CROSSCUTTING CIRCLES IN A SOCIAL DILEMMA: EFFECTS OF SOCIAL IDENTITY AND INEQUALITY ON COOPERATION

Ozan Aksoy[1,2,3]

Abstract

I investigate experimentally the effects of heterogeneity in social identities and inequality in resources on cooperative behavior in a social dilemma. The experiment also varies the overlap between social identity and resources. The results show that both heterogeneity in social identities and inequality in resources reduce cooperation. When social identity and unequal resources are crossed orthogonally, the existence of composite group memberships mitigates the negative effects of inequality, but not of heterogeneity in social identities. When social identity and unequal resources overlap, cooperation decreases substantially among the participants with low resources, in both ingroup and outgroup interactions. My results also show that participants who are assigned to advantaged roles cooperate significantly less than those who are assigned to disadvantaged roles.

Keywords: cooperation, social dilemmas, heterogeneity, inequality, crosscutting circles

[1] UCL Institute of Education, Department of Quantitative Social Science, University College London, 20 Befrord Way, WC1H 0AL, UK; email: ozan.aksoy@ucl.ac.uk.
[2] Nuffield College, New Road OX1 1NF, Oxford

INTRODUCTION

Social dilemmas are situations in which there is a conflict between individual and collective interests (Dawes 1980; Kollock 1998). Collective action problems (Olson 1965), provision of public goods (Marwell and Ames 1979), problems of trust (Gambetta 1988) and cooperation (Taylor 1987) are classical social dilemma examples. In such situations, it is individually beneficial to not contribute to the provision of a collective good while enjoying the benefits of the good provided by someone else. Yet, if such individualistic behavior is widespread, the collective good is bound to decay. The ability of a society to motivate its members to voluntarily cooperate characterizes the extent of social cohesion within that society (Coleman 1990; Roca and Helbing 2011).

Social dilemmas lie at the core of many important social phenomena. Hence, sociologists have long been conducting research into understanding factors that affect behavior in social dilemmas (recent studies include Irwin and Simpson 2013; Kuwabara et al. 2014; Harrel and Simpson 2015; Paxton and Glanville 2015; Diekmann and Przepiorka 2016; Melamed and Simpson 2016; Dijkstra and Bakker 2017; Simpson and Aksoy 2017). A particularly important open question is how social identity and inequality simultaneously affect cooperation. Research has shown that diversity in the social identities of actors involved in the social dilemma situation generally hampers cooperation (e.g., Yamagishi and Kiyonari 2000; Simpson 2006; Habyarimana et al. 2007; Aksoy 2015). Inequality between actors with respect to resources is also shown to reduce cooperation (e.g., Buckley and Croson 2006; Aksoy and Weesie 2009). The combined effect of social identity and inequality is, however, understudied.

Studying the combined effects of inequality and social identity is nontrivial. When inequality is one of the cleavages that crosscuts group identities, its effects can propagate through various channels. Firstly, inequality can serve as yet another social dimension on which identities can be formed (e.g. identification with one's social class). This, in turn, may affect cooperation just like any other form of group identity (e.g. ethnicity). Secondly, inequality can affect cooperation simply through differences in material resources (in the language of the social dilemma literature, through differences in payoffs). Disentangling these different effects of inequality, and isolating them from the effects of other forms of social identities are important but difficult to accomplish using observational data.

In this study, by inducing social identity, inequality, and the overlap between the two experimentally, I investigate the net and combined effects of the two on cooperative behavior and disentangle the social identity aspects of inequality from differences in material resources. To my knowledge this is the first experimental study that couples social identity and inequality, and varies the overlap between the two.

Sociologists have long conceptualized societies as a web of crosscutting circles. Edward Alsworth Ross (1921) conveys the core idea very vividly (also see Mullen et al. 2001):

"Suppose at a given moment there is a certain strain along the line between Christians and Jews. If now, a strain appears along a quite different line, e.g., that between employers and workmen, the religious opposition will be less intense. For Jewish bosses and Jewish workmen will be estranged; likewise Christian bosses and Christian workmen. On the other hand, Jewish and Christian capitalists will recognize that they are "in the same boat," while Jewish workers and Christian

workers will sympathize with one another as brother victims of exploitation.... Each new cleavage contributes to narrow the cross clefts so one might say that society is sewn together by its inner conflicts."

This idea of cross-cutting circles is developed further in the seminal works of Blau (1977) and Blau and Schwartz (1984) and applied to various macro-sociological contexts such as positive and negative intergroup relations. Blau (1977) and Blau and Schwartz (1984) show that the prevalence of crosscutting circles—most importantly ethnic identity and socioeconomic status—increases the frequency of intergroup relations, both in the positive (e.g., interethnic marriages and friendships) and the negative (e.g., intergroup victimization) sense. I expand this macro-sociological literature to the study of a social dilemma. Importantly, however, I also contribute to this literature by providing more accurate, empirically calibrated micro-behavioral mechanisms. For example, in their macro models, Blau and Schwartz (1984) assume that individuals favor their ingroup, and the extent of this ingroup favoritism is stable for all dimensions, be it ethnic or religious identity or social class. In addition, Blau and Schwartz (1984, p. 90) assume that "the influence of every dimension on social differentiation is partly independent of that of any other dimension". In other words, Blau and Schwartz assume that there is no interaction between different dimensions of social affiliations. While these serve as fine micro-level assumptions to start with, there are strong reasons to believe that actual micro-behavioral mechanisms in social dilemmas differ in a systematic and predictable way.

For example, social psychologists have shown that when different forms of social identities are crossed, they may suppress or accentuate the salience of each other (Mullen, Migdal, and Hewstone 2001). Social psychologists have also shown that when actors are differentiated in terms of resources, status differences may have asymmetric effects: people from

upper classes are shown to behave less prosocially than people from lower classes (Piff et al. 2012, 2015). These systematic differences in micro-behavioral motives between members of different status groups would also create systematic asymmetries at the macro-level. My study develops a set of empirically calibrated micro-behavioral mechanisms in the context of a social dilemma which may inform the study of macro-level outcomes.

THEORY AND HYPOTHESES

Social dilemmas

Diekmann and Przepiorka (2016) define social dilemmas as: "... a situation of strategic interdependence in which the decisions of individually rational actors lead to an inferior outcome for all or some parties than the decisions or "collectively rational" actors. Collective rationality means that actors, had they an opportunity to communicate and agree on a binding contract, should agree on a combination of actions leading to a welfare-enhancing outcome." (p.1311). Consider the Symmetric Prisoner's Dilemma Game (PD) in Table 1A. In this game, mutual cooperation (CC) yields an outcome of 30 for each player which is collectively the best outcome. But defecting yields a higher outcome for each player, whether the other player cooperates or defects. If both actors follow this individually selfish and rational strategy, they both defect (DD) which yields collectively the worst distribution. The difference between the outcomes produced by CC versus DD represents the tension between collective and individual rationality.[4]

---

[4] In some social dilemmas, the collective rationality dictates asymmetric outcomes, that is, some actors bear the costs of producing the public good while others free-ride (e.g. in the pure strategy equilibria of the Volunteer's Dilemma (Diekmann 1985) and the step-level Public Good Game (e.g. Dijkstra and Bakker 2017).

[Table-1-about-here]

Effects of group identity and inequality on cooperation:

Social Identity Theory (Tajfel 1970; Tajfel et al. 1971) predicts that actors categorize the world into "us", the ingroup (IG), versus "them", the outgroup (OG). This simple categorization helps actors cope with the cognitive complexity of dealing with a diverse world. Identification with the ingroup also serves as a source of self-enhancement. When the group identities of the interacting actors are made salient, members of the ingroup are perceived and treated more favorably than members of the outgroup. In ingroup interactions, the self expands to others in one's ingroup and the outcomes of ingroup members are perceived as interchangeable (Tajfel 1970; Tajfel et al. 1971; Simpson 2006). This implies that when group identities are made salient, actors behave more pro-socially toward ingroup members than toward outgroup members.[5] Thus, in a Prisoner's Dilemma, compared to outgroup interactions, in ingroup interactions there will be more actors who will follow the collective rationality path and cooperate. Of course, even in ingroup interactions there will be some actors who will follow the selfish and individually rational strategy of defection, but the proportion of such actors will be lower in ingroup interactions than in outgroup interactions. I thus predict that:

---

[5] When group identities are not salient one expects only weak forms of ingroup favouritism; in addition, whether ingroup cooperation has negative or positive externalities for the outgroup may affect the level of ingroup favoritism (Dijkstra 2013). In my design, group identities are salient and ingroup cooperation has no externalities for the outgroup.

H1a: The level of cooperation will be higher in ingroup interactions than in outgroup interactions.

This hypothesis is depicted in Figure 1A. The contrast X in Figure 1A represents the extent of ingroup bias with respect to cooperative behavior. Note that in Figure 1A the y-scale is arbitrary: at this point I do not have any prediction about the level of cooperation. All of my predictions are about *differences* between conditions.

[Figure-1-about-here]

Let's introduce another dimension of differentiation, namely socioeconomic status, and for simplicity assume that there are only two "classes": U and L. For now let's ignore social identities mentioned above. When two Ls (Us) meet, I will assume that they play the symmetric PD in Table 1A (Table 1B) with low (high) outcomes. These interactions will be referred to as "in-class" (IC) interactions. When a U and an L meet, I will assume that they play the asymmetric PD in Table 1C with U as the row player and L as the column player. Such interactions will be referred to as "out-class" (OC) interactions.[6] I predict that the level of cooperation will be lower in OC interactions than IC interactions. This is because of two reasons. *Firstly*, class can serve as another dimension of social identity, that is, people may identify with their class. Then, the same prediction in H1a will apply.

Secondly, even if I assume that identification with the class is absent, just because of the asymmetry in outcomes in the PD in Table 1C, one expects differences in cooperation rates in

---

[6] Note that all PDs in Figure 1, symmetric and asymmetric, retains the game-theoretic structure of a social dilemma. That is, there is a tension between individual and collective interests, defecting is a dominant strategy for both actors, and CC yields the Pareto optimal outcomes.

OC interactions compared with IC interactions (Aksoy and Weesie 2009; Aksoy and Weesie 2013; Diekmann and Przepiorka 2016). This is because in an asymmetric PD, the mutually cooperative outcome CC yields inequalities in outcomes (i.e., 60 vs. 30, see Table 1C). If at least some people dislike inequality, then the CC strategy combination in an asymmetric PD will be less attractive than the CC combination in a symmetric PD. Hence, cooperation will be lower in the asymmetric PDs than in the symmetric PDs.

To formalize this idea, take for instance Fehr and Schmidt's (1999) inequality aversion model. According to this model actors dislike inequality. Fehr and Schmidt distinguish two types of inequality: advantageous inequality in which self gets more than others and disadvantageous inequality in which self gets less than others. According to Fehr and Schmidt actors dislike both types of inequalities, but more so the latter type. Appendix A shows a formal analysis of the three PDs in Table 1 assuming inequality aversion a la Fehr and Schmidt. In the two symmetric PD, mutual cooperation can be equilibrium if both actors dislike advantageous inequality sufficiently strongly. In the asymmetric PD, however, mutual cooperation can never be equilibrium. The L (column) player always prefers to defect no matter what the U (row) player does (i.e. defection is a dominant strategy for L). Some U players may cooperate, even in the asymmetric PD, but these players need to have a special and highly unlikely combination of inequality aversion parameter values (see Appendix A for the exact condition). Shortly, cooperation will be lower in the asymmetric PD than in the two symmetric PDs because both U and L players, and particularly L players, will be less likely to cooperate in the asymmetric case than in the symmetric case. Hence, I predict that:

H1b: The level of cooperation will be higher in "in-class" interactions than in "out-class" interactions.

This hypothesis is also depicted in Figure 1A. The contrast Y in Figure 1A represents the extent of "in-class" bias. I do not make any prediction on whether Y will be larger or smaller than X. Whether X or Y is larger depends on how strongly actors identify with their class versus with their social group, and how large the asymmetry in the PD outcomes relative to actors' inequality aversion is.

Combined effects of group identity and inequality on cooperation:

Let's now cross group and class orthogonally. In our simple setup, there will then be four group-class combinations: actors from, say group A and B, and from class U and L, viz. AU, AL, BU, and BL. Orthogonality of group and class implies that the number of people who belong to each group-class combination will be equal. When two upper class members meet (AU vs. AU or AU vs. BU or BU vs. BU) they play the symmetric PD with high outcomes (Table 1B). When two lower class members meet (AL vs. AL or AL vs. BL or BL vs. BL) they play the symmetric PD with low outcomes (Table 1A). When an upper class and a lower class member meet (AU vs. AL or AU vs. BL or AL vs. BU or BU vs. BL), they play the asymmetric PD (Table 1C).

What do we expect in these conditions? The simplest prediction is an *additive pattern*: the effects predicted by H1a and H1b will simply be added to each other when the two dimensions are crossed. This is shown in the four right-hand bars in Figure 1A. The IG-OG and the IC-OC contrasts are simply added in the dual differentiation case. In fact, the macro-sociological models of Blau (1977) and Blau and Schwarz (1984) start with this simple prediction that different dimensions of social differentiations have independent effects on social cohesion. In the social psychological literature, this additive pattern is sometimes referred to as

the "category differentiation" model (see Mullen et al. 2001; Crisp, Walsh, and Hewstone 2006; Crisp and Hewstone 2007). This simplest prediction constitutes my Hypothesis 2a:

H2a: When group and class are crossed, their effects on cooperation will be additive (i.e. in Figure 1A: $X = X'$, $Y = Y'$, and so $X + Y = X' + Y' = K$).

To clarify what exactly H2a means, it is helpful to note that it implies all of the following. When two actors from the same identity group (say two As) meet, the difference in cooperation levels between when both actors are from the same class (say AU and AU) versus from different classes (say AU and AL) will be *the same as* the difference in cooperation levels between a situation in which actors without any salient group identity are from the same class (say U and U) versus from different classes (say U and L). Similarly, H2a implies that when two actors from the same class meet, the difference in cooperation rates when they are from the same identity group (say UA and UA) versus different identity groups (say UA and UB) will be *the same as* the difference in ingroup (say A and A) versus outgroup (say A and B) cooperation levels when inequality is not a salient division. Finally, the difference between cooperation levels when the interacting actors are from the same group and the same class (say UA and UA) versus when they are from a different group and a different class (say UA and LB) will be *the sum of* the difference between ingroup and outgroup (without any salient class), and in-class and out-class (without any salient group identity) contrasts.

While H2a serves as a natural starting point, social psychological studies on crossed group affiliations show a more complex pattern (Crisp et al. 2006; Crisp and Hewstone 2007). When two affiliations are crossed, in composite groups differentiation in one dimension will be counteracted by similarity in the second. This will weaken category distinctiveness in composite

groups which share a common dimension. Hence, bias should be lower in crossed composite groups compared to the cases in which there is only one dimension of differentiation. In addition, when an actor has to consider two distinct dimensions of differentiation instead of one, the cognitive processing related to each dimension diminishes (Hogg and Abrams, p.723). This means that the two dimensions of differentiation will suppress the effects of each other. This suppression is sometimes referred to as "category conjunction" or "composite bias" in social psychology (Mullen et al. 2001). Hence, alternative to H2a, one can predict:

H2b: When group and class are crossed, they will suppress each other's effect on cooperation (i.e., in Figure 1A: $X' < X$ or $Y' < Y$, and so $X + Y > X' + Y' = K$).

H2b implies the following. When two actors from the same identity group (say two As) meet, the difference in cooperation levels between when both actors are from the same class (say AU and AU) versus from different classes (say AU and AL) will be *lower* than the difference in cooperation levels between a situation in which actors without any salient group identity are from the same class (say U and U) versus from different classes (say U and L). Similarly, H2b implies that when two actors from the same class meet, the difference in cooperation rates when they are from the same identity group (say UA and UA) versus different identity groups (say UA and UB) will be *lower than* the difference in ingroup (say A and A) versus outgroup (say A and B) cooperation levels when inequality is not a salient division. Finally, the difference between cooperation levels when the interacting actors are from the same group and class (say UA and UA) versus when they are from a different group and a different class (say UA and LB) will be *lower than* the sum of the difference between ingroup and outgroup (without any salient class), and in-class and out-class (without any salient identity group) contrasts.

Effect of inequality: outcome or identity based?

When I formulated H1b above, I did not disentangle the two mechanisms that would predict a negative effect of inequality on cooperation: (1) identification with one's class and (2) asymmetries in payoffs. Once, however, inequality and another dimension of identification is crossed, one can disentangle the two mechanisms by invoking the suppression effect predicted by H2b. Consider Figure 1A. If contrast Y (the difference between IC and OC) is purely outcome based, that is, it occurs only because of the differences between the outcomes in the asymmetric and the symmetric PD, then we would observe the same contrast when class is crossed with group identity (i.e. in Figure 1A, Y = Y'). This is because even when class is crossed with group identity, the outcome differences between IC and OC interactions will be the same as the outcome differences between OGIC and OGOC as well as between IGIC and IGOC interactions.

Alternatively, if the contrast Y is mainly due to identification with one's class rather than being purely outcome based, then I predict that the effect of class on cooperation will reduce when it is crossed with group identity. This is due to the suppression mechanism predicted in H2b: if the effect of inequality is mainly due to identification with one's class (and H2b holds), I then expect to find that the effect of class on cooperation will be higher when class is induced alone than when it is crossed with group identity (i.e., in Figure 1A: Y > Y').

A as a third possibility, Y could be equal to Y' even though the effect of inequality is mainly due to identification. In this case, identification with one's class could override other forms of social identification, and whether it is induced alone or crossed with other forms of identities would not matter. In this case, thus, inequality works at least partially through identification with one's class but the suppression effect predicted in H2b does not work. If this is indeed the case, I will not be able to disentangle the social identity aspects of inequality from

asymmetries in payoffs. The disentanglement of the two mechanisms requires the suppression

mechanism predicted in H2b to work.[7]

Overlap between class and identity:

The degree of overlap between different dimensions of differentiation is another important factor

that may affect cooperative behavior. Blau (1977) and Blau and Schwarz (1984) make a

distinction between consolidation versus intersection. Consolidation vs. intersection refers to

concomitant variations of different dimensions, that is, the extent to which membership to

different dimensions of differentiation overlaps. In the social psychological literature, the same

distinction is referred to as convergence vs. crosscutting (Crisp and Hewstone 2007). In the

organization science literature crosscutting is referred to as crisscrossing and the existence of

crisscrossing actors in diverse teams is shown to reduce the negative effects of faultlines (Mäs et

al. 2013). We also know from observational studies that heterogeneity in social identities

decreases social cohesion particularly when identities and resources overlap (Sturgis et al. 2011;

Ziller 2015). The mechanisms that underlie this finding, however, remain unclear as discussed

next.

Using our simple setup with two groups A and B, and two classes U (upper) and L

(lower), the results of an increase in the overlap between group identity and resources can be

expressed as the following. As the overlap increases (that is, social identity and class are more

consolidated or converged), one of the groups, either A or B, will contain more Ls and fewer Us

whereas the other group will contain fewer Ls and more Us. In other words, as the overlap

increases, there will be a stronger correlation between an individual's group and his/her social

---

[7] I thank a reviewer for bringing up this possibility.

class. For instance, when the overlap between identity and resources is minimal, the proportion

of individuals belonging to the AU, AL, BU, and BL combinations will be roughly equal. When

the overlap increases there will be, say, more AUs than ALs, and fewer BUs than BLs. If

interactions between individuals take place more or less randomly, this further implies that as the

overlap increases double ingroup (IGIC) and double outgroup (OGOC) interactions will be more

prevalent, and crossed group (OGIC and IGOC) interactions will be less prevalent.[8] IGIC

interactions will be more common because AUs will be matched more frequently with other AUs

than ALs, and BLs with other BLs than BUs simply because as the overlap between group and

class increases there will be fewer ALs and fewer BUs. For the same reason, OGOC interactions

will be more common as AUs will be matched more frequently with BLs than with BUs, and

BLs with AUs than ALs.

This change in the distribution of individuals to group-class combinations may, in turn,

affect cooperation. As discussed above, an increase in the overlap between group and class

implies that OGOC interactions will be more common than IGOC and OGIC interactions. One

expects that cooperation will be lower in OGOC interactions compared to both IGOC and OGIC

interactions. Hence, the overall level of cooperation may decrease as overlap increases. At the

same time, however, as overlap increases, the prevalence of IGIC interactions will be more

common than IGOC and OGIC interactions. Because cooperation will be higher in IGIC

interactions compared to IGOC or OGIC interactions, an increase in the overlap will also affect

---

[8] See Aksoy (2015) for a case when matching is not random but endogenous. If matching is endogenous, IGIC
interactions will become even more frequent as the overlap between group and class increases.

overall cooperation levels positively. Hence, the total effect of the overlap between class and group on overall cooperation is unclear.

This can be expressed more easily using Figure 1A. Let's consider two cases: *consolidation* (or equivalently convergence) and *crosscutting* (or equivalently intersection). In the consolidation case, all members of Group A will be of U class and all members of Group B will be of L class (or all Bs will be Us and As will be Ls). In the crosscutting case, half of As and Bs will be Us and other halves will be Ls. If actors are matched approximately randomly, then in the crosscutting condition, the overall level of cooperation will simply be the average of the four bars in the most right-hand side of Figure 1A (IGIC, IGOC, OGIC, and OGOC). In the consolidation condition (assuming that the overlap between class and group *per se* does not further affect individual motives such as the level of ingroup bias) the overall level of cooperation will then be the average of the IGIC and OGOC bars. One can show formally that the average of the IGIC, IGOC, OGIC, and OGOC bars are exactly equal to the average of the IGIC and OGOC bars (both of which are equal to the height of the OGOC bar plus (X'+Y')/2). Consequently, in our simple setup one does not expect an effect of the overlap between group and class on the overall level cooperation:

H3a: a change in the overlap between group identity and class will not affect the overall level of cooperation.

Note that this hypothesis is inevitably at the macro-level, as opposed to other hypotheses which involve predictions on individual-level behavior. This is because the overlap between identity and class is a macro-level phenomenon, and so is average cooperation.

In addition to shifting the distribution of individuals to group-class combinations, the overlap between group and class will also create status differences between the groups. In our example above, as the overlap increases the group containing more Us will, on average, have higher resources. This, in turn, may diminish ingroup favoritism, particularly among the members of the low status group. Status Construction Theory (see Ridgeway 2013 for an overview, and Ridgeway and Erickson 2000) predicts that when there are differences between two social groups (e.g. men and women) with respect to tangible resources, these differences may lead to widely accepted status beliefs about these groups. These widely accepted status beliefs are formed from people's perceptions about the population composition in terms of group identities and distribution of resources. These perceptions, in turn, could be formed through personal interactions and learning from others. The extent to which these beliefs and perceptions are accurate is less of an issue once these beliefs are formed, for these beliefs and perceptions would be self-reinforcing.

These status beliefs, in turn, "blunt the in-group bias of lower status group members who are torn between sticking with their own or favoring those from high-status groups." (Ridgeway 2013, p.6). Hence, if class and group identity overlaps, one may predict that lower class members will display lower levels of ingroup favoritism. This will, in turn, reduce the level of cooperation in IGIC interactions among the disadvantaged individuals as the overlap between group and class increase:

H3b: the higher the overlap between group identity and class, the lower the level of cooperation in double ingroup interactions (IGIC) among the low class members (see Figure 1B).

To summarize the discussion leading to H3a and H3b, if the overlap between identity and resources does not further affect individual-level motives, such as the level of ingroup favoritism, it will not affect the overall level of cooperation (H3a). This is because the reduction in cooperation due to an increase in the frequency of OGOC interactions as a result of an increase in overlap will be countered by the increase in cooperation due to an increase in the frequency of IGIC interactions. If, on the other hand, an increase in the overlap between identity and resources reduce ingroup sentiments among the disadvantaged individuals, this will in turn reduce cooperation among the members of the group which has more disadvantaged members (H3b). Consequently, if H3b holds, an increase in the overlap between identity and class will reduce the overall level of cooperation.

The rich effect

A final social psychological mechanism that is expected to operate in addition to aforementioned mechanisms as soon as class is one of the dimensions of differentiation is what I will simply call the "rich effect". Research has shown that compared with people from lower classes, people from upper classes exhibit more selfish and less prosocial behaviors (Piff et al. 2015, 2012). Additionally, people from upper classes are also cognitively more self-focused, displaying lower levels of emphatic accuracy (Kraus et al. 2010). These effects are not mere spurious correlations but causal. Simple experimental (i.e. random) manipulations of perceived class is enough to create these effects (Piff et al. 2010, 2012).

These effects of belonging to an upper class are often explained by the reasoning that higher resources, status, and independence cause actors to prioritize individual material outcomes over the collective welfare and recognize greed as beneficial and positive (Piff et al. 2012). In addition, lower class actors experience lower levels personal control and seek

cooperation with others to achieve favorable outcomes (Piff et al. 2010). Consequently, lower

class actors display higher levels of attention to the context and other people around, and behave

in ways that facilitate connection with similar others (Kraus et al. 2010). In other words,

compared with members of lower classes being a member of an upper class impairs prosocial

preferences and the desire to engage in cooperative relations with others. This yields the

following hypothesis:

H4: Members of advantageous classes will cooperate less than members of

disadvantageous classes.

Note that the "rich effect" is predicted to make advantaged individuals more selfish, both

in ingroup and in-class interactions. In a PD, an increase in selfishness will automatically

decrease cooperative behavior. Hence, I expect that compared with disadvantaged individuals,

advantaged individuals will cooperate less even in ingroup and in-class interactions.

I would like to discuss a final point about the "rich effect" before moving on to methods.

This effect is related to social class. Social class is typically acquired through socialization and

prolonged exposure to high or low resources in the form of wealth or long term income, i.e.

objective indicators of class (Adler et al. 2000). Obviously, the objective elements of social class

or wealth cannot be manipulated experimentally. They could be measured but this measure will

be confounded with many other unobserved factors. Social class, however, also has a subjective

dimension. This subjective dimension is related to one's perceptions of one's social rank in the

social environment (Adler et al. 2000) and could be manipulated experimentally (e.g. Piff et al.

2010, Kraus et al. 2010). Adler et al. (2000) show that subjective social class is as important as

objective indicators of class. I will assume that being assigned to an advantageous class in the

experiment improves temporarily a subject's perceptions of their social rank. These perceptions, in turn, result in the "rich effect", by reducing the social preferences of the subjects who are assigned to the advantageous class. Indeed, past research has shown that random experimental manipulations in the form of asking subjects to compare themselves with either an upper class or a lower class individual is enough to manipulate subjective class (Piff et al. 2010, Kraus et al. 2010, Piff et al. 2012).

METHODS

Participants

191 participants from quite heterogeneous backgrounds were recruited via the online recruitment system ORSEE. ORSEE is designed specifically for organizing behavioral experiments. It helps maintain a large participant database, allows to announce experimental sessions publicly through automated mailing. It also allows to recruit a random or non-random (e.g,. selected based on some criteria such as students or non-students) sample of participants from the database. Of the recruited participants, 43 percent were female, 34 percent were non-student residents of the city of Oxford, 54 percent were UK citizens. The 46 percent non-British participants were from more than 20 countries. The mean and the standard deviation of the age of the participants were 29.9 and 13.5, respectively.

Design and Procedure

The whole experiment was implemented on computers using the z-tree software. In each of the nine sessions of the experiment, participants were randomly seated in a cubicle of the CESS lab of Nuffield College, the University of Oxford. Participants collected tokens throughout the experiment. After completing the experiment, the participants' tokens were converted to British

Pounds (20 tokens = £1 = ~1.42 USD). On average, the participants earned £13.5 (~20 USD) (min £8, max £23).

The experiment comprised three between-subjects treatments, *crosscutting*, *consolidation*, and *control,* and a within-subjects treatment, *inequality*. Table 2 shows the overall experimental design and the number of participants per treatment.

[Table-2-about-here]

The crosscutting and the consolidation conditions were identical until stage 3. In both conditions, a group identity was induced in stage 1 of the experiment using the minimal-group paradigm (e.g., Tajfel et al. 1971; Yamagishi and Kiyonari 2000) following closely the standard procedure described in Aksoy (2015).[9] Participants were shown pairs of Kandinsky's and Klee's paintings. Depending on participants' relative preferences, a half were classified as Kandinskys and the remaining half as Klees. A participant's identity as a Kandinsky or Klee remained the same throughout the experiment. In the control condition, participants were shown the same set of paintings but group identity was not induced.

In stage 2, the participants played 12 symmetric Prisoner's Dilemma Games (SPD) (see Table 1A). In the crosscutting and consolidation conditions, in six of these 12 SPDs the interaction partner was an ingroup member and in the remaining six an outgroup member. The group of the interaction partner was randomly selected in the first round and then alternated in every remaining round. Moreover, in every round a new interaction partner was randomly selected from the in- or

---

[9] While I follow some of the procedures described in Aksoy (2015), the experiment reported in this article is an independent experiment with new data.

the out-group. In the control condition there was no group, so participants played 12 SPDs, each round with a randomly selected participant.

In stage 3, inequality was induced. Participants received a series of randomly created 30-digit numbers and were instructed to report how many of a certain digit were in each of these 30-digit numbers. The task continued for two minutes and the number of correct answers was stored. In the control condition, depending of the number of correct answers, the top half of the participants were classified as Advantaged and the bottom half as Disadvantaged. In the crosscutting condition, the top half of Klees and the top half of Kandinskys were classified as Advantaged and the bottom halves of Klees and Kandinskys as Disadvantaged. In the consolidation condition, the average number of correct answers given by Klees and by Kandinskys were compared. All members of the over-performing (under-performing) group was classified as Advantaged (Disadvantaged). In other words, in the consolidation condition, the group identity (Klee versus Kandinsky) and classification (Advantaged versus Disadvantaged) were perfectly correlated.

Participants were fully informed about those procedure through which they were classified into Advantaged and Disadvantaged classes. Hence, the procedures to assign resources as well as the overlap between classification and group identities were common knowledge. In addition, in all conditions, each participant was informed about his/her score in the real-effort task. In the control condition the participant was further informed on whether his/her score was above the median score. In the cross-cutting condition, the participant was informed on whether his/her score was above the median score among Klees if participant was a Klee or among Kandinskys if participant was a Kandinsky. In the consolidation treatment, the participant was informed on whether the average scores was higher among Klees or among Kandinskys.

This particular real-effort task was chosen to induce inequality for the following reasons. It had been generally perceived that one's class was achieved rather than being immutable or assigned through ascription (Aksoy and Weesie 2009; Ridgeway 2013). I, thus, wanted to keep the notion in the experiment that a participant's class was perceived as achieved. Moreover, even when class was assigned through another (e.g. random) mechanism, advantageous participants, but not disadvantageous participants, might anyways perceive their position as an outcome of their own success, albeit in a less controlled way (Piff et al. 2012). On the other hand, I did not want to induce inequality using a task that would sort certain participants into certain classes. For instance, if I used cognitive ability or IQ to assign classes, participants in the Advantageous position would then have better cognitive abilities or higher IQs which in turn might affect cooperation. The particular real-effort task I used (counting digits) did not require any specific skill and as reported in the results section did not correlate with cooperation.

After inequality was induced, participants played another 12 PDs. In all three experimental conditions, six of these 12 PDs were played with members of the same classification (i.e. in-class) in which case the PD was a symmetric PD, and the other six were played with members of the other classification (i.e. out-class). The Disadvantaged-Disadvantaged combination played the same SPD used in stage 2 (Table 1A). The Advantaged-Advantaged combination played the same SPD used in stage 2 but all outcomes in the game was multiplied by two (Table 1B). The Advantaged-Disadvantaged (i.e. out-class) combination played the asymmetric PD shown in Table 1. In the crosscutting and consolidation conditions, six of the 12 PDs were played with an ingroup member and the other six with an outgroup member. The crosscutting condition included all possible four combinations of group and classification (i.e. ingroup & in-class, ingroup & out-class, outgroup & in-class, outgroup & out-class). Each combination was played three times. The

participants played each of the four combinations in an alternating order, and the combination in the first round was randomized. The consolidation condition allowed only two combinations (i.e. ingroup & in-class, outgroup & out-class). The participants played each of the two combinations in an alternating order and the combination in the first round was randomized.

RESULTS

Figure 2 displays the average level of cooperation when group and class are induced separately as well as when they are crossed in the crosscutting treatment (the results of the consolidation treatment are given in Figure 3).[10] Formal tests of the hypotheses are conducted using the (linear combinations) coefficients of a multilevel linear probability model with random effects for participants fitted to the experimental data.[11] This model takes into account dependency between 24 PD decisions nested in participants (Snijders and Bosker 2012), and also controls for decision round and an interaction of decision round with the inequality treatment (see Table 3). Because I predict differences between advantaged and disadvantaged participants after the inducement of inequality (H4), the parameters corresponding to the IC, OC, IGIC, IGOC, OGIC, and OGOC

---

[10] All data used in this paper and the code used to produce the results as well as experimental instructions, z-tree files, and raw experimental data could be downloaded from: https://osf.io/wkts8/.

[11] Robustness checks with a logit or probit link function or with a linear probability model with cluster-robust standard errors provide essentially the same results. Note that logit/probit parameters are more difficult to interpret. Moreover, because most of the probabilities in my experiment are between 0.2 and 0.5 (see Figure 2 and 3) a liner probability model is fine. The unadjusted mean estimates and their standard errors given in Figure 2 are based on a liner probability model with cluster-robust standard errors at the subject level which does not control for decision round. Parameter estimates from this model are given in Appendix B.

conditions are estimated separately for these two classes by including interactions of dummy indicators of class (advantaged and disadvantaged) and of the conditions. The results of this model are given in Table 3. Table 3 also includes two alternative model specifications, namely a linear probability and a probit model without random effects for subjects but with robust standard errors that take clustering within a subject into account. All p-values are based on two-tailed tests.

[Figure-2-about-here]

[Table-3-about-here]

In the baseline condition with no identity and inequality inducement, the average level of cooperation is about 35 %, which is very similar to the levels of cooperation observed in past studies (Sally 1995; Aksoy and Weesie 2014). In H1a it is predicted that cooperation will be higher in ingroup interactions compared with outgroup interactions. In line with this prediction, when group identity is induced alone (i.e. without crossing with inequality) cooperation is about 15 percentage points higher (X = .15) in ingroup interactions compared to outgroup interactions (see the IG vs. OG contrast in Figure 2) and this difference is statistically highly significant (χ^2 (1) = 71.85, p < 0.001).[12] When inequality is induced in the control condition without crossing it with group identity, as predicted in H1b, cooperation is about 14 percentage points higher (Y =

---

[12] As expected, there is no statistically significant difference between the consolidation and the crosscutting treatments before the inducement of inequality, both in terms of the average level of cooperation and in terms of the difference between ingroup and outgroup interactions (χ^2 (2) = 1.26, p = 0.53). Hence, data from these two sessions before the inducement of inequality are collapsed.

.14) in in-class interactions compared to out-class interactions (χ^2 (1) = 30.25, p < 0.001). This is a remarkably similar effect size compared to the IG-OG contrast. I have to add that this in-class vs. out-class difference is mainly due to disadvantaged participants (compare the IC-OC difference between advantaged and disadvantaged participants in Figure 2, .14 is the average of these two contrasts).[13]

In H2a an additive pattern is predicted: the effect of group identity manifested as the IG-OG contrast (X in Figure 1A) and the effect of class manifested as the IC-OC contrast (Y in Figure 1A) will remain the same when identity and class are crossed (i.e., X = X' and Y = Y' in Figure 1A). To formally test this hypothesis, I first compare the IG-OG contrast (X) with the IGIC-OGIC contrast (X') and with the IGOC-OGOC contrast (X') in the crosscutting treatment (see Figure 2). I then compare the IC-OC contrast (Y) with the IGIC-IGOC contrast (Y') and with the OGIC-OGOC contrast (Y') in the crosscutting treatment (see Figure 2). The results of those tests will effectively show whether the effect of group identity and the effect of class remain the same when they are induced alone compared to the case in which they are crossed. I perform these comparisons separately for the advantaged and disadvantaged participants.

Results show that the IG-OG difference is not statistically significantly different from the IGIC-OGIC and the IGOC-OGOC contrasts for both the advantaged (χ^2 (2) = 2.92, p = 0.23) and disadvantaged (χ^2 (2) = 1.57, p = 0.46) participants. In other words, there is not enough evidence

---

[13] The IC-OC contrast being significantly larger among disadvantaged participants than among advantaged participants does not refute H4 which predicts that advantaged participants will be less cooperative than disadvantaged participants. The finding here is on the *difference* between in-class and out-class interactions keeping class constant. H4 is about the difference in *levels* of cooperation between advantaged and disadvantaged classes.

against the prediction that the extent of ingroup bias is relatively stable when it is induced alone or when it is crossed with inequality. However, the IC-OC contrast is statistically significantly larger than the IGIC-IGOC and the OGIC-OGOC contrasts for disadvantaged participants (χ^2 (2) = 18.41, p < 0.001) which refutes H2a. This finding is in line with H2b. H2b predicts that when group and class are crossed, they will suppress each other's effect on cooperation. For disadvantaged participants, the effect of class on cooperation is significantly lower when class is crossed with group identity compared with the case in which class is induced alone. For advantaged participants, the same test yields a statistically insignificant result (χ^2 (2) = 0.28, p = 0.87).

To summarize, H2a is supported for advantaged participants, but not for disadvantaged participants. For disadvantaged participants, the alternative hypothesis H2b is supported. These findings can be seen clearly in Figure 2. The average cooperation is quite high in the IC condition and quite low in the OC condition for disadvantaged participants. However, as soon as group identity is crossed with class, this difference largely diminishes (i.e. the IGIC versus IGOC difference and the OGIC versus OGOC difference are very small compared to the IC versus OC difference for disadvantaged participants).

It is predicted in H3a that a change in the overlap between class and group will not affect the overall level of cooperation. To test this hypothesis, I compare cooperation levels in the crosscutting condition with those in the consolidation treatment. The overall level of cooperation in the crosscutting treatment is about 10 percentage points higher compared to the consolidation treatment and this difference is statistically significant (difference = 0.1, χ^2 = 4.16, p = 0.04). Hence, H3a is refuted. The overlap between identity and resources does affect cooperation significantly. Below I describe how.

H3b predicts that an increase in the overlap between class and group identity will reduce ingroup preferences among disadvantaged participants. To test this hypothesis, I compare the IGIC bar for disadvantaged participants in Figure 3 (consolidation condition) with that in Figure 2 (crosscutting condition). It is clearly seen that the cooperation rate is about *halved* in the consolidation condition in the IGIC interactions compared with the crosscutting condition (χ^2 (1) = 9.50, p = 0.002). This result strongly supports H3b. As an aside, the cooperation rate in the OGOC condition for disadvantaged participants is also significantly lower in the consolidation treatment compared to the crosscutting treatment (χ^2 (1) = 9.49, p = 0.002).[14] The cooperation rates of advantaged participants in the consolidation treatment are statistically indistinguishable from those in the crosscutting treatment (χ^2 (2) = 1.15, p = 0.56). To summarize, consolidation of group identity and class reduces significantly the overall level of cooperation. This decrease is mainly due to disadvantaged participants who reduce their cooperation when their identity also overlaps with lower class.

[Figure-3-about-here]

In H4, it is predicted that advantaged participants will cooperate less than disadvantaged participants. A meaningful and relevant comparison is looking at the cooperation rates of advantaged and disadvantaged participants in the IC condition (see Figure 2).[15] Disadvantaged

---

[14] Because cooperation with both ingroup/in-class and outgroup/out-class has decreased among disadvantaged participants, the IGIC-OGOC difference in the consolidation condition is not significantly different from the same difference in the crosscutting treatment (χ^2 (1) = 1.46, p = 0.23).

[15] In the IC condition, the PDs participants play is fully symmetric and the outcomes in the PD that advantaged participants play are just an affine transformation of the outcomes in the PD played by disadvantaged participants.

participants cooperate about more than twice as advantaged participants do in IC interactions (χ^2 (1) = 5.81, p = 0.016). In addition, as Figure 2 shows, advantaged participants cooperate generally lower than disadvantaged participants in all conditions of the control and crosscutting treatments. In fact, when I collapse the control and the crosscutting treatments after the inducement of inequality, I see that advantaged participants' average cooperation rate is about 10 percentage points lower than that of disadvantaged participants (χ^2 (1) = 4.01, p = 0.045). In the consolidation treatment, however, (see Figure 3) disadvantaged participants' cooperation rates do not differ significantly from advantaged participants' cooperation rates (χ^2 (1) = 1.36, p = 0.24).

Before concluding the results section, I would like to report a number of additional analyses. Firstly, there is no statistically significant difference between the cooperation rates of advantaged and disadvantaged *before* the inducement of inequality (difference = -0.05, χ^2 (1) = 1.30, p = 0.25). Hence, the effects of class on cooperation I document here are not due to selection, but due to the experimental manipulation. Moreover, there is no effect of a participant's score in the real-effort task (which was used to induce inequality) on cooperation, both before (coefficient = -.01, S.E. = .01, p = 0.310) and after (coeff. = -.00, S.E. = .01, p = 0.554) the inducement of inequality, conditional on class. In addition, there is no interaction between the score in the real-effort task and with the experimental treatment (χ^2 (2) = 2.53, p = 0.28).[16] Hence, I ascertain that the particular real-effort task used to induce inequality does not

---

Hence, under normal conditions no (linear) outcome based formal theory predicts any difference between advantaged and disadvantaged players in IC interactions (see Aksoy and Weesie 2013 and appendix A).

[16] One may expect that in the consolidation treatment, subjects who had a high-score in the real-effort task but became disadvantaged due to the low average performance of their group may become angry and cooperate less.

sort more or less cooperative participants to advantaged or disadvantaged roles. The treatment (control, consolidation, and crosscutting) has no significant effect on participants' scores in the real-effort task (F (2, 188) = 0.5, p = 0.6).

A final relevant result is the following. In the experiment, inequality is induced through merit with the real-effort task. A subject's score in the real-effort task is not correlated with cooperation in the PD (see above), so it does not sort more or less cooperative subjects. Nevertheless, by comparing subjects with a task score just around the median helps assess whether being assigned to the advantageous or disadvantageous classification through ascription matters. Subjects whose real-effort scores are just above and just below the median are almost equivalent in terms of merit, and falling into just above or below the cutoff is nearly random. In fact, in quite a few occasions more than one subject had a score exactly equal to the median. In those cases, the computer randomly assigned half of them to the advantageous class and the other half to the disadvantageous class. Subjects were fully informed on this procedure. Hence, a comparison of cases around the cutoff would show the effect of ascribed class on cooperation. Note, however, that comparing cases around the cutoff makes sense only in the control and the crosscutting conditions. In the consolidation condition, a subject can be in the advantageous or disadvantageous class irrespective of their score, because class is assigned based on the group means.

Figure 4 shows the association between cooperation and the score in the real-effort task. In Figure 4, the real-effort score is centered around the session and group median in the control

There is no evidence in the data for this as the interaction of the real-effort score with the consolidation treatment is statistically insignificant.

and crosscutting conditions, respectively and around the group mean in the consolidation condition. Additionally, the figure is based on data from the symmetric PDs only (i.e. in-class interactions), because in the asymmetric PDs there are payoff differences between advantaged and disadvantaged players which in itself could result in differences in cooperation rates. However, not discarding the asymmetric PDs (not-reported) yields qualitatively the same results.

[Figure-4-about-here]

The important finding in Figure 4 is the discontinuity at the cutoff in the control and crosscutting conditions (left panel). There is a strong and statistically significant drop in just above the cutoff compared to just below. In fact, if I compare cases immediately above and below the cutoff in the control and crosscutting conditions, being classified as advantaged reduces cooperation by about 38 percentage points (cluster robust standard error = 0.14, p = 0.013). The discontinuity in the consolidation condition, as expected, is not statistically significant. This shows that ascribed class, rather than merit or effort seems to affect cooperation.

CONCLUSIONS AND DISCUSSION

Sociologists have long conceptualized societies as a web of crosscutting circles (e.g. Ross 1921). Blau (1977) and Blau and Schwarz (1984) elaborated and applied this idea to various macro-sociological contexts such as positive and negative intergroup relations. In this article, I apply insights from this literature to the study of social dilemmas. There are studies that considered separately the effects of different dimensions of differentiation (such as social identity or inequality) on social dilemma behavior. But to my knowledge, this is the first study that couples group identity with inequality and varies the overlap between identity and inequality.

In my experiment, I first induce group identity and inequality separately, and then cross group and class with varying degrees of overlap. Studying inequality, group identity, and the overlap between the two simultaneously is important for the following reasons. Past research has shown that heterogeneity in social identities generally hampers social cohesion and cooperation. In most real-life situations, however, group identities and inequality overlap. For example, there are often status differences between ethnic groups and members of particular groups tend to have access to higher resources than members of other groups. Hence, in most observational studies, it is unclear whether heterogeneity in social identities or inequality between groups is worse for social cohesion, and whether different forms of heterogeneities interact. Because I manipulate inequality, identity, and the overlap between the two experimentally, I can identify their separate effects on cooperation.

My results replicate the well established findings that cooperation is much lower in outgroup interactions than in ingroup interactions, and in "out-class" interactions than in "in-class" interactions. More importantly, however, by crossing group with class I am able to show that relatively low levels of cooperation in out-class interactions is mainly due to actors' identification with their own classes, and to a lesser extent due to material differences between advantageous and disadvantageous actors.

My findings also suggest that a cross-cutting group identity may suppress the negative effect of class, particularly among disadvantaged actors. Following on Ross's quote in the introduction, in my experiment "workers" cooperate much more with other workers than with "capitalists" when there is no further cross-cutting circle. But as soon as there is an identity dimension that divides the workers and unites some workers with some capitalists, workers' "in-class" bias is significantly reduced. This gives way to the speculation that capitalists can use a

divide et impera strategy to preempt cooperation between workers by inducing arbitrary forms of identities that divide the workers. What is worse for workers, however, is that most of them also belong to the same social (e.g., ethnic) group. Rather than boosting cooperation, the consolidation of low status with a particular group identity destroys cooperation among low status actors.

Here, I would like to discuss a potential limitation of my study. In the experiment, social identity is induced with the minimal group paradigm based on subjects' rather arbitrary artistic preferences. Inequality, on the other hand, is induced with a real-effort task. Hence, class is based, at least partially, on merit. This may explain why identification with the class is particularly strong: the "achieved" nature of class may make identification with one's class "stickier" than identification with one's minimal group. One could, in principle, induce both social identity and inequality using merit. This would address the aforementioned limitation. I must add, however, that a comparison of subjects whose real-effort scores are just around the median score reveals an interesting pattern and shows that merit may not be a particularly strong element of class. Cases immediately around the median real-effort score are as-if randomly assigned to the advantaged versus disadvantaged class. A comparison of these marginal cases shows that even if there is no meaningful difference in terms of merit, being assigned to the advantaged versus disadvantaged class affects cooperative behavior significantly. This shows that even we keep merit constant, class matters for the subjects.

There is an important debate among social scientists on the potential negative consequences of diversity in social identities and inequality. This debate gained momentum as immigration into almost all Western societies increased dramatically in the last few years. While it has been shown that inequality and heterogeneity are both detrimental to social cohesion,

whether and to what extent they interact is unclear. My findings contribute to this debate. In a controlled experiment, I show, in line with previous research, that both inequality and heterogeneity in social identities between actors involved in the social dilemma reduce cooperation. When inequality and group identity are crossed, the existence of composite groups mitigates the negative effects of inequality. The consolidation of social identity and material resources (i.e., some identity groups being consistently worse off than others), however, poses the strongest threat to social cohesion. In this case, cooperation plummets among low status groups, in both ingroup and outgroup interactions, creating a state of anomy.

There are a number of caveats one should bear in mind in transporting the insights gained from the current study to the broader debate on diversity and social cohesion. In the current study, social identity is induced with the minimal group paradigm. The extent to which resource inequality interacts with real groups, such as ethnicity or religious affiliation is an open question. In addition, the subject pool the experiment is based on is not representative of a general population. Most, though not all subjects were university students. Finally, the outcome measure is cooperation in the Prisoner's Dilemma. While cooperation is an important determinant of social cohesion, social cohesion is a broader concept which has other determinants, such as generalized trust between actors and trust in institutions, network composition of actors, and the level of social capital. An experimental study such as the current one helps establish causality with tight covariate control and precise measurement. This often comes at a cost of abstracting from real-world complications, which may in turn hinder external validity. The results from this study, therefore, should be interpreted in combination with the results from studies that use other methods, such as observational analysis with representative samples. Indeed, some of the key results reported here resonates well with the results from observational studies.

An important finding of my study which resonates with a large body of observational and experimental studies is that when there are status hierarchies between actors, advantaged actors generally behave much less cooperatively, both in in-class, ingroup, outgroup, and out-class interactions, than disadvantaged actors do. This finding is in line the social psychological literature that consistently shows a negative effect of being a member of a higher socioeconomic class on pro-social behavior. My results suggest that inequality reduces the overall level of cooperation, in two ways. Firstly, it reduces cooperation between members of different classes. Secondly, it makes advantaged actors behave in a non-cooperative way. In other words, the negative effect of inequality on social cohesion is mediated partly by the antisocial behavior of the rich, not the poor.

REFERENCES

Adler, Nancy E., Elissa S. Epel, Grace Castellazzo, and Jeannette R. Ickovics. 2000. "Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy, White women." *Health Psychology* 19(6): 586.

Aksoy, Ozan. 2015. "Effects of Heterogeneity and Homophily on Cooperation. " *Social Psychology Quarterly* 78(4): 324-344.

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. 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.

Blau, Peter M. 1977. "Inequality and Heterogeneity." New York: Free Press.

Blau, Peter M., and Joseph E. Schwartz. 1984. "Cross-cutting social circles: Testing a macrosociological theory of intergroup relations." Orlando: Academic.

Buckley, E., and Croson, Rachel. 2006 "Income and wealth heterogeneity in the voluntary provision of linear public goods." *Journal of Public Economics*, 90(4): 935-955.

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

Crisp, Richard J., Judi Walsh, and Miles Hewstone. 2006. "Crossed categorization in common ingroup contexts." *Personality and Social Psychology Bulletin* 32(9): 1204-1218.

Crisp, Richard J., and Miles Hewstone. 2007. "Multiple social categorization." *Advances in Experimental Social Psychology* 39: 163-254.

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

Diekmann, Andreas. 1985. "Volunteer's dilemma." *Journal of Conflict Resolution* 29(4): 605-610.

Diekmann, Andreas, and Wojtek Przepiorka. 2016. "Take One for the Team!" Individual Heterogeneity and the Emergence of Latent Norms in a Volunteer's Dilemma." *Social Forces* 94(3): 1309–1333.

Dijkstra, Jacob, 2013. "Put your money where your mouth is: Reciprocity, social preferences, trust and contributions to public goods." *Rationality and Society* 25(3), 290-334.

Dijkstra, Jacob and Dieko M. Bakker. 2017. "Relative power: Material and contextual elements of efficacy in social dilemmas." *Social Science Research* 62: 255-271.

Fehr, Ernst, and Klaus M. Schmidt. 1999. "A theory of fairness, competition, and cooperation." *Quarterly Journal of Economics* 114(3): 817-868.

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

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.

Harrell, Ashley, and Brent Simpson. 2015. "The Dynamics of Prosocial Leadership: Power and Influence in Collective Action Groups." *Social Forces* 94(3): 1283–1308.

Hogg, M. A., & Abrams, D. (1988). *Social Identifications: A Social Psychology of Intergroup Relations and Group Processes*. London: Routledge.

Irwin, Kyle, and Brent Simpson. 2013. "Do descriptive norms solve social dilemmas? Conformity and contributions in collective action groups." *Social Forces* 91(3): 1057–1084.

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

Kraus, Michael W., Stéphane Côté, and Dacher Keltner. 2010. "Social class, contextualism, and empathic accuracy." *Psychological Science* 21(11): 1716-1723.

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* 77:344–60.

Marwell, Gerald, and Ruth E. Ames. 1979. "Experiments on the provision of public goods. I. Resources, interest, group size, and the free-rider problem." *American Journal of Sociology* 84(6): 1335-1360.

Mäs, Michael, Andreas Flache, Károly Takács, and Karen A. Jehn. 2013. "In the short term we divide, in the long term we unite: Demographic crisscrossing and the effects of faultlines on subgroup polarization." *Organization Science* 24(3): 716-736.

Melamed, David, and Brent Simpson. 2016. "Strong ties promote the evolution of cooperation in dynamic networks." *Social Networks* 45: 32-44.

Mullen, Brian, Michael J. Migdal, and Miles Hewstone. 2001. "Crossed categorization versus simple categorization and intergroup evaluations: a meta-analysis." *European Journal of Social Psychology* 31.6: 721-736.

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

Paxton, Pamela and Jennifer L. Glanville. 2015. "Is Trust Rigid or Malleable? A Laboratory Experiment." *Social Psychology Quarterly* 78(2): 194-204.

Piff, Paul K., Michael W. Kraus, Stéphane Côté, Bonnie Hayden Cheng, and Dacher Keltner. 2010. "Having less, giving more: the influence of social class on prosocial behavior." *Journal of Personality and Social Psychology* 99(5): 771-784.

Piff, P. K., Stancato, D. M., Côté, S., Mendoza-Denton, R., & Keltner, D. 2012. "Higher social class predicts increased unethical behavior." *Proceedings of the National Academy of Sciences*, *109*, 4086-4091.

Piff, P. K., Dietze, P., Feinberg, M., Stancato, D. M., & Keltner, D. 2015. Awe, the small self, and prosocial behavior. *Journal of Personality and Social Psychology, 108*, 883-899.

Ridgeway, Cecilia L. 2013. "Why status matters for inequality." *American Sociological Review* 79(1): 1–16.

Ridgeway, Cecilia L., and Kristan Glasgow Erickson. 2000. "Creating and Spreading Status Beliefs1." *American Journal of Sociology* 106(3): 579-615.

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.

Ross, Edward Alsworth. 1921. *Principles of Sociology*. Century: New York.

Sally, David. 1995. "Conversation and cooperation in social dilemmas a meta-analysis of experiments from 1958 to 1992." *Rationality and Society* 7(1): 58-92.

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

Simpson, Brent, and Ozan Aksoy. 2017. "Cumulative advantage in collective action groups: How competition for group members alters the provision of public goods." *Social Science Research* 66: 1-21.

Snijders, Tom, and Roel J. Bosker. 2012. "*Multilvel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*" 2nd ed, London: Sage.

Sturgis, P., Brunton-Smith, I., Read, S., & Allum, N. 2011. "Does ethnic diversity erode trust? Putnam's 'Hunkering Down' thesis reconsidered" *British Journal of Political Science*, 41(01): 57-82.

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.

Yamagishi, Toshio and Toko Kiyonari. 2000. "The Group as the Container of Generalized

    Reciprocity." *Social Psychology Quarterly* 63:116–32.

Ziller, Conrad. 2015. "Ethnic diversity, economic and cultural contexts, and social trust: Cross-

    sectional and longitudinal evidence from European regions, 2002–2010." *Social*

    Forces 93(3): 1211–1240.

APPENDIX

A: Inequality aversion in the symmetric and asymmetric Prisoner's Dilemma Game

Assume that actors transform the outcomes in the PD with the Fehr & Schmidt (1999) utility

function in equation 1 below where *x* is the outcome for the self and *y* is the outcome of the

other.

$$U(x, y; \beta_i, \alpha_i) = \begin{cases} x - \beta_i(x - y) \text{ if } x \geq y \\ x - \alpha_i(y - x) \text{ if } x < y \end{cases} \text{ with } 0 \leq \beta_i < 1 \text{ and } \beta_i \leq \alpha_i \quad (1)$$

In the model, it is further assumed that $0 \leq \beta_i < 1$ and $\beta_i \leq \alpha_i$. The particular PDs used

in this article and corresponding utilities transformed with the inequality aversion model are

given in Table A1.

Table A1. Prisoner's Dilemma outcomes and corresponding utilities calculated assuming Fehr

and Schmidt (1999) preferences.

A: Symmetric PD

	PD Outcomes		Transformed utilities	
	Cooperate	Defect	Cooperate	Defect
Cooperate	30, 30	15, 35	30, 30	$15 - \alpha_1 20, 35 - \beta_2 20$
Defect	35, 15	20, 20	$35 - \beta_1 20, 15 - \alpha_2 20$	20, 20

B: Symmetric PD (high)

	Cooperate	Defect	Cooperate	Defect
Cooperate	60, 60	30, 70	60, 60	$30 - \alpha_1 40, 70 - \beta_2 40$
Defect	70, 30	40, 40	$70 - \beta_1 40, 30 - \alpha_2 40$	40, 40

C: Asymmetric PD

	Cooperate	Defect	Cooperate	Defect
Cooperate	60, 30	30, 35	$60 - \beta_1 30, 30 - \alpha_2 30$	$30 - \alpha_1 5, 35 - \beta_2 5$
Defect	70, 15	40, 20	$70 - \beta_1 55, 15 - \alpha_2 55$	$40 - \beta_1 20, 20 - \alpha_2 20$

In the symmetric PD in both high and low version, mutual defection is an equilibrium.

This is because for any value of α and β U(D,D) is higher than U(C,D). In the symmetric PD (in both versions) mutual cooperation can be an equilibrium (assuming that actors know each other's β values) if:

$$\beta_i \geq 1/4 \text{ for all } i \in \{1,2\} \text{ (Condition 1)}$$

This is because when Condition 1 is satisfied, $U(CC) \geq U(DC)$ for both players.

In the asymmetric version, mutual cooperation is never an equilibrium. This is because for mutual cooperation to be equilibrium, U(CC) should be higher than U(DC) for the column player. This requires $30 - \alpha_2 30 \geq 35 - \beta_2 5$ that is, $\beta_2 - \alpha_2 6 \geq 1$ which is impossible under the assumption of $\beta \leq \alpha$. In fact, in the asymmetric game defection is a dominant strategy for the column player for $\beta \leq \alpha$. The row player cooperates if $U(CD) \geq U(DD)$. This happens if $30 - \alpha_1 5 \geq 40 - \beta_1 20$, that is:

$$\beta_1 \geq (2 + \alpha_1)/4 \quad \text{(Condition 2)}$$

Nothe that Condition 2 is much more stringent than Condition 1, because $0 \leq \beta_i < 1$ and $\beta_i \leq \alpha_i$. In other words, even the advantaged player is much less likely to cooperate in the asymmetric game than in the symmetric game.

The effect of social identity can also be captured using this inequality aversion model. For example, when both players are from the same group, β is higher and α is lower for both players.

B: Estimates of average cooperation and its cluster-robust standard error per condition

Table A2: Linear Probability Model with cluster robust standard errors predicting cooperation. A dummy for each condition is included and intercept is suppressed (which is not identified). Cluster (subject) robust standard errors in parentheses; * p (2-sided) < 0.05, ** p (2-sided) < 0.01.

	Coefficient	Robust S.E.
Baseline	0.350**	(0.042)
Inequality no identity		
IC (Ad)	0.172**	(0.048)
OC (Ad.)	0.150**	(0.043)
IC (Da)	0.402**	(0.065)
OC (Da)	0.144**	(0.044)
Identity (no inequality)		
IG	0.418**	(0.033)
OG	0.274**	(0.030)
Crosscutting (identity and inequality orthogonal)		
IGIC (Ad)	0.281**	(0.066)
IGOC (Ad)	0.240**	(0.061)
OGIC (Ad)	0.219**	(0.055)
OGOC (Ad)	0.167**	(0.049)
IGOC (Da)	0.344**	(0.072)
IGIC (Da)	0.427**	(0.076)
OGOC (Da)	0.229**	(0.062)
OGIC (Da)	0.250**	(0.066)
Consolidation (identity and inequality overlap)		
IGIC (Ad)	0.240**	(0.055)
OGOC (Ad)	0.172**	(0.047)
IGIC (Da)	0.206**	(0.055)
OGOC (Da)	0.064*	(0.031)
N(decision)	4584	
N(subject)	191	

Note: Tables and Figures are included below in the order they apper in text, i.e. Figure 1 comes as the first figure, Figure 2 comes as the second figure below etc.

TITLES OF TABLES AND FIGURES

Table 1. Symmetric and Asymmetric Prisoner's Dilemma Games. The numbers in the cells represent the outcomes for the row and column players, respectively.

Table 2. Experimental Design. SPD = Symmetric Prisoner's Dilemma, APD = Asymmetric Prisoner's Dilemma. IC = in-class, OC = out-class, IG = ingroup, OG = outgroup.

Table 3. Models predicting cooperation; Ad = Advantaged class, Da = Disadvantaged class; LPM = Linear Probability Model; + p (2-sided) < 0.1, * p (2-sided) < 0.05, ** p (2-sided) < 0.01.

Figure 1. Graphical representation of hypotheses. Panel A: hypothesis 1 and 2. Panel B: hypothesis 3b (hypothesis 3a predicts no difference between crosscutting and consolidation.). Panel A: hypothesis 4. IC = in-class, OC = out-class, IG = ingroup, OG = outgroup.

Figure 2. Average cooperation in the control group, when group and class are induced separately, and when group and class are crossed in the crosscutting treatment. Bars represent 95% confidence intervals calculated using standard errors that are robust to clustering within participants and non-normality. The dark dashed line represents the baseline average cooperation (no identity, no inequality) with fainter dashed lines representing 95% confidence interval. IC = in-class, OC = out-class, IG = ingroup, OG = outgroup.

Figure 3. Average cooperation by experimental condition in the consolidation treatment. Bars represent 95% confidence intervals calculated using standard errors that are robust to clustering within participants and non-normality. The dark dashed line represents the baseline average cooperation (no identity, no inequality) with fainter dashed lines representing 95% confidence interval. IC = in-class, OC = out-class, IG = ingroup, OG = outgroup.

Figure 4. Association between cooperation and the score in the real-effort task used to induce inequality, broken down by control and crosscutting condition on the one hand and consolidation condition on the other. Only symmetric PDs (i.e. L vs L and H vs H games) are included. The scores in the real-effort task is centered around the appropriate median or mean (i.e. session median in the control condition, group median in the crosscutting condition, and group mean in the consolidation condition). The lines are local polynomial smoothers with a bandwidth of three and its 95% confidence interval.

TABLES

Table 1. Symmetric and Asymmetric Prisoner's Dilemma Games. The numbers in the cells represent the outcomes for the row and column players, respectively.

A: Symmetric PD

	Cooperate	Defect
Cooperate	(30, 30)	(15, 35)
Defect	(35, 15)	(20, 20)

B: Symmetric PD (high)

	Cooperate	Defect
Cooperate	(60, 60)	(30, 70)
Defect	(70, 30)	(40, 40)

C: Asymmetric PD

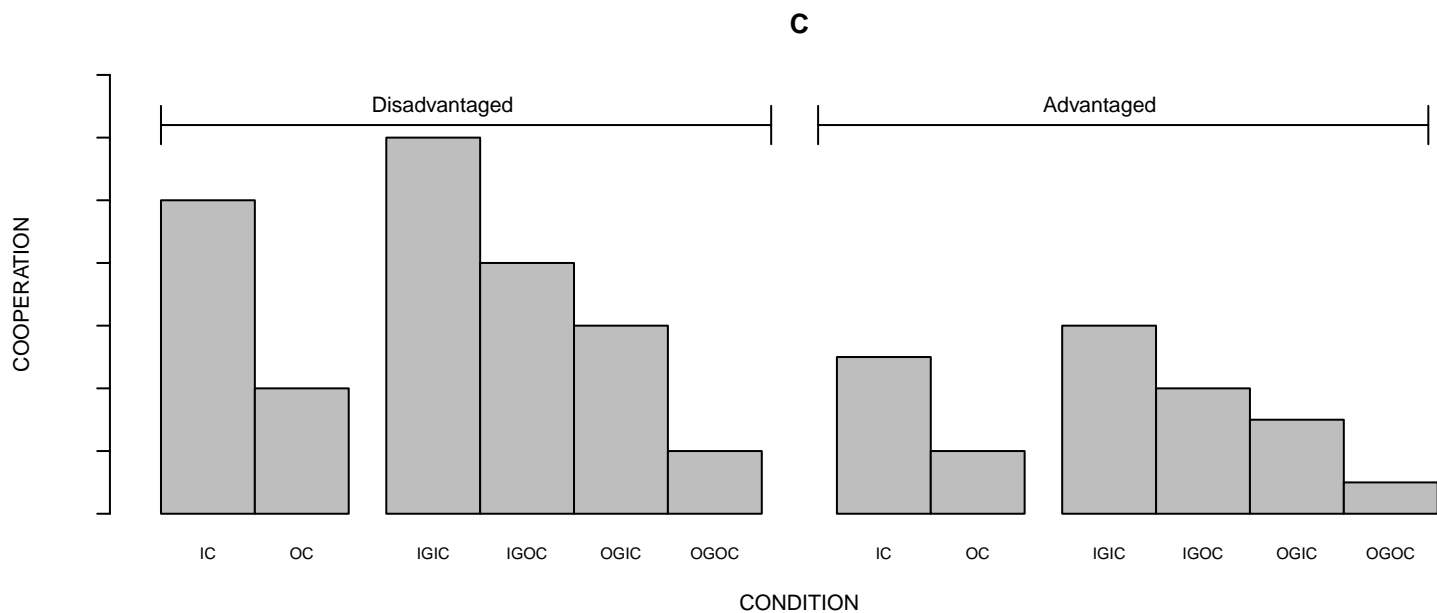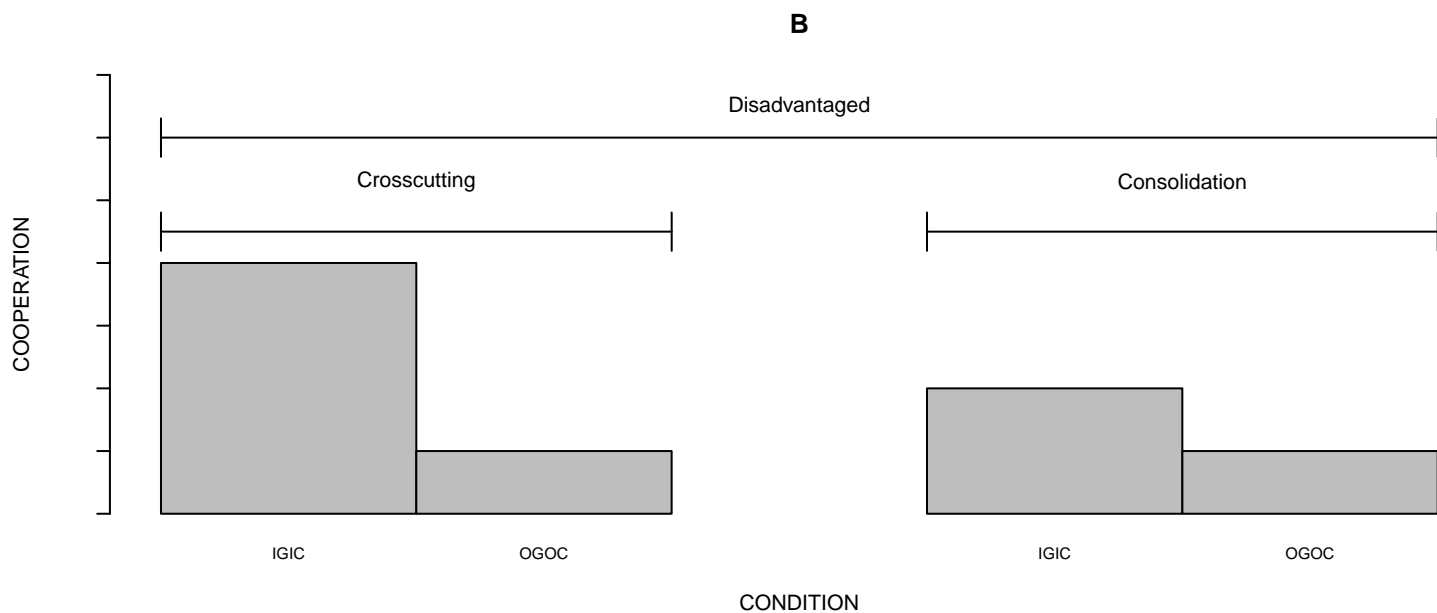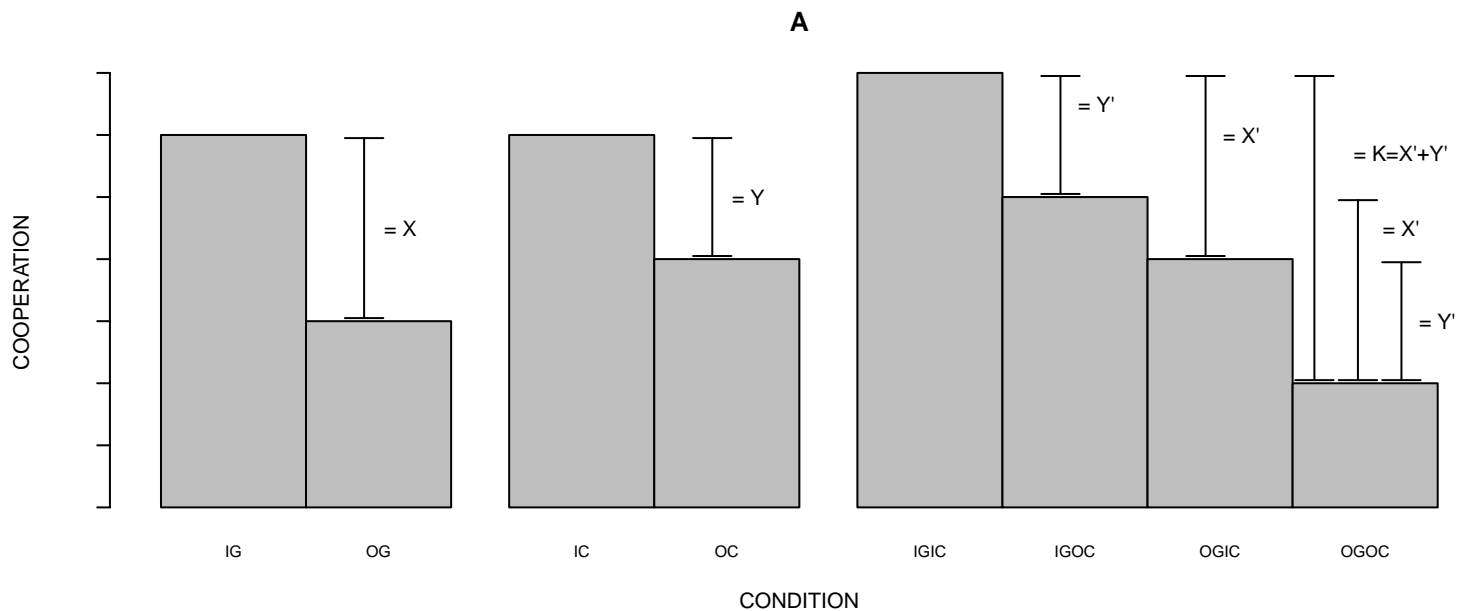|	Cooperate	Defect
Cooperate	(60, 30)	(30, 35)
Defect	(70, 15)	(40, 20)

Table 2. Experimental Design. SPD = Symmetric Prisoner's Dilemma, APD = Asymmetric Prisoner's Dilemma. IC = in-class, OC = out-class, IG = ingroup, OG = outgroup.
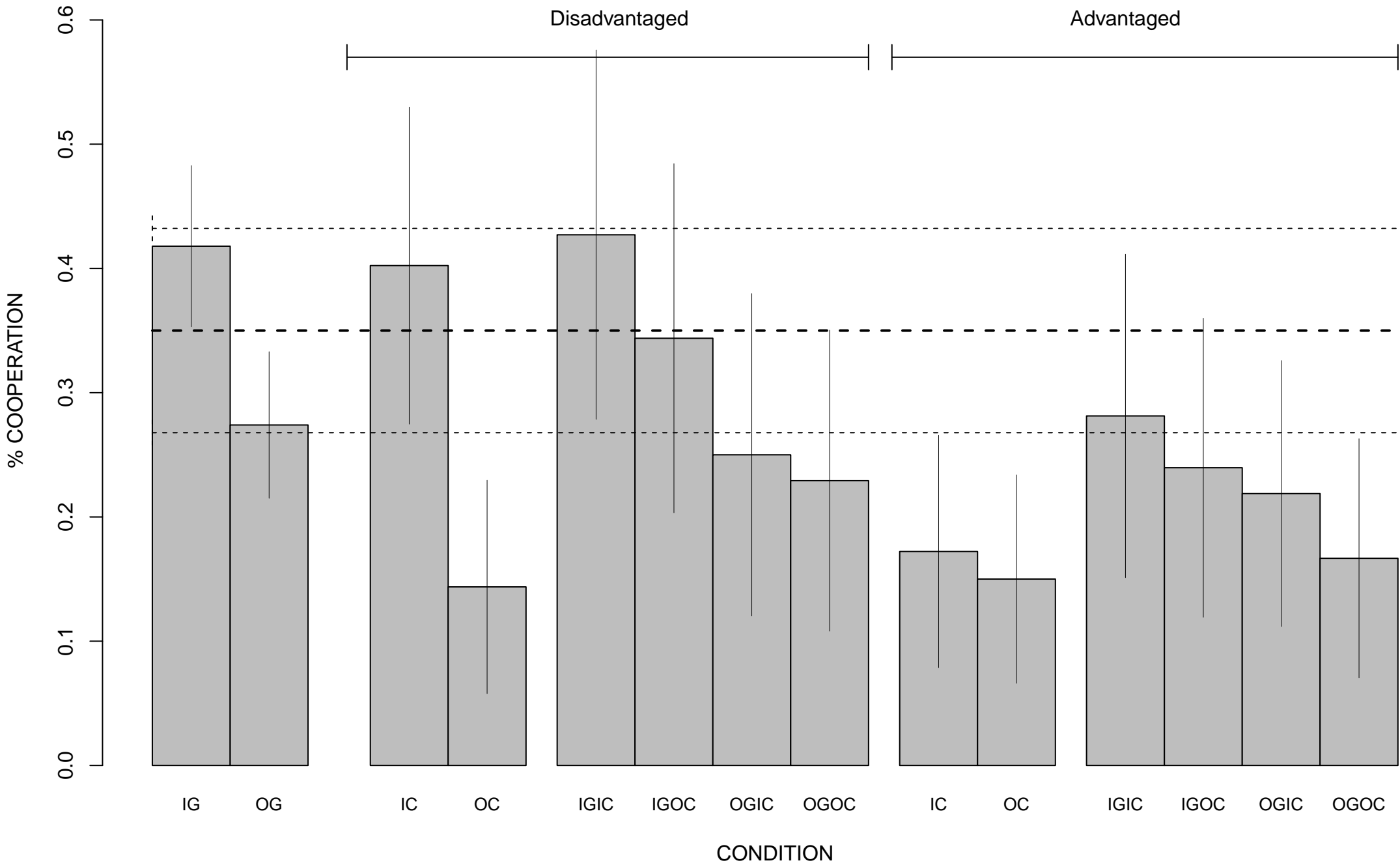
Stage	Identity		Control
	Crosscutting N = 64	**Consolidation** N = 68	N = 59
1	Inducing group identity		No group identity
2	12 SPDs: 6 with IG, 6 with OG		12 SPDs
3	Inducing inequality Group & class crosscutting	Group & class consolidation	Inducing inequality
4	12 PDs: 3 SPD with IGIC 3 APD with OGOC 3 APD with IGOC 3 SPD with OGIC	12 PDs: 6 SPD with IGIC 6 APD with OGOC	12 PDs: 6 SPD with IC 6 APD with OC

Table 3. Models predicting cooperation; Ad = Advantaged class, Da = Disadvantaged class;

LPM = Linear Probability Model; + p (2-sided) < 0.1, * p (2-sided) < 0.05, ** p (2-sided) < 0.01.

|	LPM with random effects for subjects		LPM with cluster robust st. errors		Probit with cluster robust st. errors	
	Coeff	S.E	Coeff	S.E	Coeff	S.E
Reference: Baseline (no identity no inequality)						
IC (Ad)	-0.174**	(0.036)	-0.236**	(0.051)	-0.654**	(0.179)
OC (Ad.)	-0.197**	(0.036)	-0.258**	(0.052)	-0.742**	(0.186)
IC (Da)	-0.070+	(0.037)	-0.006	(0.067)	0.071	(0.179)
OC (Da)	-0.328**	(0.037)	-0.265**	(0.056)	-0.766**	(0.209)
Identity (no inequality)						
IG	0.068	(0.046)	0.068	(0.053)	0.184	(0.145)
OG	-0.077+	(0.046)	-0.077	(0.052)	-0.221	(0.149)
Crosscutting (identity and inequality orthogonal)						
IGIC (Ad)	-0.137*	(0.061)	-0.118	(0.080)	-0.243	(0.234)
IGOC (Ad)	-0.190**	(0.061)	-0.172*	(0.078)	-0.407+	(0.241)
OGIC (Ad)	-0.217**	(0.061)	-0.198**	(0.073)	-0.496*	(0.233)
OGOC (Ad)	-0.257**	(0.061)	-0.239**	(0.067)	-0.649**	(0.234)
IGOC (Da)	-0.162**	(0.061)	-0.067	(0.085)	-0.094	(0.235)
IGIC (Da)	-0.067	(0.061)	0.028	(0.089)	0.167	(0.238)
OGOC (Da)	-0.271**	(0.061)	-0.176*	(0.077)	-0.417+	(0.244)
OGIC (Da)	-0.262**	(0.061)	-0.167*	(0.083)	-0.392	(0.255)
Consolidation (identity and inequality overlap)						
IGIC (Ad)	-0.140*	(0.055)	-0.170*	(0.072)	-0.407+	(0.220)
OGOC (Ad)	-0.204**	(0.056)	-0.234**	(0.065)	-0.630**	(0.223)
IGIC (Da)	-0.128*	(0.055)	-0.205**	(0.073)	-0.524*	(0.235)
OGOC (Da)	-0.266**	(0.056)	-0.342**	(0.056)	-1.222**	(0.286)
Round	-0.016**	(0.002)	-0.016**	(0.003)	-0.057**	(0.010)
Round × Inequa.	-0.009**	(0.003)	-0.009**	(0.003)	-0.011	(0.011)
Intercept	0.510**	(0.040)	0.510**	(0.047)	0.049	(0.124)
Random part:						
σ_u(participant)	0.272**	(0.015)				
σ_e(decision)	0.340**	(0.004)				
ρ	0.390**	(0.027)				
N (participant)	191		191		191	
N (decision)	4584		4584		4584	

Consolidation Treatment
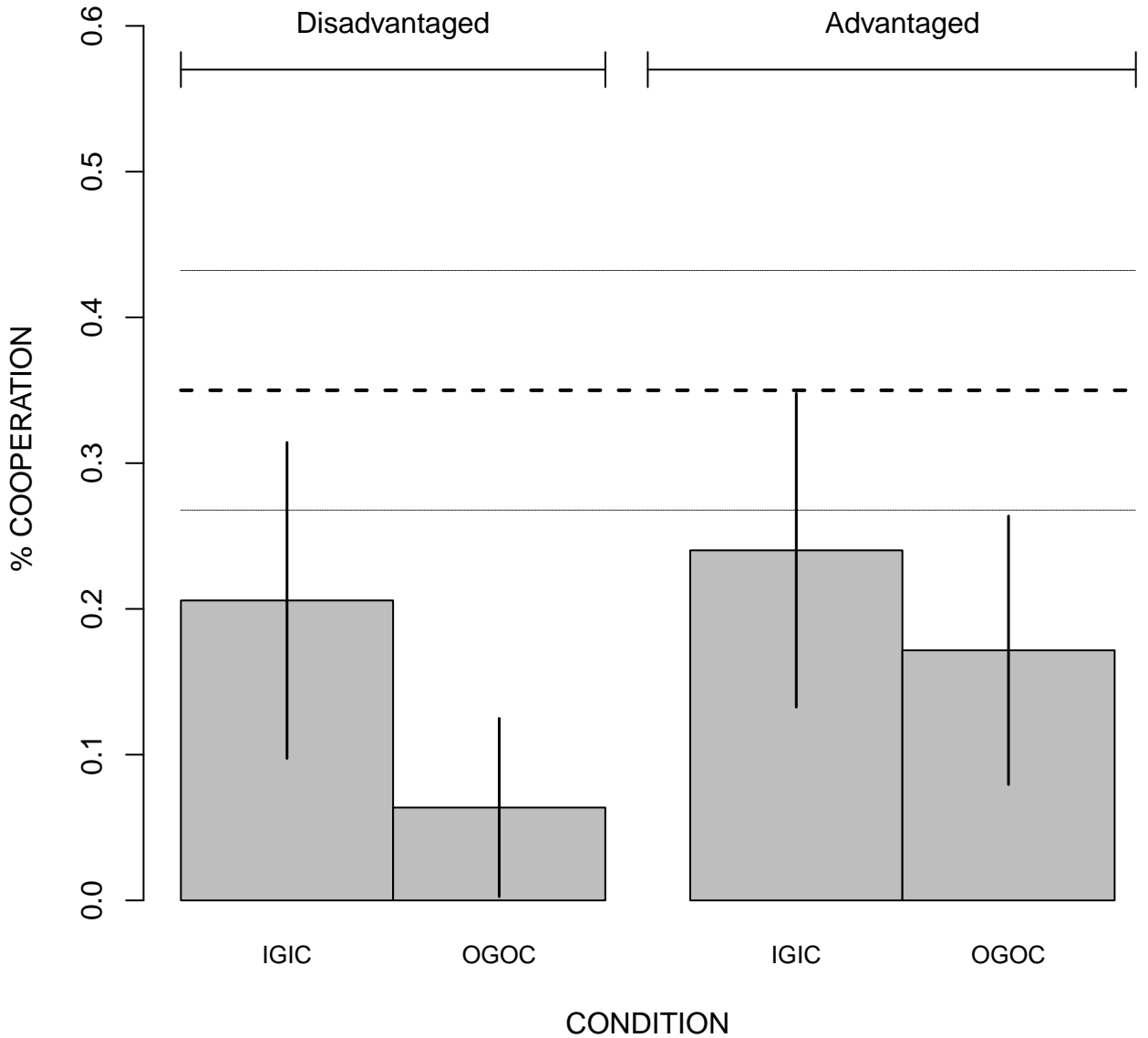
Real effort score and cooperation in symmetric PDs

Control & crosscutting

Consolidation

Mean cooperation

Score centred at the median