridge University Press.

Turner, John C. 1985. "Social Categorization and the Self-Concept: a Social Cognitive Theory of Group Behavior." *Advances in Group Processes* 2:77–122.

Uslaner, Eric M. 2010. "Segregation, Mistrust and Minorities." *Ethnicities* 10:415–34.

Van der Meer, Tom and Jochem Tolsma. 2014. "Ethnic Diversity and Its Supposed Detrimental Effects on Social Cohesion." *Annual Review of Sociology* 40:null.

Van Lange, Paul A. M. 2000. "Beyond Self-Interest: a Set of Propositions Relevant to Interpersonal Orientations." *European Review of Social Psychology* 11:297–331.

Wang, Jing, Siddharth Suri, and Duncan J. Watts. 2012. "Cooperation and Assortativity with Dynamic Partner Updating." *Proceedings of the National Academy of Sciences* 109(36):14363–8.

Yamagishi, Toshio and Toko Kiyonari. 2000. "The Group as the Container of Generalized Reciprocity." *Social Psychology Quarterly* 63:116–32.

BIO

Ozan Aksoy is a post-doctoral research fellow at the Nuffield Centre for Experimental Social Sciences and the Department of Sociology, University of Oxford. His research interests include understanding cooperation and trust as well as religious behaviour and policy using game-theoretic models, statistical tools, and experimental methods. His recent work has been published in *Social Science Research*, *Journal of Mathematical Sociology*, *Journal of Experimental Social Psychology*, and *Games*.

TABLES

Table 1 A: The Prisoner's Dilemma (PD); B: Public Goods Game (PG). The numbers in cells indicate outcomes. In PG, the first number in a cell indicates how much Player 1 earns and the remaining numbers indicate how much others earn where the right-hand side number(s) indicate the outcome(s) for non-cooperator(s).

(A) Prisoner's Dilemma

	Player 2 cooperates		Player 2 does not cooperate	
	Player 1 gets	Player 2 gets	Player 1 gets	Player 2 gets
Player 1 cooperates	30	30	15	35
Player 1 does not cooperate	35	15	20	20

(B) Public Goods Game

	Number of **other** players who cooperate			
	3	2	1	0
Player 1 cooperates	60, 60, 60, 60	45, 45, 45, 65	30, 30, 50, 50	15, 35, 35, 35
Player 1 does not cooperate	65, 45, 45, 45	50, 30, 30, 50	35, 15, 35, 35	20, 20, 20, 20

Table 2: Predictions for overall, within-group, and between-group cooperation rates: Some examples.* Overall cooperation rate in the no-choice case is calculated assuming 50% of interactions take place within and 50% between groups (i.e., random matching).

		With Partner Choice				Without Partner Choice		
		% Cooperation			% Intergroup contact	% Cooperation		
Mean(θ^I)	Mean(θ^O)	Overall	Within	Between		Overall	Within	Between
.3	−.1	48	49	38	8	26	46	7
.3	.0	50	51	43	15	29	46	13
.3	.1	54	55	50	25	34	46	22
.3	.2	59	60	58	37	39	46	33
.2	−.2	34	35	24	8	18	33	4
.2	−.1	36	37	29	15	20	33	7
.2	.0	39	41	35	25	23	33	13
.2	.1	44	45	42	37	27	33	22

*(θ^I, θ^O) is normally distributed with $SD(\theta^I) = SD(\theta^O) = .3$ and $Corr(\theta^I, \theta^O) = .3$.

Table 3: Experimental design. Columns are between-subjects rows are within-subjects treatments. PD=Prisoner's Dilemma, PG=Public Goods Game.

Stage		Mixed	Choice	Isolated	Control
1		Inducing group identity			No group identity
2	No-feedback	10 PDs 5 PDs with ingroup 5 PDs with outgroup	10 PDs Partner's group based on choice	10 PDs All PDs with ingroup	10 PDs
3		5 Mixed PGs	5 Mixed PGs	5 Mixed PGs	5 PGs
4	Aggregate feedback	10 PDs 5 PD with ingroup 5 PD with outgroup	10 PDs Partner's group based on choice	10 PDs All PDs with ingroup	10 PDs
5		5 Mixed PGs	5 Mixed PGs	5 Mixed PGs	5 PGs
6		Post-experiment survey			

40

Table 4: Multilevel logistic regression models with random intercepts for subjects predicting cooperative behavior (1=Cooperate, 0=Defect) in the PD and PG.

Treatment	PD Coeff.	S.E.	PG Coeff.	S.E.
Control (reference)				
Isolated	−.73	(.76)	1.27	(.91)
Choice			−.45	(.87)
Choice ingroup	−.76	(.74)		
Choice outgroup	−1.39	(.77)		
Mixed			−1.19	(.86)
Mixed ingroup	−.65	(.74)		
Mixed outgroup	−2.17*	(.74)		
NoFeedback (NF)	−1.39*	(.26)	.17	(.42)
Isolated × NF	.36	(.30)	−1.47*	(.52)
Choice × NF			−1.27*	(.48)
Choice_ingr × NF	−.18	(.33)		
Choice_ougr × NF	.75	(.44)		
Mixed × NF			−.14	(.45)
Mixed_ingr × NF	.36	(.33)		
Mixed_ougr × NF	.73*	(.35)		
Constant	1.84	(.94)	3.24*	(1.13)
Control variables				
Period	−.28*	(.03)	−.49*	(.09)
Period×NF	.15*	(.04)	.32*	(.12)
Female	.86	(.51)	1.34*	(.60)
Right-wing	−.22*	(.11)	−.46*	(.13)
Age	.04*	(.02)	.035	(.02)
Random				
Ln(sig_u^2)	2.26*	(.17)	2.56*	(.20)
N(decision)	3240		1620	
N(Subject)	162		162	
Log-likelihood	−1400.60		−636.33	

* $p < .05$ for two-tailed tests.

Table 5: Tests of hypotheses as linear combinations (LC) of coefficients in Table 4.

Predicted contrast	LC of coefficients in Table 4 to be tested	Calculation of LC	(S.E.) and [z] of LC	Odds ratio: exp(LC)	Predicted sign of LC
PD					
Mixed ingr. vs. mixed ougr. (Feedback)	mixed ingr. − mixed ougr. =: A	−.65 − −2.17 = **1.53***	(.28) [5.41]	4.61	+ (H1)
Mixed ingr. vs. mixed ougr. (No Feedback)	(mixed ingr + mixed_ingr × NF) − (mixed ougr + mixed_ougr × NF) =: B	(−.65 + .36) − (−2.17 + .73) = **1.16***	(.27) [4.24]	3.18	+ (H1)
Choice ingr. vs. Choice ougr. (Feedback)	choice ingr. − choice ougr. =: C	−.76 − −1.39 = **.63***	(.34) [1.86]	1.88	+ (H2b)
Choice ingr. vs. Choice ougr. (No Feedback)	(choice ingr + choice_ingr × NF) − (choice ougr + choice_ougr × NF) =: D	(−.76 + −.18) − (−1.39 + .75) = **−.30**	(.36) [-.83]	.74	+ (H2b)
(Choice ingr. − Choice ougr.) vs. (Mixed ingr. − mixed ougr.) (Feedback)	$C − A$	.63 − 1.53 = **−.89***	(.44) [− 2.03]	.41	− (H2c)
(Mixed ingr. − mixed ougr.) vs. (Choice ingr. − Choice ougr.) (No Feedback)	$D − B$	−.30 − 1.16 = **−1.46***	(.46) [−3.21]	.23	− (H2c)
PG					
Isolated vs. Mixed (Feedback)	Isolated − Mixed	1.27− −1.19 = **2.46***	(.91) [2.70]	11.70	+ (H3)
Isolated vs. Choice (Feedback)	Isolated − Choice	1.27 − −.45 = **1.72***	(.91) [1.89]	5.58	+ (H3)
Choice vs. Mixed (Feedback)	Choice − Mixed	−.45 − −1.19 = **.74**	(.84) [.88]	2.10	+ (H3)

* $p < .05$ for one-tailed tests.

FIGURES

Figure 1: Average cooperation by treatment. The error bars indicate 95% confidence intervals obtained using robust standard errors. In the choice condition the average percentage of subjects who choose to play with an ingroup vs. an outgroup member is also indicted.