# Do Economic Inequalities Affect Long-Run Cooperation & Prosperity?*

Gabriele Camera
Chapman University
University of Basel

Cary Deck
University of Arkansas
Chapman University

David Porter
Chapman University

June 8, 2017

**Abstract**

We experimentally investigate if inequalities in past economic opportunities affect long-run cooperation. Groups of participants played an uncertain number of helping games each time in a random anonymous pair where a coin flip determined who could help whom. This provided exogenous variation in past opportunities, while ensuring equality of future opportunities. Full cooperation is the efficient equilibrium. Theoretically, variation in past opportunities should not affect the incentive structure and behavior. Empirically, variation in past opportunities weakened norms of mutual support and reduced coordination on efficient play. Participants conditioned choices on own past opportunities and, with inequalities made visible, discriminated against those who were better-off, while becoming more tolerant of defections.

Keywords: cooperation, coordination, equilibrium multiplicity, indefinitely repeated games, social dilemmas.
JEL codes: C70, C90, D03, E02

## 1 Introduction

Inequality in opportunity, income, and wealth looms large in the mind of people across the world because of the view that it undermines the long-run prosperity of a nation (NYT, 2015; Pew Research Center, 2014; Stiglitz,

* We thank Nat Wilcox for helpful conversations, Kladji Bregu for help running the experiments and seminar participants at Chapman University and the Cleveland Fed. G. Camera acknowledges partial research support through the NSF grant CCF-1101627. Correspondence address: Gabriele Camera, Economic Science Institute, Chapman University, One University Dr., Orange, CA 92866; e-mail: camera@chapman.edu.

2012). Broadly speaking one can identify two channels through which inequality harms prosperity. One is conventional: inequality induces economic inefficiency by creating distortions in capital and labor markets, which result in resource misallocation (e.g., Aghion and Williamson, 1998; Piketty, 2014). A second channel is behavioral: inequality erodes cohesion and trust in society (Putnam, 2000), which harms prosperity because it distorts the structure of incentives for cooperation, especially in societies where economic interaction is impersonal (Kimbrough et al., 2008; North, 1991). Much less is known about this second channel, compared to the first. This study contributes to filling this gap by means of an experiment.

We construct societies of "strangers," where subjects cannot exploit reputational mechanisms to build trust and cohesion in their group. The catch is that though cooperation is beneficial and, in fact, necessary to maximize ex-ante payoffs, it cannot guarantee income equality ex-post. We ask: does this kind of variation affect overall cooperation and the group's ability to prosper? Theory says it should not if agents are expected income maximizers. Field data offers ambiguous evidence on this point because many institutional and environmental factors co-vary with economic inequality. Prosperity may reflect changes in the capital market's structure, not wealth inequality; or, reduction in social cohesion may stem from migration, not income gaps. Inequality itself may stem from a mix of factors (choice, luck, power, ability) and may affect the return from cooperation. Our experiment controls for these kinds of confounding factors.

In our experiment, a group of four subjects plays an indefinite sequence of helping games in ever-changing pairs. This offers the simplest setup to think about long-run cooperation because the interaction amounts to a random sequence of individual decision problems. In each pair there is a recipient, and a donor who has the option to provide a large benefit to the recipient (cooperation). As an alternative, the donor may choose a smaller benefit for herself (defection). This cost of cooperation implies that donors may have a short-run temptation to free ride, receiving help from others while giving none. This gives rise to a social dilemma where per-capita income increases with cooperation. A key design feature is that a virtual coin flip determines who can help whom in each pair. This random role assignment amounts to an uncontrollable shock to earning opportunities. It ansures equal opportunity *ex-ante* but not *ex-post* as the realized role sequences are inherently heterogeneous. Someone will be a donor more often than others and vice versa. Theoretically, this heterogeneity in past opportunities is payoff-irrelevant and should not affect the structure of incentives. Full cooperation maximizes expected payoffs for everyone, independently of how roles have been distributed up to that point. Under an efficiency criterion, full cooperation is thus a natural outcome.

In this indefinite horizon environment, cooperative strategies such as "tit-for-tat" cannot support cooperation because players interact as strangers that are constantly rematched, not as partners in fixed pairs.[1] Here, opportunistic temptations can be removed by exploiting information about the actions taken in the group (Abreu et al., 1990; Kandori, 1992). A social norm based on a grim strategy supports full cooperation as one of multiple equilibria. Although full cooperation maximizes ex-ante payoffs, it cannot guarantee equal payoffs *ex-post* because of the random role assignment. Some players may have the advantage of having enjoyed more of the benefits while shouldering less of the cost from cooperation (frequent recipients). Others will have the disadvantage of having shouldered most of the burden of cooperation (occasional recipients)

[1]Strangers are rematched at random in each round, cannot identify opponents and can neither build a reputation nor engage in relational contracting (Camera et al., 2013).

and in extreme cases may even end up earning less that the full-defection payoff, ex-post. Theoretically, these inequalities should not alter the structure of incentives for payoff-maximizing players because the expected return from cooperation is independent of past opportunities. This "payoff irrelevance" of inequality implies that choices should not depend on role histories they can only depend on past conduct, so variation in *past* opportunities should not empirically induce a variation in behavior.

Our first contribution is to document that uncontrollable variation in past opportunities hindered coordination on efficient play in the experiment. There is evidence that donors conditioned their choices on *their own* role history. Those who were at a disadvantage with few past opportunities to get help and many to offer it cooperated significantly less than other participants. The message is that intra-group economic inequalities that should be theoretically inconsequential, may in practice present an obstacle to group cooperation.

As a second contribution, we demonstrate that cooperation is elastic to information about inequality. In a treatment, donors could compare their past opportunities to the recipient's, prior to choosing. These disclosures made salient inequalities in past roles without theoretically altering the structure of incentives compared to Baseline; the disclosures neither reveal past conduct or future intentions, nor alter the expected return from cooperation. Empirically, these disclosures affected behavior. Players who had seen no defection in their group cooperated less than in Baseline. Disadvantaged donors were less likely to help when the recipient was known to have been advantaged in her past opportunities. As a result, cooperation and coordination on efficient play both suffered compared to Baseline where inequalities remained hidden.

Overall, not all defections were equal. Some are evidence of free-riding, some of punishment, and yet others reveal that participants defected to counteract unfavorable past opportunities. Inequality disclosures led to discrimination against more fortunate counterparts but, interestingly, the average donor also became more tolerant of defections. These findings are robust to disclosing inequalities in wealth, in treatments where donors could compare their past earnings to the recipient's, before taking an action. The evidence that donors were less cooperative even if no one else had defected, used the information to discriminate against more fortunate players, and were more tolerant of defections, suggests that inequality or fairness concerns (e.g., Fehr and Schmidt, 1999) might have driven behavior when inequality was made salient.

Three features set our design apart from other experiments about cooperation and inequality. First, not only is full cooperation a Nash equilibrium, but it is only one of many possible equilibria. By contrast, in previous designs the efficient outcome is not an equilibrium and oftentimes the inefficient outcome is the unique Nash equilibrium (e.g., Andreoni and Varian, 1999). Second, interaction has a long-run horizon and is *impersonal*, which is unlike previous partners designs (e.g., Nishi et al., 2015). This prevents reciprocity or reputation-building, and implies that cooperation cannot be incentivized by individual punishment of defectors but must rely on collective punishment. The study of cooperation and inequality in such a setup is still largely unexplored in the experimental literature on supergames. Finally, we isolate the behavioral impact of inequality by filtering away possible effects of differential economic incentives. In our design, inequality is theoretically irrelevant because cooperation has the same expected return for all group members, unlike in previous experiments with heterogeneous benefits (e.g., Gangadharan et al., 2015). Section 2 discusses in more detail the related experimental literature. Section 3 describes the design. Section 4 presents the theory. Section 5 reports the main results and Section 6 offers some final considerations.

## 2 Related studies

Our work is mainly related to experimental studies of cooperation with repeated play and, in particular, to indefinitely repeated dilemmas (Palfrey, 1994). These games support a richer set of equilibria compared to games that are one-shot or with a commonly known number of rounds. In the typical experimental design of indefinitely repeated dilemmas, the matching protocol involves fixed pairs of subjects, who take an action in every round, in a symmetric game (e.g., see the survey in Crawford, 2016). That design not only allows for reciprocity mechanisms to support cooperation, but it also ensures that earnings are equal under full cooperation.

Recent experiments have considered a design that rules out reciprocity, and where full cooperation does not guarantee equal earnings (Camera and Casari, 2014; Camera et al., 2013). In their strangers design, pairs are randomly re-matched each round and subjects do not take an action in every round; they either have the opportunity to give a benefit to someone else, or to receive a benefit from another subject at randomly alternating points in time. The running total of earnings dynamically evolves according to random elements as well as the actions of counterparts. In these experiments subjects are neither informed of their position in the distribution of earnings nor of the realized distribution of opportunities to give or receive benefits. Our experiment manipulates this informational condition, to determine how, if at all, such information impacts cooperation and realized efficiency.

There is a vast economics literature about fairness and inequality aversion and how to incorporate it into economic models (e.g., Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999; Rabin, 1993) and although our study is related to it, it is not an experiment about testing these models in the lab (e.g., Blanco et al., 2011; Deck, 2001; Kagel and Wiley-Wolfe, 2001). The experimental literature suggests that individuals do care about equality in outcomes, at least to some degree, as inequality affects behavior. Some experiments have focused on how wealth inequality affects behavior in one-shot or finitely repeated social dilemmas. The results are mixed. Andreoni and Varian (1999) use the canonical trust game but provided varying show-up payments to subjects to induce inequality. They find no consistent effect of induced inequality. However, Greiner et al. (2012) find that initial inequality leads to greater trust in a repeated trust game with anonymous rematching because higher wealth is a clearer signal of previous untrustworthiness when initial conditions are equal. Nishi et al. (2015), finds that players who are informed about partners' past behavior, in a networked public goods game, cooperate less when they are informed about the wealth of others. Gangadharan et al. (2015) find a negative impact of inequality on efficiency in a linear public good game where subjects can communicate with and reward others. (Dannenberg et al., 2010) find that participants in non-strategic games show considerable aversion to inequality, be it advantageous or disadvantageous. In all of these studies, cooperating by investing own wealth to bestow benefits on others (i.e., social fungibility) is not part of a Nash equilibrium for a self-interested, rational player unless one explicitly considers heterogeneity or introduces social components in preferences. By contrast, we study a game where full cooperation is a Nash equilibrium even if players are homogeneous and self-interested wealth is not socially fungible, and others' past conduct is kept private.

There is also mixed evidence on how externally-imposed payoff inequality affects behavior. In a real effort task, (Ku and Salmon, 2012) show that exogenously imposed wage inequality leads to discouragement for disadvantaged players but does not encourage advantaged players. In strategic settings externally-imposed payoff inequality affects efficiency and the division of surplus. Goeree and Holt (2000) find that differences in fixed payments that should not theoretically alter behavior induce in laboratory bargaining games offers that are not consistent with Nash equilibrium but are consistent with a fair division of final payments. In a dynamic public goods experiment, Sadrieh and Verbon (2006) find no clear link between cooperation among partners and the exogenous distribution of rights to the surplus generated. Instead, Tavoni et al. (2011) shows that endowment inequality hampers the ability of people to coordinate on a cooperative outcome in a threshold public goods game. Just as in Goeree and Holt (2000), our design introduces heterogeneity that should theoretically be irrelevant, but instead empirically affect behavior. As in Tavoni et al. (2011), we find evidence that this heterogeneity significantly affects cooperation.

Fairness and inequality have also been investigated in non-strategic distributive choice experiments where a disinterested third-party (a spectator) must select a division of resources between two others. Some of this research finds support for "luck egalitarianism," the notion that actions are meant to smooth out outcome differences due to uncontrollable factors. In Konow (2000) spectators overwhelmingly choose an equal split between two agents when the allocation of resources is uncontrollable, and instead condition the division on subjects' effort when effort affects outcomes. However, other research finds that both controllable and uncontrollable factors are taken into account in distributive choices. In Mollerstrom et al. (2015), spectators condition the division of resources based on subject's risky choices even if these choices did not cause resource differences. In our study there are no disinterested spectators but, as in these experiments, the role assignment is a form of uncontrollable inequality. Frequent donors have limited earning opportunities but abundant opportunities to create wealth in the group, while frequent recipients face the opposite situation. Egalitarian subjects might thus attempt to address these uncontrollable differences by conditioning their choices on their role history as the game progresses, even if past role assignments do not affect the subjects' future economic prospects and in theory should not affect choices of payoff-maximizing players.

Finally, our study is situated in an experimental literature about how information influences the efficiency of outcomes. This literature is vast, as it straddles several research agendas, from the study of reputation (Camera and Casari, 2009; Bolton et al., 2005; Schwartz et al., 2000), to the study of transparency and communication (Ellingsen and Östling, 2010; Huck et al., 2000; Isaac and Plott, 1981), from the impact of payoff asymmetries on cooperation (Andreoni and Varian, 1999; Chen and Gazzale, 2004) to the study of the role of information in market and strategic experiments (Kagel and Levin, 1986; Nagel, 1995; Roth and Malouf, 1979). Although there are elements of commonality with all of these research themes, our most direct contribution is to the last strand of research. First, we adopt a design where subjects cannot build a reputation in any treatment they remain strangers in all interactions and can never establish reciprocal relationships. Second, all treatments are designed so that the past conduct of an individual always remains opaque and players have no ability to communicate. Third, in all treatments cooperation symmetrically benefits players because there is always equal opportunity and players who face identical decisional situations have equal payoff matrices. When we manipulate the amount of information across treatments we find that less, not more, information is beneficial. This result is related to similar findings in market experiments and strategic bargaining games (see Smith, 1994, p. 119). Unlike those settings, however, we focus on information that is payoff-irrelevant, cannot disclose past conducts, nor can be used to build reputations. We find that the less informed players are about the distribution of past earning opportunities, the easier it is for them to coordinate on the efficient equilibrium. This pattern is not predicted by the standard application of folk theorem-type results because the efficient outcome is equally attainable in all informational settings. Yet, providing this information alters behavior in the experiment as subjects who have had few recipient roles, and thus have fallen behind in their earnings, tend to be less cooperative when they meet someone who appears to be ahead of them.

# 3 Experimental design

In our experiment, four subjects faced an indefinite sequence of "helping games." We adopt a 2x2 design where one factor is knowledge of relative past roles and the other is knowledge of relative past earnings. In all treatments: the socially optimal and selfish outcomes can be supported in equilibrium; subjects cannot build reputations or engage in reciprocity; past behavior cannot impact future opportunities or feasible outcomes. Table 1 provides a brief summary by treatment, while the specific details are discussed below.

Table 1: Sessions and treatments

| | **Treatment** | | | |
|---|---|---|---|---|
| **Variable** | *Baseline* | *Roles* | *Wealth* | *History* |
| Blue Index | No | Yes | No | Yes |
| Earnings Index | No | No | Yes | Yes |
| Subjects/Sessions | 64/4 | 64/4 | 64/4 | 64/4 |
| Superg./Avg. rounds | 80/18.5 | 80/18.4 | 80/18.5 | 80/19.6 |
| Salient \$ Earnings | | | | |
| average | 26.38 | 25.94 | 25.00 | 30.08 |
| min, max | 8.75, 54.00 | 6.50, 55.50 | 9.25, 54.00 | 10.00, 54.00 |

**Interaction in a round:** In each round of the experiment subjects face a helping game. Each game consists of a "donor" who is endowed with a good and a "recipient" who values the good more than the donor. The donor faces an individual decision problem: she can transfer the good to the recipient (*Help*), or she can consume the good (*Do nothing*). The recipient has no endowment and no action to take. All framing in the experiment was neutral. The structure of the game is in Table 1, while screenshots and instructions can be found in Appendix C.

Table 2: Payoff Matrix

| | **Donor** | |
|---|---|---|
| | *Help* | *Do nothing* |
| **Recipient** | $g, 0$ | $d - l, d$ |

**Notes:** In the experiment $g$=25, $d$=6, $l$=2 points. 1 point=$0.20.

If the donor helps, then the recipient earns $g$ while the donor earns nothing. Otherwise, both subjects earn a default payoff, which is higher for the donor; $d$ denotes the donor's default payoff and $d - l$ denotes the recipient's default payoff, with $g > 2d - l > 0$. There is *cooperation* in a meeting when the donor chooses to help the recipient; otherwise, there is *defection*. Given the payoffs, the donor's dominant action is to do nothing. Cooperation is *not* mutually beneficial but it is (socially) optimal as it maximizes surplus in the pair. In the experiment $d = 6$ points, $l = 2$ points, $g = 25$ points, where points is the experimental currency unit. The surplus from cooperation is $g - (2d - l) = 15$ points. The cost of cooperation to a donor is the difference in the two possible payoffs $0 - d = -6$ points; the benefit of cooperation to a recipient is her surplus $g - (d - l) = 21$ points; hence the benefit/cost ratio is 3.5.

**Baseline session and supergame:** A session involved 16 subjects in the lab at the same time, all exposed to the same treatment, which was chronologically divided into five distinct supergames. In a supergame, subjects interacted for an indefinite number of rounds in fixed matching groups of size four. Hence, there were four concurrent supergames being played in a session. In each group, subjects were randomly re-matched into pairs at the start of each round, so there was a 1/3 probability of meeting the same person in two consecutive rounds. Subjects did not know with whom they were paired nor did they know who was in their matching group in any supergame.

Every round, in each pair the computer randomly assigned the recipient role to one subject ("blue," in the experiment), and the donor role to the other ("red"), with equal probability. Hence, in every round half the subjects where recipients and half were donors. The random assignment of roles is a shock that affects the subject's earning potential for the round because recipients have a superior earning potential (25 points vs 6 points). This shock ensures equal economic opportunity going forward because payoff matrix and role assignment process are fixed. As a result, in a fully cooperative outcome future earning prospects are identical across individuals and rounds, and are completely unaffected by differences in past roles. Yet, the random assignment of roles provides an exogenous source of variation in cumulative earnings, and so it is likely to generate unequal economic results over the course of the supergame. As the supergame progressed, some participants could be recipients more often than others, thus having more chances of getting the higher payoff of 25 if cooperation occurred.

The duration of the supergame was uncertain because it was determined by a random continuation rule (as in Roth and Murnighan, 1978). A supergame began with 15 fixed rounds after which successive rounds occurred with probability $\beta = 0.75$. This continuation probability can be interpreted as the discount factor of a risk-neutral subject. *A priori*, the expected duration of a supergame was 18 rounds because from round 15, in each round the supergame is expected to last 3 more rounds. At the end of each round a computer drew an integer number between 1 and 100 with equal probability, which was then revealed to all subjects. A draw equal to or below 75 informed subjects that the supergame would continue (otherwise, it would end). [2]

At the end of each round, subjects observed whether or not the outcomes were identical in both pairs of their group. This form of anonymous public monitoring allows public detection of deviations from a social norm, and it could also simplify coordination tasks, but it does not allow agents to identify opponents (see screenshots from instructions in Appendix C). Hence, because individual histories remained private, subjects could neither build a reputation nor engage in relational contracting. Public monitoring ensured that the minimum discount factor supporting full cooperation in equilibrium was invariant across treatments (see next section).[3]

Supergames terminated simultaneously for all concurrent groups. After each of the first four supergames, subjects were placed into new four person matching groups and began playing another supergame. Matching groups were constructed so that no one was ever in a group with someone else more than once. Subjects were aware of this fact and, as a result, we have twenty unique groups per session. At the conclusion of a session, one supergame was selected randomly (Sherstyuk et al., 2013) and subjects were paid based upon their earnings in that supergame at the rate of \$0.20 per point.

[2]This number could also serve as a public coordination device, at the group level.

[3]Subjects had access to information about past outcomes of every match in which they were involved. Each subject had pen and paper at their station.

**Roles treatment:** In this treatment donors observed explicit information about inequality before making their choice. At the start of any round after the first round in a supergame, one can measure the proportion of past rounds in which a subject was a recipient (we call this the *recipient rate*). Unequal recipient rates give rise to inequality in past earnings, especially when cooperation rates are high given the greater spread in points.

In the Roles treatment, before making a choice, donors observed the normalized recipient rate for each group member. This information was called the "blue index" as it conveyed information about how often players had been in the blue role. The donor observed her blue index, the paired recipient's index, and the index of the two others in a random order. To facilitate comparisons, the average relative frequency of 0.50 was normalized to 100, so a value of $100 + x$ indicated a $x\%$ departure from the average.

Adding this index neither expands the action set relative to Baseline, nor affects payoffs in the stage game. The index expands the strategy set, as donors can condition their choice on the provided information, in rounds $t \geq 2$. Otherwise, the treatment is identical to Baseline. In particular, index values (i) neither yielded points nor could be redeemed for points or dollars, and (ii) masked the identity of donors and preserved anonymity because they were not associated with individual identifiers and were unobservable to recipients.

**Wealth and History treatments:** In the Wealth treatment, donors received information about the distribution of "wealth" defined as the subject's running total of points earned as the supergame progressed. Since the mean running total varied from round to round, this information was presented in relative form, through an index with the round mean normalized to 100 ("earnings index" in the experiment). A donor would observe her own relative wealth, the relative wealth of the matched recipient, and of the other two people in her group. In the History treatment, donors saw both the "blue" and the "earnings" indices. Recipients could never observe these indices.

**Procedural Details:** We recruited a total of 256 subjects through announcements at the University of Arkansas. All subjects recruited had no previous experience with this type of game.[4] After giving informed consent, subjects were seated at private terminals. Neither communication nor eye contact was possible among subjects at any time during the session. The experimenter publicly read the paper instructions at the start of the experiment, which were then left on the subjects' desks. The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007). On average, a session lasted 94 rounds for a running time of approximately 120 minutes including instructions, a paid post-instruction comprehension quiz, and post-experiment payment. Average earnings were \$27.00 per subject ($\text{min} = \$6.50, \text{max} = \$55.50$) excluding a \$5 fixed participation payment and an average of \$2.10 ($\text{min} = \$.75, \text{max} = \$2.50$) from providing correct answers to the post-instruction comprehension quiz (\$0.25 for each of 10 questions). Only one randomly selected supergame from the session was paid.

# 4 Theoretical considerations

In our setup, players benefit from cooperating fully. The experiment can shed light if full cooperation is ever achieved and, if not, what efficiency level is attained. Here we show that in every treatment, groups can theoretically attain multiple Pareto-ranked equilibria, which range from full defection (no

[4]About 55% of subjects were males, and the rest female. The subject pool is composed of about 90% undergraduate students with the remainder being primarily graduate students although some faculty, staff, and non-university associated people are in the pool. We include sex differences as controls in the econometric analysis.

donor ever helps) to full cooperation (every donor always helps). Defection is the unique Nash equilibrium in a one-shot interaction because help is costly to a donor (0 instead of 6 points). It follows that *Full defection* is always a sequential Nash equilibrium because it consists of an indefinite repetition of the one-shot Nash equilibrium. Payoffs in the supergame are minimized under *Full defection* and are maximized only under *Full cooperation*, which is also an equilibrium because subjects could observe whether or not choices differed in their group, in each round. This form of anonymous public monitoring can be exploited to build a rule of cooperation supported by a punishment convention that is triggered if the rule gets broken. Specifically, full cooperation can be supported as a (sequential) equilibrium if a subject helps from the beginning of the game whenever she is a donor, but switches to defection forever after someone defects (Kandori, 1992, Proposition 1). When everyone adopts this strategy, then we say that cooperation is a *social norm*. Here, any defection is punished with permanent defection by the entire group.

Cooperation is an equilibrium when two conditions apply: in equilibrium, every donor prefers to help; out of equilibrium *no* donor prefers to help. The latter condition is immediately verified: once someone moves off equilibrium, that deviation is publicly observed. Hence, every donor defects thereafter and there is no longer an incentive to cooperate. The first condition requires checking that a donor cannot improve her payoff by moving off equilibrium (unimprovability criterion). In Appendix A we prove that this is the case as long as players are sufficiently patient, if $\beta \geq \beta^* := \dfrac{2d}{g+l}$.

**Proposition 1.** *In our experiment, full defection and full cooperation can be supported as an equilibrium.*

The threshold value $\beta^*$ is the ratio between the cost of cooperation for a donor $d$ and the surplus difference expected next round, amounting to $\frac{g+l}{2}$. The parameter $\beta$ is the continuation probability of the game, 0.75. The condition $\beta \geq \beta^*$ is necessary and sufficient for the existence of a cooperative equilibrium. Based on the experimental design we have $\beta^* = 4/9$, so cooperation is an equilibrium in every treatment. Many other equilibria exist in all treatments, with efficiency degrees below that attainable under full cooperation and above that attainable under full defection. Equilibrium multiplicity gives rise to strategic uncertainty and equilibrium selection is an open question (see the discussion in Blonski et al., 2011). Full cooperation is Pareto dominant and so is the natural equilibrium for players to coordinate on, if efficiency is the selection criterion.[5] Full defection, however, is evolutionary stable (Camera et al., 2013). Hence, there is no guarantee that full cooperation is realized instead of a lower-efficiency equilibrium.

Three comments are in order. First, there is "equal opportunity." In each round, players' ex-ante earnings *potential* is governed by a payoff matrix and a role assignment process that are identical across players and fixed independent of past roles and past frequencies of cooperation.

Second, full cooperation supports income inequality ex-post because the realized sequences of donor and recipient roles inherently vary across subjects, as the game progresses. Disparities in individual histories of roles are an exogenous source of earnings variation in the experiment.[6] Although this factor

[5]A version of the concept of risk dominance in Blonski et al. (2011) can be adapted here to show that cooperation is risk dominant. The coefficient $\beta^*$ also depends on the assumption of linear preferences. One can show that with CRRA preferences of the type $u^{1-\gamma}/(1-\gamma)$ full cooperation remains an equilibrium if $\gamma \leq 0.37$. Estimates of CRRA coefficients vary widely depending on many factors; in experiments with a fixed recruitment fee, such as ours, we find a coefficient of 0.34 (Harrison et al., 2009), while in experiments with low stakes 60% of subjects lay below 0.41 (Holt ad Laury, 2002). Hence, full cooperation cannot be ruled out as an equilibrium even under empirically reasonable risk aversion.

[6]In a cooperative group it would be the only source of earnings variation. Using the roles

induces inequality in realized cumulative earnings, it is uncontrollable and does not alter the structure of incentives because it does not affect continuation payoffs in the efficient equilibrium.

Third, neither income inequality that is realized ex-post, nor the underlying factor that generates it, can alter the power structure in the game because high-income participants have no greater control over the earnings of others than low-income participants. By design ex-post inequality of opportunities is irrelevant for future payoffs and theoretically should not affect the structure of incentives. Given these considerations, we put forward a first testable hypothesis:

**Hypothesis 1.** *Players do not condition actions on their own history of roles.*

A key feature of our experimental design is that revealing role or earnings histories does not change important properties of the game. We state this explicitly in the following Proposition.

**Proposition 2.** *Revealing the distribution of past roles or earnings neither eliminates the equilibria that are possible in Baseline, nor alters the expected payoff in the efficient equilibrium or the parameter set supporting it.*

The proof of this claim is simple. In Role, Wealth and History treatments donors can condition their help on the information provided by the blue and earnings indices. Adding these indices increases the set of available strategies compared to Baseline, but does not expand the payoff set because the efficient outcome is an equilibrium in *all* treatments. Hence, even if the possibility to condition behavior on either the blue or the earning index alters the equilibrium set, this does not alter the Pareto efficiency frontier. Moreover, the Role,

sequences realized in the experiment, a counterfactual simulation reveals that full cooperation in Baseline would have generated a mean payoff 12.50 points per round, with a standard deviation of ±3 points amounting to an income gap of approximately ±56 points for the average supergame.

Wealth and History treatments do not eliminate *any* of the equilibria that are possible in Baseline; as players are not forced to use the information provided, they can always rely on strategies that ignore either index.

The second part of the statement follows from observing that the use of conditional strategies is neither necessary nor sufficient to sustain full cooperation. It is not necessary because defections are publicly revealed, so the efficient outcome can be attained in all treatments by conditioning choices on the outcomes seen in the group. It is not sufficient because either index masks the identity of counterparts, cannot be used to signal a cooperative intention, and does not reveal individual past conduct.

Summing up, the structure of incentives remains unaltered as we add the payoff-irrelevant information offered by the blue and earnings indices: the efficient outcome remains an equilibrium in all treatments and the return from cooperation is unchanged, also. This suggests that there is no obvious reason to expect significantly different behavior across treatments. Conditioning actions on information about past roles or earnings does not enhance prospective outcomes. Payoff maximizing players can easily coordinate on efficient play in all treatment by relying on public monitoring of defections in their four-person group. Conditioning actions on a (privately observed) statistic about past roles or earnings does not go in the direction of facilitating coordination on cooperation, and in fact may only contribute to increase coordination complexity. While payoff-maximizing players could employ such strategies there is no clear benefit from doing so and hence no reason to expect that subjects will use them, if available. We thus put forward two additional testable hypotheses:

**Hypothesis 2.** *Players do not condition actions on role histories of others.*

**Hypothesis 3.** *Players do not condition actions on earnings histories of others.*

As noted above, the experimental literature suggest that individuals may be driven by a mix of motives, including fairness or aversion to inequality in outcomes. Thus one may ask: would social norms exist that support *some* cooperation if players were driven to some extent by inequality or fairness concerns? Unfortunately, there is very little work on cooperation in indefinite-horizon games among heterogeneous, anonymous and randomly matched players. One related study is Camera and Gioffré (2017), which demonstrates that "asymmetric" social norm equilibria exist in which players defect when they are at a relative disadvantaged, and cooperate otherwise. In that study, there is an infinite sequence of PD games with payoff matrices that stochastically vary over time and can be occasionally asymmetric. As the norm calls for occasional defections, a simple public monitoring of actions cannot consent an accurate detection of *deviations*, hence the model loses the nice recursive structure of a full cooperation norm. Asymmetric cooperation must thus be supported by a contagious punishment process that is triggered by a privately observed deviation. It is conceivable that in our setting with observable past roles or wealth a similar scheme could support partial cooperation equilibria in which defections are tolerated under certain circumstances (e.g., being a frequent recipient) but should trigger a switch to a punishment mode in others. If a norm of this kind is indeed adopted, then adding blue or earnings indices should reduce average cooperation rates and coordination on full cooperation relative to Baseline. Moreover, defections should be less likely to trigger future punishment relative to Baseline because not every defection is a deviation. We thus put forward this hypothesis:

**Hypothesis 4.** *If inequality or fairness concerns drive behavior, then disclosing inequalities should result in declines in efficiency and full cooperation, but greater tolerance of defections.*

# 5 Results

This section is dividend into three parts. We start by documenting that when inequalities remained hidden (Baseline), donors conditioned their choices on their own past roles in the preceding rounds. Subsequently, we provide evidence that, when we revealed inequalities in past roles (Roles treatment), donors discriminated against advantaged players, which further hindered coordination on efficient play compared to the Baseline treatment. Finally, we discuss the robustness of these findings to explicitly revealing wealth inequalities as in the Wealth and History treatments.

## 5.1 Inequalities remain hidden

The Baseline experimental economies struggled to achieve the efficient outcome.

**Result 1** (Baseline). *Cooperation and efficiency increased over the course of a session, but seldom reached 100%. No group coordinated on the inefficient equilibrium, and 12.5% of groups coordinated on the efficient equilibrium.*

Support for this result is provided in Tables 3 and 4. By design, (realized) efficiency in a group corresponds to the mean cooperation rate of the four subjects composing that group (N=80 per treatment). A subject's cooperation rate corresponds to the proportion of cooperative choices the subject took as a donor in the supergame (N=320 per treatment).

Table 3: Realized Efficiency: Average and Thresholds

| | **Supergame** | | | | | | | **Realized efficiency** | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Treatment** | 1 | 2 | 3 | 4 | 5 | *Overall* | t=1 | ≤20% | ≥80% | 100% |
| Baseline | 0.44 | 0.50 | 0.61 | 0.67 | 0.60 | *0.56* | 0.63 | 9 | 21 | 10 |
| Roles | 0.30 | 0.50 | 0.52 | 0.52 | 0.52 | *0.47* | 0.59 | 14 | 11 | 2 |
| Wealth | 0.38 | 0.51 | 0.49 | 0.56 | 0.58 | *0.51* | 0.53 | 12 | 12 | 4 |
| History | 0.42 | 0.50 | 0.57 | 0.55 | 0.49 | *0.51* | 0.62 | 10 | 13 | 1 |

**Notes:** The unit of observation is a four-person group (N=16 per supergame, per treatment). Each cell reports the average proportion of cooperative choices in a supergame. The *t=1* column reports round 1 averages across supergames. The *Realized efficiency* columns report the number of groups that attained a given cooperation level (N=80 per treatment). Only one group achieved 0% cooperation in supergame 1 of the Wealth treatment.

The first row in Table 3 reports average realized efficiency by supergame. For Baseline data, this value lies between 44 and 67 percent; the average value is significantly different from 100 percent (one-tailed t-test, p-value $< 0.001$, $N = 80$). As a comparison, the mean cooperation rate in the first round of a supergame lies between 31 and 81 percent.

Models 1 and 2 in Table 4, reports marginal effects on the mean cooperation rate from a regression that pools data from all treatments.[7] The regression includes controls for treatment effects, as well as a standard set of individual and other controls (e.g., subject's self-reported sex and duration of the supergame). Model 2 includes a supergame regressor, while Model 1 includes a dummy variable for each supergame after the first, which is the base level. This allows us to trace how experience with the task affects cooperation.

[7] For a continuous variable, the marginal effect measures the change in the likelihood to cooperate for an infinitesimal change of the independent variable. For a dummy variable, the marginal effect measures the change in the likelihood to cooperate for a discrete change of the dummy variable from its base level (0).

Table 4: Realized Efficiency: Marginal Effects

| Dep. variable: | Cooperation rate | | | | =1 if 100% efficiency | |
|---|---|---|---|---|---|---|
| | Model 1 | | Model 2 | | Model 3 | |
| Treatment dummies | | | | | | |
| Roles | -0.121** | (0.057) | -0.099* | (0.054) | -0.097** | (0.044) |
| Wealth | -0.044 | (0.072) | -0.058 | (0.115) | -0.068 | (0.050) |
| History | -0.033 | (0.027) | -0.074 | (0.059) | -0.117*** | (0.046) |
| Supergame | | | 0.057*** | (0.011) | 0.059*** | (0.020) |
| Supergame dummies | | | | | | |
| Supergame 2 | 0.091*** | (0.033) | | | | |
| Supergame 3 | 0.185*** | (0.064) | | | | |
| Supergame 4 | 0.232*** | (0.037) | | | | |
| Supergame 5 | 0.214*** | (0.053) | | | | |
| Controls | Yes | | Yes | | Yes | |
| N | 320 | | 320 | | 320 | |

**Notes:** One observation is a group in a supergame (N=80 per treatment). Models 1-2: GLM Regression; the dependent variable is the relative frequency of cooperation in a group (N=80 per treatment); robust standard errors (S.E.) adjusted for clustering at the session level. Model 3: Logit regression; the dependent variable = 1 if group attained 100% cooperation, 0 otherwise. All regressions include interaction terms between treatment and supergame (or supergame dummies); *Controls* include supergame duration, current and previous (set to 18 rounds, in supergame 1), sex, two measures of understanding of instructions (response time and wrong answers in the quiz), and a self-reported measure of risk attitudes. Marginal effects are computed at the mean value of regressors of continuous variables. Symbols $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively.

There is evidence that cooperation significantly increased as subjects gained experienced with the task. The supergame regressor in Model 2 is positive and highly significant. All *Supergame* dummies in model 1 are positive and significant and there is evidence that cooperation significantly increased as subjects gained experienced with the first two supergames, and then stabilized.[8] This evidence is in accordance with what emerges from other studies on indefinitely repeated social dilemmas among strangers (Camera and Casari, 2009), and in contrast with the dynamics of cooperation observed under deterministic hori-

[8]This comes from a series of two-sided Wald test on the estimated coefficients. The coefficient on supergame 2 is statistically smaller than the other coefficients (p-values 0.031, <0.001 and <0.001 respectively). All other pairwise comparisons show coefficients that are similar to each other (p-value ranges from 0.365 to 0.585).

zons, as in that case cooperation tends to fall as subjects gain experience with the game (e.g., see Dal Bó, 2005; Palfrey, 1994).

As our design admits multiple Pareto-ranked equilibria, it is possible that the efficiency loss in Baseline is the result of different groups coordinating on different equilibria. As an example, if 45 out of 80 groups coordinate on the efficient outcome (100% cooperation) and the rest coordinate on full defection, then we obtain 56 percent average cooperation, and an efficiency loss of 44 percent. Baseline data do not support this conjecture and, in fact, reveal considerable heterogeneity in individual cooperation rates.

First, no group ever coordinated on the inefficient equilibrium: 7 groups out of 80 are below 20% efficiency, but no group reached 0%. On the other hand, 21 groups reached at least 80% efficiency, 10 of which attained full efficiency; see Table 3. This suggest that a sizable portion of subjects attempted to coordinate on the efficient outcome but failed because behavior was heterogeneous. As evidence, consider that 14.0% of subjects never cooperated, while 32.5% fully cooperated (N=320). As only 10 groups attained full efficiency, clearly some subjects always cooperated even when others in their group did not. This kind of heterogeneity has been observed in previous experiments with a similar design (Camera and Casari, 2014; Camera et al., 2013).

We report that heterogeneous behavior can be partly explained by inequalities due to the random role assignment. Occasional recipients were less likely to cooperate compared to frequent recipients.

**Result 2** (Baseline)**.** *Donors conditioned their choices on their own role history. The cooperation probability of occasional recipients is 11 percentage points lower than frequent recipients.*

We trace a subject's role history in round $t$ of a supergame using the *recipient frequency* $r_t = 0, \ldots, t-1$. It corresponds to the frequency of the subject's past recipient roles in that supergame. Hence, $r_t = 0$ for a donor who had not yet been a recipient, while for $t > 1$ we have $r_t = t-1$ for someone who just became a donor.[9] In this manner, for each round we can classify a donor according to her position in the distribution of the donors' recipient frequency for that round (N=160 per treatment). Denote as *occasional recipients* those in the bottom half, and as *frequent recipients* those in the top half.

The mean cooperation rate of donors declines from 0.62 to 0.50 as we move from frequent recipients to occasional recipients (Table 5). We ran a panel logit regression to determine if these differences in cooperation are statistically significant. We considered all rounds after the initial one. The dependent variable takes value 1 if the subject cooperated as a donor in a meeting, and is 0 if she defected. The panel variable is a subject in a session and the regression includes individual controls (see notes to Table 6). Supergame effects are soaked up using dummies, while round fixed effects are controlled for using a series of dummies.

Table 5: Average cooperation conditional on past roles

| | **Donor's Past Recipient Roles** | |
|---|---|---|
| **Treatment** | Occasional | Frequent |
| Baseline | 0.50 | 0.62 |
| Roles | 0.44 | 0.49 |
| Wealth | 0.44 | 0.57 |
| History | 0.46 | 0.54 |

**Notes:** One obs.=one donor in a round > 1 ($N = 160$ per round, per treatment). Each cell reports the average proportion of cooperative choices in the treatment.

[9] The recipient rate for the supergame can thus be normalized to 1 (0) for someone who was always (never) a recipient. The distribution of this statistic in Baseline data has mean and median equal to the expected value, 0.5, and standard deviation 0.12. Half of subjects lay between .41 and .59. The rate ranges between .19 and .87, meaning that in a supergame of 19 rounds, some were recipients in as few as 3 rounds or in as many as 16 rounds.

In Baseline donors *could not* observe any element of the history of others, but could see if someone had defected in the group. As discussed in Section 4, according to theory, a subject should stop cooperating once a defection is observed. To capture this trigger strategy, the regression includes six dummy variables that determine the impact of observing a defection in the group on the subsequent choices to help. The variable *Punishment grim trigger* takes value 1 starting the round after the subject suffers or observes a defection by another member of the four-person group for the first time (and is 0 otherwise). As reported in Camera and Casari (2009, 2014), subjects might delay their punishment response, thus we include the *choice n* dummy variables. They take the value 1 if the subject has already made $n-1 = 0, \dots, 4$ choices after the initial defection (and are 0 otherwise). The sum of the coefficients on the *Punishment grim trigger* and each *choice n* dummy identifies the average donor's reaction when it was the $n^{th}$ time she could make a choice after observing the defection. The coefficients on the *Punishment grim trigger* dummy itself captures the long-run response.[10] To determine if a donor's own role history impacts behavior, we include the dummy variables *occasional recipient*, which takes value 1 if it classifies the subject (and 0 otherwise). In this manner, *frequent recipient* is taken to be the base in the regression. Table 6 reports the marginal effects on the donor's probability of cooperating.

[10]Subjects could take choices at random points in time so these regressors allows us to trace the individual's behavior on the first five occasions in which she can make a choice, after suffering or observing an initial defection by another member of the four-person group. Given the random alternation of roles, these five regressors allow us to study how the reaction to an observed defection evolves in the long run because, empirically, the first opportunity to react to an observed defection occurs, on average, in round 4, and the following opportunities occur after an average of two additional rounds. For a detailed discussion on this econometric technique see Camera and Casari (2009, 2014).

Table 6: Past roles & cooperation: marginal effects

| Dep. variable: | Marginal effects | | |
|---|---|---|---|
| *=1 if donor helps* | Coeff. | | S.E. |
| *Own role history* | | | |
| Occasional recipient | -0.065 | *** | (0.014) |
| *Punishment regressors* | | | |
| Punishment grim trigger | -0.330 | *** | (0.035) |
| Choice 1 | 0.155 | *** | (0.024) |
| Choice 2 | 0.115 | *** | (0.033) |
| Choice 3 | 0.088 | *** | (0.022) |
| Choice 4 | 0.053 | ** | (0.022) |
| Choice 5 | 0.069 | *** | (0.020) |
| Controls | Yes | | |
| N | 2672 | | |

**Notes:** Logit panel regression with random effects. Baseline data only. Dependent variable = 1 if donor helps, 0 otherwise. One observation = choice of a donor in a round $> 1$ (N=160 for rounds $\leq 15$, and $8 \leq N \leq 160$ otherwise as not all supergames ended in the same round). Base case = donor was a *frequent recipient* in the period, i.e., in the top half of the distribution of recipient frequency in that period. *Controls* include dummies for supergames 2-4, round fixed effects through a series of dummy variables (a single dummy variable for rounds 19 and above), duration of previous supergame (set to 18 rounds, in supergame 1), two measures of understanding of instructions (response time and wrong answers in the quiz), a self-reported measure of sex and of risk attitudes. Marginal effects are computed at the mean value of regressors of continuous variables. Symbols $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively.

The results are broadly consistent with the notion that the average subject adopted a social norm of the kind identified in Section 4. The average donor conditioned her choice on the choices observed in the group, switching to a punishment mode after observing a defection. The *Punishment grim trigger* regressor is negative and highly significant. Each of the five sums of this coefficient and a *choice n* regressor remains negative and highly significant (Wald tests, p-values $< 0.001$). In contrast with the theory, we find that donors based their choices on *their own* role history. Those with infrequent opportunities to receive help in the past cooperated significantly less than the rest (see the *occasional recipient* coefficient).

An interpretation is that subjects acted to reduce their own exposure to unfavorable past earning shocks brought about by the random role assignment. They cooperated less when they had few past opportunities to receive help. This behavior is inconsistent with expected payoff maximization because, as we have seen, subjects cooperate much less after observing a defection. Hence, basing cooperation on inequalities in past roles can only reduce the chances to coordinate on high-payoff equilibria. We offer more evidence on this point in the next section, where we investigate what happened when donors could compare their role history to that of others, before making a choice.

## 5.2 Inequalities in past roles are observable

In the Roles treatment donors saw the counterparts' relative frequency of past roles (their *blue index*) before making a choice. The blue index is simply the recipient frequency, normalized so that in each round the mean is 100. These disclosures made salient inequalities in past role assignments, and allowed an easy interpersonal comparison in relative positions. However, these disclosures do not theoretically alter the structure of incentives compared to Baseline because they neither reveal the counterparts' past conduct, nor their future intentions. Consequently, payoff-maximizing players should not condition their actions on the relative blue indices in the pair (Hypothesis 2).

A first observation is that group cohesion suffered when we disclosed the relative positions in the distribution of past roles.

**Result 3** (Roles)**.** *Average efficiency fell 11 percentage points, and the probability of coordination on the efficient outcome fell 9 percentage points compared to Baseline.*

Support comes from Tables 3-4 and Figure 1. Realized efficiency is lower in Roles relative to Baseline in each supergame (see Table 3). Overall, the difference is 9 percentage points (0.56 vs. 0.47). The *Roles* dummy in Table 4 suggests that this difference is significantly different from zero (p-value=0.022).

What lies behind this efficiency decline? This result cannot be ascribed to differences in the allocation of roles in the two treatments. We can reject the hypothesis that the distribution of realized earning opportunities  randomly assigned by a computer program  differed across any of the treatments we ran. A test for equality of distribution functions finds no statistically significant difference between the underlying distributions of subjects' recipient rates in a comparison between Baseline and Roles (Epps and Singleton test, p-value=0.852, N=320 per treatment, one observation is one subject in a supergame). Table B2 in Appendix B reports statistics on inequality in opportunities experienced by donors, and its evolution over the supergame.

A second conjecture is that donors acted more uncooperatively in Roles partly out of a desire to reduce income inequality in their group. Coordinating on full defection, which minimizes per-capita income, would have eliminated almost all income inequality because recipients and donors' round earnings are similarly low (4 vs. 6). Yet, no group did so in the Roles treatment, suggesting that if fairness or inequality concerns drive behavior in the experiment, then subjects did not exploit the available information about differences in past opportunities to reduce income inequality in their group. Supporting evidence comes from analyzing differences in the Gini coefficients for income data across the two treatments. The average Gini coefficients for income data are statistically similar across the two treatments; see Table 7.[11]

[11] The average Gini coefficients for income are 0.121, 0.143, 0.118 and 0.117 respectively in Baseline, Roles, Wealth and History. One observation is the Gini measure for one group in a supergame (N=80 per treatment); income is measured as the average payoff of a subject in a supergame. In all treatments, income exhibits a higher degree of inequality than in counterfactual simulations were roles alternate as in the experiment but choices are imposed (varies from 0.012 to 0.019 across treatments). The average Gini for income drops

Table 7: Gini coefficient: Marginal Effects

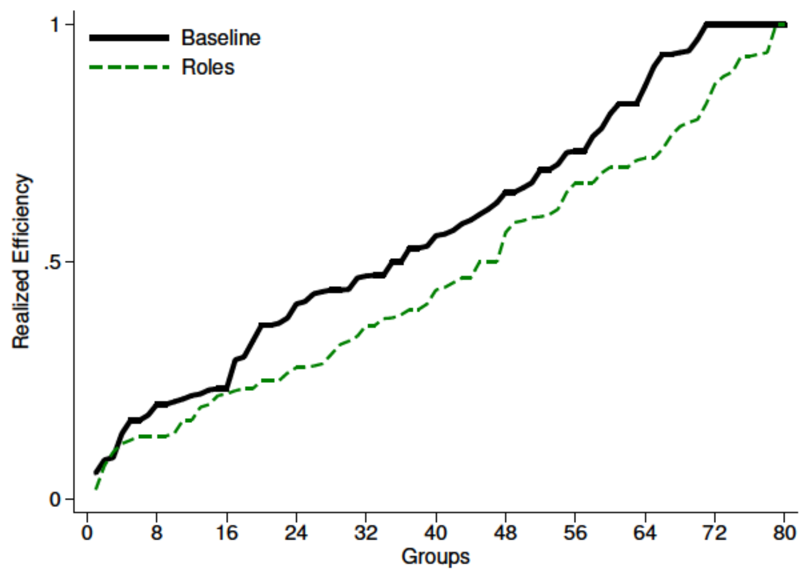
| Dep. variable: | Estimate | S.E. |
|---|---|---|
| Treatment dummies | | |
| Roles | -0.009 | (0.019) |
| Wealth | -0.022 | (0.027) |
| History | -0.005 | (0.019) |
| Supergame dummies | | |
| Supergame 2 | -0.006 | (0.023) |
| Supergame 3 | -0.013 | (0.015) |
| Supergame 4 | -0.018 | (0.025) |
| Supergame 5 | -0.012 | (0.013) |
| Controls | Yes | |
| N | 320 | |

**Notes:** One observation is a group in a supergame (N=80 per treatment). GLM Regression; dependent variable = Gini coefficient for income in a group; robust standard errors (S.E.) adjusted for clustering at the session level. Interaction terms and *Controls* as in Table 4 (only the duration variable is statistically significant). Marginal effects are computed at the mean value of regressors of continuous variables.

A third conjecture is that disclosing extant inequalities in the past assignment of roles influenced behavior by making more salient the differences in past economic opportunities. If players were driven at least to some extent by inequality or fairness concerns, then donors might have conditioned their cooperation on the recipient's index knowing that some defections would be tolerated. This would have impaired coordination on efficient play without triggering a global switch to punishment. Table 3 and Figure 1, lend some support to this view.

to between 0.021 and 0.024 (all treatments) in the counterfactual full-defection outcome, and to between 0.106 and 0.119 in the counterfactual full-cooperation outcome (donors always help); the average Gini coefficient for simulated data where choices are assumed random varies between 0.112 and 0.127.

Figure 1: Realized Efficiency (1 obs.=one group)




Recall that efficiency in a group is proportional to cooperation, therefore subjects interested in maximizing per-capita income should coordinate on full cooperation. The distribution function of realized efficiency in Baseline is first-order stochastically dominated by the Roles treatment (Figure 1). A pairwise treatment comparison rejects the hypothesis that the distributions of realized efficiency are similar (Epps and Singleton test, p-value<0.069, N=80 per treatment). Coordination on the efficient outcome is also less frequent in Roles compared to Baseline (2 vs. 10 groups), a statistically significant difference (Mann-Whitney test, p-value=0.017, N=80 per treatment).[12]

To uncover the source of this coordination decline we study individual

[12] We have also estimated a logit regression where the dependent variable takes value 1 if the group attained 100% cooperation (and 0 otherwise). The regression includes the same treatment dummies and controls in Table 3, and a supergame regressor. The coefficient on the Roles dummy is negative and significant (p-value=0.07). The result is robust to lowering the coordination threshold to 90% cooperation.

behavior. Donors conditioned actions on their own role histories similarly to Baseline (see Table 5). A panel logit regression reveals that donors who were occasional recipients reacted similarly to Baseline donors.[13] We do, however, find evidence that donors discriminated against recipients who had better luck in their past opportunities. They were more likely to defect when recipients had a relative advantage in past economic opportunities.

**Result 4** (Roles). *Donors conditioned their cooperation on the recipient's history of past economic opportunities. The cooperation probability fell 14 percentage points when recipients had a relative advantage in past opportunities, compared to when they did not.*

Evidence comes from the logit panel regression in Table 4. The dependent variable equals 1 if a donor cooperated (0 otherwise) in a round $> 1$. The regression includes the punishment dummies, and the controls discussed earlier. However, unlike Baseline, now donors could observe the role history of others before making their choice.

To capture the effect of this observable information on choices each donor-recipient pair is categorized based on their relative recipient frequencies, which is the exogenous source of variation in the experiment. This is done using the *blue index*. The index can be calculated in any treatment, even when not shown to the subjects (as in Baseline). An "advantaged" ("disadvantaged") subject has a blue index at or above average (strictly below average) and is denoted A (D). The four possible classifications are AA, AD, DA, and DD where the first letter indicates the donor's circumstances (and the second the recipient's).

[13]The econometric model is identical to the one used for the analysis of Baseline data in Table 6. Marginal effects are reported in Table B1 Appendix B. The punishment regressors indicate a highly significant, and negative response to observation of a defection. The *Occasional recipient* regressor is negative, significant but statistically smaller than the corresponding coefficient in Baseline ( two-sided Wald test on the estimated coefficients from a stacked regression; p-value=0.092).

Based on this classification, we have four possible dummy variables. The regressor AD=1 if the donor was advantaged and the recipient was not, i.e., the donor had a *relative* advantage. Conversely, the donor has a relative disadvantage when the regressor DA=1. The donor is on equal footing with the recipient if DD=1 or if AA=1, which we take as the base level. Table 8 reports the distribution of the cases AA, AD, DA, and DD for all treatments pooled together (as it is nearly identical across treatments); the evolution in a supergame is reported in Figure B2 in Appendix B.

Table 8: Average cooperation conditional on past roles

| | **Classification of Pairs** | | | |
|---|---|---|---|---|
| **Rounds in supergame** | DD | DA | AD | AA |
| Rounds $\leq 9$ | 0.15 | 0.30 | 0.30 | 0.25 |
| Rounds $\geq 10$ | 0.14 | 0.27 | 0.28 | 0.31 |
| All rounds | 0.15 | 0.29 | 0.29 | 0.28 |

**Notes:** One obs.=one donor in a round $> 1$ (N=10,928). Each cell reports the average proportion of the four possible classifications are AA, AD, DA, and DD in the treatment. AA observations include the case in which both subjects have an index of 100, which are increasingly more frequent as the supergame evolves because heterogeneity subsides.

Table 9 reports the marginal effects on the probability of observing cooperation in a pair for Baseline and, separately, Roles data. Two observations stand out. First, donors actively discriminated against advantaged recipients. The coefficients on the DA and DD regressors are negative in both treatments but are both significant only in Baseline, while in Roles the DD regressor is not (p-value=0.459). This is evidence that, when they could, donors conditioned choices on information about role inequalities in the pair.[14]

[14]DA and DD are statistically similar in Baseline (p-value=0184), which is what we should expect given that donors could not observe the counterpart's role history.

Table 9: Inequality & cooperation: marginal effects.

| Dep. variable=1 if donor helps | Baseline | | Roles | |
|---|---|---|---|---|
| AD | -0.000 | | -0.001 | |
| | (0.019) | | (0.019) | |
| DA | -0.090 | *** | -0.071 | *** |
| | (0.020) | | (0.021) | |
| DD | -0.057 | ** | -0.022 | |
| | (0.025) | | (0.025) | |
| Punishment regressors | Yes | | Yes | |
| Controls | Yes | | Yes | |
| N | 2672 | | 2680 | |

**Notes:** Logit panel regression with random effects. Dependent variable = 1 if donor helps, 0 otherwise. One observation is a choice in a round > 1 in which the subject was a donor. Base case = donor and recipient have a blue index $\geq 100$. A=advantaged subject with a blue index equal to or above average, D=disadvantaged subject with blue index below average (=100). Punishment dummies include the same variables in Table 6. *Controls* include dummies for supergames 2-4, round fixed effects through a series of dummy variables (a single dummy variable for rounds 19 and above), duration of previous supergame (set to 18 rounds, in supergame 1), two measures of understanding of instructions (response time and wrong answers in the quiz), a self-reported measure of sex and of risk attitudes. Marginal effects are computed at the mean value of regressors of continuous variables. Symbols $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively.

Since DD is not significant in Roles, one could interpret this as evidence that providing information on past roles made things better, not worse. However, this is not so for a couple of reasons. First, pairs of disadvantaged players are less likely to be formed than other pairs (see Table 8, as they are in a recipient role less than average) so the lack of a decline in cooperation in DD matches observed in Roles did little to improve coordination on cooperation over Baseline. Second, the regression reveals that the converse behavior is not true: donors did not act more cooperatively with recipients known to be at a relative disadvantage. The AD coefficient is statistically insignificant in both treatments, and statistically similar. One may have expected that advantaged players would have lowered cooperation with disadvantaged players, as a way to punish for discriminating them. Our design hinders this kind of negative reciprocity. While donors saw the recipient's index, the converse is not true, which is probably a reason why advantaged donors did discriminate against the disadvantaged.

Summing up, disadvantaged donors indiscriminately reduced their cooperation in Baseline, but instead specifically targeted advantaged recipients in Roles. These findings support the view that revealing information about the distribution of past opportunities hindered the group's ability to coordinate on efficient play due to inequality or fairness concerns driving behavior. The behavior of donors before and after they suffered or observed a defection, provides additional pieces of evidence in support of this view.

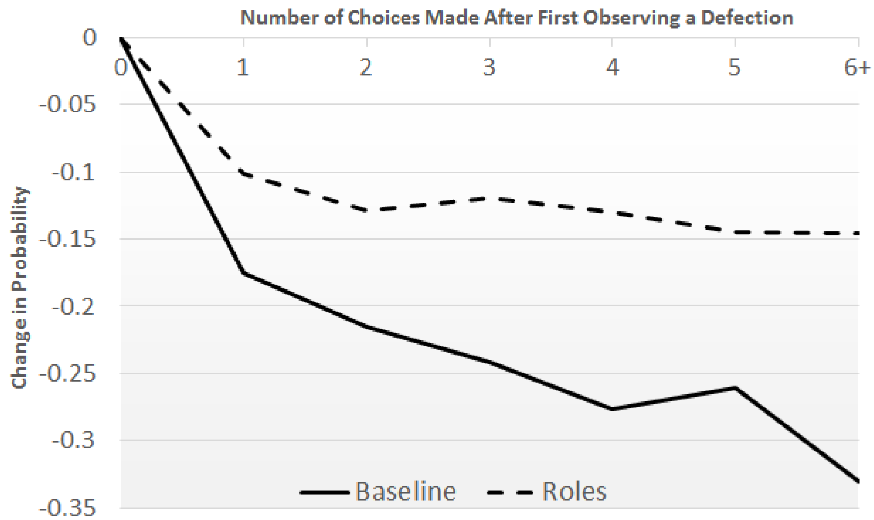
**Result 5** (No defection observed)**.** *Donors who neither suffered nor observed a defection cooperated less in Roles as compared to Baseline.*

**Result 6** (Tolerance of defections)**.** *Suffering or observing a defection triggered a persistent decline in cooperation for the average donor. This decline is smaller in Roles as compared to Baseline.*

In support of Result 5, the average cooperation rate of donors who have neither suffered nor seen a defection in their group (other than their own, possibly) is 0.77 in Baseline and it significantly falls to 0.61 in Roles (p-value$<$0.001, see regression in Table B5 in Appendix B). This is evidence that revealing the inequalities in past opportunities reduced the average subject's cooperativeness, even if the other three components of the group had always cooperated.

Figure 2 provides support for Result 6. It traces the estimated change in cooperation probability of the average donor, the $n^{th}$ time she was a donor after suffering or observing a defection; it is based on the stacked regression in TableB1, Appendix B.

Figure 2: Cooperation Decline After a Defection: Marginal Effects




**Notes:** Change in cooperation probability of the average donor on the $n^{th}$ time she chose after suffering or observing a defection. To each point $n = 1, 2, \ldots, 5$ corresponds the sum of the estimated coefficient on *Punishment grim trigger* regressor and *Choice n* regressor; the point "6+" is associated with the *Punishment grim trigger* regressor coefficient. A series of Wald tests performed point by point, allows us to reject only the hypothesis that the (sums of) the coefficients are equal. The corresponding stacked regression is in TableB1.

Two elements stand out. First, subjects who suffered a defection or saw others defect permanently reduced their cooperation. The first choice made (point 1 on the horizontal axis) shows a significant drop in the cooperation probability for each treatment. We see no reversion to the original cooperation levels even after the subject had five opportunities to make a cooperative choice (10 periods on average from the initial defection). The *Punishment grim trigger* regressors are all negative and highly significant (Table B1 in Appendix B). Second, when inequalities in past opportunities were disclosed donors became more "tolerant" of defections. The initial cooperation decline in Roles is significantly smaller than in Baseline (cross-treatment pairwise

Wald tests on the sum of the regression coefficients "Choice 1" and "Public Grim Trigger"). Moreover, moving right on the figure, the magnitude of the cooperation decline increases significantly only in Baseline, while in Roles the curve remains statistically flat (Table B1 Appendix B).

Did the discriminatory behavior vary as rounds went by? One can imagine that donors react less at the beginning of a game either because an initial disadvantage can be overturned in the continuation game, or because "bad luck" must be persistent to become salient. We only have evidence that in Roles disadvantaged donors cooperated less in rounds 14 and above compared to the supergame's earlier rounds (see Model 2 in Table B4 Appendix B). Therefore, we cannot exclude that disadvantages in past opportunities could have been more salient later in the supergame when there was little time left for the opportunities' balance to overturn.

Summing up, the Roles treatment confirms that frequent donors cooperated significantly less than frequent recipients, as happened in Baseline. However, when inequalities in past opportunities were disclosed, donors cooperate less even if no one else had defected in their group, discriminated against the more fortunate players, and did not sanction defections as sharply as in Baseline. This evidence is consistent with the view that inequality or fairness concerns might have driven behavior. This kind of behavior does grant an immediate redistribution of earnings (6 points instead of 0 for the donor, 4 points instead of 25 for the recipient), but ultimately backfires because it damages the subject's earning prospects in the continuation game since uncooperative actions triggered a long-lasting negative response. Moreover, the greater tolerance of defections did not help support higher overall cooperation in Roles compared to Baseline because cooperation was lower even when player had not witnessed anyone else defecting. Given this, basing cooperation on inequalities in past roles simply reduced the chances to coordinate on equilibria that deliver high expected payoffs.

### 5.3 Robustness: inequalities in earnings are observable

The Roles treatment reinforces the view that subjects might have sometimes acted uncooperatively to reduce their exposure to unfavorable realizations of past economic opportunity. Disparities in past opportunities are a proxy for cumulative earnings (or, wealth), as they strongly correlate with subjects' wealth in the data; the correlation between wealth and recipient frequency is 0.377 in Baseline, 0.372 in Roles (one observation is one subject in a supergame, N=320 per treatment). If so, then revealing inequalities in past earnings should generate a behavioral response similar to the one observed in Roles.

We test this hypothesis through the Wealth and History treatments, which make salient inequalities in wealth. In the first, donors observe relative past earnings of everyone in the group (earnings index). The History treatment combines this information with that about relative positions in past roles (blue index). The data offer evidence that information about economic inequalities again obstructed long-run cooperation.

**Result 7** (Wealth & History)**.** *Average realized efficiency declined and coordination on the efficient outcome was less frequent compared to Baseline.*

Table 3 suggests that realized efficiency fell compared to Baseline. Models 1 and 2 in Table 4 offer mixed evidence on the significance of this decline. Though the coefficients on Wealth and History are not statistically significant, they are statistically similar to the Roles coefficients but for one comparison (Model 1, Roles vs History, p-value=0.054).

Coordination on the efficient outcome is less frequent in Wealth and History compared to Baseline: 4 and 1 groups vs. 10 groups, respectively (see Table

3). These differences are significant according to Mann-Whitney tests (p-values 0.094 and 0.051, respectively, N=80 per treatment). We have also estimated a logit regression where the dependent variable takes value 1 if the group attained 100% cooperation (and 0 otherwise); see Model 3 in Table 4. The coefficient on the Wealth and History dummies are each negative, and although only the second is significant (p-values 0.156 and 0.016, respectively) the three treatment coefficients are all statistically similar (according to Wald tests on according to the three pairwise comparisons).[15]

Result 7 reinforces the view that disclosing economic inequalities in the group had an adverse effect on efficiency and reduced the chance of coordination on efficient outcomes.

**Result 8** (Wealth & History)**.** *Donors with a disadvantage in past economic opportunities cooperated less and discriminated against advantaged recipients.*

Evidence comes from logit regressions. The econometric model treats the blue index (which is a source of exogenous wealth variation) as an instrument for the earning index (which is correlated with the donor's past actions). Marginal effects are reported in Table 10. As seen earlier, the behavior of advantaged donors is not affected by information about economic inequalities. The AD coefficients are statistically insignificant, and statistically similar across all treatments (six pairwise Wald tests of coefficients estimated through a stacked regression not reported). However, disadvantaged donors cooperated less and reacted to information about the recipient.[16]

[15] We cannot reject the hypothesis that the distribution of realized efficiency in Baseline is similar to that in Wealth or History (Epps and Singleton tests, p-value= 0.665 and 0.248, respectively, N=80 per treatment; see Figure B1 in the Appendix B).

[16] We also ran logit regressions, based on the specification in Table 6. Donors who were *occasional recipients* were significantly less cooperative than donors who were not, in both Wealth and History (marginal effects are in Table B1, Appendix B). The coefficient in Baseline is statistically similar to those in Wealth and History (Wald tests of stacked regression coefficients, p-values are 0.499 and 0.257, respectively).

Table 10: Inequality & cooperation: marginal effects

| Dep. variable =1 if donor helps | **Wealth** | | **History** | |
|---|---|---|---|---|
| *Role history: donor & recipient* | | | | |
| AD | 0.026 | | 0.023 | |
| (0.019) | | (0.017) | | |
| DA | -0.115 | *** | -0.056 | *** |
| | (0.021) | | (0.019) | |
| DD -0.082 | *** | -0.010 | | |
| | (0.025) | | (0.023) | |
| Punishment regressors | Yes | | Yes | |
| Controls | Yes | | Yes | |
| N | 2720 | | 2856 | |

**Notes:** See notes to Table 9.

Disadvantaged donors discriminated against advantaged recipients: they cooperated less when the recipient was seen to be disadvantaged. The evidence is as follows. The cooperation probability fell if donors had a relative disadvantage (case DA), but did not fall or fell by less otherwise (case DD). We can reject the hypothesis that the DA coefficients are similar in Wealth and History (Wald tests from a stacked regression, p-value=0.035) and that the DD coefficients are similar (p-value=0.024). In History, the DD coefficient is insignificant and different from the DA coefficient (Wald test, p-value=0.033); it is also statistically similar to the coefficient associated with Roles data (Wald test for regression estimates from stacked regression). In Wealth (where the blue index of the recipient was hidden), the coefficient on the DD regressor is negative, significant but we cannot reject the hypothesis that is different from the coefficient on the DA regressor (Wald test, p-value=0.150).

This last finding for the Wealth treatment requires a bit of additional scrutiny because on the one hand the DD coefficient is not statistically different from the DD regressor in Roles (Wald test for regression estimates from stacked regression, p-value=0.148); and on the other hand donors could not clearly differentiate between advantaged and disadvantaged recipients, as the earning index is positively but not perfectly correlated with the blue index. This might explain the weaker contrasts in the discriminatory strategy adopted by disadvantaged donors in Wealth. To sort this out, we ran additional regressions to examine the behavior of donors classified as being disadvantaged. The regressions provide evidence that donors did discriminate against the advantaged also in Wealth. For all treatments except Baseline, we can reject the hypothesis that disadvantaged donors cooperated identically when the recipient was advantaged; see Table B4 in Appendix B. Finally, Result 5 and Result 6 also holds for Wealth and History.[17]

To conclude there is evidence that disadvantaged donors indiscriminately reduce their cooperation in the Baseline treatment. Instead, in the treatments where advantaged and disadvantaged participants can be differentiated, participants who have an unlucky streak of past economic opportunities follow discriminatory strategies, specifically targeting recipients who are more fortunate.Though it is possible that this discriminatory strategy was used out of a desire to reduce income inequality in the group, the regression in Table 7 does not support this hypothesis; the average Gini coefficients for income data are statistically similar across all treatments.

These findings are consistent with the view that revealing information about inequality in past opportunities and wealth made inequality more salient, which in turn hindered the group's ability to coordinate on efficient play due

[17]In Appendix B, see Table B5 for the equivalent of Result 5. The average donor who has neither suffered nor seen a defection in their group (other than their own, possibly) cooperates 12 to 10 percentage points less in Wealth and History compared to Baseline (p-value=0.084 and 0.014, respectively). For the equivalent of Result **??** see Figure B3 and Table B1 in Appendix B. The point-by-point pairwise comparison of the reaction curves in Figure B3 is statistically significant for points 4 and above in Baseline vs Wealth, and for points 2 and above in Baseline vs History.

to inequality or fairness concerns driving behavior. Though this behavior is inconsistent with maximization of ex-ante payoffs, it may be seen as a rational attempt to maximize ex-post payoffs. When economic opportunities are random, some subjects may have greater *ex-post* payoffs under full defection than full or partial cooperation. Though the data supports this intuition, it also shows that few would have benefitted and those that would benefit would only do so by a small amount. We calculated counterfactual payoffs for each subject, under full cooperation and full defection, using their empirical sequence of roles. Full defection would have generated a slight ex-post payoff increment (0.8 points/round) for a few instances (132/1280)  those with especially long donor sequences; however, this is dwarfed by the 4.4 average loss in ex-post payoff for the remaining 90% of subjects.

# 6 Discussion

The experiment induced inequality of opportunities ex-post, but not ex-ante. Though this inequality is theoretically inconsequential, it reduced cooperation and coordination on efficient play in groups of four participants. This suggests that economic inequalities present a *behavioral* obstacle to a society's cohesion and prosperity.

In every group, players could maximize their present-valued expected payoff by helping one another. Random shocks  over which players had no control  ensured equal *future* earning potential, while inducing variation in *past* earning opportunities, and thus realized earnings. Inequalities in past shocks could neither alter the power structure in the group, nor the future earning potential or the expected return from cooperation. Hence, these inequalities are theoretically neutral if we consider the structure of incentives of payoff-maximizing players. However, we find a clear influence on behavior in the data.

Players conditioned their choices on their own past opportunities, acting uncooperatively when they had few favorable economic opportunities in the past. Inequality disclosures compounded the problem, giving rise to discriminatory behavior: players took to defecting with counterparts known to have had better past opportunities. Simply put, the average participant in the experiment became less cooperative as a result of repeated random assignment to the disadvantaged role. An interpretation is that people acted with the apparent intent to counteract unfavorable past economic opportunities, in relation to others in their group.

This type of backward-looking non-strategic behavior has not been documented before in indefinite-horizon dilemmas and it is consistent with experimental results from finite-horizon settings (Loch and Wu, 2008; Sonnemans at al., 1999). It is remarkable because it cannot improve the player's earning potential in the continuation game and, in fact, it is likely to backfire. First, if past opportunities cannot alter payoff matrix and future assignment of opportunities (as in our experiment), then earning potential and return from cooperation in the continuation game are theoretically unaffected by variation in past opportunities. Hence, there is no economic incentive to alter behavior based on past luck. Second, reacting to bad luck by defecting can only lower the future earning potential if free-riding is sanctioned with future defections; empirically, the average participant reacted to an observed defection by cooperating less in the continuation game, which is consistent with the use of a norm of community punishment. Third, there is no reason to worry about past opportunities off-equilibrium, as defection is the theoretical best response.

A possible behavioral explanation is that participants were unwilling to follow a norm of mutual support when the associated benefit did not reflect *their* contribution to the prosperity of others. In our setup frequent donors had limited earning opportunities but abundant chances to increase others' fortunes. The opposite holds true for frequent recipients. Subjects might have thus attempted to smooth these differences by conditioning their choices on their past roles as the game progressed. One could also interpret this behavior as consistent with the "luck egalitarianism" observed in non-strategic settings (Konow, 2000; Mollerstrom et al., 2015).

# References


Abreu, D., D. Pierce and E. Stacchetti. 1990. Toward a Theory of Discounted Repeated Games with Imperfect Monitoring. *Econometrica*, 58, 1041-1063.

Aghion, P. and J. Williamson. 1998. *Growth, Inequality, and Globalization.* New York: Cambridge University Press.

Anderson, Lisa R., Jennifer M. Mellor, and Jeffrey Milyo. 2006. Induced heterogeneity in trust experiments. *Experimental Economics* 9, 223-235

Andreoni, James and Hal Varian. 1999. Preplay contracting in the Prisoners' Dilemma. *Proceedings of the National Academy of Sciences of the United States of America* 96(19) 10933-10938.

Blanco, Mariana, Dirk Engelmann and Hans Theo Normann. 2011. A within-subject analysis of other-regarding preferences. *Games and Economic Behavior* 72(2), 321-338

Blonski M., P. Ockenfels and G. Spagnolo. 2011. Equilibrium Selection in the Repeated Prisoner's Dilemma: Axiomatic Approach and Experimental Evidence. *American Economic Journal: Microeconomics*, 3(3), 164-92.

Bolton, G.E, Ockenfels, A., 2000. ERC: a theory of equity, reciprocity and competition. *American Economic Review* 90, 166-193.

Bolton, Gary E., Elena Katok, and Axel Ockenfels. 2005. Cooperation among strangers with limited information about reputation. *Journal of Public Economics* 89 (2005) 1457-1468.

Camera, G. and M. Casari. 2009. Cooperation among strangers under the shadow of the future. *The 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., Casari, M., and Bigoni, M. 2013. Money and trust among strangers. *Proceedings of the National Academy of Sciences of the United States of America*, 110(37), 14889-14893.

Camera, G., and A. Gioffré. 2017. Asymmetric social norms. *Economics Letters*, 152, 27-30.

Chen, Yan and Robert Gazzale. 2004. When Does Learning in Games Generate Convergence to Nash Equilibria? The Role of Supermodularity in an Experimental Setting. *American Economic Review* 94(5), 1505-1535.

Crawford, Vincent P. 2016. New Directions for Modelling Strategic Behavior: Game-Theoretic Models of Communication, Coordination, and Cooperation in Economic Relationships. *Journal of Economic Perspectives* 30 (4), 131-150.

Dal Bó, P. 2005. Cooperation under the Shadow of the Future: Experimental Evidence from Infinitely Repeated Games. *American Economic Review* 95(5), 1591-1604.

Dannenberg, A., Sturm, B., and Vogt, C. 2010. Do Equity Preferences Matter for Climate Negotiators? An Experimental Investigation. *Environ Resource Econ*, 47(1), 91-109

Deck, C. 2001. A Test of Game Theoretic and Behavioral Models of Play in Exchange and Insurance Environments. *American Economic Review* 91 (5), 1546-1555.

Ellingsen, Tore, and Robert Östling. 2010. When Does Communication Improve Coordination? *American Economic Review* 100, 1695-1724

Fehr, Ernst and Klaus M. Schmidt. 1999. A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics* 114(3), 817-868

Gangadharan, L., N. Nikiforakis and M. C. Villeval. 2015. Equality Concerns and the Limits of Self-Governance in Heterogeneous Populations. Manuscript.

Goeree, Jacob K., and C. A. Holt. 2000. Asymmetric inequality aversion and noisy behavior in alternating-offer bargaining games. *European Economic Review* 44(4), 1079-1089.

Greiner, B., A. Ockenfels, and P. Werner. 2012. The Dynamic Interplay of Inequality and Trust   An Experimental Study. *Journal of Economic Behavior and Organization* 81, 355-365.

Harrison, Glenn W. , Morten I. Laub, and E. Elisabet Rutström. 2009. Risk attitudes, randomization to treatment, and self-selection into experiments. Journal of Economic Behavior & Organization 70(3), 498-507

Holt, Charles. A, and Susan K. Laury. 2002. Risk Aversion and Incentive Effects. *American Economic Review* 92(5), 1644-1655

Huck, S., H.T. Normann, and J. Oechssler. 2000. Does Information about Competitors' Actions Increase or Decrease Competition in Experi mental Oligopoly Markets? *International Journal of Industrial Organization* 18, 39-57.

Isaac, R. Mark, and Charles R. Plott. 1981. The Opportunity for Conspiracies in Restraint of Trade. *Journal of Economic Behavior and Organization* 2, 1-31.

Kagel, John H. and Dan Levin. 1986. The Winner's Curse and Public Information in Common Value Auctions. *American Economic Review*, 76(5), 894-920

Kagel, John H., and K. Willey-Wolfe. 2001. Tests of fairness models based on equity considerations in a three-person ultimatum game. *Experimental Economics* 4(3), 203-219.

Kandori, Michihiro. 1992. Social norms and community enforcement. *Review of Economic Studies*, 59, 63-80.

E. O. Kimbrough, V. Smith and B. J. Wilson. 2008. Historical Property Rights, Sociality, and the Emergence of Impersonal Exchange in Long-Distance Trade. *Am. Econ. Rev.*, 98(3), 1009-1039

Konow, James, 2000. Fair shares: accountability and cognitive dissonance in allocation decisions. *American Economic Review* 90, 1072-1091.

Ku, Hyejin, Timothy C. Salmon. 2012. The Incentive Effects of Inequality: An Experimental Investigation. *Sourthern Economic Journal*, 79(1), 46-70

Loch, Christoph H. and Yaozhong Wu. 2008. Social Preferences and Supply Chain Performance: An Experimental Study. *Management Science* 54(11), 1835-1849,

Mollerstrom J., B. Reme, and Erik Sørensen. 2015. Luck, choice and responsibility An experimental study of fairness views. *Journal of Public Economics* 131, 33-40

Nagel, Rosemarie. 1995. Unraveling in Guessing Games: An Experimental Study. *American Economic Review*, 85(5), 1313-1326

Nishi, A. H. Shirado, D. G. Rand, and N. A. Christakis. 2015. Inequality and visibility of wealth in experimental social networks. *Nature*.

D.C. North, 1991. Institutions. *J. Econ. Persp.*, 5(1), 97-112.

New York Times. Inequality Troubles Americans Across Party Lines. By Noam Scheiber and Dalia Sussman, June 3, 2015

Palfrey, T.R. and Rosenthal, H., 1994. Repeated Play, Cooperation and Coordination: An Experimental Study. *The Review of Economic Studies*, 61(3), 545-565.

Pew Research Center. 2014. Emerging and Developing Economies Much More Optimistic than Rich Countries about the Future.

Piketty, Thomas. 2014. *Capital in the Twenty-First Century.* Cambridge, MA: Harvard University Press.

Putnam, Robert. 2000. *Bowling Alone: The Collapse and Revival of American Community.*

Rabin, Matthew. 1993. Incorporating Fairness into Game Theory and Economics. *American Economic Review* 83, 1281-1302.

Roth, Alvin E. and Michael W. Malouf. 1979. Game-theoretic models and the role of information in bargaining. *Psychological Review*, 86(6), 574-594.

Roth, Alvin E., and Keith Murnighan. 1978. Equilibrium behavior and repeated play of the prisoner's dilemma. *Journal of Mathematical Psychology*, 17, 189-98

Sadrieh, A., and H.A. Verbon. 2006. Inequality, cooperation, and growth: an experimental study. *European Economic Review* 50, 1197-1222.

Schwartz, Steven T., Richard A. Young, and Kristina Zvinakis. 2000. Reputation without Repeat Interaction: A Role for Public Disclosures. *Review of Accounting Studies*, 5(4), 351-75.

Sherstyuk, Katerina, Nori Tarui, and Tatsuyoshi Saijo. 2013. Payment schemes in infinite-horizon experimental games. *Experimental Economics*16 (1), 125-153.

Smith, V. 1994. Economics in the Laboratory. *Journal of Economic Perspectives* 8(1), 113-131.

Sonnemans, Joep, Arthur Schram, Theo Offerman. 1999. Strategic behavior in public good games: when partners drift apart. *Economics Letters* 62, 35-41

Stiglitz, Joseph E. 2012. *The Price of Inequality: How Today's Divided Society Endangers Our Future*. W. W. Norton & Company, New York.

Tavoni, Alessandro, Astrid Dannenberg, Giorgos Kallis, and Andreas Löschel. 2011. Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *Proceedings of the National Academy of Sciences of the United States of America*, 108 (29), 11825-11829.

# Appendix A: Proof of Proposition 1

This analysis is based on the existence of equilibrium proof in Camera et al. (2013). In each round $t = 0, 1, 2\ldots$ individuals in the group are matched in pairs, with uniform probability of selection. In each pair, a computer randomly determines who is the donor and who is the recipient (with equal probability). If cooperation (=Help) is the outcome, then $g$ is the payoff to the recipient and for generality let $a$ denote the payoff to the donor. If defection (=Do nothing) is the outcome, then $d$ is the payoff to the donor and $d - l$ to the recipient. Round payoffs are geometrically discounted at rate $\beta \in (0, 1)$ starting from round $n > 0$.

The equilibrium payoff (=expected lifetime utility) at $t = 0$ is

$$v(n) := (n+1) \times \frac{g+a}{2} + \sum_{j=1}^{\infty} \beta^j \times \frac{g+a}{2} \;=\; \frac{g+a}{2} \times \left(n + \frac{1}{1-\beta}\right).$$

A player is a donor or a recipient with equal probability in each round, hence expects to earn $\frac{g+a}{2}$ in each round. The payoff $v(n)$ is increasing in $n$ because payoffs are discounted by $\beta$ in rounds $t \geq n$.

The equilibrium payoff in the continuation game starting on any date $t \geq 0$, *before* any uncertainty is resolved, corresponds to

$$V_t = \begin{cases} v(n-t) & \text{if } \; t < n \\ v^* := \dfrac{g+a}{2(1-\beta)} & \text{if } \; t \geq n. \end{cases}$$

The equilibrium payoff of a donor at the start of any date $t$ is

$$V_{dt} = \begin{cases} a + v(n-t-1) & \text{if } \; t < n \\ a + \beta v^* & \text{if } \; t \geq n. \end{cases}$$

We must check that in equilibrium donors have no incentive to defect; out of equilibrium, donors have no incentive to cooperate.

Defection is the dominant action off-equilibrium; i.e., it is always individually optimal to punish after a defection from equilibrium play is made public. To see this suppose a donor deviates by helping off equilibrium. She would earn $a$ instead of $d$ but her continuation payoff would not improve since everyone else keeps defecting as prescribed by the rule of punishment. Since $d > a$, it is optimal to punish off equilibrium.

In equilibrium, cooperation is a best response in every round $t = 0, 1, \ldots$, if $V_{dt} \geq \hat{V}_{dt}$. The left-hand-side denotes the payoff to a donor who cooperates;

the right-hand-side denotes the donor's payoff when she moves off equilibrium under a one-time deviation. Such deviation is publicly observed, hence  when everyone follows the cooperative strategy  every donor will always defect in the future. The payoff to the deviator is thus

$$\hat{V}_{dt} = \begin{cases} \hat{v}(n-t) := d + (n-t)\dfrac{2d-l}{2} + \beta\dfrac{2d-l}{2(1-\beta)} & \text{if } \ 1 \le t < n \\ \hat{v}^* := d + \beta\dfrac{2d-l}{2(1-\beta)} & \text{if } \ t \ge n \end{cases}$$

Now define

$$\Delta_t = V_{dt} - \hat{V}_{dt} = a - d + \frac{g+a-2d+l}{2} \times \begin{cases} n - t + \dfrac{\beta}{1-\beta} & \text{if } \ t < n \\ \dfrac{\beta}{1-\beta} & \text{if } \ t \ge n \end{cases}$$

The minimum value of $\Delta_t$ is achieved for $t \ge n$. The implication is that cooperation is individually optimal in all rounds $t$ whenever

$$\beta \ge \beta^* := \frac{2(d-a)}{g+l-a}.$$

# Appendix B: Supplementary Tables & Figures (not for publication)

Figure B1: Realized Efficiency: all treatments (1 obs.=one group)

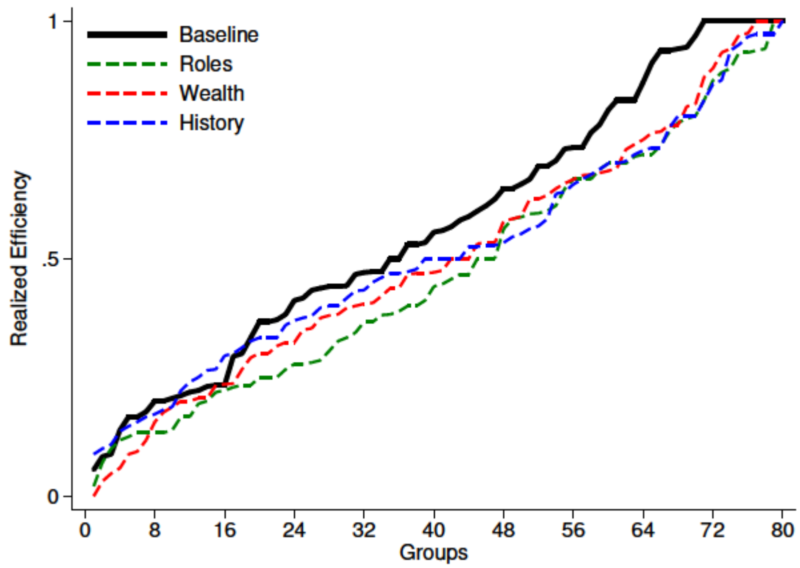


Table B1: Past roles & cooperation: marginal effects (all treatments)

| Dep. variable: *=1 if donor helps* | Baseline | Roles | Wealth | History |
|---|---|---|---|---|
| *Own role history* | | | | |
| Occasional recipient | -0.065*** | -0.042*** | -0.090*** | -0.040*** |
| | (0.014) | (0.014) | (0.015) | (0.012) |
| *Punishment regressors* | | | | |
| Punishment grim trigger | -0.330*** | -0.146*** | -0.250*** | -0.249*** |
| | (0.035) | (0.033) | (0.038) | (0.034) |
| Choice 1 | 0.155*** | 0.045 | 0.122*** | 0.134*** |
| | (0.024) | (0.029) | (0.026) | (0.024) |
| Choice 2 | 0.115*** | 0.017 | 0.098*** | 0.123*** |
| | (0.023) | (0.028) | (0.025) | (0.022) |
| Choice 3 | 0.088*** | 0.027 | 0.068*** | 0.095*** |
| | (0.022) | (0.025) | (0.024) | (0.021) |
| Choice 4 | 0.053** | 0.016 | 0.062*** | 0.067*** |
| | (0.022) | (0.024) | (0.022) | (0.020) |
| Choice 5 | 0.069*** | 0.001 | 0.074*** | 0.066*** |
| | (0.020) | (0.023) | (0.020) | (0.019) |
| Controls | Yes | Yes | Yes | Yes |
| N | 2672 | 2680 | 2720 | 2856 |

**Notes:** Logit panel regressions with random effects. Dependent variable = 1 if donor helps, 0 otherwise. One observation is a choice of a subject who was a donor in a round > 1 of a supergame (N=160 per treatment if rounds $\leq 15$, and $8 \leq N \leq 160$ in subsequent rounds since not all supergames ended in the same round). Base case = donor was a *frequent recipient* in the period, i.e., in the top half the distribution of recipient frequency of all donors in that treatment, in that period. *Controls* include dummies for supergames 2-4, round fixed effects through a series of dummy variables (a single dummy variable for rounds 19 and above), duration of previous supergame (set to 18 rounds, in supergame 1), two measures of understanding of instructions (response time and wrong answers in the quiz), self-reported measures of sex and of risk attitudes. Marginal effects are computed at the mean value of regressors of continuous variables. Marginal effects are computed at the mean value of regressors of continuous variables. Symbols $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively.
A series of Wald tests performed on the *Occasional recipient* coefficients on these stacked regressions allows us reject only the hypothesis that the coefficients are equal for the pairwise comparison Roles vs Wealth (p-value=0.023), and History vs. Wealth (p-value=0.013).

Table B2: Donors: distribution of past recipient roles & role persistence

| | Percentile | | | | |
|---|---|---|---|---|---|
| **Donor statistics** | 10 | 25 | 50 | 75 | 90 |
| Past recipient roles (N) | | | | | |
| Rounds 1-9 | 0 | 1 | 2 | 3 | 4 |
| Rounds $\geq$10 | 4 | 5 | 7 | 8 | 10 |
| Consecutive donor roles (N) | | | | | |
| Rounds 1-9 | 1 | 1 | 1 | 2 | 3 |
| Rounds $\geq$10 | 1 | 1 | 2 | 3 | 4 |

**Notes:** One observation = one donor in a round > 1 ($N = 160$ per round). First two rows: each cell reports the number of past rounds in which the subject had a recipient role. Last two rows: each cell reports the number of consecutive rounds in which the donor has been in that role (1=just switched role, 2=switched to being a donor in the previous round, etc.). No appreciable difference exist across treatments.

Figure B2: Evolution of Classification of Pairs (1 obs.=one group in a period)

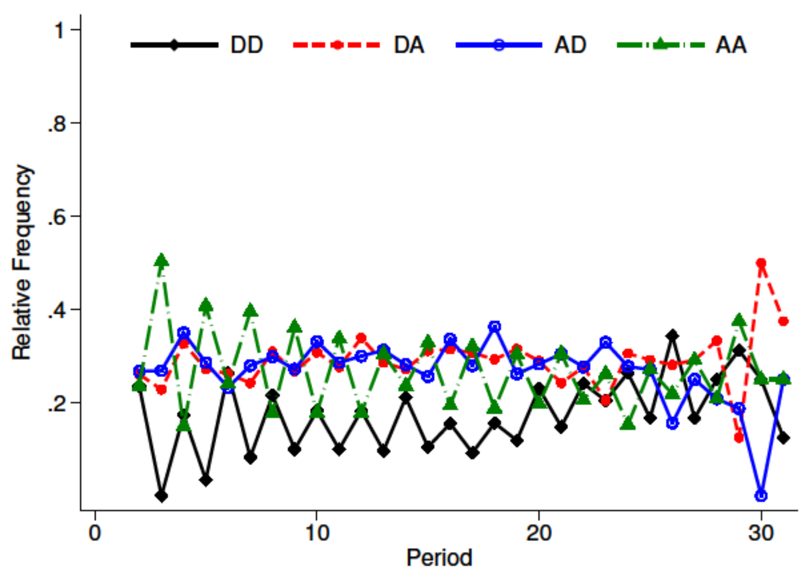


Table B3: Inequality & cooperation: marginal effects (all treatments).

| Dep. variable: *=1 if donor helps* | Baseline | Roles | Wealth | History |
|---|---|---|---|---|
| *Role history: donor & recipient* | | | | |
| AD | -0.000 | -0.001 | 0.026 | 0.023 |
| | (0.019) | (0.019) | (0.019) | (0.017) |
| DA | -0.090*** | -0.071*** | -0.115*** | -0.056*** |
| | (0.020) | (0.021) | (0.021) | (0.019) |
| DD | -0.057** | -0.022 | -0.082*** | -0.010 |
| | (0.025) | (0.025) | (0.025) | (0.023) |
| Punishment regressors | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes |
| N | 2672 | 2680 | 2720 | 2856 |

**Notes:** Logit panel regression with random effects. Dependent variable = 1 if donor helps, 0 otherwise. One observation is a choice in a round > 1 in which the subject was a donor. Base case = donor and recipient have a blue index $\geq$ 100. A=advantaged subject with a blue index equal to or above average, D=disadvantaged subject with blue index below average (=100). *Controls* include dummies for supergames 2-4, round fixed effects through a series of dummy variables (a single dummy variable for rounds 19 and above), duration of previous supergame (set to 18 rounds, in supergame 1), two measures of understanding of instructions (response time and wrong answers in the quiz), self-reported measures of sex and of risk attitudes. Marginal effects are computed at the mean value of regressors of continuous variables. Marginal effects are computed at the mean value of regressors of continuous variables. Symbols $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively.

Table B4: Disadvantaged donors: marginal effects

| Dep. variable: | Baseline | | Roles | | Wealth | | History | |
|---|---|---|---|---|---|---|---|---|
| *=1 if donor helps* | coeff. | | coeff. | | coeff. | | coeff. | |
| **Model 1** | | | | | | | | |
| Advantaged recipient | -0.033 | | -0.038 | * | -0.035 | * | -0.069 | *** |
| | (0.026) | | (0.021) | | (0.020) | | (0.026) | |
| Rounds regressors | Yes | | Yes | | Yes | | Yes | |
| Punishment regressors | Yes | | Yes | | Yes | | Yes | |
| Controls | Yes | | Yes | | Yes | | Yes | |
| **Model 2** | | | | | | | | |
| Advantaged recipient | -0.025 | | -0.039 | * | -0.034 | * | -0.067 | *** |
| | (0.024) | | (0.021) | | (0.020) | | (0.026) | |
| *Rounds regressors* | | | | | | | | |
| rounds 6-9 | 0.071 | * | 0.047 | | -0.030 | | 0.087 | ** |
| | (0.037) | | (0.035) | | (0.036) | | (0.041) | |
| rounds 10-13 | 0.075 | * | -0.005 | | 0.020 | | 0.091 | * |
| | (0.041) | | (0.040) | | (0.043) | | (0.049) | |
| rounds $\geq$14 | 0.054 | | -0.084 | ** | -0.046 | | 0.047 | |
| | (0.045) | | (0.040) | | (0.045) | | (0.052) | |
| Punishment regressors | Yes | | Yes | | Yes | | Yes | |
| Controls | Yes | | Yes | | Yes | | Yes | |
| N | 1146 | | 1160 | | 1216 | | 1185 | |

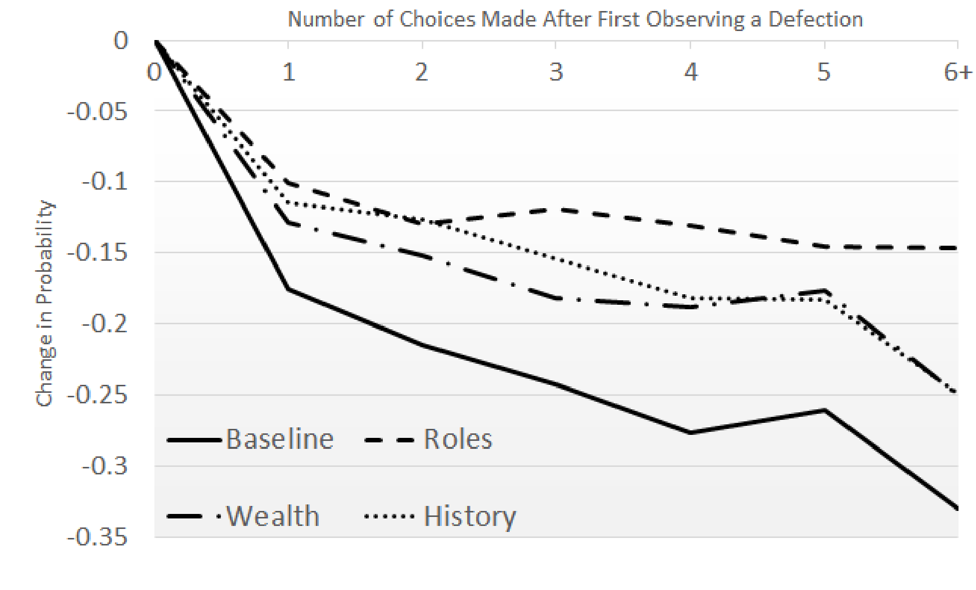
**Notes:** Logit panel regression with random effects. Dependent variable = 1 if donor helps, 0 otherwise. One observation is a choice in a round > 1 in which the subject was a Disadvantaged donor (a subject with blue index below average 100). Model 1 includes the same round regressors as the model in Tables 9-10 while models 2 and 3 differ in the size of the intervals. *Advantaged*=1 if donor meets an advantaged recipient with a blue index (visible only in Roles and History) equal to or above average (0, otherwise). The *rounds* regressors take value 1 for rounds in the specified interval (0 otherwise); a supergame lasted about 18 rounds on average: it included 15 rounds for sure plus 3 expected additional rounds, in each round that followed. Hence we include three dummies for 4-round time intervals (2-5, 6-9, 10-13) and one for the interval comprising rounds 14 and above; the three initial intervals capture approximately 24% of the observations each. Controls include dummies for supergames 2-4 and the other standard controls. Marginal effects are computed at the mean value of regressors of continuous variables. Symbols $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively.

Table B5: Choice before observing or experiencing a defection: marginal effects

| Dep. variable: | Marginal effects | | |
|---|---|---|---|
| *=1 if donor helps* | Coeff. | | S.E. |
| *Treatment dummies* | | | |
| Roles | -0.619 | *** | (0.147) |
| Wealth | -0.226 | * | (0.131) |
| History | -0.318 | ** | (0.129) |
| *Supergame dummies* | | | |
| Supergame 2 | -0.147 | | (0.101) |
| Supergame 3 | 0.087 | | (0.079) |
| Supergame 4 | 0.064 | | (0.076) |
| Supergame 5 | 0.086 | | (0.078) |
| *Rounds regressors* | | | |
| Rounds 6-9 | -0.057 | * | (0.033) |
| Rounds 10-13 | -0.040 | | (0.039) |
| Rounds 14+ | -0.123 | *** | (0.043) |
| Controls | Yes | | |
| N | 1704 | | |

**Notes:** Logit panel regression with random effects. Dependent variable = 1 if donor helps, 0 otherwise. One observation is a choice of a donor in a round. Only data about rounds where subject has not yet suffered or observed a defection in the supergame (she might have defected herself). In the econometric model the supergame dummies (Base case = supergame 1) and the 5-round interval dummy variables (Base case = rounds 1-5) are interacted with the three treatment dummies. *Controls* include duration of previous supergame (set to 18 rounds, in supergame 1), two measures of understanding of instructions (response time and wrong answers in the quiz), a self-reported measure of sex and of risk attitudes. Marginal effects are computed at the mean value of regressors of continuous variables. Symbols $***$, $**$, and $*$ indicate significance at the 1%, 5% and 10% level, respectively.

Figure B3: Cooperation Decline After a Defection: Marginal Effects




**Notes:** The figure traces the change in cooperation probability of the average donor on the $n^{th}$ time she could make a choice after observing the defection. Each point $n = 1, 2, \dots, 5$ on the horizontal axis is associated with the sum of two coefficients: the estimated coefficient on the *Punishment grim trigger* regressor and the *Choice n* regressor; the point "6+" is associated with the *Punishment grim trigger* regressor only. A series of Wald tests performed on the *Occasional recipient* coefficients on these stacked regressions allows us reject only the hypothesis that the (sums of) the coefficients are equal. The estimated coefficients for each treatment are reported in TableB1.