

جامعة نيويورك أبوظبي

 NYU | ABU DHABI



Cooperation Under the Shadow of Inequality

James Bland, Olivier Bochet, Nikos
Nikiforakis and Huanren Zhang

Working Paper # 0093

December 2023

Division of Social Science Working Paper Series

New York University Abu Dhabi, Saadiyat Island P.O Box 129188, Abu Dhabi, UAE

<http://nyuad.nyu.edu/en/academics/academic-divisions/social-science.html>

Cooperation under the shadow of inequality*

James Bland^a, Olivier Bochet^{b,c}, Nikos Nikiforakis^{b,c} and Huanren Zhang^d

Abstract

Cooperation often entails an unequal distribution of benefits. We study how inequality concerns affect the willingness to cooperate with others in an indefinitely repeated prisoner's dilemma. The experimental treatments vary the equality of payoffs resulting from mutual cooperation, the expected duration of an interaction, and whether the inequality remains constant throughout an interaction. At the aggregate level, we find that cooperation rates across treatments are accurately predicted by a model that assumes players solely care about their pecuniary payoffs. At the individual level, we find evidence that individuals care about treating others fairly, but not about inequality per se.

Keywords: cooperation, inequality, infinitely repeated prisoner's dilemma

JEL Codes: C72, C73, C91, C92

* The authors gratefully acknowledge the research assistance of Alexandros Agiostratitis and Hamna Kahn. We wish to thank Marco Casari, Tim Cason, Aurélie Dariel, Guillaume Fréchette, Ernesto Reuben and John Wooders for helpful comments, as well as participants in the 12th International Conference of the French Association of Experimental Economics in Lyon, France, and the 2022 International Symposium for Experimental Economics at NYU Abu Dhabi, UAE. Olivier Bochet and Nikos Nikiforakis gratefully recognize financial support by Tamkeen under the NYU Abu Dhabi Research Institute Award CG005. The project was approved by the IRB at NYU Abu Dhabi (#016-2019).

^a Department of Economics, University of Toledo, 2801 Bancroft St, Toledo, Ohio 43606, USA. James.Bland@utoledo.edu;

^b Division of Social Science, New York University Abu Dhabi, P.O. Box 129188, Abu Dhabi, UAE.

^c Center for Behavioral Institutional Design, New York University Abu Dhabi, P.O. Box 129188, Abu Dhabi, UAE. ob16@nyu.edu, nikos.nikiforakis@nyu.edu.

^d Department of Economics, University of Southern Denmark, Odense, Denmark. huanren@sam.sdu.dk

1. Introduction

The theory of infinitely repeated games suggests that the prospects of cooperation improve as the expected duration of an interaction increases. This is because the long-term benefits derived from cooperating can outweigh the short-term gains obtained from defection (Friedman 1971, Fudenberg & Maskin 1990, Fudenberg et al. 1994). Empirical evidence from laboratory experiments supports this proposition showing that the presence of a credible threat of punishment under the “shadow of the future” helps limit opportunistic behavior (Aoyagi & Fréchette 2009, Camera & Casari 2009, Dal Bo 2005, Dal Bo & Fréchette 2011, Duffy & Ochs 2009, Normann & Wallace 2012).

Previous studies, both theoretical and experimental, have predominantly focused on the specific scenario of symmetric prisoner's dilemmas in which mutual cooperation generates equal gains for all players (Dal Bo & Fréchette 2018). An open question is whether inequality concerns can undermine cooperation prospects within infinitely repeated interactions. This question is of obvious importance considering that cooperation frequently leads to unequal distributions of benefits in everyday life. Notable examples include countries enjoying disparate gains from alliances and treaties, and firms deriving different gains from forming cartels.

This paper aims to fill this gap in the literature by examining how inequality concerns impact cooperation in indefinitely repeated games. The influence of inequality on cooperation rates is a priori unclear. Generally, if individuals have a strong aversion to inequality or believe that others do, the “shadow of inequality” could diminish cooperation rates. Indeed, ample experimental evidence from one-shot games suggests that some individuals exhibit a dislike for unequal payoffs (e.g., Bolton & Ockenfels 2000, Fehr & Schmidt 1999). Additionally, evidence indicates that payoff inequality influences equilibrium selection in one-shot interactions (Bland & Nikiforakis 2015; Chmura et al. 2005). Evidence from other studies, however, suggests that inequality concerns can carry less weight in interactions of indefinite duration.

Experimental findings indicate that efficiency concerns tend to dominate those for equality when the two are at odds (Balafoutas et al. 2012, Cabrales et al. 2010, Charness & Rabin 2002, Engelmann & Strobel 2004, Faravelli et al. 2013, Fisman et al. 2007). Therefore, it is possible that inequality concerns play a diminished role when the expected duration of interactions is sufficiently high as efficiency concerns loom larger. Similarly, individuals seem to be less concerned about inequality when it is exogenously determined (Falk et al. 2008). Fairness preferences also do not appear to predict individual choices in symmetric indefinitely repeated prisoner dilemmas when cooperation can be supported in equilibrium (Dreber et al. 2014, Davis et al. 2016).

To the best of our knowledge, ours is the first experiment to explore the evolution of cooperation under the “shadow of inequality” in indefinitely repeated games.¹ Research in *finitely* repeated games indicates that cooperation rates can be lower in treatments in which cooperation leads to unequal payoffs (Gangadharan et al. 2017, Nikiforakis et al. 2012, Reuben and Riedl 2013).² Apart from involving multiple players and allowing for different degrees of cooperation that can affect behavior (Gangadharan & Nikiforakis 2009), these experiments do not allow one to determine whether the reduction in inequality is due to inequality concerns *per se* or due to reduced incentives for the “disadvantaged” players to cooperate. Further, as the theory of infinitely repeated games highlights, incentives to cooperate differ substantially in finitely and indefinitely repeated interactions. It is therefore unclear whether inequality concerns undermine cooperation in indefinitely repeated games.

In the next section, we present the experimental design, consisting of seven treatments varying the equality of payoffs resulting from mutual cooperation, the expected duration of interactions, and whether the inequality remains constant throughout a given interaction or varies across rounds. In Section 3, we derive predictions for the impact of our experimental treatments on cooperation. The predictions differ if players are assumed to care about inequality in payoffs or not. In Section 4, we present our experimental findings from a sample of 770 participants. We show that, on aggregate, behavior across treatments is accurately predicted by the model in which individuals are assumed to care solely about their pecuniary payoffs. Individual-level analysis reveals a more nuanced picture. Specifically, individuals revealed to care more about others in a modified dictator game are less likely to defect on others, irrespective of whether mutual cooperation implies unequal earnings or not. That is, individuals appear to care about fairness, but not about inequality *per se*. Finally, Section 5 concludes with a discussion of topics for future research.

2. The experiment

2.1 Experimental design

The experiment consists of seven treatments in a between-subjects design. Six of these treatments explore all possible combinations of varying (*i*) the payoffs of the stage game (EQ-H vs. EQ-L vs. UNEQ), and (*ii*) the probability with which an interaction ends ($\delta = 0.1$ and $\delta = 0.8$). For these six treatments, participants are randomly assigned roles at the start of the experiment (either the row or column player), which they retain for the duration of the experiment.

¹ Camera et al. (2020) show that *past* inequalities can affect an individual’s willingness to behave prosocially towards another player in an indefinitely repeated helping game with random re-matching in every round. In their setting, since roles are reassigned in each round, in expected terms, future earnings are equal.

² See Fischbacher et al. (2018) for an exploration on the impact of inequalities on conditional cooperation in one-shot public good games.

Figure 1 presents the payoff matrices used for the stage games in each of the experimental treatments. As shown, treatment EQ-H (EQ-L) corresponds to the stage game in which mutual cooperation results into equally high (low) payoffs for individuals. While EQ-H and EQ-L are symmetric prisoner’s dilemma games, treatment UNEQ introduces payoff asymmetries between the column and row players. Specifically, in UNEQ mutual cooperation results into unequal earnings for the row and column player, whereas players receive the same payoffs as the row player and the column player as in EQ-H and EQ-L otherwise.

Figure 1. Payoff matrices for treatments Equal-High (EQ-H), Unequal (UNEQ) and Equal-Low (EQ-L)

	C	D		C	D		C	D
C	9,9	0,10	C	9,6	0,10	C	6,6	0,10
D	10,0	3,3	D	10,0	3,3	D	10,0	3,3
	EQ-H			UNEQ			EQ-L	

We also explore behavior in a seventh treatment, UNEQ-Alt, in which participants indefinitely play the UNEQ stage game with a continuation probability of $\delta = 0.8$ but in which, unlike in the other treatments, they swap roles between the row and column players in every round of a match. Specifically, participants are informed that swapping is deterministic, with the row and column player changing roles in every round, except in the first round in which the assignment to row or column is randomly determined. Note that in the UNEQ-Alt treatment, mutual cooperation leads to unequal payoffs in any given round – specifically, the row player will earn 50% more than the column player, as in UNEQ – but (roughly) equal earnings if the game lasts long enough.

At the start of an experimental session, the subjects are randomly grouped into 10-person “silos”. Participants are randomly re-matched before each match of indefinite duration within their respective silos. To allow for learning, subjects play 10 matches of an indefinitely repeated prisoner’s dilemma games with a group member randomly assigned to them at the start of each match. A 20-sided die is rolled in front of all the participants at the end of each round to determine the length of each match. The number drawn from the die determines whether the matched participants will get to play an additional stage game.

In order to explore the channels through which inequality may impact cooperation in the experiment, we use a method developed by Blanco et al. (2011) to obtain individual measures of a participant’s inequality aversion. Specifically, at the start of the experiment, participants play an Ultimatum game and a Modified Dictator Game which allow us to identify a person’s aversion to

disadvantageous and advantageous inequality (Fehr & Schmidt 1999).³ The order of the ultimatum and dictator game is randomly determined, but both games always preceded the indefinitely repeated prisoner dilemma. The justification is that we have no reason to anticipate that these measurements will affect the ranking of treatments in the main experiment, irrespective of individuals' actions. On the other hand, if these games followed the prisoner dilemma, the anticipated differences across treatments could have impacted behavior in the ultimatum and dictator games through wealth effects. Instructions for each game are distributed only after the previous game is finished.

In total, we recruited 770 participants, that is, 110 subjects in each of the seven treatments. Participants were all university students. The laboratory sessions were conducted at Purdue University (Vernon Smith Experimental Economics Laboratory) and the University of Valencia (Lineex laboratory). Although we do not observe any substantial or significant differences in cooperation across locations for any of the treatments, we also add location fixed effects in our regression analyses. On average, each session lasted 40 minutes ($\delta = 0.1$) or 70 minutes ($\delta = 0.8$). The average payment was around \$18.

2.2 Behavioral hypotheses

The standard approach for analyzing the prospects of cooperation in indefinitely repeated interactions is to calculate the critical value of the probability with which a bilateral interaction continues such that the “shadow of the future” i.e., the expected gains from future cooperation, equal (or exceed) the short run benefit from always defecting. We will follow this approach here too, but relax the behavioral assumptions to allow for the possibility that, apart from their pecuniary payoffs, some individuals may care about payoff equality.

To evaluate the impact of inequality on cooperation prospects, we will assume that individuals have Fehr-Schmidt preferences (Fehr and Schmidt, 1999). Letting x_i be the pecuniary payoff obtained by agent i and x_j be the pecuniary payoff obtained by agent j , preferences over outcomes can be written as shown below:

$$U_i(x_i, x_j) = \begin{cases} x_i - \beta_i(x_i - x_j) & \text{if } x_i \geq x_j \\ x_i - \alpha_i(x_j - x_i) & \text{if } x_i < x_j. \end{cases}$$

³ In the Ultimatum Game, the proposer is asked to allocate 20 points between himself and the responder, while the responder must decide the minimal offer she is willing to accept. If the offer is accepted, the proposed split is implemented. If it is rejected, then both players receive 0 points. In the Modified Dictator Game, the dictator is presented with 21 allocation problems. In each of the problems, the dictator needs to choose between either an unequal distribution (20 points for oneself, 0 point for another passive player) and an equal distribution (s points for self and the passive player). The equal distributions increase in an increment of 1 point from (0,0) to (20,20). One of the 21 cases is randomly chosen to determine the payment.

Variable α_i can be understood as an “envy” parameter for individual i arising from her disadvantageous financial position, and β_i can be thought of as a “guilt” parameter arising from her advantageous financial position in a given round of the interaction. Fehr and Schmidt (1999) assume that $\alpha_i \geq \beta_i$ and $1 \geq \beta_i \geq 0$. Using this formulation, we see, for example, that, when both players cooperate in UNEQ, the column player receives a utility of $(6 - 3\alpha_i)$. Intuitively, this indicates that, *all else equal*, incentives to cooperate are weaker if $\alpha_i > 0$ as players dislike having different payoffs than their counterpart. On the other hand, if $\beta_i > 0$, the prospects of cooperation *improve* as individuals suffer disutility when they defect on someone who cooperates. This is reflected in the minimum continuation probability required for “Grim Trigger” (a strategy that cooperates until the other defects) to be a subgame perfect equilibrium in UNEQ, which, as a function of α_i and β_i is:

$$\delta_{SPE}^{UNEQ} = \frac{4 + 3\alpha_i - 10\beta_i}{7 - 10\beta_i}.$$

Note that the critical δ in UNEQ is that of the *column* player which is always higher than that for the row player. The expression is increasing in α_i and decreasing in β_i . For comparison, the equivalent expression for EQ-L is:

$$\delta_{SPE}^{EQ-L} = \frac{4 - 10\beta_i}{7 - 10\beta_i}.$$

Comparing the two expressions one can see that, if $\alpha_i > 0$, then the requirements on δ to support cooperation is greater in UNEQ. However, if $\alpha_i = 0$, then $\delta_{SPE}^{UNEQ} = \delta_{SPE}^{EQ-L}$, for all β_i . These insights will come in handy later when analyzing the data.

Table 1. Critical continuation probabilities

	$(\alpha, \beta) = (0, 0)$		$(\alpha, \beta) = (1.2, 0.5)$	
	δ_{SPE}	δ_{RD}	δ_{SPE}	δ_{RD}
EQ-H	0.14	0.40	0.00	0.65
EQ-L	0.57	0.70	0.00	0.82
UNEQ	0.57	0.70	1.00	1.00
UNEQ-Alt	0.44	0.59	0.48	0.89

Note: Critical continuation probabilities for cooperation to be supported in equilibrium (δ_{SPE}) and for Grim Trigger to be a risk-dominant strategy (δ_{RD}), when individuals do not care about inequality $(\alpha, \beta) = (0, 0)$, and when they do $(\alpha, \beta) = (1.2, 0.5)$; the latter are median estimates of α and β from Blanco et al. (2011).

Table 1 presents the critical continuation probabilities for each treatment, for different values of α , β .⁴ The columns with δ_{SPE} present the minimum continuation probabilities that can support cooperation as a subgame perfect equilibrium, whereas those with δ_{RD} depict the minimum continuation probabilities for which “Grim Trigger” is risk dominant compared to Always Defecting. As mentioned above, the critical δ for UNEQ and UNEQ-Alt is that of the *column* player, i.e., the player who experiences the lowest pecuniary payoff from mutual cooperation. Indeed, it is this player who has the weakest incentives to cooperate.

Table 1 offers clear predictions about how inequality will affect the relative prospects of cooperation in our experimental treatments. In doing so, it shows how our design enables us to identify the impact of inequality on cooperation. Below, we present three hypotheses under the assumption that individuals care sufficiently about inequality in payoffs.

Hypothesis 1: *Cooperation rates will be lower in UNEQ than in EQ-L, irrespective of the expected duration of the interaction.*

The intuition behind Hypothesis 1 is as follows. In UNEQ, the column player has the same pecuniary incentives to cooperate as players in EQ-L, but will be less willing to cooperate if she dislikes the unequal earnings associated with mutual cooperation, i.e., if their α (envy) is sufficiently high. However, as mentioned in the introduction, there are different reasons why inequality concerns may play a diminished role in indefinitely repeated prisoner dilemmas. If we observe similar rates in EQ-L and UNEQ, this will be evidence that individuals do not care about inequality per se in our setting. We can also not rule out the possibility that cooperation rates are greater in UNEQ than they would be in EQ-L. Since this is not in line with concerns for own payoff maximization or with concerns for payoff equality, it could suggest that concerns for efficiency could affect preferences or players’ beliefs.

Our second hypothesis involves the prospects of cooperation as the expected duration of an interaction increases.

Hypothesis 2: *If individuals sufficiently dislike payoff inequality, cooperation rates will not change with the expected duration of interactions in UNEQ, but will increase in EQ-L (as well as in EQ-H).*

As can be seen in Table 1, if individuals dislike inequality sufficiently strongly, there is no continuation probability that can support Grim Trigger as an equilibrium in UNEQ. This is not the case for EQ-L and EQ-H where there is no inequality (or for UNEQ-Alt where inequality washes away as the expected duration of the interaction increases). Stated differently, if individuals care

⁴ Table A1 in the appendix displays δ_{SPE} and δ_{RD} as functions of α_i and β_i for all treatments.

strongly for equality, the “shadow of inequality” in UNEQ will outweigh the “shadow of the future”. On the other hand, if individuals do not care about inequality, i.e., $(\alpha_i, \beta_i) = (0, 0)$, we should observe the same increase in cooperation rates in EQ-L and UNEQ.

Our final hypothesis deals with cooperation in UNEQ-Alt, when the roles of row and column are swapped in each round, relative to the other treatments.

Hypothesis 3: *Cooperation rates will be highest in EQ-H followed by EQ-L and UNEQ-Alt.*

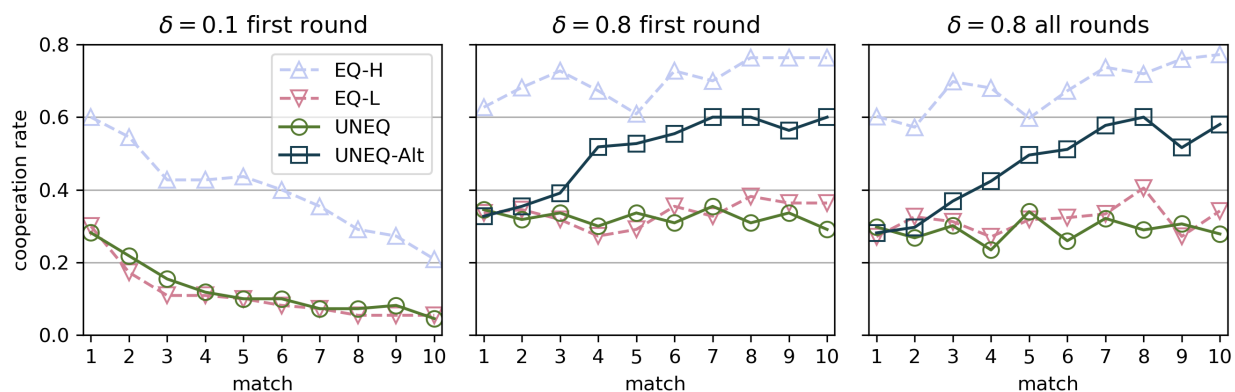
In Table 1, we see that, if individuals dislike inequality, cooperation is predicted to be lower in UNEQ-Alt than in EQ-L. On the other hand, if individuals care solely about their pecuniary payoff, then we predict that cooperation rates will be highest in EQ-H followed by UNEQ-Alt and EQ-L. The reason is that the pecuniary returns to cooperation are weakest in EQ-L, followed by those in UNEQ-Alt.

3. Experimental Results

We divide the analysis of the data into three parts. In the first part, we present tests of our behavioral hypotheses. In the second part, we explore the impact of concerns for inequality at the *individual* level. Finally, in the third part, we investigate how the presence of payoff inequalities affects the strategies used by participants.

Unless otherwise stated, we follow past studies on indefinitely repeated prisoner’s dilemma that focus on the decision to cooperate at the first round of each match (Dal Bo & Fréchet 2011, Dal Bo & Fréchet 2019, Dreber et al. 2014, Fudenberg et al. 2012). This approach facilitates the comparison of choices across matches and treatments as behavior in the first round is unaffected by the actual length of a given match or the behavior of a specific opponent.

Figure 2. Average first-round cooperation over time for each treatment



3.1 Testing the behavioral hypotheses

Result 1: *Cooperation rates are indistinguishable in EQ-L and UNEQ.*

SUPPORT: Figure 2 presents the average first-round cooperation rate over time for all the treatments. As can be seen, cooperation rates are indistinguishable in EQ-L and UNEQ, at every given point of the experiment, irrespective of the continuation probability, i.e., whether $\delta = 0.1$ or 0.8 . Mann-Whitney tests fail to reject the hypothesis that cooperation rates are the same in the two treatments both when $\delta = 0.1$ (p -value = 0.89, $m = n = 11$, two-tailed) and when $\delta = 0.8$ (p -value = 1.00, $m = n = 11$, two-tailed).⁵ ■

Given the striking similarity of cooperation rates in UNEQ and EQ-L, an explicit test of Hypothesis 2 appears to be redundant. Nevertheless, for symmetry and completeness, we provide formal statistical support.

Result 2: *The increase in cooperation rates when moving from $\delta = 0.1$ to $\delta = 0.8$ is indistinguishable in treatments EQ-L and UNEQ.*

SUPPORT: Cooperation rates increase from 12 % when $\delta = 0.1$ to 32% in UNEQ when $\delta = 0.8$, and from 11% to 34% in EQ-L. The difference-in-difference in cooperation rates is not statistically significant (p -value=0.86, mixed-effects regression clustered at both the group and subject levels, controlling for trend and location fixed effects). The increase in cooperation rates in EQ-H is also similar (40% when $\delta = 0.1$, and 70% when $\delta = 0.8$) and significantly greater than that in UNEQ (p -value<0.01 from mixed-effects regression) and EQ-L (p -value<0.01). ■

Result 3: *Cooperation rates are highest in EQ-H followed by UNEQ-Alt and EQ-L.*

SUPPORT: As seen in Figure 2, when $\delta = 0.8$, across matches, cooperation rates are 0.70 in EQ-H, 0.50 in UNEQ-Alt, and 0.34 in EQ-L. The difference is statistically significant between EQ-H and UNEQ-Alt (p -value = 0.06, $m = n = 11$, Mann-Whitney, two-tailed), between UNEQ-Alt and EQ-L (p -value = 0.03, $m = n = 11$, Mann-Whitney, two-tailed), and between EQ-H and EQ-L (p -value < 0.01, $m = n = 11$, Mann-Whitney, two-tailed).⁶ ■

⁵ Mixed-effects regressions clustered at both the group and subject levels, controlling for time trends and location fixed effects, provides the same conclusions (p -value=0.58 for $\delta = 0.1$, and p -value=0.86 for $\delta = 0.8$).

⁶ Mixed-effects regressions clustered at both the group and subject levels, controlling for time trends and location fixed effects, provides the same conclusions (p -value<0.01 for all three comparisons).

The results above contradict Hypothesis 1, Hypothesis 2 and Hypothesis 3 which we formed under the assumption that individuals care about inequality in payoffs. Taken together, Result 1, Result 2, and Result 3 suggest that inequality concerns do not affect cooperation in our experiment. The striking similarity in cooperation rates in EQ-L and UNEQ, irrespective of the expected duration of interactions, is clearly at odds with the prediction obtained under the assumption that individuals care about inequality as in Fehr-Schmidt (1999). Also at odds is the fact that cooperation is higher in UNEQ-Alt than in EQ-L, the fact that cooperation is maintained in UNEQ when $\delta = 0.8$, and also that the increase in cooperation rates as the “shadow of the future” increases are unaffected by the presence of inequality. This and the other patterns seen in Figure 2 are in line with the predictions derived under the assumption that participants do not care about payoff inequalities. Specifically, when $\delta = 0.1$, cooperation unravels in all treatments, whereas cooperation is maintained when $\delta = 0.8$. One minor inconsistency would appear to be the fact that cooperation rates are higher in EQ-H than in the other treatments when $\delta = 0.1$. However, we note that the implemented continuation probability ($\delta = 0.1$) is quite close to the critical continuation probability for supporting cooperation in this treatment ($\delta = 0.14$, Table 1).

3.2 A closer examination of the determinants of cooperation

Table 2 presents the results from linear mixed-effects regressions exploring the determinants of cooperation in the first round of each match. Mixed effects models extend the random effects models by introducing multiple levels of dependence, allowing for random slopes as well as random intercepts. Given that individuals interact within “session silos” and make several decisions each, standard errors are clustered at both the individual and silo levels. Models (I) and (II) consider behavior from all experimental rounds when $\delta = 0.1$; models (III) and (IV) do the same for treatments with $\delta = 0.8$. Models (V) to (VIII) re-estimate the coefficients in the first four models using only data from the last 5 matches. A comparison between the estimates in the first and the last four models, therefore, provides us with information about whether the effects of certain variables change over time.

The independent variables include treatment dummies, *EQ-H*, *UNEQ* and *UNEQ-Alt* (the reference treatment in all models is *EQ-L*), a variable to capture overall trends in cooperation (*Match Number*), individual-level measures of α (envy) and β (guilt), and location fixed effects. We obtain similar results if, instead of controlling for individuals with multiple switching points in the elicitation task, we drop them from the analysis.

The first thing to note in Table 2 is that the estimates lend additional support to Results 1, 2 and 3 in the previous subsection. In all models, cooperation rates are similar and statistically indistinguishable between EQ-L and UNEQ (Result 1). As the coefficient for UNEQ is similar across models, it follows that the increase in cooperation rates as the continuation probability

increases from 0.1 to 0.8 is also indistinguishable across EQ-L and UNEQ (Result 2). By contrast, cooperation rates are notably higher in EQ-H than those in either EQ-L or UNEQ (Result 3). Also, in line with the theoretical analysis, the coefficient for *Match Number* indicates that there is a significant decline in cooperation when the critical continuation probability is higher than the implemented continuation probability ($\delta = 0.1$), but not when it is lower ($\delta = 0.8$). In fact, in the latter case, we observe a slight (but significant) increase in cooperation, though the level stabilizes in the last five matches.

Table 2. Mixed-effects regressions of first-round cooperation rates

	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)
<i>EQ-H</i>	0.33*** (0.06)	0.34*** (0.06)	0.49*** (0.10)	0.53*** (0.10)	0.28*** (0.07)	0.29*** (0.07)	0.52*** (0.12)	0.56*** (0.11)
<i>UNEQ</i>	0.01 (0.04)	0.02 (0.04)	-0.01 (0.07)	0.00 (0.07)	0.01 (0.05)	0.02 (0.05)	-0.04 (0.08)	-0.03 (0.07)
<i>UNEQ-Alt</i>			0.29*** (0.10)	0.31*** (0.10)			0.36*** (0.12)	0.39*** (0.11)
<i>Match number</i>	-0.03*** (0.00)	-0.03*** (0.00)	0.01*** (0.00)	0.02*** (0.00)	-0.02*** (0.00)	-0.02*** (0.00)	0.00 (0.00)	0.01 (0.00)
α_i (envy)		-0.01 (0.01)		0.00 (0.01)		-0.00 (0.01)		-0.00 (0.01)
β_i (guilt)		0.15*** (0.06)		0.09 (0.07)		0.12** (0.06)		0.16* (0.08)
Constant	0.22*** (0.05)	0.16** (0.07)	0.15* (0.09)	0.08 (0.10)	0.19*** (0.07)	0.13 (0.08)	0.19* (0.11)	0.12 (0.11)
Observations	3,300	2,540	4,400	3,350	1,650	1,270	2,200	1,675
Number of groups	33	33	44	44	33	33	44	44
δ	0.10	0.10	0.80	0.80	0.10	0.10	0.80	0.80
Matches	All	All	All	All	Last 5	Last 5	Last 5	Last 5

Note: Standard errors in parentheses, which are clustered at both the group and the subject levels. All regressions include location fixed effects. Models (II), (IV), (VI) and (VIII) drop observations from individuals with multiple switching points in the modified dictator game as we cannot calculate their β_i . *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Moving beyond treatment comparisons, Table 2 reveals an association between an individual's attitudes towards "guilt" β_i and cooperation, which Result 4 discusses.

Result 4: *Individuals with a greater sensitivity towards "guilt" are generally more likely to cooperate. There is no association between an individual's sensitivity towards "envy" and their willingness to cooperate.*

SUPPORT: Table 2 shows that β_i has a sizable coefficient (from 0.09 to 0.16) which is statistically significant in three out of the four regressions. On the other hand, the coefficient for α_i is essentially 0 in all specifications. ■

A few comments concerning Result 4 appear to be in order. First, we must ask ourselves how we can reconcile the fact that some individuals appear to care about factors other than their pecuniary payoffs, and yet, find that, on aggregate, cooperation rates are accurately predicted by a model that assumes players solely care about their pecuniary payoffs. The answer has to do with the fact that β_i affects the willingness to cooperate in the same manner in all treatments (see Section 2.2., and Table A1 in the Appendix). What matters for the comparative statics is whether $\alpha_i > 0$. If $\alpha_i = 0$, the predictions are identical irrespective of the value of β_i .

The insignificance of α_i in Table 2 may seem surprising given that it is often assumed that $\alpha_i \geq \beta_i \geq 0$ (Fehr and Schmidt, 1999). Indeed, looking into behavior in our elicitation task, this condition is satisfied for 76% of our participants, with the median α and β being 1.50 and 0.53, respectively. So, why does α_i have no effect? As shown in Section 2.2., α_i can play a critical role in *undermining* cooperation in the presence of payoff inequalities. (Note that the coefficient for α_i is also statistically insignificant if we run a separate regression for the column player in UNEQ.) One explanation is that α_i , as measured in an ultimatum game, captures a concern for reciprocity (Blanco et al. 2011). This concern is moot in UNEQ where the inequality associated with mutual cooperation is *exogenously* determined and not the result of participants' choices. By contrast, β_i can *bolster* cooperation by reducing individual's desire to defect on cooperators. Defecting when the other is expected to cooperate *endogenously* creates a pay inequality between players. Subjects who care about treating others fairly, e.g., due to a preference for conditional cooperation, would suffer from defecting in this case.

Notice that the positive association between β_i and the tendency to cooperate is broadly in line with the findings in Dreber et al. (2014). Specifically, Dreber et al. found that, in *symmetric* prisoner's dilemmas, subjects exhibiting stronger guilt were more inclined to cooperate but only when cooperation was *not* an equilibrium strategy. In our experiment, we see evidence of such a relationship even when cooperation is an equilibrium strategy, but the association appears to be less robust in the latter case.

3.3 Strategy frequency estimation

To further investigate the mechanisms behind cooperation in the $\delta = 0.8$ treatments, we explore the strategies used by participants in our experiment. Following the method proposed by Dal Bo and Fréchet (2011), we estimate the fraction of subjects using one of five strategies that have received particular attention in the literature: Grim Trigger ("Grim"), Tit-For-Tat ("TFT"), Tit-

For-Two-Tats (“TF2T”), Always Defect (“AllD”), and Suspicious Tit-For-Tat (“D-TFT”).⁷ For the UNEQ and UNEQ-Alt treatments, we estimate strategy frequencies separately by player type. We use subscripts h and l to distinguish between the strategies used by players receiving the high and low payoff under mutual cooperation, respectively.

Table 3. Strategy frequency estimation for $\delta = 0.8$ treatments

	EQ-H	UNEQ _h	UNEQ-ALT _h	UNEQ _l	UNEQ-Alt _l	EQ-L
Grim	0.23*** (0.06)	0.04* (0.03)	0.05** (0.03)	0.00 (0.01)	0.10** (0.06)	0.14*** (0.04)
TFT	0.33*** (0.06)	0.09** (0.04)	0.32*** (0.06)	0.12*** (0.05)	0.21*** (0.06)	0.14*** (0.04)
TF2T	0.24*** (0.05)	0.15*** (0.06)	0.25*** (0.05)	0.20*** (0.06)	0.28*** (0.05)	0.09*** (0.03)
AllD	0.12*** (0.03)	0.43*** (0.08)	0.26*** (0.03)	0.54*** (0.07)	0.23*** (0.03)	0.42*** (0.06)
D-TFT	0.09*** (0.03)	0.29*** (0.07)	0.11*** (0.03)	0.14*** (0.06)	0.18*** (0.04)	0.22*** (0.05)
Cooperative	0.79*** (0.04)	0.28*** (0.06)	0.62*** (0.03)	0.32*** (0.07)	0.59*** (0.03)	0.36*** (0.05)
Forgiving	0.66*** (0.06)	0.53*** (0.07)	0.69*** (0.05)	0.46*** (0.07)	0.67*** (0.06)	0.45*** (0.06)
Lenient	0.24*** (0.05)	0.15*** (0.06)	0.25*** (0.05)	0.20*** (0.06)	0.28*** (0.05)	0.09*** (0.03)

Note: Data is from the last five matches. In treatments with unequal payoffs, we use subscripts h and l to distinguish between the strategies used by players receiving the high/low payoff under mutual cooperation, respectively. As TF2T is the only lenient strategy, the row Lenient is identical to the row of TF2T. Bootstrapped standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The top panel in Table 3 presents the estimated frequency with which each strategy is used in the experiment. With the exception of Grim Trigger, there are similarities in the strategies employed in EQ-L and UNEQ by the column player (UNEQ_l). The similarities are easier to identify in the bottom panel in Table 3 which reports the total fraction of “cooperative” (starts by

⁷ To determine which strategies to be included in our estimation, we first run our estimation based on the strategies used by Fudenberg et al. (2012) and eliminated those strategies that never account for a significant proportion across treatments.

playing C), “forgiving” (can return to C after punishment), and “lenient” (not retaliating after the first defection).

The estimates in the bottom panel of Table 3 show that cooperative strategies are about twice as likely to be used in EQ-H and UNEQ-Alt treatments (79-59%) than they are in the EQ-L and UNEQ treatments (36-32%). The strategies used in EQ-L are similar to those used by both player types in UNEQ: the modal strategy is AllD, with the next most common strategy being D-TFT. The majority of players begin a match by playing Defect in these treatments, and then a large fraction never considers the possibility of cooperation thereafter. Furthermore, the strategies of TFT and TF2T, which are both cooperative and forgiving, are important drivers of prolonged cooperation, accounting for approximately half of players’ strategies in the treatments where we see the most cooperation (EQ-H and UNEQ-Alt).

4. Discussion

Our study helps fill a gap in the literature by exploring the impact of inequality concerns on cooperation in indefinitely repeated games. We find that cooperation rates are well predicted by a model in which individuals are assumed to solely care about their own pecuniary payoff without regard for others. We also find that the strategies employed by individuals are similar across treatments with and without inequalities in payoffs, all else equal. A closer inspection of the data, however, reveals a nuanced role for other-regarding concerns. Specifically, we observe that individuals revealed to care more about others in a modified dictator game are less likely to defect in the indefinitely repeated prisoner’s dilemma. As this concern affects equilibrium predictions equally across treatments, the comparative statics are identical to those obtained from the model in which individuals are assumed to care solely about their monetary payoff.

In their authoritative review of the literature of cooperation in symmetric indefinitely repeated prisoner dilemmas, Dal Bo and Fréchet (2018) write: “infinite repetition seems to reduce the importance of other-regarding preferences.” Our findings suggest that the “judges are still out”. On the one hand, we do find evidence that our measure of “guilt” predicts the willingness to cooperate even when cooperation is not supported in equilibrium (see Table 2). On the other hand, the effect appears to be less robust than when cooperation is not supported in equilibrium. More work is needed to identify to what extent other-regarding preferences play less of a role in indefinitely repeated games. More work is also needed to test whether the “shadow of inequality” may loom larger in other settings. While inequality concerns *per se* appear not to matter in our setting, it seems possible that the player receiving the lowest payoff for cooperating would be less willing to cooperate if the row player got to determine the payoff profile, thus activating reciprocity concerns. These would appear to be interesting topics for future research.

References

- Aoyagi, M., & Fréchette, G. (2009). Collusion as public monitoring becomes noisy: Experimental evidence. *Journal of Economic Theory*, 144(3), 1135–1165.
- Balafoutas, L., Kerschbamer, R., & Sutter, M. (2012). Distributional preferences and competitive behavior. *Journal of Economic Behavior & Organization*, 83(1), 125-135.
- Blanco, M., Engelmann, D., & Normann, H. T. (2011). A within-subject analysis of other-regarding preferences. *Games and Economic Behavior*, 72(2), 321-338.
- Bland, J., & Nikiforakis, N. (2015). Coordination with third-party externalities. *European Economic Review*, 80, 1-15.
- Bolton, G. E., & Ockenfels, A. (2000). A theory of equity, reciprocity, and competition. *American Economic Review*, 90(1), 166-193.
- Cabrales, A., Miniaci, R., Piovesan, M., & Ponti, G. (2010). Social preferences and strategic uncertainty: an experiment on markets and contracts. *American Economic Review*, 100(5), 2261-78.
- Camera, G., & Casari, M. (2009). Cooperation among strangers under the shadow of the future. *American Economic Review*, 99 (3), 979–1005.
- Camera, G., Deck, C., & Porter, D. (2020). Do economic inequalities affect long-run cooperation and prosperity?. *Experimental Economics*, 23(1), 53-83.
- Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, 117(3), 817-869.
- Chmura, T., Kube, S., Pitz, T., & Puppe, C. (2005). Testing (beliefs about) social preferences: Evidence from an experimental coordination game. *Economics Letters*, 88(2), 214-220.
- Dal Bó, P. (2005). Cooperation under the Shadow of the Future: Experimental Evidence from Infinitely Repeated Games. *American Economic Review*, 95(5), 1591–1604.
- Dal Bó, P., & Fréchette, G. R. (2011). The Evolution of Cooperation in Infinitely Repeated Games. *American Economic Review*, 101(1), 411–429.
- Dal Bó, P., & Fréchette, G. R. (2018). On the Determinants of Cooperation in Infinitely Repeated Games: A Survey. *Journal of Economic Literature*. 56(1), 60-114.
- Dal Bó, P., & Fréchette, G. R. (2019). Strategy choice in the infinitely repeated prisoner's dilemma. *American Economic Review*, 109(11), 3929–3952.
- Davis, D., Ivanov, A., & Korenok, O. (2016). Individual characteristics and behavior in repeated games: An experimental study. *Experimental Economics*, 19(1), 67-99.
- Dreber, A., Fudenberg, D., & Rand, D. G. (2014). Who cooperates in repeated games: The role of altruism, inequity aversion, and demographics. *Journal of Economic Behavior & Organization*, 98, 41–55.

- Duffy, J., & Ochs, J. (2009). Cooperative behavior and the frequency of social interaction. *Games and Economic Behavior*, 66(2), 785–812.
- Engelmann, D., & Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review*, 94(4), 857-869.
- Falk, A., Fehr, E., & Fischbacher, U. (2008). Testing theories of fairness—Intentions matter. *Games and Economic Behavior*, 62(1), 287-303.
- Faravelli, M., Kirchkamp, O., & Rainer, H. (2013). The effect of power imbalances on incentives to make non-contractible investments. *European Economic Review*, 61, 169-185.
- Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3), 817-868.
- Fischbacher, U., Schudy, S., & Teyssier, S. (2014). Heterogeneous reactions to heterogeneity in returns from public goods. *Social Choice and Welfare*, 43(1), 195–217.
- Fisman, R., Kariv, S., & Markovits, D. (2007). Individual preferences for giving. *American Economic Review*, 97(5), 1858-1876.
- Friedman, J. W. (1971). A Non-cooperative Equilibrium for Supergames. *The Review of Economic Studies*, 38(1), 1–12.
- Fudenberg, D., & Maskin, E. (1990). Evolution and cooperation in noisy repeated games. *American Economic Review*, 80(2), 274-279.
- Fudenberg, D., Levine, D., & Maskin, E. (1994). The Folk Theorem with Imperfect Public Information. *Econometrica*, 62(5), 997–1039.
- Fudenberg, D., Rand, D. G., & Dreber, A. (2012). Slow to Anger and Fast to Forgive: Cooperation in an Uncertain World. *American Economic Review*, 102(2), 720–749.
- Gangadharan, L., Nikiforakis, N., & Villeval, M. C. (2017). Normative conflict and the limits of self-governance in heterogeneous populations. *European Economic Review*, 100, 143–156.
- Gangadharan, L., & Nikiforakis, N. (2009). Does the Size of the Action Set Matter for Cooperation? *Economics Letters* 104, 115-117.
- Normann, H. T., & Wallace, B. (2012). The impact of the termination rule on cooperation in a prisoner’s dilemma experiment. *International Journal of Game Theory*, 41(3), 707-718.
- Nikiforakis, N., Noussair, C. N., & Wilkening, T. (2012). Normative conflict and feuds: The limits of self-enforcement. *Journal of Public Economics*, 96(9-10), 797-807.
- Reuben, E., & Riedl, A. (2013). Enforcement of contribution norms in public good games with heterogeneous populations. *Games and Economic Behavior*, 77(1), 122–137.

APPENDIX

The critical continuation probabilities for each treatment, as a function of the parameters in the Fehr-Schmidt model are given below. δ_{SPE}^T is the minimum continuation probability that can support cooperation as a subgame perfect equilibrium in treatment T, and δ_{RD}^T is the minimum continuation probability that makes cooperation the risk-dominant strategy in treatment T if one's opponent is playing either the "Grim trigger" or the "Always defect" strategy.

$$\delta_{SPE}^{EQ-H} = \frac{1-10\beta_i}{7-10\beta_i}, \quad \delta_{RD}^{EQ-H} = \frac{4+10\alpha_i-10\beta_i}{10+10\alpha_i-10\beta_i}$$

$$\delta_{SPE}^{UNEQ} = \frac{4+3\alpha_i-10\beta_i}{7-10\beta_i}, \quad \delta_{RD}^{UNEQ} = \frac{7+13\alpha_i-10\beta_i}{10+10\alpha_i-10\beta_i}$$

$$\delta_{SPE}^{EQ-L} = \frac{4-10\beta_i}{7-10\beta_i}, \quad \delta_{RD}^{EQ-L} = \frac{7+10\alpha_i-10\beta_i}{10+10\alpha_i-10\beta_i}$$

Notice that when $\beta > 0.4$ for EQ-L, $\beta > 0.1$ for EQ-H, $\beta > 1/7$ for UNEQ-H, and $\beta > 0.4 + 0.3\alpha$ for UNEQ-L, the stage game becomes a coordination game and cooperation is therefore always an equilibrium.

ONLINE Appendix

A. Instructions

General Instructions

You are now taking part in an experiment. If you read the following instructions carefully, you can, depending on your and other participants' decisions, earn a considerable amount of money. It is therefore important that you take your time to understand the instructions. Please do not communicate with the other participants during the experiment. Should you have any questions, please raise your hand and an experimenter will come and answer your questions.

Participants in this experiment have been randomly assigned to several clusters, **each cluster with 10 participants. You will only interact with participants in your cluster.**

The experiment consists of three different parts. In each part you will be asked to make one or more decisions. You will have to make your decisions without knowing other participants' decisions in the previous parts. Note further that the other participants will also not know your decisions.

You will receive specific instructions for each part, only once the previous part has been completed.

Once the experiment has finished, your earnings will be paid to you in cash. You will then be allowed to leave the lab and no one except you will know either your earnings or your decisions.

Note that, at the end of the experiment, the computer will randomly determine whether the actions for Part 1 or Part 2 will count towards your final earnings. In addition, you will be paid for your earnings in Part 3.

For the rest of these instructions, payments will be expressed in "points". In the instructions specific to each part you will be told how these points will be converted into euros.

[In the instructions shown here, the Modified Dictator Game appears first as Part 1. The order the Modified Dictator Game and the Ultimatum Game in the actual experiment is balanced across sessions.]

Instructions Part 1

**Points in Part 1 will be exchanged for euros at a rate of:
1 point = € 0.25**

In this part, the situation is as follows:

*Person A is asked to **choose between two possible distributions of money** between herself and Person B in twenty-one different decision problems. Person B knows that A has been called to make those decisions, and there is nothing he can do but accept them.*

The roles of Person A and Person B will be randomly determined at the end and will remain anonymous.

Before making your decisions please read carefully the following paragraphs.

The decision problems will be presented in a chart. Each decision problem will look like the following:

Person A's Payoff	Person B's Payoff	Decision	Person A's Payoff	Person B's Payoff
20	0	Left Right	5	5

You will have to decide as Person A; hence if in this particular decision problem you choose Left, you decide to keep the 20 points for you so Person B's payoff will be 0 points. Similarly, if you choose Right, you and the Person B will earn 5 points each.

You will need to choose one distribution (Left or Right) in each of the twenty-one rows you will have in the screen. If this part is chosen at the end of the experiment, the computer will randomly choose one of the twenty-one decisions. The outcome in the chosen decision will then determine your earnings from this part.

The computer will randomly pair you with another participant in your cluster and will assign the roles. The matching and roles assignment will remain anonymous.

Please note that you will make all decisions as Person A but the computer might assign you Person B's role.

If you are assigned the role of A, you will earn the amount that you have chosen for Person A in the relevant situation and the person paired with you will earn the amount that you have chosen for Person B.

In the case that you are assigned the role of Person B, you will earn the amount that Person A with whom you are paired has chosen for Person B (i.e., you) in the relevant situation.

Instructions Part 2

Points in Part 2 will be exchanged for euros at a rate of:
1 point = € 0.25

In this part, the situation is as follows:

Person A (the proposer) will propose a distribution of 20 points between Person A and Person B. Person B (the responder) will state a minimum offer that he is willing to accept. If the proposed distribution offers greater than or equal to this minimum, the offer is accepted. If the proposed distribution offers less than this minimum, the proposal is rejected.

In the case that Person B accepts A's proposed distribution, that will be implemented. If B rejects the offer, both receive nothing.

The roles of Person A and Person B will be randomly determined by the computer and will remain anonymous.

Before making your decision please read carefully the following paragraphs.

In the case that this section is selected to determine your earnings, the computer will randomly pair you with another participant in your cluster and will assign the roles. The matching and roles assignment will remain anonymous.

You will have to make decisions **as if you were Person A** and **also as if you were Person B**. Note that Person B will have to make a decision about the offer he is willing to accept *before* he can see Person A's actual offer. To do this Person B will be prompted to state the minimum amount he is willing to accept from Person A. Note that if the minimum offer that Person B is willing to accept is greater than the offer of Person A then both Person A and Person B will have zero earnings from this part of the experiment. Therefore, make sure you take your time to make your decisions, and that you state the minimum amount you would be willing to accept and not the amount you would like Person A to offer to you.

If you are assigned the role of Person A you will earn the payoff you chose for yourself if the Person B that you are paired with accepts your offer. Otherwise, both will earn nothing.

If you are assigned the role of Person B, you will earn the payoff that Person A that you are paired with chose for B, only if you had accepted that particular offer. Otherwise, you both earn nothing from this part.

Instructions Part 3 [UNEQ80]

Points in Part 3 will be exchanged for euros at a rate of:
1 point = € 0.07

In this part, the situation is as follows:

The participants are divided in two groups: A and B. A and B participants will be matched together to interact in the following way. A participants can choose between U(p) or D(own). The B participant can choose between L(ef) and R(ight). The payoffs of A and B depending their choices are shown in the following table:

		B's choice	
		L	R
A's choice	U	9, 6	0, 10
	D	10, 0	3, 3

In this table, the first number in each cell represents A's payoff, and the second number is B's payoff. For example:

If A chooses U and B chooses R, A earns 0 points and B participant earns 10 points.

If A chooses D and B chooses L, A earns 10 points and B participant earns 0 points.

If A chooses D and B chooses R, A earns 3 points and B participant earns 3 points.

On the computer screen, the points of the A participant are indicated in red, and the B participant points are indicated in blue. In addition, the screen will show on the right hand side the result of previous rounds of the current match.

Part 3 will consist of **10 matches**. In each match every A participant is paired with a B participant. Your role as A or B is randomly determined at the start of Part 3 and will remain the same throughout Part 3. At the start of each new match, **you will be randomly matched to a participant** with a different role in your cluster.

Each match will consist of a number of rounds. The exact number is randomly determined by a **twenty-sided die** which the experimenter will roll in public at the end of each round. If the numbers 1-16 appear, the match will continue for another round. If the numbers 17-20 appear, the match ends. Therefore, each match continues until a number equal to or greater than 17 appears.

We will first go through one practice match to get you familiarized with the session and the computer program. You are *not* paid for the practice match.

Once we are finished with the practice match, you will participate in 10 matches, each match randomly paired with a participant in your cluster.

Instructions Part 3 [UNEQ_ALT]

Points in Part 3 will be exchanged for euros at a rate of:
1 point = € 0.07

In this part, the situation is as follows:

The participants will be assigned the role of A and B and interact in the following way. A participants can choose between U(p) or D(own). The B participant can choose between L(eft) and R(ight). The payoffs of A and B depending their choices are shown in the following table:

		B's choice	
		L	R
A's choice	U	9, 6	0, 10
	D	10, 0	3, 3

In this table, the first number in each cell represents A's payoff, and the second number is B's payoff. For example:

- If A chooses U and B chooses L, A earns 9 points and B participant earns 6 points.
- If A chooses U and B chooses R, A earns 0 points and B participant earns 10 points.
- If A chooses D and B chooses L, A earns 10 points and B participant earns 0 points.

On the computer screen, the points of the A participant are indicated in red, and the B participant points are indicated in blue. In addition, the screen will show on the right hand side the result of previous rounds of the current match.

Part 3 will consist of **10 matches**. At the start of each new match, **you will be randomly matched to a participant** in your cluster.

Each match will consist of a number of rounds. The exact number is randomly determined by a **twenty-sided die** which the experimenter will roll in public at the end of each round. If the numbers 1-16 appear, the match will continue for another round. If the numbers 17-20 appear, the match ends. Therefore, each match continues until a number equal to or greater than 17 appears.

At any point in a given match you will have one of two roles: A or B. The role will alternate across rounds:

- If you are assigned role A in Round 1, you will be a role B in Round 2, a role A in Round 3, and so on so long as the match continues.
- If you are assigned role B in Round 1, you will be a role A in Round 2, a role B in Round 3, and so on so long as the match continues.

We will first go through one practice match to get you familiarized with the session and the computer program. You are *not* paid for the practice match. Once we are finished with the practice match, you will participate in 10 matches, each match randomly paired with a participant in your cluster.