6-15-2021

Group-identity and Long-run Cooperation: An Experiment

Gabriele Camera
Chapman University, camera@chapman.edu

Lukas Hohl
University of Basel

Group-identity and Long-run Cooperation: An Experiment

Comments

ESI Working Paper 21-10

This paper later underwent peer review and was published as:

Camera, G. & Hohl, L. (2021). Group-identity and long-run cooperation: an experiment. *Journal of Economic Behavior & Organization, 188*, 903-915. https://doi.org/10.1016/j.jebo.2021.06.017

Group-identity and long-run cooperation: an experiment*

Gabriele Camera
ESI, Chapman University
DSE, University of Bologna

Lukas Hohl
University of Basel

June 15, 2021

Abstract

We stress-test the limits of the power of group identity in the context of cooperation by constructing laboratory economies where participants confront an indefinitely repeated social dilemma as strangers. Group identity is artificially induced by random assignment to color-coded groups, and reinforced by an initial cooperation task played in-group and in fixed pairs. Subsequently subjects interact in-group and out-group in large economies, as strangers. Indefinite repetition guarantees full cooperation is an equilibrium. Decision-makers can discriminate based on group affiliation, but cannot observe past behaviors. We find no evidence of group biases. This suggests that group effects are less likely to emerge when players cannot easily observe and compare characteristics on which to base categorizations and behaviors.

Keywords: large groups, indefinitely repeated game, social norms.

JEL codes: C70, C90, D03

---

* The authors thank two anonymous reviewers for many helpful comments. We acknowledge partial research support through the SNSF grant No. 100018-172901. Correspondence address: Gabriele Camera, Economic Science Institute, Chapman University, One University Dr., Orange, CA 92866, USA; e-mail: camera@chapman.edu. Lukas Hohl, Faculty of Business and Economics, University of Basel, Peter Merian-Weg 6, 4052 Basel, Switzerland; e-mail: lukas.hohl@unibas.ch.

1 Introduction

Social identity theory elaborated in Tajfel et al. (1971) and Tajfel and Turner (1979) asserts that an individual's sense of self gets strengthened when she is affiliated with a high status group. Behavioral economists have incorporated group membership (or, group-identity) factors into theories of decision making (Akerlof and Kranton, 2000) and have measured their impact in the lab. There is mounting evidence that group-identity affects behavior in dictator games (Chen and Li, 2009; Fatas et al., 2018), public goods games (Eckel and Grossman, 2005), trust games (Hargreaves Heap and Zizzo, 2009), and coordination games (Brooks et al., 2018), to give some examples. A minimal categorization (see Chen and Li, 2009) is generally enough to generate ingroup favoritism and outgroup hostility.

At the heart of this finding is a human tendency to use observable differences to categorize selves and others into groups, and then condition actions based on this differentiation.[1] Group identity experiments generally make it quite easy to observe and compare group characteristics and behaviors, facilitating the task of making inter-group comparisons—an important mechanism behind ingroup biases (Tajfel and Turner, 1979). Our experiment offers an attempt to stress-test the limits of the power of group identity in the context of cooperation. We ask: do group effects emerge in strategic settings where it is *not* easy to observe and compare characteristics on which to base categorizations and behaviors?

To do so, we artificially induce group identity in laboratory economies populated by twelve strangers who are artificially divided into three groups. These

[1] We give credit to an anonymous referee for making this point and for providing a useful commentary, which is partly reflected in this paragraph.

strangers meet in random pairs an indefinite number of times (a supergame), which is unlike the typical finite-duration experimental design.[2] Each encounter consists of a helping game, in which one randomly selected person (the "producer") has the option to suffer a small cost to bestow a large benefit upon the other (the "consumer"). That is: the producer is free to cooperate or defect. This choice can be based on the consumer's group affiliation, but not on their identity, past behavior, or group's behavior—all of which are unobservable. This setting precludes behavior based on reciprocity and reputation because the history of play of individuals and the group they belong to always remain shrouded. An outcome is efficient if all producers always cooperate, independent of group affiliations, i.e., if there is *full cooperation*. Players are notified if cooperation is less than full in a period.

The indefinite repetition design ensures that full cooperation is a sequential equilibrium because players can adopt a trigger strategy based on the social norm proposed in Kandori (1992). It consists of a "rule of cooperation" (always cooperate) to be used in equilibrium, and an irreversible "rule of punishment" (always defect) that is triggered if someone breaks the first rule. It is this threat of permanent, economy-wide punishment that supports efficient play. We are interested in assessing if a minimal categorization design can lead strangers to engage in discriminatory behavior, acting differently with insiders than outsiders. In particular, do group effects emerge from the start of the supergame? If so, do they persist as the game progresses? Does going beyond the minimal group categorization by adding an economic inequality aspect

[2] For example, subjects in Chen and Li (2009) played between seven and twelve games (a mix of dictator and ultimatum games), nine decisions in Fatas et al. (2018) (a mix of dictator games and risky choices), fifteen games in Eckel and Grossman (2005), six games in Hargreaves Heap and Zizzo (2009), and ten rounds of a stag-hunt game in Brooks et al. (2018). We found just one group-identity experiment involving an indefinitely repeated game, the two-player prisoner's dilemma in Li and Liu (2017).

influence ingroup/outgroup cooperation?

In our design, group affiliation is artificially induced by randomly assigning subjects to three color-coded groups. The identity induced by this categorization is then reinforced by having subjects interact ingroup in fixed pairs for two consecutive supergames. In these partners' economies subjects can attain high cooperation by leveraging reciprocity and reputation mechanisms. This payoff-relevant stage should facilitate the creation of trust among insiders, meaning that subjects should come to expect a high frequency of cooperation from ingroup participants. We then mix subjects into economies of twelve strangers (four per group) for two more supergames. Here, there is random rematching into pairs in each round of play. In these meetings, producers *can* see the consumer's color before choosing, so can discriminate based on group affiliation. Yet, the past conduct of individuals and groups is not observable. Hence, these strangers do not have observable characteristics other than the counterpart's color on which to base their discrimination. Intergroup biases cannot depend on reputation or reciprocity mechanisms at the individual or group level.

In the Equal treatment group affiliation is payoff irrelevant—all players face the same payoff matrix and—when they interact as strangers—they are randomly assigned to a series of ingroup and outgroup meetings. We use this within-subject variation to investigate if a minimal categorization design supports group effects. To determine if going beyond the minimal group categorization influences behavior we use instead a between-subjects design. In the Unequal treatment the gains from cooperation vary across groups. Adding this payoff-relevant aspect to the induced categorization introduces inequality in economic opportunity. This increases the observable characteristics on which players can base their categorizations and discriminatory behaviors, potentially affecting the psychological basis for group identity effects over and above the minimal categorization.

The analysis of the data reveals no evidence of group biases in our economies of strangers. We do not detect discriminatory behavior in the initial rounds of play, meaning that if psychological group biases exist, they are not sufficiently strong to affect interaction at the start of the economy. Group effects are also not detected in later rounds of the supergame, as we do not observe choices being based on the insider/outsider distinction. This applies when we restrict attention only to meetings in which the decision maker had not suffered from a defection earlier in the supergame, or to meetings after suffering a defection. In the first kind of meetings decision makers have no explicit reason to sanction others by defecting. In the second kind of meetings subjects did lower their cooperation frequency—as a way to sanction the earlier defection—and were as likely to defect with insiders as well as outsiders. In fact, once defections occur in economies of strangers cooperation simply declines without recovering, as evidence also in other experiments with strangers (e.g., Camera and Casari, 2009). This suggests that group effects are less likely to emerge in settings where players cannot easily find observable characteristics, other than group affiliation, on which to base their categorizations and discriminatory behaviors.

The paper proceeds as follows. In section 2 we review the relevant experimental literature, in section 3 we describe the experimental design. Section 4 presents the theory and offers testable research hypotheses. Section 5 contains the results and section 6 provides a final discussion.

2 Related literature

A large economics literature investigated the influence of naturally-existing or artificially induced group-identity in the lab in a variety of experimental tasks. In a few experiments participants are grouped according to their naturally-existing social identities, such as political or social affiliation for instance (Brooks et al., 2018). In the majority of experiments, instead, the experimenter induces group-identity artificially (see Goette et al., 2012). There are two ways to accomplish this task. One is to divide subjects into groups, either at random (e.g., by randomly assigning a color label) or on the basis of some payoff-irrelevant element of commonality (e.g., aesthetic preferences). The available evidence suggests that this mechanism of artificially-inducing group-identity is rather weak and does not significantly affect behavior (Charness et al., 2007; Eckel and Grossman, 2005). A preferable alternative is to make subjects undergo a group task, where same-group members share common experiences, that precedes the experimental task proper; this task can be payoff-irrelevant. There is some evidence that artificially-inducing group-identity in this second manner is more likely to affect behavior, although the empirical evidence is inconclusive.[3]

In-group favoritism is a typical finding, meaning that subjects act more cooperatively with members of one own's group (in-group members) as compared to subjects with a different identity (out-group members) (e.g., see the meta-analysis in Lane, 2016).

[3]For example, Eckel and Grossman (2005, p. 373) report that cooperation increase significantly when group identity is enhanced by having group members cooperate on achieving an unrelated goal before playing a public good game, as compared to treatments without no prior interaction among team members. On the other hand, the differences between treatments with and without common experiences are minimal in Chen and Li (2009, p.450), who report that a problem-solving stage (an online chat) tended to increase self-reported group attachment but did not have a strong effect on behavior.

For example, in Chen and Li (2009) subjects that are matched to an in-group member are more likely to choose actions that maximize social-welfare, to reward good behavior and less likely to punish misbehavior, as compared to subjects that are matched to out-group members. Morita and Servàtka (2013) find that group-identity strengthens individuals' altruistic preferences towards in-group members and thus can help to resolve the hold-up problem. Group-identity increases contributions in a public goods game (Eckel and Grossman, 2005), in homogeneous and heterogeneous groups (Weng and Carlsson, 2015), and improves coordination in the Minimum Effort coordination game (Chen and Chen, 2011). Charness et al. (2007) find, that group-identity affects the subjects' behavior both in a Battle of the Sexes game as well as in a Prisoner's dilemma. Daskalova (2018) reports the effect of group-identity on a joint-decision-making task. There is also evidence that individuals display preference conformism in non-strategic settings, as subjects tend to base their decisions on those of peers (Fatas et al., 2018).

A first aspect of these designs is their focus on interaction that is either one-shot (e.g., Charness et al., 2007), or of short duration and with a restricted horizon (e.g., Weng and Carlsson, 2015). An open question is whether artificially induced group-identity biases are permanent or if they are short-lived, and dissipate as subjects share common experiences over the long-haul. A second aspect of existing studies of group-identity is a focus on decisional situations in which there are only a few players (often two), who can easily differentiate group behavior, and therefore can reciprocate both discrimination and favoritism. In this scenario, biases can be self-reinforcing. The open question is whether ingroup biases emerge when reciprocation is impossible because group conduct is opaque. The answer to these questions matters from an external validity perspective because interaction in society is generally open-ended, and it often has an anonymous structure that precludes reputation or reciprocity.

To this end, we design interaction as taking place among many strangers—so that neither individuals nor groups can develop reputations—and over an open-ended horizon. As noted earlier, we are aware of just one group-identity experiment that involves an indefinitely repeated game, Li and Liu (2017). The study reports evidence of ingroup favoritism in a Prisoner's Dilemma where two partners exclusively interact ingroup or, alternatively, outgroup, for an average of 3 to 4 rounds. Instead, we contribute to the debate about group identity in indefinitely repeated games by studying twelve-person economies of strangers who interact *both* ingroup and outgroup for long horizons, 20 rounds on average. In our design, reciprocation (direct or indirect) is impossible because group-affiliation and IDs of decision-makers remain hidden. As a result, neither favoritism nor discrimination can be equilibria, in contrast with designs of supergames among partners. Removing individual reputation is important because the experimental literature suggests that individual reputation dominates artificially induced group-identity biases in finite-horizon games. In that case, subjects can self-select into groups of like-minded individuals by developing individual reputations.[4] The open question is thus if group-identity biases play a role when self-selection is difficult or impossible, as is often the case in modern industrialized societies. Our experiment can provide an answer because individuals' past conduct remains unobservable for the duration of the

[4]For example, in a public goods game where subjects can be excluded from the group, Charness et al. (2014) find little evidence that artificially induced and subsequently strengthened group-identity determines exclusion choices, because the probability of exclusion primarily depends on the subjects' past contributions. This result is in line with experiments on endogenous group formation without group identity, where the ability to identify and isolate free riders, supports the formation of cooperative groups; see Ahn et al. (2009), Cinyabuguma et al. (2005), Güth et al. (2007), or Maier-Rigaud et. al. (2010).

supergame, and subjects can neither self-select into a desired group nor can be individually targeted for exclusion.[5]

3 Experimental design

We start by describing the EQUAL treatment, which is our baseline. A session in the experiment comprises twenty-four subjects who play a sequence of four supergames where the size and composition of the interaction group is exogenous. At the start of the session, subjects are randomly assigned to an eight-person set that is color-coded (green, red, and blue). This random assignment, which is permanent for the session, is a weak form of artificially inducing group-identity (Eckel and Grossman, 2005), which we aim to strengthen through a payoff-relevant task carried out in the first two supergames, as we next explain.

A supergame: Each supergame consists of 18 rounds plus an indefinite number of additional rounds. After 18 rounds, another round is played with probability $\beta = 0.75$ and, otherwise, the supergame ends for everyone in the session.[6] This ensures the same minimum length of 18 rounds for all supergames, and an expected duration of 21 rounds with a tight standard deviation of 3.5 rounds.

In each round, each subject faces a "helping game" in a pair composed

[5] There are other difference with Li and Liu (2017). In particular, the number of groups (two vs. three in ours) and the mechanism used to form group identity after the initial random color assignment. The group manipulation protocol in Li and Liu (2017) involves pre-play communication about a payoff-irrelevant task as in Chen and Li (2009). By contrast, we use a group task that is payoff-relevant, does not involve communication, and is part of an initial phase of the experiment—subjects interact in-group, and in fixed pairs.

[6] This probability was common knowledge in the session. Because supergames started and ended simultaneously for all participants in a session, all participants in the session played exactly the same number of rounds.

of a "producer" and a "consumer." Roles are randomly assigned in the first round of the supergame and then deterministically alternate so that in each round half of the subjects of any color are producers and half are consumers. In each pair, the consumer has no choice to make; see Table 1. The producer can choose Cooperate or Defect. By choosing Defect she earns 6 points, while the consumer earns 3. By choosing Cooperate she earns 0 points, and the consumer $13 + a$ points.

Table 1: Payoffs in a consumer-producer pair

	Producer's choice	
	Defect	*Cooperate*
Payoff to Consumer	3	$13 + a$
Payoff to Producer	6	0

Notes: In the experiment $a = 0$ for partners and $a = 3$ for strangers.

The parameter a depends on whether the counterpart is fixed or randomly changes during the supergame. Specifically, there are two kinds of economies, "partners" and "strangers." We set $a = 0$ for partners and $a = 3$ for strangers. Under partners, subjects interact in-group and with a fixed counterpart so the economy has 2 participants. Under strangers, we form 12-person matching sets by selecting four subjects from each color, who are randomly rematched into 6 pairs in each round. Here, economies have 12 participants who interact both in-group and out-group depending on the outcome of matching process; subjects are strangers because past actions and identities remain hidden. Producers can see the consumer's color so can discriminate based on group affiliation, but consumers cannot see the producer's color (see the screenshots Figs. B1-B2 in Appendix B) so the color-coded groups cannot develop a reputation.[7]

[7]This implies that decision-makers cannot act based on the observed correlation between

Summing up, a partners setting offers the possibility to easily attain moderate gains, while a strangers setting offers larger gains from cooperation without the possibility to leverage reputation or reciprocity mechanisms.

At the end of each round strangers can see if all pairs of their economy attained the same outcome or not. This information allows subjects to determine if every producer in the economy cooperated or not.[8]

A session: A session consists of four separate supergames, and two phases; see Fig. 1. Phase 1 is an identity-strengthening phase that comprises supergames 1-2. Here, subjects interact as partners to facilitate cooperation through reciprocity and reputation. Phase 2 comprises supergames 3-4, when subjects interact as strangers, both in and outside their group. To the extent that partners cooperate more than strangers, Phase 1 can foster a sense of trust in those who share the player's same color, thus strengthening group-identity. To minimize spillover effects from earlier supergames to subsequent supergames, economies are constructed so that counterparts from previous supergames cannot be met in subsequent supergames.[9]

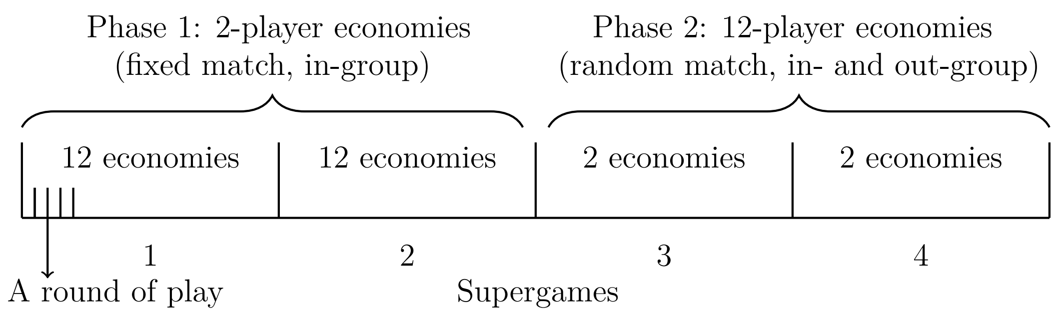
---

group-affiliation and cooperation, thus breaking possible feedback effects reinforcing an initial group bias. This is necessary to determine if color-based discrimination is primarily stemming from in-group biases, or if it is the result of self-reinforcing feedback effects that rely on forms of history-dependent, selective punishments (e.g., defect out-group because the player saw that out-group producers defected more often than in-group producers).

[8]See the column "Same Outcome In All Pairs" at the bottom left of the screen in Figs. B1-B2 in Appendix B. This allows subjects in 12-person economies to monitor defections occurring elsewhere in the economy, not only their meeting, while maintaining anonymity. For example, if the outcome is C in the subject's meeting and the column "Same Outcome In All Pairs" reports NO, then this means that some producer in the economy selected D this round. This is equivalent to raising a red flag warning that someone defected, which can facilitate coordination on a punishment mechanism in large groups. In fact, it ensures that the condition sufficient to support full cooperation as an equilibrium is independent of the economy's size—as demonstrated in the next Section .

[9]Running two consecutive supergames with each economy configuration minimizes possible hurdles to subjects' proper comprehension of the two different economic environments. Indeed, the 12-player economies constitute a rather complex decisional environment as

Figure 1: A session



Treatments: Our baseline treatment is described above. In the Unequal treatment, we manipulate the importance of group affiliation by making consumers' cooperation payoffs color-dependent. This is done using a symmetric mean-preserving spread of Equal payoffs: green consumers earn 11 points, red 13, and blue 15. Neither the parameter a, nor producers' payoffs are manipulated. This gives rise to inequality of economic opportunity: everyone can gain from cooperation, but some players can gain more while others less. In this case, group affiliation has a concrete consequence for prospective payoffs. By comparing the Equal to the Unequal case, we can determine if the payoff (ir)relevance of group affiliation affects the emergence or strength

compared to the fixed partnerships of Phase 1. This project is part of a larger research agenda about whether inequality—payoff relevant or not—might affect cooperation and the endogenous integration into larger, heterogeneous economies. For this reason, each session included a fifth supergame with an endogenous size and composition of the interaction set—either fixed pairs as in Phase 1 or larger economies as in Phase 2, but of a larger size. The instructions (see Appendix B) informed participants that they would have an opportunity to alter size and composition of their interaction set in supergame 5, without providing specific details until the start of supergame 5 in order not to influence behavior in supergames 1-4. Given the scope of this paper this last segment of the session is not part of the present analysis. Camera et al. (2021) discusses an experiment that studies if and how size and distribution of *potential* gains from integrating groups into a large and potentially quite profitable economy influences individuals' inclination to integrate and to cooperate to reap the benefits it can offer.

of group-identity biases.

Experimental procedures: The experiment was conducted at the Economic Science Institute's laboratory at Chapman University. We recruited 192 undergraduate students between 2/2017 and 11/2017. We ran 4 sessions per treatment, each with 24 participants. Subjects were informed that at the end of the experiment one of the five supergames would be randomly selected, via public randomization. The points the subjects earned in this supergame were converted into dollars (1 point is worth USD 0.18). On average, subjects received about USD 21 in salient earnings, a USD 7 show-up fee, and about USD 2 from a post-instruction incentivized quiz (included in Appendix B). One session lasted about 1 hour and 40 minutes. Instructions are included in Appendix B, were recorded in advance and played aloud at the beginning of a session, while participants could follow on individual copies. For the instructions neutral language was used. The experiment was programmed using the software z-Tree (Fischbacher, 2007). No eye contact was possible between participants. Demographic data was collected in an anonymous survey at the session's end. Given our design, the experimental data indicates that the smallest detectable value of the treatment effect size is 0.037 with a power of 0.8 (significance level of 0.05), and 0.0428 when the power is 0.9.[10]

[10] We thank an anonymous referee for suggesting this calculation. The idea is to find out if we can detect economically meaningful group effects, if there were any. We ran power calculations to estimate the minimum group effect size we can detect with our design. The power analysis test is run for paired means, using the mean and standard deviation taken from the experimental data for the alternative hypothesis. We use the test for paired data because we have measurements on the same economy for in-group and out-group interaction, so the test accounts for the dependence between the two groups of observations. Fixing one 12-player economy as our unit of observation, we have $N = 32$ observations each for in-group and out-group interaction (both treatments pooled). In the data, outgroup cooperation is 0.507 as compared to 0.527 ingroup, with standard deviations of, respectively, 0.204 and 0.203. The null hypothesis is that mean cooperation observed in-group is identical to out-group. The alternative hypothesis is that of a group

4 Theoretical considerations and hypotheses

According to standard theory, participants could attain maximum profits in every experimental economy. We summarize this in the following:

Proposition 1. *Full cooperation is a sequential equilibrium in every economy and in every treatment of the experiment.*

The proof, which can be found in Appendix A, relies on tacitly coordinating on a "grim" strategy that, in economies of strangers, leverages the availability of anonymous public monitoring. This strategy works as follows: all participants should start the supergame by following the strategy "always cooperate" with the tacit threat of switching to "always defect," if any producer defects at any point. If all players adopt the grim strategy, then full cooperation is an equilibrium when the probability of continuation β is sufficiently high.[11] In that case, no producer, of any color, has an incentive to deviate in equilibrium—defecting instead of cooperating. Moreover, no producer has an incentive to deviate from following the grim sanction off equilibrium because "always defect" is an equilibrium of the original game (a best response to "always defect" by everyone else).

Following the discussion in Kandori (1992), we say that a social norm of cooperation emerges when the grim strategy is adopted by everyone in the economy. This norm is based on two distinct and complementary rules of conduct. The first one is a "rule of cooperation," according to which everyone

identity effect. We use the outgroup mean 0.507 as our alternative "pre-treatment" mean, and the ingroup mean of 0.527 as our alternative "post-treatment" mean. The standard deviation of the difference is 0.072. The results of the power test suggest that our design has enough power to detect economically meaningful group effects, as it can detect quite small differences in cooperation, up to a value of about 0.04.

[11] The threshold is 0.55, approximately. The continuation probability in the experiment is 0.75. Hence, assuming risk neutrality, full cooperation is a sequential equilibrium in all supergames of all treatments.

should cooperate when they are a producer, starting from the very beginning of the supergame. The second component is a "rule of punishment"—stop cooperating forever—which should be followed only if someone breaks the first rule, and, hence, should never be observed in equilibrium.

Because consumers cannot observe producer's group affiliation, and players cannot see statistics about the relative cooperativeness of the three groups, we have an additional result.

Corollary 1. *Neither ingroup discrimination nor favoritism is part of an equilibrium in economies of strangers.*

The proof is obtained by deriving a contradiction. Suppose the following discriminatory equilibrium exists in Phase 2 economies: players always cooperate with insiders, and defect with outsiders. Since decision-makers' colors are hidden, it is optimal for a producer to deviate with insiders because her color remains undetected and she cannot be individually punished. If so, deviating to D in insider meetings is optimal, contrary to our conjecture. The same holds true for behavior favoring outsiders, instead of insiders. It follows that to support a cooperative equilibrium players must sanction a defection by defecting in every future meeting. This means that if group effects emerge in the experiment they are likely due to a social identity explanation and not epistemic considerations.

What is central to our design is that the grim strategy can be used equally effectively in all economies. However, Phase 1 economies can also rely on reciprocity and reputation to support cooperation (e.g., tit-for-tat). Previous experiments suggest that this supports high cooperation.[12] If so, Phase 1 interaction should promote cooperation, trust in the group's affiliates, which

[12] Cooperation rates are higher among partners than strangers, even when public monitoring is available (Camera and Casari, 2009). Economies of strangers suffer from strategic uncertainty that reduces trust (see the discussion in Bigoni et al., 2019)

should strengthen group identity. We rely on Phase 2 to uncover possible in-group (or, intergroup) biases, by determining if producers condition their cooperation on the consumer's group affiliation. Of particular interest is the possibility that producers cooperate more in-group than out-group meetings, something that emerges in one-shot or finitely repeated experiments where group-identity is payoff-irrelevant (Charness et al., 2007; Eckel and Grossman, 2005) and also payoff-relevant (Weng and Carlsson, 2015). We thus hypothesize that group effects should emerge in both treatments, form the start of a supergame. Recalling that half of the subjects make their initial choice in round 1, and the rest in round 2, we formulate the following.

H 1. *In the first two rounds of a Phase 2 supergame, in-group meetings exhibit higher cooperation than out-group meetings.*

To formulate hypotheses about actions past the initial one, consider that Li and Liu (2017) uncover in-group favoritism in two-player indefinite horizon Prisoner's Dilemmas where group-identity is payoff-irrelevant. This suggest that in-group/out-group effects should also be observed in our experiment, once players move past their initial choice.

H 2. *In the average round of a Phase 2 supergame, in-group meetings exhibit higher cooperation than out-group meetings.*

Finally, note that strangers who interact indefinitely can base their choices on their personal experience in the supergame (e.g., see Camera et al., 2013). The norm discussed above suggests that suffering a defection should trigger a permanent decline in cooperation, i.e., a permanent switch to a punishment mode. A reasonable hypothesis is that social identity should also emerge during the punishment phase because if subjects have a motive to discriminate against outsiders under normal circumstances, then this motive should be reinforced when there is evidence of free riding in the community.

H 3. *After suffering a defection in Phase 2, subjects' actions are consistent with group identity effects.*

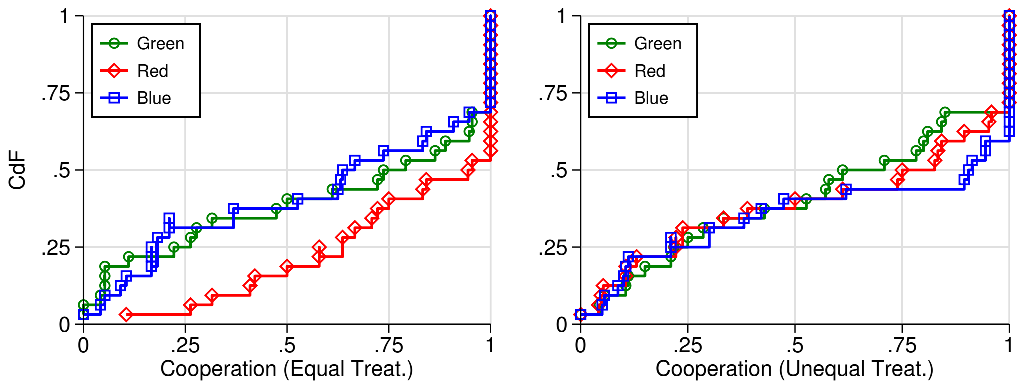
5 Results

We start by reporting outcomes in Phase 1, when subjects interacted in-group, in fixed pairs.

Result 1. *Cooperation in Phase 1 was similar across treatments and across groups within a treatment.*

Evidence is provided by Fig. 2 and Table 2. Fig. 2 reports the distribution of cooperation rates in Phase 1, by treatment. One observation is one economy in a supergame, which in Phase 1 consisted of a fixed pair of homogeneous color. Each marker reports the average cooperation rate in the supergame, separated by group affiliation (color). Overall, average cooperation in Phase 1 was 62% and 67% in, respectively, Equal and Unequal. Fig. 2 reveals that in each treatment about 1/3 of economies achieved full cooperation (0.36 and 0.34, respectively), while the share of economies that coordinated on full defection was close to zero (0.03 in each treatment). In other words, Phase 1 was a similarly highly cooperative phase, in both treatments.

Figure 2: Cooperation in Phase 1 Economies



Notes: One obs.=one economy in a supergame of Phase 1 ($N = 32$ per color, per treatment). Each marker reports the average cooperation rate in the economy during the supergame; this measure ranges from 0 when there was full defection (D was selected in each round of the supergame) to 1 when there was full cooperation (C was selected in each round of the supergame).

Fig. 2 suggests that group affiliation did not affect behavior when colors were payoff-relevant in the UNEQUAL treatment (right panel). In that treatment Green players had the least economic benefit from cooperating (11 points as consumers), while Blue players had the greatest incentive (15 points as consumers). The figure does not reveal a positive association between potential gains from cooperation and cooperation rates; except for a few observations, the markers show substantial overlap. Interestingly, when colors were payoff irrelevant (left panel), Blue and Green players behaved similarly while Red players cooperated more than others.

To establish the statistical significance of these observations we exploit the longitudinal structure of the data, which gives us repeated choices for each subject. Table 2 reports the marginal effects from panel logit regressions with random effects, ran separately for each treatment (first four columns), and pooled (last two columns). Robust standard errors are adjusted for clustering at the session level to account for possible dependencies within a session. The panel variable is a subject in Phase 1. The dependent variable takes value 1 if the subject cooperated in a round in which she was a producer, and is 0 otherwise. The factor variable *Affiliation* identifies the group affiliation of the decision maker (the base case is Green), which is the exogenous source of variation in the experiment.

To determine if and how suffering D affected the subsequent probability of cooperating, we include the *Suffered D* and *Lag n* indicator variables. *Suffered D* takes value 1 starting the round after the first instance in which the subject suffered D as a consumer (0 otherwise). Because the subject might react with delay or revert back to cooperate after a while, we include five *Lag n* indicator variables, for $n = 1, 2, 3, 4, 5$, each of which takes the value 1 for a choice made $2n - 1$ rounds after suffering D (0 otherwise).[13] The *Choice #* regressor captures possible time trends, as it takes value n at the n^{th} choice of the subject[14], the *Supergame 2* takes value 1 in the second supergame (else, 0, the base case), while individual controls consist of two measures of understanding of instructions (response time and wrong answers in the post-instruction quiz), and sex.

[13] This econometric technique is used in Camera and Casari (2009). The sum of the coefficients *Suffered D+Lag n* identifies the average reaction $2n - 1$ rounds after the event (subjects made choices every other round). With these five sums we can trace the subjects' response between 1 and 9 rounds after suffering the initial D, while the *Suffered D* coefficient captures the response 10+ rounds after suffering D, i.e., the long-run effect on the probability of cooperating after falling victim to a defection.

[14] Half of the subjects made a choice in odd rounds and half in even rounds.

Table 2: Cooperation in Phase 1: Marginal Effects

Dep. var.	(1) Equal		(2) Unequal		(3) Pooled	
=1 if C (0 if D)	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Affiliation						
Red	0.076	(0.071)	0.014	(0.037)		
Blue	-0.026	(0.083)	0.002	(0.051)		
Unequal treatment					-0.046	(0.054)
Punishment						
Suffered D	-0.406***	(0.019)	-0.502***	(0.042)	-0.437***	(0.037)
Lag 1	0.068***	(0.016)	0.125***	(0.040)	0.085***	(0.011)
Lag 2	0.023	(0.045)	0.078***	(0.012)	0.044**	(0.022)
Lag 3	0.050**	(0.024)	0.082***	(0.023)	0.057***	(0.016)
Lag 4	0.033	(0.024)	0.063***	(0.018)	0.041***	(0.012)
Lag 5	0.016	(0.012)	0.060***	(0.014)	0.033***	(0.010)
Supergame 2	0.014	(0.016)	0.054*	(0.029)	0.031**	(0.013)
Choice #	0.019**	(0.008)	0.033***	(0.005)	0.023***	(0.004)
Controls	Yes		Yes		Yes	
N	1884		1944		3828	

Notes: Logit panel regression with random effects (for the intercept) at the individual level and robust standard errors (SE) adjusted for clustering at the session level. One observation = one producer in a round of supergames 1 and 2. Dependent variable = 1 if C chosen, 0 otherwise. Base case = Green player, supergame 1. *Controls:* sex and two standardized measures of understanding of instructions from the incentivized post-instruction quiz (response time and wrong answers). Marginal effects are computed at the regressors' mean value (at base levels for factor variables). Symbols $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively.

Group affiliation does not significantly affect the probability to choose cooperation in Unequal (panel 2). In Equal (panel 1) it partially affected Red subjects who cooperated more than Blue (Wald test, p-value=.0454) but not Green subjects. Cooperation was also statistically similar across treatments; see the *Unequal* coefficient in panel (3).[15]

Three other points can be made. First, subjects switched from a cooper-

[15]Table B1 in Appendix B reports similar results when we use a more conservative GLM specification, where the unit of observation is one economy of Phase 1, and the dependent variable is average cooperation in a supergame.

ative to a long-lasting punishment mode if their partner defected; the coefficient on *Suffered D* is a large and highly significant negative number. We can reject the null hypothesis that the sum *Suffered D+Lag n* is zero for each $n = 1, 2, 3, 4, 5$ (Wald tests, p-value < 0.001 for all tests.) Second, in Phase 1 subjects learned to cooperate with their partner (the *Choice #* coefficient is positive and significant), an outcome that is visually apparent in Fig. B3, in Appendix B. Third, repeating the interaction a second time with a different partner did not decrease cooperation and sometimes significantly increased it (the coefficient on *Supergame 2*). Taken together, this evidence suggests that subjects learned to trust and to cooperate with their partner in the long-run, in Phase 1. This view is reinforced by considering the first action taken in a supergame. We say that a participant has an *uncooperative* inclination if her first choice is D despite not having seen that choice in that supergame (an initial producer in round 1, or an initial consumer in round 2 who experienced C in round 1). The share of uncooperative participants is similar across treatments in supergame 1, and declined by more than 50% in the second supergame (0.43 and 0.42 in supergame 1 of Equal and Unequal, and 0.15 and 0.2 in supergame 2).

Overall, these findings suggest that Phase 1 of the experiment contributed to build trust in the subject's group affiliates, in both treatments. To uncover possible intergroup biases, we now focus on Phase 2, which is when subjects participated in 12-person economies, with random rematching, in- or out-group, in each round. We start by considering the first choice of subjects in these mixed-affiliation economies.

Result 2. *In Phase 2, producers did not condition their initial choice on the consumer's group affiliation.*

Evidence is provided in Table 3, which reports the marginal effects from logit regressions with random effects at the individual level. Robust standard errors are adjusted for clustering at the session level. One observation corresponds to the first choice of a subject in Phase 2 supergames, i.e., the round 1 choice of initial producers and round 2 choice of initial consumers. The dependent variable is set to 1 if C is selected (0, if D). The *In-Group* indicator takes the value 1 for in-group meetings (else 0, the base case) and is interacted with the *Affiliation* factor variable identifying the color of the decision maker (Green is the base case).

As variation in Phase 1 experience might affect initial choices in Phase 2, we follow the modeling technique used in Bigoni et al. (2019) and include the indicator *Full C in Phase 1.* This fixed effect takes value 1 if the subject experienced full cooperation in at least one Phase 1 economy (else 0, the base case). Full cooperation means that in the supergame C was chosen in every round—Fig. 2 reveals that several pairs achieved it.

The behavior of the initial donor might also influence the first choice of initial consumers (pooling both treatments, D occurred in 61 out of 192 round 1 meetings). We thus add the categorical variable *Round 2* which we set to 0 for a choice taken in round 1 (the base case), and to 1 (resp. 2) for a choice taken in round 2 after observing C (resp. D). We also include an indicator for supergame 4 (the second in Phase 2), and standard controls at the individual level consisting of the subject's sex and our two standardized measures of understanding of instructions.

Table 3: Subject's Initial Choice in Phase 2: Marginal Effects

Dep. var.	Equal		Unequal	
=1 if C (0 if D)	Coeff.	S.E.	Coeff.	S.E.
In-Group	-0.060	(0.084)	-0.055	(0.043)
Affiliation				
Red	-0.019	(0.121)	-0.278***	(0.071)
Blue	0.199***	(0.069)	-0.163*	(0.084)
Round 2, C	0.043	(0.168)	0.121	(0.143)
Round 2, D	-0.049	(0.187)	0.028	(0.087)
Full C in Phase 1	0.261**	(0.112)	0.290***	(0.057)
Supergame 4	-0.139*	(0.082)	0.006	(0.051)
Controls	Yes		Yes	
N	192		192	

Notes: Logit panel regression with random effects (random intercept) at the individual level and robust standard errors (SE) adjusted for clustering at the session level. One observation = one producer in rounds 1-2 of supergames 3-4. Dependent variable = 1 if C chosen, 0 otherwise. Base case = Green producer, out-group, round 1, supergame 3, full cooperation never experienced in Phase 1. *Controls:* sex, duration of the previous supergame (standardized), and two standardized measures of understanding of instructions from the incentivized post-instruction quiz (response time and wrong answers). Marginal effects are computed at the regressors' mean value (at base levels for factor variables). Symbols $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively.

We find no evidence that initial choices are conditioned on counterparts' group affiliation. The *In-Group* coefficients are small and statistically not different from zero. Thus we can reject H1 because there is no evidence that social identity effects emerge from the beginning of the supergame. The regression reveals that the subject's experience in Phase 1 was the main factor driving initial choices in Phase 2. Those who experienced full cooperation at least once in Phase 1, are significantly more likely to select C as their first action as compared to those who never coordinated on full cooperation earlier. The odds of selecting C are significantly higher (almost 30% higher) for a subject that experienced full cooperation in Phase 1 as compared to someone who did not.

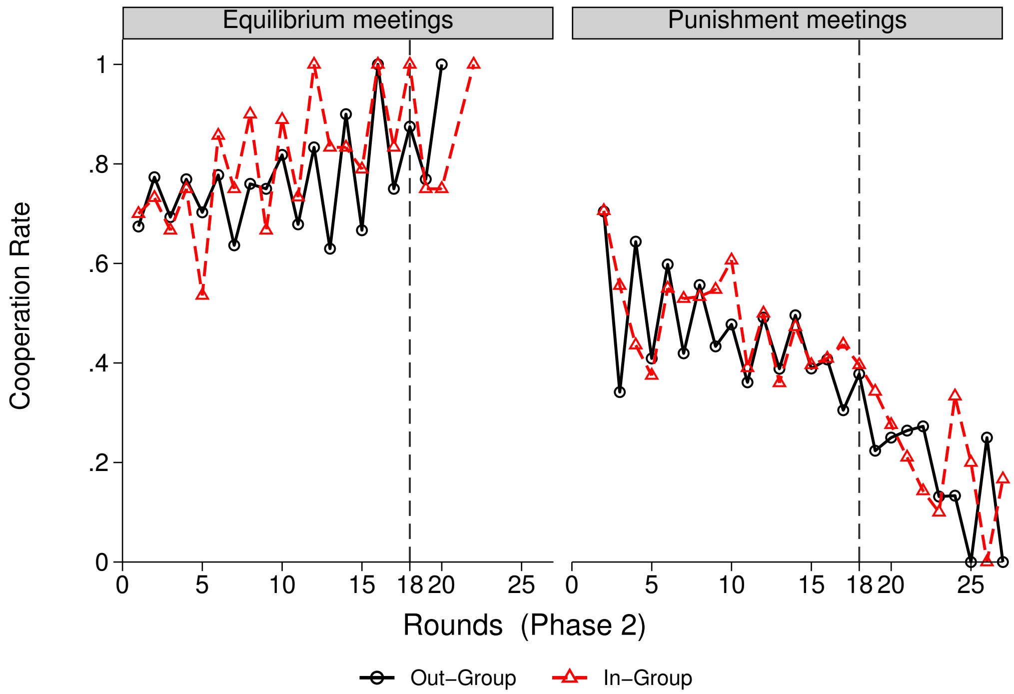
It is possible that subjects started to condition their choices on counterparts' group affiliation later in the supergame, even if they did not do so initially. To investigate this possibility, it is convenient to study the data in two kinds of meetings. Those in which the decision-maker had not yet suffered a defection in the supergame, and those in which they had suffered at least one defection. We refer to the first type as "equilibrium meetings" and the second type as "punishment meetings" because, according to theory, here the producer had a motive to choose D. About 71% of all choices was taken in a punishment meeting (N=2804, both treatments pooled).

Result 3. *Group effects neither emerged in equilibrium meetings nor in punishment meetings of Phase 2.*

Fig. 3 and Table 4 provide evidence. Fig. 3 reports average cooperation rates separately for in- and out-group meetings in each period of Phase 2 supergames (both treatments pooled). One marker corresponds to the average choice selected by producers in that period, with circles identifying out-group and triangles in-group meetings. The left panel considers only "equilibrium meetings,"' while the right panel only "punishment meetings." Equilibrium meetings mostly comprise the few initial periods of the supergame because defections quickly increased as the game progressed, unlike Phase 1; overall, by round 3 about 45% of players had suffered a defection, a percentage that doubles by round 11.[16]

[16] This implies that we have progressively fewer (more) observations as the supergame progresses in the left (right) panel; it also explains we do not have observations past round 22 in the left panel. The vertical dashed line identifies round 18, which is when the random stopping process started in the supergame.

Figure 3: Cooperation in- and out-group, in Phase 2



Notes: One obs.=one producer in a period of supergames 3 and 4 (both treatments pooled). Each marker reports the cooperation rate averaged across all observations for the period. A circle marker reports the average cooperation rate in out-group meetings, while a triangle refers to in-group meetings. Left panel: meetings in which the producer did not suffer D earlier in the supergame. Right panel: meetings in which the producer suffered D earlier in the supergame.

On the left panel, cooperation rates for in-group and out-group meetings both increases as the supergame progresses, and mostly overlap. The overlap is evident also in the right panel as is the negative trend in cooperativeness. Overall, Fig. 3 does not reveal a pattern consistent with social identity effects.[17]

[17]Fig. B4 in Appendix B reports a similar illustration that considers all meetings, separating the data by the producer's group affiliation. There is a clear decreasing trend for all colors.

To assess the significance of these visual observations we use a panel logit regression with random effects. The dependent variable takes value 1 if the subject cooperated as a producer in a meeting, and is 0 otherwise. The panel variable is a subject in a session. As Phase 2 economies include all three colors, we interact the *In-group* indicator variable with the factor variable *Affiliation* capturing the color of the decision maker (Green is the base). As before, we control for Phase 1 positive experience, using the indicator variable *Full C in Phase 1.* In addition, we also add the indicator *Previous Outcome* taking value 1 if the outcome was C in the previous round—when the subject was a consumer—and 0 otherwise (the base case). This allows us to soak up the effect of recent experience when studying the behavior of producers who suffered a defection in an earlier round, i.e., in "punishment meetings."

Table 4 reports the marginal effects on the probability of cooperating, by treatment. Panel (a) considers choices in equilibrium meetings, while panel (b) considers choices in punishment meetings. All coefficients on the *In-Group* dummy are statistically close to zero. This suggests that individuals did not generally discriminate based on group affiliation. We can thus reject H2 and H3 for both treatments.[18]

[18] Here, we relied on a within-subject design to identify group effects in a treatment. As suggested by a anonymous Referee, an alternative would be a between-subjects design with a treatment without group identities to be used as benchmark for group effects.

Table 4: Cooperation in Phase 2: Marginal Effects

Dep. var.	**(a) Equil. meet.**		**(b) Punish. meet.**	
=1 if C (0 if D)	Equal	Unequal	Equal	Unequal
In-Group	-0.123	-0.013	0.024	-0.029
	(0.099)	(0.026)	(0.038)	(0.027)
Affiliation				
Red	0.020	-0.130	0.192***	-0.026
	(0.173)	(0.290)	(0.069)	(0.090)
Blue	0.055	-0.017	0.056	-0.008
	(0.167)	(0.147)	(0.067)	(0.035)
Punishment				
Lag 1			-0.000	0.136*
			(0.033)	(0.075)
Lag 2			-0.063	0.018
			(0.045)	(0.079)
Lag 3			-0.031	0.043
			(0.038)	(0.069)
Lag 4			-0.060*	0.035
			(0.033)	(0.041)
Lag 5			-0.024	0.006
			(0.026)	(0.046)
Previous Outcome			0.145***	0.206***
			(0.011)	(0.046)
Full C in Phase 1	0.258***	0.268	0.145***	0.246***
	(0.054)	(0.616)	(0.046)	(0.065)
Choice #	-0.021***	-0.006	-0.040***	-0.014
	(0.005)	(0.010)	(0.008)	(0.012)
Supergame=4	-0.159***	-0.043	-0.117***	-0.034
	(0.026)	(0.051)	(0.011)	(0.031)
Controls	Yes	Yes	Yes	Yes
N	475	645	1433	1371

Notes: Logit panel regression with random effects (random intercept) at the individual level and robust standard errors (SE) adjusted for clustering at the session level. One observation = one producer in a round of supergames 3 and 4. Dependent variable = 1 if C chosen, 0 otherwise. Base case = Green player, out-group, previous outcome = D, supergame 3. *Controls:* sex and two standardized measures of understanding of instructions from the incentivized post-instruction quiz (response time and wrong answers). Panel (a): equilibrium meetings; Panel (b): punishment meetings. Marginal effects are computed at the regressors' mean value (at base levels for factor variables). Symbols $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively.

It is interesting to compare the trend in cooperation in Phase 2 economies relative to Phase 1. Recall that in Phase 1 average cooperation exhibited a positive time trend, evidence that subjects learned to coordinate on efficient play when they interacted with a fixed parter. Instead, in Phase 2 economies average cooperation exhibits a significant decreasing trend, evidence that subjects did not manage to learn how to coordinate on efficient play when they interacted with random parters; all *Choice n* regressors are negative and significant in Table 4, even when we consider only equilibrium meetings.

6 Discussion

Current thinking in social identity theory suggests that categorization alone can affect norms of cooperation by inducing social identity or group effects. Reinforcing group identity via ingroup cooperative tasks might further affect behavior. We constructed an indefinitely repeated social dilemma experiment where subjects interact as strangers. The initial categorization was reinforced with an ingroup cooperative task, and yet we do not detect an impact on group behavior. Subjects cooperated similarly with insiders and outsiders—they did not condition their actions on the counterpart's group affiliation. This holds true both when group affiliation is payoff irrelevant and when it is not.

What explains this finding? Several possibilities exists. First, earlier studies document that weak forms of group-identity do not influence cooperation Eckel and Grossman (2005). It is conceivable that our design did not create a sufficiently strong categorization and Phase 1 interactions did not sufficiently reinforce group identity. Cooperating in fixed pairs before moving to larger mixed-color economies may be too weak a form of artificially reinforcing group-identity. Introducing free-form communication within a group might prove to be a more effective way to create a stronger form of group identity, although the evidence on this is mixed (Chen and Li, 2009). It is also possible that the number of groups, three in our design instead of the customary two groups, might have blurred group boundaries. In other words, increasing the number of groups might progressively weaken the effect of group identity.

An alternative explanation is the impossibility of making group comparisons as subjects could not differentiate the conduct of insiders from those of outsiders. It is conceivable that group comparison is necessary to make groups sufficiently meaningful to individuals, when they interact as strangers. In our design participants faced a random mix of anonymous encounters with outsiders and insiders. Producers could see the consumers' group affiliation but were blind to their track records and of the group to which they belonged. Consumers were blind to the producer's color so could not determine how choices depended on group affiliation. This allowed discrimination based on group affiliation, but prevented its detection. This removed the incentive to follow a discriminatory strategy—even if group identity was indeed strong—for two reasons.

On the one hand, by precluding group reputation our design breaks the feedback effect that likely reinforces pre-existing discriminatory behavior. Being able to assess cooperation at a group level allows punishment to be conditioned on group affiliation. Here, pre-existing biases find a fertile ground to flourish, as an initial discriminatory action that confirms these biases is likely to trigger retaliation toward that group. In a way, initial ingroup biases may trigger out-group punishment, which ignites a self-reinforcing spiral of group effects. The informational opaqueness in our design rules out this self-reinforcing mechanism.

On the other hand, hiding the decision-maker's group affiliation equalizes the temptation to defect with insiders and outsiders. It also reduces the attractiveness of engaging in ingroup favoritism because that, too, can backfire. To explain, someone suffering a defection has insufficient information to justify sanctions limited only to outsiders. If so, defecting outgroup induces a cooperation decline in *all* future meetings, ingroup and outgroup, without distinction. Individuals that care about their own group thus have an extra incentive to avoid discrimination due to spillover effects that would damage her own group.

These two considerations help reconcile our results with those in Li and Liu (2017), which finds ingroup favoritism when fixed pairs play an indefinitely repeated PD game. Instead, we work with groups of twelve individuals who interact with random insiders and outsiders, as strangers. Decision makers can discriminate based on group affiliation, but unlike fixed pairs they cannot compare outsiders' to insiders behavior. It is possible that precluding group comparison prevents the creation of a sufficiently strong ingroup bias, or the reinforcement of pre-existing biases. In fact, this informational opaqueness ensures that acting in a discriminatory manner towards outsiders *can* have spillover effects on insiders. Cheating an outsider can trigger a cascade of defections that reverberates throughout the economy, lowering payoffs of outsiders and insiders without distinction. The message is that if relative group performance can be made hard to assess, then this mitigates the impact on behavior of possible psychological identity biases.

References

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

Akerlof, G, and R. Kranton. (2000). Economics and identity. *The Quarterly Journal of Economics*, 115(3), 715-753.

Bigoni, M., G. Camera, and M. Casari. (2019). Partners or Strangers? Cooperation, monetary trade, and the choice of scale of interaction. *American Economic Journal: Microeconomics*, 11(2), 195-227.

Brooks, B. A., K. Hoff, and P. Pandey. (2018). Cultural impediments to learning to cooperate: An experimental study of high-and low-caste men in rural India. *Proceedings of the National Academy of Sciences*, 115(45), 11385-11392.

Camera, G., and M. Casari. (2009). Cooperation among strangers under the shadow of the future. *American Economic Review* 99(3), 979-1005.

Camera, G., and M. Casari. (2014). The Coordination Value of Monetary Exchange: Experimental Evidence. *American Economic Journal: Microeconomics*, 6(1), 290-314.

Camera, G., M. Casari, and M. Bigoni. (2013). Money and trust among strangers. *Proceedings of the National Academy of Sciences*, 110(37), 14889-14893.

Camera, G., L. Hohl, and R. Weder. (2021). *Inequality as a Barrier to Economic Integration: an Experiment.* Unpublished manuscript.

Charness, G., R. Cobo-Reyes, and N. Jiménez. (2014). Identities, selection, and contributions in a public-goods game. *Games and Economic Behavior*, 87, 322-338.

Charness, G., L. Rigotti, and A. Rustichini. (2007). Individual behavior and group membership. *American Economic Review*, 97(4), 1340-1352.

Chen, Y., and S.X. Li. (2009). Group-identity and social preferences. *American Economic Review*, 99(1), 431-57.

Chen, R., and Y. Chen. (2011). The potential of social identity for equilibrium selection. *American Economic Review*, 101(6), 2562-89.

Cinyabuguma, M., T. Page, and L. Putterman. (2005). Cooperation under the threat of expulsion in a public goods experiment. *Journal of Public Economics* 89(8), 1421-1435.

Daskalova, V. (2018). Discrimination, social identity, and coordination: An experiment. *Games and Economic Behavior*, 107, 238-252.

Eckel, C., and P. Grossman. (2005). Managing diversity by creating team identity. *Journal of Economic Behavior & Organization*, 58(3), 371-392.

Fatas, E., S. Hargreaves Heap, and D. Rojo Arjona. 2018. Preference conformism: An experiment. *European Economic Review* 105, 71-82.

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

Goette, L., D. Huffman, and S. Meier. (2012). The impact of social ties on group interactions: Evidence from minimal groups and randomly assigned real groups. *American Economic Journal: Microeconomics*, 4(1), 101-15.

Güth, W., V. Levati, M. Sutter, and E. Van der Heijden. (2007). Leading by example with and without exclusion power in voluntary contribution experiments. *Journal of Public Economics*, 91(5), 1023-1042.

Hargreaves Heap, S., D. J. Zizzo. (2009). The value of groups. *American Economic Review* 99 (1), 295-323 .

Kandori, M. (1992). Social Norms and Community Enforcement, *Review of Economic Studies*, 59(1), 63-80.

Lane, T. (2016). Discrimination in the laboratory: A meta-analysis of economics experiments. *European Economic Review*, 90, 375-402.

Li, S. X., and T. X. Liu. (2017). *Group-identity and cooperation in infinitely repeated games.* Mimeo.

Maier-Rigaud, F., Martinsson, P., and Staffiero, G. (2010). Ostracism and the provision of a public good: experimental evidence. *Journal of Economic Behavior & Organization*, 73(3), 387-395.

Morita, H., and M. Servàtka. (2013). Group-identity and relation-specific investment: An experimental investigation. *European Economic Review*, 58, 95-109.

Tajfel, H., M. Billig, R. Bundy, and C. Flament. (1971). Social categorization and intergroup behaviour. *European journal of social psychology*, 1(2), 149-178.

Tajfel, H., and J. Turner. (1979). An integrative theory of intergroup conflict. *The social psychology of intergroup relations*, 33(47), 74.

Weng, Q., and F. Carlsson. (2015). Cooperation in teams: The role of identity, punishment, and endowment distribution. *Journal of Public Economics*, 126, 25-38.

A Appendix

A.1 Proof of proposition 1

Here we prove that full cooperation is a sequential equilibrium in every group and treatment. We say that a norm of cooperation is being followed in the group whenever all players adopt the trigger strategy discussed in Section 4. For convenience let the defection payoffs be, respectively, d and $d - l$ to a producer and a consumer. Let $k_i := 9 + 2i$ denote the cooperation payoff to a consumer of type (color) $i = 1, 2, 3$ under fixed pairs. Indeed we have $k_i = 13$ for all colors in the Equal treatment, while $k_1 = 11$, $k_2 = 13$ and $k_3 = 15$ in the Unequal treatment. Given this notation, a necessary and sufficient condition for full cooperation to be an equilibrium is reported in the following lemma:

Lemma 1. *Fix an economy. Let $\underline{k} + a$ denote the smallest cooperation payoff in that economy. If the continuation probability*

$$\beta \geq \beta^* := \frac{d}{a + \underline{k} - d + l} \in (0, 1),$$

then full cooperation is a sequential equilibrium.

A type i player alternates between producer and consumer roles, having the opportunity to earn $k_i + a$ every other round, in a cooperative equilibrium. Let $s = 0, 1$ denote the role of the player at the start of a round, where 0 is for a producer and 1 for a consumer. The type of counterparts does not affect the player's payoff—only the counterparts' cooperation rate. In cooperative equilibrium the player nets

$$v_0 := \frac{\beta(a + k_i)}{1 - \beta^2} \qquad \text{and} \qquad v_1 := \frac{a + k_i}{1 - \beta^2},$$

while off-equilibrium there is full defection so the payoff is type-invariant and corresponds to the one associated to infinite repetition of the static Nash equilibrium, denoted

$$\hat{v}_0 := \frac{d + \beta(d - l)}{1 - \beta^2} \qquad \text{and} \qquad \hat{v}_1 := \frac{d - l + \beta d}{1 - \beta^2}.$$

It is immediate that off-equilibrium a producer has no incentive to deviate from the sanctioning rule, because defecting is the unique best response to every other producer defecting in every round. Hence, we only need to show that $v_0 \geq \hat{v}_0$, i.e., in equilibrium the player has no incentive to defect as a producer, by refusing to help some consumer.[19] This inequality can be rearranged as $\beta \geq \beta_i^* := \dfrac{d}{a + k_i - d + l}$ and the Lemma automatically follows. Note that $\beta^* < 1$ because $a + k_i - (2d - l) > 0$ for all player types.

The lowerbound probability β consistent with cooperation is a decreasing function of the player's return from cooperation $a + k_i$. Hence β^* is the largest when k_i is the lowest. Proposition 1 follows from observing that in the experiment $\beta = 0.75$ and the most stringent requirement comes from fixed pairs ($a = 0$) composed of players who earn $k = 11$ points as consumers if the producer cooperates. In this case $\beta_1^* = 0.75$ represents the smallest lowerbound threshold.

[19] In the experiment discounting starts on round $T = 18$, when the random termination rule started. One can prove that the incentives to cooperate monotonically decline until round T is reached and then they remain constant. Hence, studying the incentives to cooperate in equilibrium in round T ensures those incentives are satisfied in all $t < T$. In round $t = T$ payoffs correspond to v_s above. The details are in Bigoni et al. (2019).

B Appendix: Supplementary Materials for Online Publication

Figure B1: Input Screen, 12-player economies, Unequal treatment

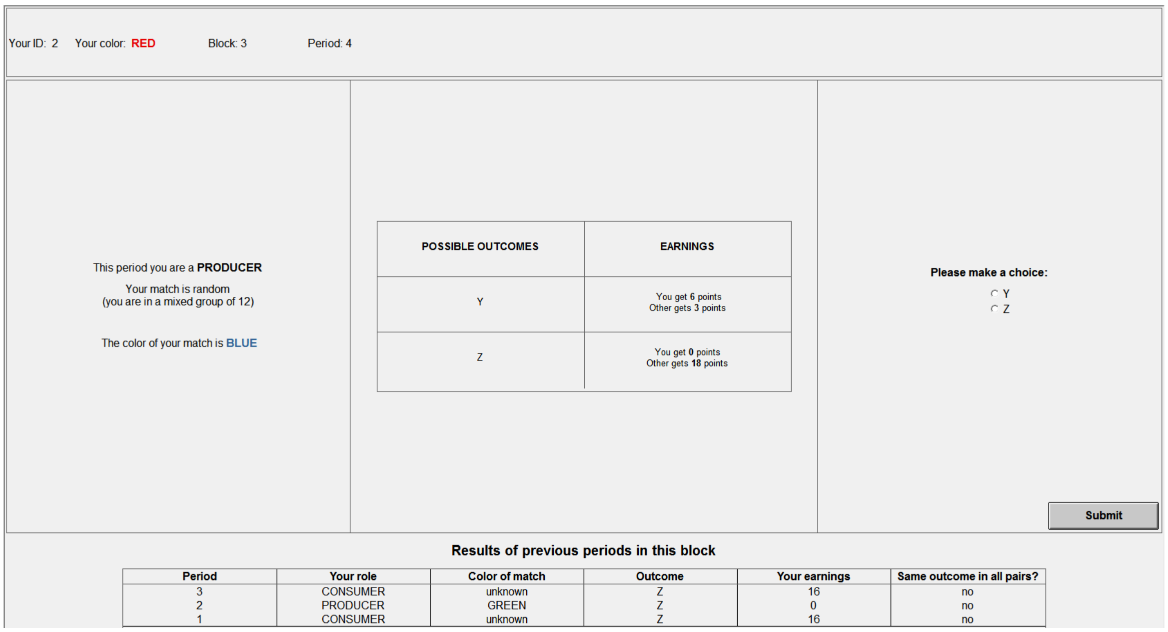


Figure B2: Results Screen, 12-player economies, Unequal treatment

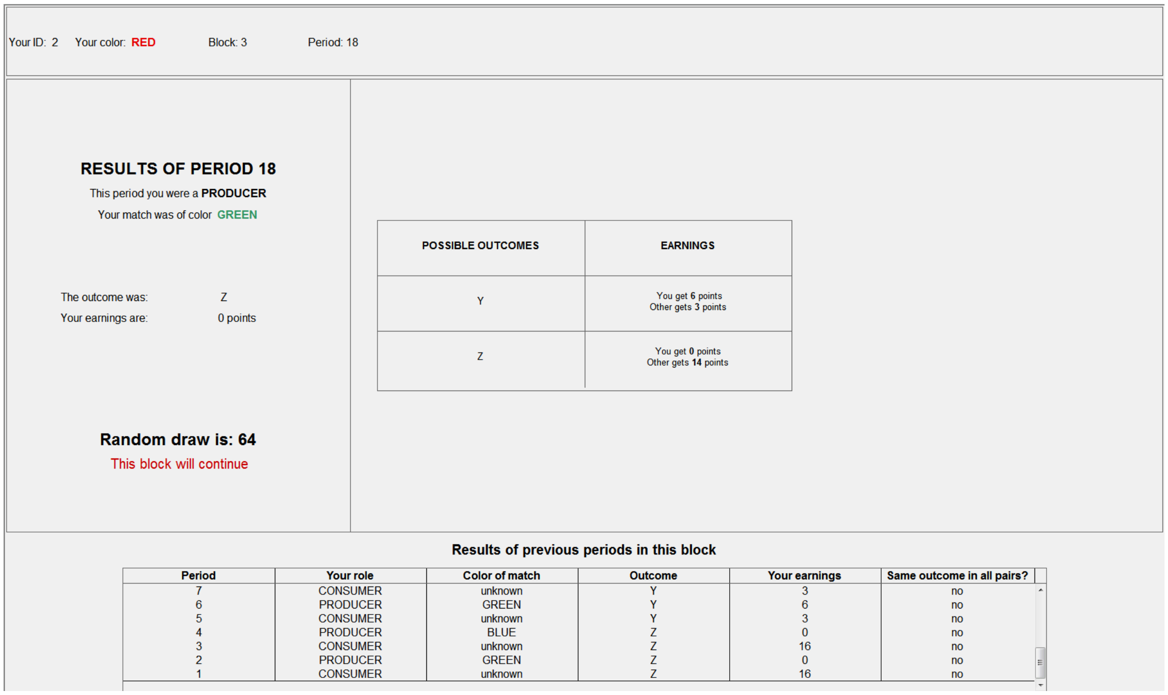
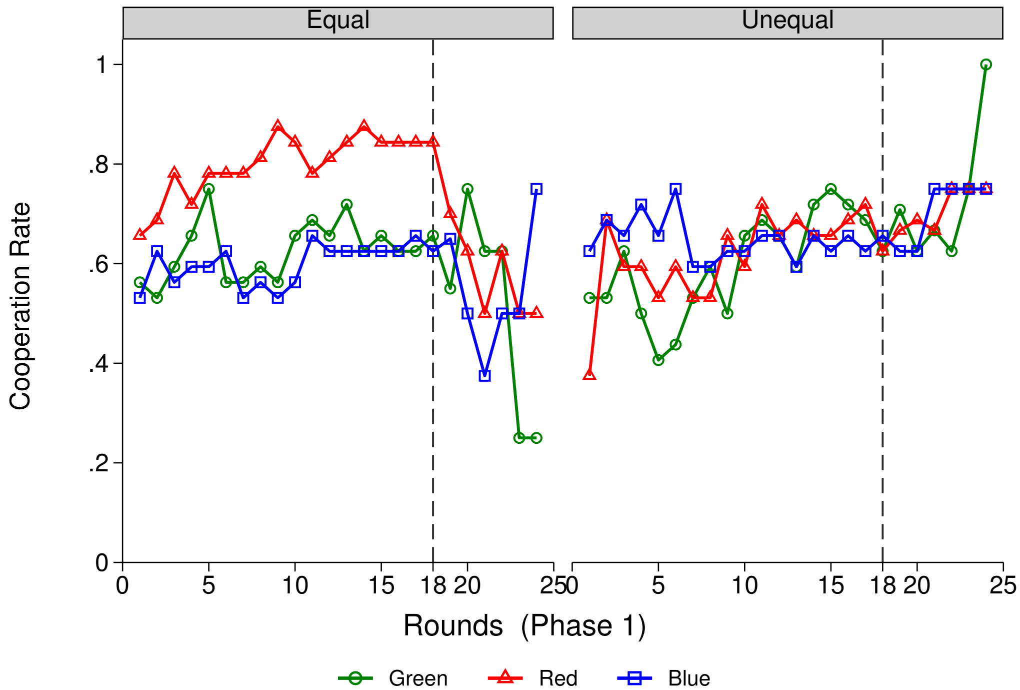


Table B1: Cooperation in Phase 1: Marginal Effects

Dep. var.: avg. cooper.	Equal		Unequal		Pooled	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Group affiliation						
Red	0.207	(0.133)	0.051	(0.075)		
Blue	-0.012	(0.138)	0.053	(0.060)		
Controls						
Unequal					-0.020	(0.069)
Supergame 2	0.179***	(0.053)	0.175***	(0.036)	0.158***	(0.030)
Male	-0.006	(0.106)	0.139***	(0.053)	0.130	(0.093)
Duration	-0.023	(0.019)	0.013	(0.032)	0.000	(0.019)
Duration lagged	0.005	(0.007)	0.043*	(0.023)	0.032**	(0.016)
Response Time	0.034	(0.040)	0.024	(0.089)	0.046	(0.047)
Incorrect Answers	-0.153***	(0.036)	0.022	(0.032)	-0.064*	(0.036)
N	96		96		192	

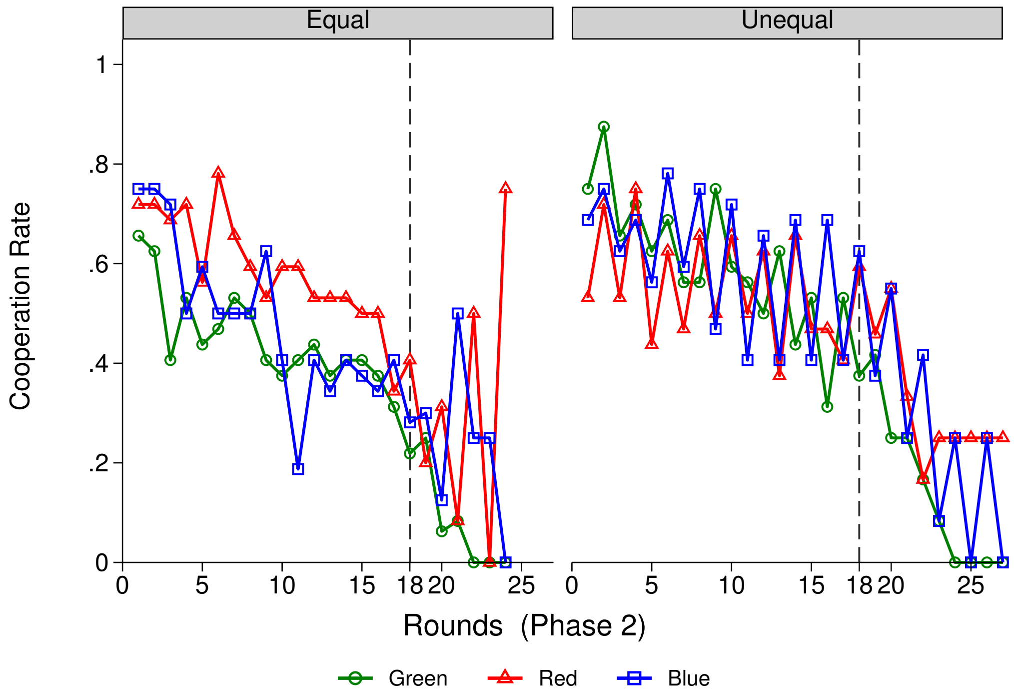
Notes: GLM regression on the average cooperation rate in a supergame. One obs.=one economy in supergames 1 and 2. Robust standard errors in parentheses are adjusted for clustering at the session level. Duration lagged is set to 18 in supergame 1. Marginal effects are computed at the regressors' mean value (at base levels for factor variables). Symbols $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively. When considering the EQUAL treatment, the coefficient on *Red* is different from *Blue* (Wald test, p-value=0.0339).

Figure B3: Cooperation in Supergames of Phase 1



Notes: One obs.=one producer in a period of supergames 1 and 2. Each marker reports the cooperation rate averaged across all observations for the period. All economies had a common minimum duration of 18 rounds. A circle marker reports the average cooperation rate of Green producers, a triangle refers to Red producers, and a square refers to Blue producers. Left panel: Equal treatment. Right panel: Unequal treatment.

Figure B4: Cooperation in Supergames of Phase 2



Notes: One obs.=one producer in a period of supergames 3 and 4. Each marker reports the cooperation rate averaged across all observations for the period. All economies had a common minimum duration of 18 rounds. A circle marker reports the average cooperation rate of Green producers, a triangle refers to Red producers, and a square refers to Blue producers. Left panel: Equal treatment. Right panel: Unequal treatment.