Experimental Economics (2025), **28**, 22–42
doi:10.1017/eec.2024.11

ORIGINAL PAPER

Social preferences under the shadow of the future

Felix Kölle[1], Simone Quercia[2] and Egon Tripodi[3]

[1]Department of Economics, University of Cologne, Albertus Magnus Platz, 50923 Cologne, Germany
[2]Department of Economics, University of Verona, Via Cantarane 24, 37129 Verona, Italy
[3]Hertie School, Friedrichstraße 180, 10117 Berlin, Germany
Corresponding author: Felix Kölle; Email: felix.koelle@uni-koeln.de

(Received 5 February 2024; Revised 6 November 2024; Accepted 15 November 2024)

Abstract
Social interactions frequently take place under the shadow of the future. Previous literature explains cooperation in indefinitely repeated prisoner's dilemma as driven predominantly by self-interested strategic considerations. This paper provides a causal test of the importance of social preferences in such contexts. In a series of pre-registered experiments, we show that high levels of cooperation can be sustained when prosocial individuals interact in segregated groups. By comparing their behavior with that of mixed and selfish groups, we highlight the conditions under which other-regarding motivations matter in repeated interactions.

Keywords: Cooperation; Experiment; Indefinitely repeated games; Prisoner's dilemma; Social preferences
JEL Codes: C73; C91; C92

1. Introduction

Social dilemmas are ubiquitous in nature and exist at all levels of human society, ranging from team production and collaborations among firms to the maintenance of natural resources and the provision of public goods. The unifying element behind these examples is the fundamental tension between individual and collective interest. The simplest and most commonly used game to study this tension is the prisoner's dilemma, which raises continued interest across the social sciences. The individual interest that can determine cooperation is the threat of future punishment; the Folk theorem identifies sharp conditions on how large the shadow of the future needs to loom for cooperation to be compatible with self-interest (Fudenberg & Maskin, 1986). Despite being broadly in line with this theory, the empirical evidence is more nuanced and substantial variation in cooperation remains unexplained (see Dal Bó & Fréchette, 2018, for a recent overview).

One natural candidate to explain this variation is the heterogeneity in people's social motivations, which is pervasive across different demographics, cultures, and societies (Croson & Gneezy, 2009; Falk et al., 2018; Fisman et al., 2015; Henrich et al., 2001). Social preferences have further been shown to explain behavioral patterns across a large variety of contexts, both in the lab (Cooper & Kagel, 2016) and in the field (Fehr & Charness, 2023). Yet, despite their widespread influence on behavior, social preferences are surprisingly perceived as unimportant for cooperation in indefinitely repeated games. In a comprehensive review of the literature, Dal Bó & Fréchette (2018, p. 88) write: "It is interesting that altruistic and trusting tendencies (as captured by the dictator and trust games) do not seem to play an important role in indefinitely repeated games." This is especially noteworthy as studies on

© The Author(s), 2025. Published by Cambridge University Press on behalf of Economic Science Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/eec.2024.11 Published online by Cambridge University Press

one-shot and finitely repeated cooperation games show that many empirical findings can be explained by social preferences (Fehr & Schurtenberger, 2018; Fischbacher & Gächter, 2010; Gintis, 2000).

In this paper we provide a novel test of social preferences as a driver of cooperation in indefinitely repeated games that goes beyond correlating measures of social preferences with cooperation outcomes and thereby overcomes a recurring challenge in this literature. While it is empirically reasonable to expect a positive correlation whenever social preferences *are* a driver of cooperation, two key challenges require attention. First, indefinitely repeated games often produce a multiplicity of equilibria such that even self-interested agents may be able to sustain cooperation for strategic motives. Second, composition effects matter: for example, they can induce self-interested players to strategically cooperate if aware of the presence of individuals with social preferences, and they can induce prosocial individuals who are conditional cooperators to not cooperate if they believe others do not cooperate. To overcome these challenges, our theory-guided experiment is designed to manipulate both the equilibrium outcomes that are attainable depending on the preference composition of participants and to exogenously alter the composition of groups in terms of the prevalence of social preferences. This allows us to take into account both heterogeneity at the individual level and heterogeneity in terms of group composition (which may matter on top of individual heterogeneity, as recently shown by Proto et al. (2019, 2022) in the context of intelligence and personality traits).

Our empirical strategy is based on a series of pre-registered laboratory experiments in which we first elicit players' revealed preferences for cooperation using a sequential prisoner's dilemma game. Based on their decisions, we classify individuals as either *selfish* or *prosocial*. Using these elicited preferences, in the second part of our experiment, we sort players into different groups and let them play an indefinitely repeated version of the game. The groups thereby differ with regard to the composition of player types. We consider three types of groups: (i) *selfish groups* that only consist of participants classified as selfish, (ii) *prosocial groups* that only consist of participants classified as prosocial, and (iii) *mixed groups* that consist of a combination of both prosocial and selfish types.

Our findings reveal that pairing prosocial types with like-minded individuals has a strong and persistent effect on cooperation compared to groups of selfish individuals. We find this result across different experimental conditions. In our main experiment, we consider a situation where (i) participants are informed about the group composition and (ii) the continuation probability, δ, is sufficiently low such that under standard assumptions of material self-interest cooperative equilibria do not exist. These two conditions ensure that, from a theoretical point of view, full cooperation (and many other equilibrium outcomes) can be achieved in groups of prosocial players, while full defection is the only equilibrium outcome among groups of purely self-interested players. In line with this prediction, we find stark differences in cooperation rates across *prosocial groups* (72%) and *selfish groups* (18%).

In two additional experiments, we demonstrate the robustness of this effect when relaxing either condition (i) or (ii). In particular, when relaxing (i) by not announcing the composition of groups, we find similarly strong differences as in our main experiment with *prosocial groups* achieving a cooperation rate of 68% compared to 22% in *selfish groups*. Likewise, when relaxing condition (ii), by considering a situation in which δ is increased to a level at which mutual cooperation can be sustained as an equilibrium outcome even among self-interested players, we still find pronounced differences in cooperation across groups. While *selfish groups* now achieve considerably higher cooperation rates of 40%, they only cooperate half as often as *prosocial groups* (85%).

To test the validity of our measure of social preferences, as elicited in the first part of our experiment, we link it to a series of individual characteristics that previous literature has associated with (preferences for) cooperation. Our results reveal a strong association between our type classification and an incentivized measure of norm-following (Kimbrough & Vostroknutov, 2018), supporting the notion that (conditional) cooperation is related to norm adherence (Fehr & Schurtenberger, 2018; Kölle & Quercia, 2021). In line with previous evidence, we further find that prosocial and selfish players differ along important personality dimensions, such as agreeableness (Proto et al., 2019;

https://doi.org/10.1017/eec.2024.11 Published online by Cambridge University Press

Table 1 Monetary payoffs of the prisoner's dilemma game

		Player 2	
Player 1		*Cooperate*	*Defect*
	Cooperate	R, R	S, T
	Defect	T, S	P, P

Volk et al., 2012). Finally, we find no differences across types with regard to risk attitudes or intelligence, ruling out alternative explanations for our findings.

Our paper contributes to the growing experimental literature on indefinitely repeated games, studying the conditions that favor the emergence of cooperation (see e.g., Aoyagi et al., 2019; Aoyagi & Fréchette, 2009; Arechar et al., 2017; Bigoni et al., 2015; Camera & Casari, 2009; Dal Bó, 2005; Engle-Warnick & Slonim, 2006; Fréchette & Yuksel, 2017; Fudenberg et al., 2012; Ghidoni & Suetens, 2022; Palfrey & Rosenthal, 1994). Most related to our paper are two studies by Dreber et al. (2014) and Davis et al. (2016) who investigate the role of social preferences in indefinitely repeated games by correlating cooperation behavior with donation decisions elicited ex-post. The evidence provided in these papers is mixed, leading Dreber et al. (2014) to conclude that altruism does not play a major role in explaining heterogeneity of play in repeated games. Further related is a study by Reuben & Suetens (2012) who use a sequential version of the repeated prisoner's dilemma in which both players can condition their choice on whether the current round of the interaction is the last one or not. They find strong *end-game* effects that point to cooperation being strategically motivated. Our paper is also related to a recent paper by Kartal & Müller (2022) who propose a model with heterogeneous preferences for cooperation. They show that even when preferences are private information, there exist Bayesian Nash equilibria in which people with strong social preferences play cooperative strategies while other players defect. Their theory shows, in line with our findings, that the equilibrium level of cooperation depends on the prevalence of social preferences in a group. An additional indication that group composition could matter for cooperation comes from existing evidence that being able to choose a partner who cooperates in a repeated gift-exchange game improves cooperation (Bernard et al., 2018). Relative to the existing work, a key contribution of our paper is to *causally* identify the role of social preferences by exogenously manipulating the composition of groups. As our results demonstrate, group composition effects are crucial for social preferences to matter in repeated contexts, which can explain why previous studies that correlated a measure of social preferences with repeated game play have found no conclusive evidence. Our findings have implications for the formation of social groups and the design of institutions to foster efficiency (see Section 6 for a discussion).

The remainder of the paper is structured as follows. Section 2 discusses the theoretical implications of social preferences in the repeated prisoner's dilemma game. Section 3 describes the experimental design and hypotheses. Sections 4 and 5 report our experimental results. Section 6 concludes.

2. Theoretical considerations

Consider the monetary payoff matrix represented in Table 1 for a Prisoner's Dilemma game with $T > R > P > S$. Assuming pure self-interest and common knowledge of rationality, *Defect* constitutes a dominant strategy and, thus, {*Defect, Defect*} is the unique Nash Equilibrium (NE) in the stage game. Through the logic of backward induction, the same prediction holds when the game is repeated a *finite* number of times.

In *indefinitely repeated* contexts, in contrast, the Folk Theorem (Fudenberg & Maskin, 1986) predicts that if agents are sufficiently patient, {*Cooperate, Cooperate*} can be supported as a Nash equilibrium even among completely self-interested individuals. In particular, if both players play grim

https://doi.org/10.1017/eec.2024.11 Published online by Cambridge University Press

trigger, that is, start with cooperation, continue to cooperate until a player defects, and then defect forever, for each player mutual cooperation can be supported as a subgame perfect Nash equilibrium and yields a higher monetary payoff than unilaterally deviating to defection if:

$$\sum_{t=0}^{\infty} \delta^t R > T + \sum_{t=1}^{\infty} \delta^t P$$

or, rearranging,

$$\delta \geq \hat{\delta}^{SPE} = \frac{T-R}{T-P}$$

Hence, if players are only interested in their own monetary payoff, mutual cooperation can be sustained in a subgame perfect equilibrium (SPE) only if the shadow of the future is sufficiently long, that is, if $\delta \geq \frac{T-R}{T-P}$.

The role of social preferences becomes apparent when the shadow of the future is *not* sufficiently long. Let us call *prosocial* those agents who have sufficiently strong social preferences that their $u_i(Cooperate, Cooperate) > u_i(Defect, Cooperate)$.[1] If both players are *prosocial*, and this is commonly known, the Prisoner's Dilemma game turns into a coordination game with two Pareto-ranked equilibria.[2] Under these conditions, *prosocial* agents can sustain mutual cooperation as an equilibrium in an indefinitely repeated prisoner's dilemma for *any* level of δ. This contrasts with the theoretical predictions for agents with $u_i(Cooperate, Cooperate) < u_i(Defect, Cooperate)$, which we call *selfish*. These agents should either never cooperate (if their preferences correspond to *homo oeconomicus*), or cooperate only under certain circumstances (asDuffy & Munoz-Garcia, 2012, show in a fairness model where even moderate degrees of social preferences can make cooperation an equilibrium outcome for *some* levels of $\delta < \frac{T-R}{T-P}$).

To test the causal role of social preferences, our experiment is primarily designed to observe game play when $\delta < \frac{T-R}{T-P}$, in groups where each agent (knows that each agent) is *prosocial*, and to contrast this with game play where each agent (knows that each agent) is *selfish*.

3. The experiment

Our experiment consists of three parts. In the first two parts, participants play different variants of the Prisoner's Dilemma game as displayed in Table 2. We set the payoff for mutual cooperation to $R = 15€$, the payoff for mutual defection to $P = 10€$, the temptation payoff to $T = 25€$, and the sucker payoff to $S = 0€$. These parameters are chosen to reproduce normalized payoffs identical to Dreber et al. (2014). In Part 1, we elicit a proxy for individuals' social preferences using a sequential one-shot version of the game. In Part 2, participants play an indefinitely repeated version of the stage game. Given the parameters of the game, mutual cooperation is a possible equilibrium outcome for selfish players if $\delta \geq \hat{\delta}^{SPE} = 0.67$. In Part 3, we elicit a series of individual characteristics to assess the validity of our proxy for social preferences. In the following, we explain each part in detail.

3.1. Experimental design

Part 1: Eliciting preferences for cooperation. In Part 1, participants play a sequential one-shot version of the Prisoner's Dilemma game as shown in Table 2. Based on the design by

[1]Such utility could be consistent with many prominent models of social preferences, such as inequity concerns (Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999), feelings of guilt (Battigalli & Dufwenberg, 2007), the willingness to reward kind actions of others (Dufwenberg & Kirchsteiger, 2004; Falk & Fischbacher, 2006; Rabin, 1993), or the propensity to follow norms (Fehr & Schurtenberger, 2018; Götte & Tripodi, 202).

[2]Considerations based on the interplay of social preferences and strategic uncertainty are introduced in Section 3.3 and further developed in Appendix A.

https://doi.org/10.1017/eec.2024.11 Published online by Cambridge University Press

Table 2 Monetary payoffs in the Prisoner's Dilemma game

		Player 2	
Player 1		*Cooperate*	*Defect*
	Cooperate	15€, 15€	0€, 25€
	Defect	25€, 0€	10€, 10€

Fischbacher et al. (2001), we elicit an individual's willingness to cooperate as a function of the other player's action. To this end, participants are asked to make one *unconditional* and two *conditional* decisions. In the unconditional (first-mover) decision, participants are simply asked to choose one of the two options, cooperate or defect. In the conditional (second-mover) decisions, participants are asked to make a decision contingent on the other player's unconditional decision. Using the strategy method (Selten, 1967), we ask them (i) whether they want to cooperate or defect in case the other player defects, and (ii) whether they want to cooperate or defect in case the other player cooperates. To guarantee incentive compatibility of all choices, at the end of the experiment, in each pair, a random mechanism selects one player as the first-mover and the other player as the second-mover. For the first-mover, the unconditional decision is implemented, while for the second-mover the corresponding conditional decision (depending on the first-mover's unconditional decision) is implemented.[3]

We use a participant's responses in the conditional decisions as a proxy for their social preferences. Previous studies have shown that there is pronounced heterogeneity in individuals' preferences for cooperation. The vast majority of individuals can be classified into one out of two types: free-riders who choose to defect irrespective of the other player's decision, and conditional cooperators who are willing to cooperate if others do so too (Chaudhuri, 2011; Fischbacher et al., 2001; Gächter et al., 2017). While these studies have typically used the entire strategy profile to classify individuals, for our purpose it is sufficient to only consider a participants's revealed preference when responding to the other's cooperation. This is because the condition that guarantees the existence of a mutual cooperation equilibrium is that individuals prefer to cooperate when their matched counterpart cooperates. Therefore, we classify an individual as prosocial if she responds to the other's choice to cooperate with cooperation, and as selfish if she responds to other's cooperation with defection.

An important feature of our design is that we elicit participants' other-regarding concerns in Part 1 in a task that is closely related to the strategic decision situation that participants face in the repeated game. This approach allows us to obtain a revealed preference measure of the willingness to cooperate given others' cooperation (see Section 2), which separates the participants for whom the stage game in utilities is a coordination game (*prosocial*) from those for whom the game is a Prisoner's Dilemma (*selfish*). This parsimonious measure can be easily communicated to participants to effectively shape their perceived type distribution of the group to which they will be assigned in a way that keeps our test as close as possible to our theoretical considerations (see also Sections 2 and 3.2). This would not be possible with a more unrelated measure, such as the commonly used measure based on the dictator game, which also fails to capture the nature of social preferences at play in strategic settings.

Part 2: Indefinitely repeated game. In Part 2 of the experiment, participants play the same stage game as in Part 1 for an indefinite number of times. We use a random continuation rule: after each round, given a fixed and known continuation probability δ, a random device determines whether the game goes on for another round or stops. We fix the continuation probability at $\delta = 0.6 < \hat{\delta}^{SPE} = 0.67$, such that under narrow self-interest, mutual cooperation cannot be supported as an equilibrium outcome in the repeated game. Every time the game stops, a supergame ends and a new

[3]The description of both the unconditional and conditional decisions and the mechanism to determine the payoff-relevant choices are described in the instructions that participants read before taking their decisions.

https://doi.org/10.1017/eec.2024.11 Published online by Cambridge University Press

one begins. Participants play a total of twenty supergames. Participants remain matched with the same counterpart for all rounds within a supergame, but are randomly re-matched (in matching groups of ten participants) with a new counterpart at the beginning of a new supergame.

The crucial feature of our experimental design is that we manipulate the composition of the groups that participants are assigned to. We create three different types of groups: prosocial groups, selfish groups, and mixed groups. Groups differ with regard to the composition of types as determined in Part 1 of the experiment. Specifically, prosocial groups consist only of participants who in their conditional decision chose to cooperate if the other player cooperates; selfish groups consist only of participants who chose to defect if the other player cooperates; mixed groups consist of a combination of both these types.

Before the beginning of the repeated game, participants were informed about the type of group they were assigned to. In particular, at the beginning of Part 2 we communicated the exact decision from Part 1 that was used for the matching. In prosocial groups, we explained that all participants in the matching group chose to cooperate in the conditional decision in case their counterpart chose to cooperate, while in selfish groups participants were told that everyone chose to defect in this case. In mixed groups, we explained that the group was composed of some participants who had chosen to cooperate and some who had chosen to defect in response to other's cooperation.[4] Importantly, to avoid that participants strategically distort their revealed social preferences in Part 1 of the experiment, participants were informed about the details of Part 2 and the type of group they were assigned to only before the start of the repeated game.

A critical issue when performing experiments where participants are matched according to their behavior in previous parts of the experiment is the possibility of deceiving participants by withholding potentially payoff-relevant information. To tackle this issue, in the instructions of Part 1 we informed participants that we would use their decisions for Part 2. In particular, we told participants that in Part 2 they would either interact with players who in Part 1 made the same or different choices than themselves. This statement was true irrespective of the own type, as participants were always either placed in a segregated or mixed group. Given the uncertainty on the group matching and the fact that we pay only one out of the two parts at random (see below), it is unlikely that individuals distorted their choices in Part 1 to affect the matching of Part 2. It is further worth noting that if strategic considerations would have a strong effect on participants' responses in Part 1, for example, if many selfish individuals would pretend to be cooperative in order to be matched with and exploit prosocial types in Part 2, this should reduce potential differences between our two segregated groups and, thus, work against our hypothesized effect.

Part 3: Eliciting individual characteristics. Given recent evidence highlighting the importance of personal characteristics for outcomes in repeated interactions (Proto et al., 2019, 2022), and to provide insights into the "behavioral validity" of our social preference measure elicited in Part 1, in Part 3 of our experiment we elicit a series of individual characteristics that have been previously associated with cooperation. In particular, given the importance of personality and intelligence (Proto et al., 2019), we implement a big-five personality inventory (Schupp & Gerlitz, 2008) and elicit a measure of IQ using a 10-item Raven's progressive matrices test (Raven, 2000). We further elicit participants' general risk attitudes (Dohmen et al., 2011), gender, and their propensity to follow norms. Eliciting norm-following behavior is interesting, because (conditional) cooperative behavior has long been argued to be associated with (the willingness to follow) social norms (see, e.g., Camerer & Fehr, 2004;

[4]The exact wording for the typical session with 30 participants was as follows: "In each sequence you are paired with a participant randomly drawn from a group of 9 people. The group of 10 people (including yourself) has been determined according to one of the conditional decisions of Part 1." For prosocial [selfish] groups the instructions continued as "In particular, all participants in your group (including you) chose to play A [B] in case the other player chooses A." For mixed groups the instructions continued as "In particular, some participants in your group chose A in case the other player chooses A and some chose B in case the other player chooses A" (see Appendix E for the full instructions). For sessions with fewer than 30 participants, we kept the group size constant at one third of the session size and adjusted the wording of the instructions accordingly.

https://doi.org/10.1017/eec.2024.11 Published online by Cambridge University Press

Fehr & Fischbacher, 2004; Fehr & Schurtenberger, 2018; Kölle & Quercia, 2021). To test for this, we implement an incentivized norm-following task as introduced by Kimbrough & Vostroknutov (2018). In this task, participants are asked to allocate 50 balls between two urns. The blue urn pays 0.02€ per ball placed and the yellow urn pays 0.04€ per ball. The instructions specify that "the rule is to place the balls in the blue urn" (see Appendix F for the instructions and further details). As shown by Kimbrough & Vostroknutov (2018), the number of balls placed in the blue urn constitutes a proxy for individuals' propensity to follow norms.

3.2. Hypotheses

We start our discussion on the expected levels of cooperation in selfish groups. Recall that in these groups, we match together individuals who, in Part 1 of our experiment, have revealed a preference for defection when the other player cooperates. As discussed in Section 2, when $\delta < \hat{\delta}^{SPE}$, the unique equilibrium outcome among individuals with no social preferences is full defection. Moderate cooperation could still occur, for example, as a result of strategic mistakes (Selten & Stoecker, 1986) or moderate levels of social preferences (Duffy & Munoz-Garcia, 2012; Kartal & Müller, 2022) in selfish groups. These considerations lead us to formulate our first hypothesis that:

Hypothesis 1 Selfish groups exhibit negligible levels of cooperation.

Next, we consider the predictions for the difference between prosocial and selfish groups. As discussed in Section 2, the conditions for the existence of cooperative equilibria in the prosocial groups are that (i) all participants prefer to cooperate rather than defect when the other player cooperates and (ii) there is common knowledge that all group members have such preferences. Our experimental setup closely matches these conditions.[5] It further provides a tight test of the main proposition of Kartal & Müller (2022), where the prediction that a group cooperates at a higher rate than another requires that the distribution of social preferences in the former group first order stochastic dominates the one in the latter (see Proposition 1 (iv)) — which holds true in our setup even when allowing for a moderate level of social preferences in selfish groups.

Hypothesis 2 Prosocial groups achieve higher levels of cooperation than selfish groups.

Finally, we turn to the prediction for mixed groups, in which some members exhibit a strong degree of social preferences, while others are mainly self-interested. As shown theoretically by Kartal & Müller (2022), in mixed groups Bayesian Nash equilibria exist where, depending on their degree of social preferences, some players play cooperative strategies (such as Grim Trigger or Tit-for-Tat), and some other players always defect. These equilibria exist in addition to the full defection equilibrium. Hence, the set of possible equilibria in mixed groups is larger and includes more cooperative equilibria than in selfish groups.[6] In contrast with prosocial groups, the incomplete information about the type of other group members featuring mixed groups makes mutual cooperation much harder to achieve (Duffy & Munoz-Garcia, 2012). Additional insights regarding our mixed groups can be derived from previous studies on indefinitely repeated games, which have paired participants at random and have recruited participants from student subject pools similar to ours. Most of these studies have found that when $\delta < \hat{\delta}^{SPE}$ cooperation reaches very low levels, especially after some time (Dal Bó & Fréchette, 2018). This leads to our third hypothesis:

[5] Because we cannot announce group composition publicly, it is in principle possible that when informed of the composition of their group agents believe that not everyone in their group receives such information. Under such skeptical beliefs, the weaker notion of common knowledge that we induce is often referred to as *mutual knowledge* (Liang, 2023).

[6] Similar to Hypothesis 2, the distribution of social preferences in mixed groups first order stochastically dominates the one in selfish groups.

https://doi.org/10.1017/eec.2024.11 Published online by Cambridge University Press

Hypothesis 3 Cooperation in mixed groups will be weakly higher than in selfish groups but lower than in prosocial groups.

3.3. Additional experiments

Further to our main experiment, we provide complementary tests of our hypotheses in four additional experiments.

Robustness experiment: Random matching. This experiment closely resembles our main experiment except that after the elicitation of preferences (in Part 1), participants are matched at random into three equally sized matching groups. This experiment serves two main purposes. First, it allows us to assess whether behavior in groups in which players are told that the composition of types is mixed is comparable to behavior in groups in which players are simply matched at random without further announcements, as done in most previous studies. Second, it allows us to study how naturally occurring variation in the share of prosocial types in groups affects cooperation.

Robustness experiment: Long-run. This experiment also resembles our main experiment except that in Part 2, participants interact for 40 rather than 20 supergames. This experiment gives even greater scope for the behavior of participants to converge to their equilibrium strategies and allows us to study the persistence of group composition effects.

Additional experiment: Unannounced matching. In our main experiment participants are (i) matched into segregated groups, and (ii) made aware of the group composition. This way we can fix beliefs about the distribution of types and obtain sharp theoretical predictions on the set of achievable equilibria (see Section 2). However, announcing group composition may also come with some disadvantages. Specifically, announcing the group composition may affect behavior via channels different from (social) preferences. For instance, the announcement may shift beliefs about others' actions that could help prosocial groups to coordinate on high levels of cooperation.[7] Moreover, perfect knowledge of group composition may have little relevance empirically, as in real world settings this information is often unavailable (although people may form some impressions of others based on (imperfect) signals). To control for these issues, we ran an additional experiment in which, as in our main experiment, participants were matched based on their Part 1 choices, but without any mention of how the groups were formed and that their previous choices mattered (see Sabater-Grande & Georgantzis, 2002; Proto et al., 2019, 2022, for a similar approach).[8] By comparing the results of this experiment to those from our main experiment, we can disentangle the effect of group composition from potential confounding effects of the group composition announcement, and thus also contribute to a recent literature that examines the effects of revealing information about personality data on prosocial behavior (Drouvelis & Georgantzis, 2019; Lambrecht et al., 2024).

Additional experiment: High delta. Our main experiment was designed to test stark predictions on cooperation differences between *prosocial* and *selfish* groups. To this end, we relied on a situation in which the game does not have cooperative equilibria for self-interested agents. While such a design helps with identification, at the same time, it studies the importance of social preferences only under restrictive conditions. Therefore, to test for the role of social preferences more broadly, in an additional experiment we study a situation where mutual cooperation is a subgame-perfect Nash equilibrium of the game even for self-interested agents. Studying such a situation is interesting not only because it is the one most often studied in previous literature (Dal Bó & Fréchette, 2018, p. 74), but also because social preferences are no longer necessary to sustain cooperation. Evidence that they

[7] Alternatively, the announcement may create a positive group identity that can arise when being told to be "playing with the good guys," or affect cooperation through social norms whereby selfish types don't feel bad by defecting because it is announced that the predominant social norm in their group is to defect.

[8] Chaudhuri (2011) show that omitting information about the use of data in later parts of the experiment is perceived as a milder form of deception compared to the others analyzed and that several participants in both their students and researchers samples do not consider this practice to be deception as long as it does not involve explicit lying.

https://doi.org/10.1017/eec.2024.11 Published online by Cambridge University Press

do matter even in such contexts would support the view that social preferences matter for cooperation more generally. The design of this additional experiment is identical to our main experiment, except that in Part 2 we increase the continuation probability to $\delta = 0.8 > \hat{\delta}^{SPE} = 0.67$.

At such δ, both mutual cooperation and mutual defection are possible equilibria for selfish, mixed, and prosocial groups. Thus, a relevant question concerns equilibrium selection. A widely used criterion for predicting the stability of cooperation under strategic uncertainty is to use the size of the basin of attraction (henceforth, SizeBAD), that is, the maximum probability of the other player following the grim strategy such that playing Always Defect is optimal (Dal Bó & Fréchette, 2018). The larger SizeBAD is, the less robust cooperation is to strategic uncertainty.

To explore the interplay between social preferences and strategic uncertainty, in Appendix A we report an analysis of SizeBAD under two models of social preferences that have been shown to predict many stylized facts in the cooperation literature, a norm-following model of conditional cooperation (Fehr & Schurtenberger, 2018) and the inequity aversion model by Fehr & Schmidt (1999). We derive two main results. First, both models predict cooperation to be more stable among *prosocial* groups as δ increases from 0.6 to 0.8 as SizeBAD is decreasing in δ. Second, while the norm-following model predicts that cooperation will be more stable in *prosocial* groups than in *selfish* groups even if we assume mild levels of social preferences in the *selfish* groups, for the inequity aversion model the prediction is ambiguous as it depends on the exact parameters. As a result, depending on the type of social preferences considered, equilibrium selection considerations based on strategic uncertainty yield that cooperation rates may be either higher or lower in prosocial groups than in selfish groups. The results of our high delta experiment will speak to the question of which case holds empirically.

3.4. Procedures

The sessions of our experiments were conducted at the Cologne Laboratory for Economic Research (CLER) at the University of Cologne and at the Decision Lab of the Max Planck Institute for Research on Collective Goods in Bonn, using oTree (Chen et al., 2016). In total, we ran 38 sessions with $N = 1,074$ participants sorted in 108 matching groups (see Appendix Table B1 for an overview). Participants were students from various disciplines recruited from the subject pools of the universities in Bonn and Cologne. Our main experiment and our additional experiments were pre-registered on the AsPredicted platform (#18979, #100729 and #28887).

At the beginning of each session, participants were informed about the three-part nature of the experiment. They then received instructions explaining the general decision situation of the Prisoner's Dilemma game including some examples (see Appendix E for an English copy of the instructions). After that, participants read the instructions for Part 1 of the experiment, followed by control questions designed to ensure participants' understanding of the game. Only after each participant answered all the questions correctly, Part 1 started. Upon completion of Part 1, participants received instructions for Part 2. The instructions were again followed by a set of control questions, testing participants' understanding. As before, the experiment continued only after all the questions were answered correctly by each participant. In our main experiment as well as the long-run and the high delta experiments, before the beginning of Part 2, participants were informed about the type of group they were assigned to. In the random matching and the unannounced matching experiments, in contrast, participants were simply told that they interact in fixed matching groups, but nothing was said about the composition of groups. The length of each supergame was determined randomly within each session. That implies that the total number of rounds played is the same across all matching groups within a session, but differs across sessions. For $\delta = 0.6$, the length of the supergames ranged between 1 and 13 rounds with an average of 2.39. For $\delta = 0.8$, the length of the supergames ranged between 1 and 24 rounds, with an average of 4.93.

https://doi.org/10.1017/eec.2024.11 Published online by Cambridge University Press

After finishing Part 2, but before learning about their earnings from the experiment, participants were introduced to Part 3, containing the norm-following task, the IQ test, the big-five personality inventory, the risk question, and demographic questions. At the end of Part 3, we randomly selected either Part 1 or Part 2 (with equal chance) to determine participants' earnings. If Part 1 was selected, participants were paid either according to their unconditional or their conditional decision as described above. If Part 2 was selected, the computer randomly selected one supergame and paid the last round of that supergame. As shown by Sherstyuk et al. (2013), paying only for the last instead of all rounds has the advantage that it is theoretically robust to different risk attitudes (see also Dal Bó & Fréchette, 2018, for a further discussion of different payment methods). We further chose to pay only one of the two parts in order to avoid spillover effects due to, for example, income effects, and to limit strategic incentives for participants to distort their choices in Part 1. In addition, participants received their earnings from the norm-following task, ranging between 1€ and 2€. On average, participants earned 17.27€ for sessions that lasted between one and one and a half hours (see Appendix Table B1 for a more detailed breakdown).

4. Results

We organize the discussion of our results as follows. We start by describing the results of the strategy method in Part 1 of our experiment. After that, we describe how grouping different types of participants into mixed and segregated groups affects cooperation in the indefinitely repeated game in Part 2.

4.1. Cooperation types

In the unconditional decision of Part 1, 37% of participants chose to cooperate. In the conditional decisions, cooperation rates amount to 7% when choosing conditional on the partner's choice to defect and to 44% when choosing conditional on the partner's choice to cooperate.[9] As explained in Section 3.1, we use the latter choice to distinguish between participants with high and low degrees of social preferences. Based on this, we classify 44% of participants as prosocial and 56% as selfish.[10] Using this classification, we then formed the different groups as described above.[11]

4.2. The effects of social preferences on cooperation when group composition is known

Using the data from our main experiment, we provide a first test of our three hypotheses from the previous section. As pre-registered, in our analyses we mainly focus on first round cooperation because supergames may have different lengths and cooperation rates may depend on histories within these supergames. We show that all our findings are robust to using data from further rounds.

[9]Recall that in our announced matching treatments, participants were told that their Part 1 decisions would matter for Part 2, whereas in the random matching and unannounced matching treatments, participants were unaware of this fact. We find no evidence that this systematically affected Part 1 decisions. The cooperation rate in the unconditional decision in both cases amounts to 37% ($\chi^2 - test, p = 0.793$), and in the conditional decisions it amounts to 7% and 6%, respectively, when the other player defects ($\chi^2 - test, p = 0.243$) and to 45% and 43%, respectively, when the other player cooperates ($\chi^2 - test, p = 0.645$). These results show that knowing whether decisions matter for the subsequent part did not distort behavior.

[10]When we use both conditional choices to classify people into types as in previous studies, we find 4% unconditional cooperators, 40% conditional cooperators, 53% unconditional defectors (free-riders), and 3% mis-matchers (see Appendix Table B2 for a full breakdown of choices). The distribution of types is thereby very similar to the ones reported by Nosenzo & Tufano (2017) and Miettinen et al. (2020), for slightly different parameters of the stage game.

[11]Note that because types were determined endogenously within each session, it was not always possible to form all three types of groups. In particular, in total we had three sessions without a prosocial group and four sessions without a mixed group, leading to a total of 32 prosocial groups, 31 mixed groups, and 42 selfish groups. In addition, we had 9 groups with complete random matching. Further note that due to some no-shows, some sessions were run with fewer than 30 participants.

https://doi.org/10.1017/eec.2024.11 Published online by Cambridge University Press

Figure 1 gives a complete overview of the main findings. The left panel shows, separately for each group type, the evolution of first round cooperation rates across supergames, while the right panel shows total averages over all supergames (see Figure B1 in Appendix B for a breakdown of cooperation rates by matching group). We observe pronounced differences in cooperation rates across the three types of groups. In line with Hypothesis 2, we observe much higher cooperation rates in prosocial compared to selfish groups. Furthermore, while in prosocial groups we find very high levels of cooperation throughout the entire game, in selfish groups we find cooperation rates to converge to very low levels towards the end of the game, which is in line with Hypothesis 1.[12] Averaged over all supergames, first round cooperation rates amount to 72% in prosocial groups, four times higher than in selfish groups (18%). Very similar differences are obtained when considering data from all rounds (prosocial groups: 58%, selfish groups: 14%) or when looking at cooperation rates in the last round of each supergame excluding those supergames that only lasted one round (see Appendix Table D1). This indicates that the differences in cooperation we observe do not only arise at the beginning of a supergame, but are maintained through repeated interactions.[13]

Mixed groups, which on average are composed of 50% prosocial types and 50% selfish types, start off with intermediate levels of cooperation that are higher than in selfish groups but lower than in prosocial groups. As the game proceeds, however, cooperation rates decrease to similar levels as in selfish groups. Averaged over all supergames first round cooperation rates amount to 35%, which is only slightly higher than the levels observed in selfish groups, but much lower than in prosocial groups. This is in line with Hypothesis 3.

To test the significance of these results, we use probit regressions with the decision to cooperate as the dependent variable, and dummy variables for the different group types as independent variables. In all regressions we cluster standard errors at the matching group level. The results are shown in Table 3, reporting the cooperation levels for the first, the last ten, and all supergames combined, along with the significance levels from each pairwise comparison. Columns 2–4 show cooperation levels using first round data, while Columns 5–7 report the results using data from all rounds.

The results confirm the visual impressions from Figure 1. In particular, they reveal that the difference between prosocial groups and the other two group types are not only large in size but also statistically significant right from the start of the game. Furthermore, the differences remain statistically significant and similar in size when considering only the last ten supergames or when looking at cooperation rates from all rounds. This is reassuring as it indicates that experience or learning effects do not diminish the group composition effects that we observe. Finally, the table shows that while mixed groups cooperate at significantly higher levels than selfish groups at the beginning of the game, this difference becomes insignificant both when looking only at the last ten supergames or all supergames combined. We summarize these findings in our first result:

[12]Note that even in prosocial groups we observe a slight negative trend in cooperation. Given this time trend, one interesting question is whether the differences in cooperation rates would eventually disappear if the game is played long enough. We tackle this question in two ways. First, we follow the approach by Kartal & Müller (2022) who, based on a technique proposed by Noussair et al. 1995 and Barut et al. (2002), estimate asymptotes of cooperation rates. The results from this analysis reveal that cooperation rates in prosocial groups would not have converged to zero in the long run, but would have stayed significantly higher than in the other two types of groups (see Appendix Table B3). Second, we can rely on the data from our long-run experiment (see Section 3.3) in which participants interacted for forty instead of twenty supergames. The results, shown in Appendix Figure B2 and Table B6, reveal that cooperation rates in prosocial groups remain significantly higher than the ones observed in selfish groups even if we just focus on the last twenty supergames. Average first round cooperation rates in supergames 21–40 amount to 0.61 in prosocial groups and 0.22 in selfish groups, a difference that is highly significant ($p = 0.006$). Similar results hold when looking at data from all rounds (0.49 vs. 0.16, $p = 0.015$).

[13]One might expect that selfish groups should converge to full defection, because for self-interested agents the only equilibrium of the repeated game is mutual defection. However, in line with Section 2, this might not be the case for several reasons: participants in selfish groups may still have moderate levels of social preferences, our design may fail to induce common knowledge of the group composition, there may be measurement error in Part 1 of the experiment.

https://doi.org/10.1017/eec.2024.11 Published online by Cambridge University Press

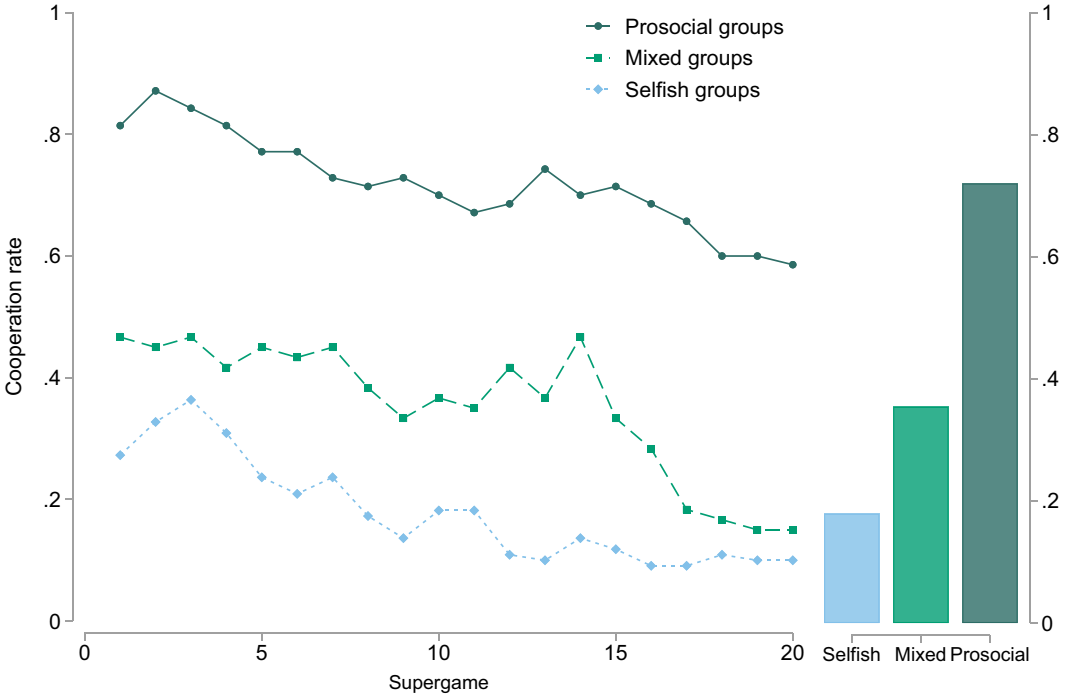


Fig. 1 First round cooperation by group type

Table 3 Cooperation rates across supergames and group type

	First round			All rounds		
	1	11-20	All	1	11-20	All
Prosocial groups	0.81	0.66	0.72	0.73	0.48	0.58
Selfish groups	0.27	0.11	0.18	0.21	0.09	0.14
Mixed groups	0.47	0.29	0.35	0.41	0.20	0.29
H_0: Prosocial = Selfish	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
	$q < 0.001$	$q < 0.001$	$q < 0.001$	$q < 0.001$	$q < 0.001$	$q < 0.001$
H_0: Prosocial = Mixed	$p < 0.001$	$p = 0.017$	$p = 0.009$	$p = 0.006$	$p = 0.032$	$p = 0.042$
	$q = 0.002$	$q = 0.026$	$q = 0.019$	$q = 0.019$	$q = 0.039$	$q = 0.042$
H_0: Mixed = Selfish	$p = 0.007$	$p = 0.185$	$p = 0.144$	$p < 0.001$	$p = 0.301$	$p = 0.208$
	$q = 0.020$	$q = 0.250$	$q = 0.250$	$q < 0.001$	$q = 0.301$	$q = 0.250$

*Notes:*Differences between group types are tested using probit regressions with standard errors clustered at the matching group level. In addition to p-values for the six tests of each hypothesis, we calculate Benjamini & Hochberg (1995) q-values to correct for multiple hypotheses testing.

Result 1 Grouping prosocial types together in segregated groups has a strong positive effect on cooperation rates.

To put our results into perspective, we can compare the achieved levels of cooperation in the different groups to those observed in previous studies. A recent survey by Dal Bó & Fréchette (2018) reports nine studies with $\delta < \delta^{SPE}$ and a minimum of seven supergames and finds that first round

https://doi.org/10.1017/eec.2024.11 Published online by Cambridge University Press

Table 4 Estimated strategy frequencies

	Prosocial groups	Mixed groups	Selfish groups
Always defect (AD)	0.264***	0.679***	0.704***
	(0.093)	(0.141)	(0.075)
Always cooperate (AC)	0.091	0.033	0.003
	(0.071)	(0.031)	(0.013)
Grim trigger (GT)	0.199***	0.095	0.061
	(0.079)	(0.075)	(0.044)
Tit-for-tat (TFT)	0.395***	0.166**	0.056
	(0.107)	(0.078)	(0.038)
Suspicious Tit-for-tat (STFT)	0.051	0.027	0.177***
	(0.038)	(0.041)	(0.074)
Gamma	0.483***	0.493***	0.436***
	(0.052)	(0.055)	(0.056)
Frequency of cooperative strategies	0.736	0.321	0.296
Observations	70	60	110

Notes: Estimates from maximum likelihood based on all rounds of all supergames. Bootstrap standard errors in parentheses. Statistical significance is assessed using Wald tests in which the null hypothesis is that a strategy frequency equals zero. *$p < 0.10$; **$p < 0.05$; ***$p < 0.01$. *Gamma* is a positive coefficient that captures the amount of noise in the decision making. The frequency of cooperative strategies is equal to 1 minus the share AD. See Appendix Tables D4 and D5 for robustness checks.

cooperation rates in the seventh supergame vary between 2% and 42%. The results from our selfish and mixed groups broadly fall within that range, amounting to 24% and 45%, respectively. The levels observed in our prosocial groups, in contrast, are by far the highest ever reported in this type of context, amounting to 72%, more than three times higher than the median of 21% of these earlier studies.

To shed some further light on the observed differences in cooperation rates across the different group types, and to investigate to what extent the different cooperation types (as elicited in Part 1 of the experiment) follow different strategies, we estimate repeated-game strategies using the Strategy Frequency Estimation Method (SFEM) as proposed by Dal Bó & Fréchette (2011). Based on previous literature (Dal Bó & Fréchette, 2019), we use maximum likelihood to estimate the prevalence of the following five strategies: Always Defect (AD), Always Cooperate (AC), Grim trigger (GT), Tit-for-Tat (TFT), and suspicious Tit-for-Tat (STFT). The only non-standard strategy is STFT, which starts by defecting and subsequently matches what the other player did in the previous round (see Appendix C for a detailed description of all strategies and estimation procedures).

Table 4 reports the estimates of the proportion for each strategy, separately for each group type.[14] Table 4 reveals several interesting patterns. First, in line with previous experiments where δ is not conducive to cooperation (Dal Bó & Fréchette, 2011), we find that in mixed and selfish groups AD is the most common strategy, amounting to 68% and 70%, respectively. These patterns change dramatically when considering prosocial groups. Here, a large majority of decisions (74%) are best described by cooperative strategies. Among the set of cooperative strategies, TFT is the most popular amounting to 40%, followed by GT with 20%. The unconditional cooperative strategy AC, in contrast, is chosen in only 9% of the cases. The fact that prosocial types predominantly play conditionally cooperative

[14]One challenge of estimating strategies is that supergames that end after just one round do not allow to distinguish between cooperative strategies. As a robustness check, in the appendix we rerun our analysis using data from only those supergames with more than one round of interactions. We find that this slightly improves the efficiency of the estimation and reduces the trembling probability *Gamma*, but that this does not qualitatively affect our point estimates (see Appendix Table D4). As an additional robustness check, we re-run our analysis by excluding the data from the first ten supergames, thus focusing only on data from those supergames in which participants already have gained experience in the game. The results, reported in Table D5 in Appendix D, are qualitatively very similar to the ones reported in the main text.

https://doi.org/10.1017/eec.2024.11 Published online by Cambridge University Press

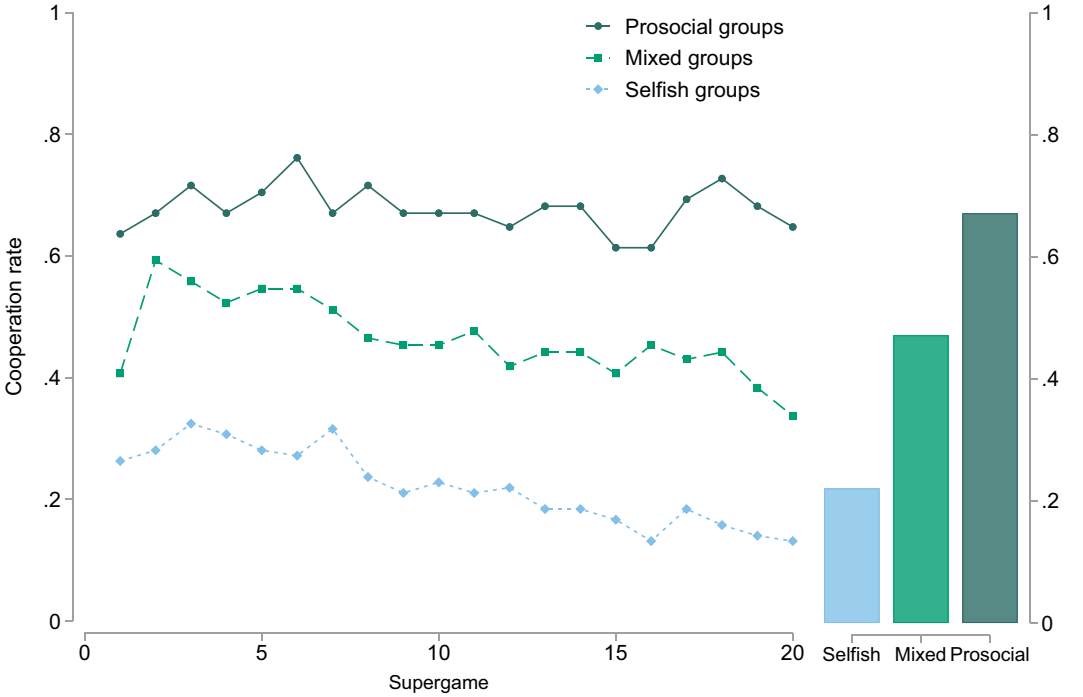


Fig. 2 First round cooperation by group type ($\delta = 0.6$, unannounced matching)

strategies is consistent with our findings from the first part of the experiment in which the large majority of prosocial types revealed that they are willing to cooperate only conditionally on others doing so, too.

In our analyses above, we have argued that behavior in mixed groups can be taken as a proxy for previous findings from indefinitely repeated experiments. The only difference with regard to the matching is that participants in our experiment were explicitly told that they would interact in mixed groups. To ascertain whether such announcement had any behavioral effect per se, we now turn to the data from our random matching experiment. We find that cooperation rates in this experiment are not significantly different from the ones observed in mixed groups in our main experiment, both when looking at first round cooperation rates ($p = 0.261$) and when using data from all rounds ($p = 0.200$). At the same time, in line with our results above (compare Table 3), we find cooperation to be significantly lower than in prosocial groups (both $p < 0.001$) and not significantly different from selfish groups (both $p > 0.555$). Overall, these results indicate that the results from mixed groups in our main experiment are a good proxy for findings from previous literature that have matched participants at random.

4.3. The effects of social preferences on cooperation when group composition is unknown

As a next step in our analysis, we move to the results from our first additional experiment in which group composition was not announced. This allows us to disentangle the effect of group composition from potential confounding factors due to its announcement.

Our main findings are shown in Figure 2, displaying average first round cooperation rates both overall and across supergames (see Appendix Figure B3 for disaggregated data at the matching group level). Differences in cooperation across the different group types remain stark as in the main experiment. Prosocial groups achieve the highest levels of cooperation (0.68), followed by mixed (0.46)

https://doi.org/10.1017/eec.2024.11 Published online by Cambridge University Press

and selfish groups (0.22). As shown in Appendix Table B4, the differences in cooperation rates are all statistically significant, both when considering data from only first or all rounds. Because differences between groups emerge already in the first supergame and remain stable as the game proceeds, we conclude that learning or experience effects do not mitigate our group composition effects.

In line with our main experiment, estimations on repeated game strategies reveal that while a large majority of decisions in prosocial groups (77%) are best described by cooperative strategies (with TFT being the modal one), the predominant strategy in selfish and mixed groups is AD with an estimated frequency of 80% and 42%, respectively (see Appendix Table B7 for further details). We summarize these findings in our second result:

Result 2 Grouping prosocial types together in segregated groups has a strong positive effect on cooperation rates even when the group composition is unknown.

These results are remarkable because, as laid out in Section 2, segregation of types and common knowledge thereof are both important for the existence of a cooperative equilibrium in prosocial groups. Here, we show that prosocial groups manage to coordinate on very high levels of cooperation even when the group composition is unknown. Moreover, comparing these results to our main experiment suggests that the announcement per se has a very minor effect on overall levels of cooperation. Specifically, when group composition is known, first round cooperation rates in prosocial groups amount to 0.72, compared to 0.68 when group composition is not announced ($p = 0.711$). Similar small differences are observed for selfish (0.18 vs. 0.22, $p = 0.516$) and mixed groups (0.35 vs. 0.46, $p = 0.385$). Taken together, these results demonstrate that group composition announcements have negligible effects relative to the effects of group composition itself.

As a final test for the importance of group composition for cooperation, we take a closer look at mixed groups in our random matching and unannounced matching experiments. As in previous experiments in the literature, in both of these settings participants were not informed about how groups were composed. Even though we have relatively few observations, we can use these to test whether the natural variation in the fraction of prosocial individuals in the group had an impact on cooperation. Our data reveals that this is indeed the case. Specifically, we find a strong positive correlation between the fraction of prosocial types and overall cooperation rates (Pearson's correlation, first round: $\rho = 0.41$, $p = 0.088$, all rounds: $\rho = 0.46$, $p = 0.055$, see also Appendix Figure B4). These results are consistent with one propositions in the model of Kartal & Müller (2022), which predicts that the number of cooperative equilibria increases in the number of cooperators in a group (see, in particular, Proposition 1 (iv)).

4.4. The role of social preferences when cooperation can be sustained also among self-interested players

As a final step in our analysis, we turn to the results of our second additional experiment, in which we set $\delta = 0.8 > \hat{\delta}^{SPE} = 0.67$. How do we expect the different group types to respond to the increase in δ? Recall from Section 3 that, according to the Folk Theorem, for such high δ, mutual cooperation can be sustained as an equilibrium outcome even among purely self-interested players. Compared to our main experiment, we should therefore expect an increase in cooperation for both *selfish* and *mixed* groups. For *prosocial* groups, however, there is no difference in the set of achievable equilibria, since mutual cooperation was already achievable at low levels of δ. Of course, an increase in δ might matter for equilibrium selection. To shed light on this question, in Appendix A, we provide a detailed analysis of the interplay of social preferences and strategic uncertainty. Our analyses yield two important insights. First, we show that cooperation is predicted to be higher even for *prosocial* groups as δ increases because strategic uncertainty decreases. Second, with respect to the expected level of cooperation across groups (recall that we now have a situation where all group types share

https://doi.org/10.1017/eec.2024.11 Published online by Cambridge University Press

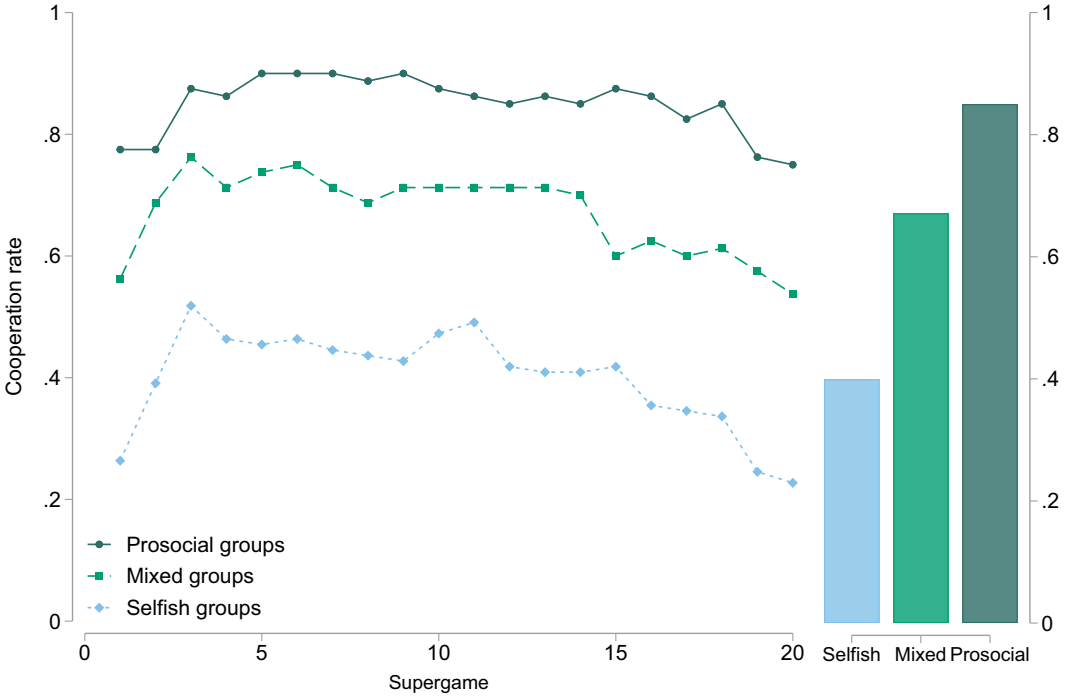


Fig. 3 First round cooperation by group type ($\delta = 0.8$)

the same set of equilibria) we show that some models of social preferences, such as norm-following as in Fehr & Schurtenberger (2018), predict cooperation to be more stable in *prosocial* than in *selfish* groups, while according to others, such as the one by Fehr & Schmidt (1999), the relative stability of cooperation across groups may be ambiguous.

Aggregate results from this experiment are shown in Figure 3 (see Appendix Figure B5 for disaggregated data at the matching group level). Once again, they reveal a strong positive effect of social preferences in promoting cooperation. Averaged over all supergames, cooperation rates amount to 85% in prosocial groups and 40% in selfish groups. As we show in Appendix Table B5, this difference is highly significant, also when considering only the last ten supergames or when using data from all rounds. As before, we find cooperation rates in mixed groups to lie in between these two extremes. Different from the case with $\delta = 0.6$, however, cooperation levels in mixed groups are now closer to those observed in prosocial groups, although still significantly lower.

The results from estimations on repeated game strategies help explain these results. As shown in Appendix Table B8, we find that, despite the fact that cooperation is a possible equilibrium outcome, in selfish groups cooperative strategies account for only about 51% of chosen strategies. The remaining 49% of choices are best described by AD, which explains why selfish groups largely fail to coordinate on the efficient outcome of mutual cooperation. In mixed and prosocial groups, in contrast, cooperative strategies account for 79% and 90% of choices, respectively (the modal cooperative strategy is TFT, followed by GT). This, in turn, explains why these groups are successful in achieving high levels of cooperation.

Comparing the observed cooperation rates with those from our main experiment, we find that, in line with evidence from previous literature (Dal Bó & Fréchette, 2018), increasing the continuation probability δ (from 0.6 to 0.8) has an overall strong positive effect on cooperation. On average, first round cooperation rates increase from 0.38 to 0.61 ($p = 0.005$). The increase in cooperation rates is

https://doi.org/10.1017/eec.2024.11 Published online by Cambridge University Press

particularly pronounced in mixed (from 0.35 to 0.67, $p = 0.028$) and selfish groups (from 0.18 to 0.40, $p = 0.005$), for which, according to the Folk Theorem (see above), mutual cooperation is now also a possible equilibrium outcome. Consistent with the prediction that increasing δ reduces strategic uncertainty (see Appendix A), cooperation rates also increase in prosocial groups, but to a lesser extent and not significantly (from 0.72 to 0.85, $p = 0.200$).

Overall, these results reveal that even in situations in which all groups can sustain mutual cooperation as an equilibrium outcome, grouping together individuals with prosocial attitudes has a strong positive effect on cooperation. This constitutes our third result:

Result 3 Even when cooperation can be sustained in equilibrium among self-interested players, a large gap in cooperation remains between prosocial and selfish groups.

5. Validity of our type classification

Our findings above have revealed strong differences in cooperation rates across prosocial and selfish groups, both when δ is conducive to cooperation for self-interested players and when it is not, and when group composition is known or not. We have argued that these differences are caused by differences in social preferences, as elicited by the strategy method. Previous research has demonstrated that the strategy method is a behaviorally valid instrument to elicit participants' attitudes (Brandts & Charness, 2011; Fischbacher & Gächter, 2010; Fischbacher et al., 2012; Gächter et al., 2017). In this section we use our post-experimental questionnaire to provide further evidence in support of our interpretation, using data from all $N = 1,074$ participants. We show that our type classification is associated with tendencies to follow norms as well as with personality traits, but does not relate to cognitive ability.

Norm following. Kimbrough & Vostroknutov (2016, 2018) have shown that behavior in the norm-following task predicts social behavior in a variety of contexts, including dictator-game giving and second-mover behavior in a trust game. Here, we demonstrate that behavior in this task is strongly correlated to participants' revealed preferences for cooperation, as elicited in Part 1 of the experiment. This is shown in Appendix Figure B6, depicting the distribution of balls that were put in the prescribed blue urn, separately for participants being classified as prosocial and selfish. The figure reveals pronounced differences across the two samples, with prosocial types being significantly more likely to follow the rule of putting the balls in the blue urn than selfish types (Kolmogorov-Smirnov test, $p < 0.001$). The average number of balls that were put in the prescribed urn amounts to 23.2 for prosocial and 15.3 for selfish types (t-test, $p < 0.001$). This result demonstrates that cooperation is indeed associated with willingness to follow social norms, as recently argued by Fehr & Schurtenberger (2018) and shown by Kölle & Quercia (2021).

Personality traits. We measure personality using the big-five personality inventory (Schupp & Gerlitz, 2008). Previous literature has provided a link between personality traits and cooperativeness. Consistent with this previous evidence (e.g. Becker et al., 2012; Proto et al., 2019; Volk et al., 2012), we find a strong positive correlation between agreeableness and the likelihood of being classified as prosocial. The evidence on the relationship between other personality traits and social preferences is more mixed. For example, with respect to conscientiousness, some previous studies found negative effects on cooperation (Proto et al., 2019), while others report either no systematic (Becker et al., 2012; Volk et al., 2012) or even a positive relationship (Dohmen et al., 2008). Similarly, some studies have found openness to be positively related to reciprocity (Becker et al., 2012; Dohmen et al., 2008), while others report negative effects (Volk et al., 2012). Our data indicate a significant negative relationship between conscientiousness and cooperativeness and a significant positive relationship between openness and cooperativeness.

Cognitive ability. Cognitive ability is measured as performance in a 10-item Raven's progressive matrices test (Raven, 2000). Previous literature has highlighted the importance of cognitive skills for

https://doi.org/10.1017/eec.2024.11 Published online by Cambridge University Press

a variety of economic primitives such as risk aversion, patience, and rationality (Benjamin et al., 2013; Borghans et al., 2008; Burks et al., 2009; Dohmen et al., 2010; Frederick, 2005; Gill & Prowse, 2016; Heckman et al., 2006; Oechssler et al., 2009). Moreover, intelligence has been shown to foster cooperation in indefinitely repeated interactions (Proto et al., 2019). Our results reveal no systematic relationship between prosociality and performance in the Raven's test: prosocial types solve on average 4.43 tasks correctly compared to 4.49 for selfish types (t-test, $p = 0.640$; see also Figure B7 in Appendix B). This result indicates that differences in cognitive abilities are unlikely to be the explanation for the different cooperation rates that we observe between the different group types.[15]

6. Conclusions

Social interactions predominantly take place under the shadow of the future. Previous literature on indefinitely repeated games has emphasized the primary role of self-interest in explaining outcomes. In this paper, we present a series of experiments highlighting the significance of social preferences. Our findings demonstrate that high levels of cooperation can persist when prosocial individuals interact in segregated groups. These conclusions hold true across various conditions, including the announcement or non-announcement of group composition, as well as favorable or unfavorable continuation probabilities for cooperation. Our results are important because they provide novel insights into the role of other-regarding motivations in repeated interactions, which can help explain some of the unexplained variation in cooperative behavior observed in previous literature (Dal Bó & Fréchette, 2018). Of course, our study only begins to explore the interplay between social preferences and strategic uncertainty. We believe that the robust role of social preferences in determining cooperative outcomes opens up new questions for future research, such as how these preferences can help refine theoretical predictions in situations involving multiple equilibria.

Our study also offers more general insights into the importance of a common level of prosociality in groups, highlighting its implications for the success of organizations and society as a whole. One direct implication is that organizations can benefit from creating homogeneous groups in terms of social preferences. Specifically, based on the data from our main experiment, our results indicate that forming two segregated groups (one comprising prosocial individuals and the other comprising selfish individuals) instead of two mixed groups can enhance group performance in terms of achieved cooperation by 30 % (from 0.35 to 0.45, compare Table 3). We note, however, that the effectiveness of this approach depends on various factors such as the length of interaction as well as people's information about their peers' types, as shown by our additional experiments. Our findings further emphasize the value of attracting prosocial individuals and fostering a culture of prosociality within organizations. For instance, when recruiting new employees, firms can utilize screening devices, such as socially beneficial commitments, to create incentives for individuals to self-sort themselves into different groups (Brekke et al., 2011; Grimm & Mengel, 2009; Hauge et al., 2019). Additionally, organizations can employ other strategies to encourage prosocial behavior among existing groups. These may include promoting integration (Goette et al., 2006), shaping people's identity (Akerlof & Kranton, 2000, 2005), investing in a socially-minded culture (Ashraf & Bandiera, 2017), or favoring the formation of (new) institutions (see e.g. Dal Bó et al., 2018; Dannenberg & Gallier, 2020; Kosfeld et al., 2009).

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/eec.2024.11.

[15] In Appendix Table B9, we analyze the relative importance of norm-following, personality, and cognitive abilities for the likelihood of being categorized as prosocial. This regression analysis shows that norm following has the strongest positive impact, while conscientiousness has the biggest negative impact on the likelihood of being a prosocial type. The results further reveal that gender and the general willingness to take risks are unrelated to participants' willingness to reciprocate others' cooperation.

https://doi.org/10.1017/eec.2024.11 Published online by Cambridge University Press

Acknowledgements. The project was approved by the Ethics Committee at the University of Cologne. Both our main experiment (#18979) and our follow-up experiments (#28887, #100729) were pre-registered on the AsPredicted platform. Research funding from the Reinhardt Selten Institute and the Center for Social and Economic Behavior (C-SEB) is gratefully acknowledged. We further gratefully acknowledge the financial support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1 390838866. We are especially grateful to Maria Bigoni, Pedro Dal Bó, Guillaume Fréchette, David K. Levine, Daniele Nosenzo, and Andis Sofianos for very constructive feedback. We thank Clara Barrocu, Eleonora Guseletova, Aenne Läufer, Carina Lenze, Jae Youn Nam, Margarita Radkova, Vincent Selz, Valerie Stottuth, Luis Wardenbach, and Rosa Wolf for research assistance. Support by Deutsche Forschungsgemeinschaft through CRC TRR 190 (project number 280092119) is gratefully acknowledged. The replication material for the study is available at https://osf.io/xv72k/. All errors remain our own.

References

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, 115, 715–753.

Akerlof, G. A., & Kranton, R. E. (2005). Identity and the Economics of Organizations. *Journal of Economic Perspectives*, 19, 9–32.

Aoyagi, M., Bhaskar, V., & Fréchette, G. R. (2019). The impact of monitoring in infinitely repeated games: Perfect, public, and private. *American Economic Journal: Microeconomics*, 11, 1–43.

Aoyagi, M., & Fréchette, G. (2009). Collusion as public monitoring becomes noisy: Experimental evidence. *Journal of Economic Theory*, 144, 1135–1165.

Arechar, A. A., Dreber, A., Fudenberg, D., & Rand, D. G. (2017). I'm just a soul whose intentions are good: The role of communication in noisy repeated games. *Games and Economic Behavior*, 104, 726–743.

Ashraf, N., & Bandiera, O. (2017). Altruistic capital. *American Economic Review*, 107, 70–75.

Barut, Y., Kovenock, D., & Noussair, C. N. (2002). A comparison of multiple-unit all-pay and winner-pay auctions under incomplete information. *International Economic Review*, 43, 675–708.

Battigalli, P., & Dufwenberg, M. (2007). Guilt in games. *American Economic Review*, 97, 170–176.

Becker, A., Deckers, T., Dohmen, T., Falk., A., & Kosse, F. (2012). The relationship between economic preferences and psychological personality measures. *Annual Review of Economics*, 4, 453–478.

Benjamin, D. J., Brown, S. A., & Shapiro, J. M. (2013). Who is 'behavioral'? Cognitive ability and anomalous preferences. *Journal of the European Economic Association*, 11, 1231–1255.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57, 289–300.

Bernard, M., Fanning, J., & Yuksel, S. (2018). Finding cooperators: Sorting through repeated interaction. *Journal of Economic Behavior & Organization*, 147, 76–94.

Bigoni, M., Casari, M., Skrzypacz, A., & Spagnolo, G. (2015). Time horizon and cooperation in continuous time. *Econometrica*, 83, 587–616.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, 90, 166–193.

Borghans, L., Duckworth, A. L., Heckman, J. J., & Ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of Human Resources*, 43, 972–1059.

Brandts, J., & Charness, G. (2011). The strategy versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics*, 14, 375–398.

Brekke, K. A., Hauge, K. E., Lind, J. T., & Nyborg, K. (2011). Playing with the good guys. A public good game with endogenous group formation. *Journal of Public Economics*, 95, 1111–1118.

Burks S. V., Carpenter, J. P., Goette, L., & Rustichini, A. (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proceedings of the National Academy of Sciences*, 106, 7745–7750.

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

Camerer, C. F., & Fehr, E. (2004). Measuring social norms and preferences using experimental games: A guide for social scientists. *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*, 97, 55–95.

Charness, G. A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics*, 14, 47–83.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics*, 14, 47–83.

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree–An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9, 88–97.

Cooper, D. J., & Kagel, J. (2016). Other-regarding preferences. *The Handbook of Experimental Economics*, 2, 217.

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic literature*, 47, 448–474.

https://doi.org/10.1017/eec.2024.11 Published online by Cambridge University Press

Dal Bó, E., Dal Bó, P., & Eyster, E. (2018). The demand for bad policy when voters underappreciate equilibrium effects. *The Review of Economic Studies*, 85, 964–998.
Dal Bó, P. (2005). Cooperation under the shadow of the future: experimental evidence from infinitely repeated games. *American Economic Review*, 95, 1591–1604.
Dal Bó, P., & Fréchette, G. R. (2011). The evolution of cooperation in infinitely repeated games: Experimental evidence. *American Economic Review*, 101, 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, 60–114.
Dal Bó, P., & Fréchette, G. R. (2019). Strategy choice in the infinitely repeated prisoner's Dilemma. *American Economic Review*, 109, 3929–52.
Dannenberg, A., & Gallier, C. (2020). The choice of institutions to solve cooperation problems: a survey of experimental research. *Experimental Economics*, 23, 716–749.
Davis, D., Ivanov, A., & Korenok, O. (2016). Individual characteristics and behavior in repeated games: an experimental study. *Experimental Economics*, 19, 67–99.
Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2008). Representative trust and reciprocity: prevalence and determinants. *Economic Inquiry*, 46, 84–90.
Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review*, 100, 1238–60.
Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, 9, 522–550.
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.
Drouvelis, M., & Georgantzis, N. (2019). Does revealing personality data affect prosocial behaviour? *Journal of Economic Behavior & Organization*, 159, 409–420.
Duffy, J., & Muñoz-García, F. (2012). Patience or fairness? Analyzing social preferences in repeated games. *Games*, 3, 56–77.
Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, 47, 268–298.
Engle-Warnick, J., & Slonim, R. L. (2006). Learning to trust in indefinitely repeated games. *Games and Economic Behavior*, 54, 95–114.
Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., & Sunde, U. (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics*, 133, 1645–1692.
Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, 54, 293–315.
Fehr, E., & Charness, G. (2023). Social preferences: fundamental characteristics and economic consequences. *CESifo Working Paper No. 10488*.
Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences*, 8, 185–190.
Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114, 817–868.
Fehr, E., & Schurtenberger, I. (2018). Normative foundations of human cooperation. *Nature Human Behaviour*, 2, 458.
Fischbacher, U., & Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review*, 100, 541–56.
Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71, 397–404.
Fischbacher, U., Gächter, S., & Quercia, S. (2012). The behavioral validity of the strategy method in public good experiments. *Journal of Economic Psychology*, 33, 897–913.
Fisman, R., Jakiela, P., Kariv, S., & Markovits, D. (2015). The distributional preferences of an elite. *Science*, 349, aab0096.
Fréchette, G. R., & Yuksel, S. (2017). Infinitely repeated games in the laboratory: Four perspectives on discounting and random termination. *Experimental Economics*, 20, 279–308.
Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19, 25–42.
Fudenberg, D., & Maskin, E. (1986). The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, 54, 533–554.
Fudenberg, D., Rand, D. G., & Dreber, A. (2012). Slow to anger and fast to forgive: Cooperation in an uncertain world. *American Economic Review*, 102, 720–49.
Gächter, S., Kölle, F., & Quercia, S. (2017). Reciprocity and the tragedies of maintaining and providing the commons. *Nature Human Behaviour*, 1, 650–656.
Ghidoni, R., & Suetens, S. (2022). The effect of sequentiality on cooperation in repeated games. *American Economic Journal: Microeconomics*, 14, 58–77.
Gill, D., & Prowse, V. (2016). Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis. *Journal of Political Economy*, 124, 1619–1676.
Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology*, 206, 169–179.

https://doi.org/10.1017/eec.2024.11 Published online by Cambridge University Press

Goette, L., Huffman, D., & Meier, S. (2006). The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups. *American Economic Review*, 96, 212–216.
Goette, L., & Tripodi, E. (2021). Social influence in prosocial behavior: evidence from a large-scale experiment. *Journal of the European Economic Association*, 19 (4), 2373–2398.
Grimm, V., & Mengel, F. (2009). Cooperation in viscous populations—Experimental evidence. *Games and Economic Behavior*, 66, 202–220.
Hauge K. E., Brekke, K. A., Nyborg, K., & Lind, J. T. (2019). Sustaining cooperation through self-sorting: the good, the bad, and the conditional. *Proceedings of the National Academy of Sciences*, 116, 5299–5304.
Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics*, 24, 411–482.
Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., & McElreath, R. (2001). In search of homo economicus: behavioral experiments in 15 small-scale societies. *American Economic Review*, 91, 73–78.
Kartal, M., & Müller, W. (2022). A new approach to the analysis of cooperation under the shadow of the future: Theory and experimental evidence. *Available at SSRN 3222964*.
Kimbrough, E. O., & Vostroknutov, A. (2016). Norms make preferences social. *Journal of the European Economic Association*, 14, 608–638.
Kimbrough, E. O., & Vostroknutov, A. (2018). A portable method of eliciting respect for social norms. *Economics Letters*, 168, 147–150.
Kölle, F., & Quercia, S. (2021). The influence of empirical and normative expectations on cooperation. *Journal of Economic Behavior & Organization*, 190, 691–703.
Kosfeld, M., Okada, A., & Riedl, A. (2009). Institution formation in public goods games. *American Economic Review*, 99, 1335–1355.
Lambrecht, M., Proto, E., Rustichini, A., & Sofianos, A. (2024). Intelligence disclosure and cooperation in repeated interactions. *American Economic Journal: Microeconomics*, 16 (3), 199-231.
Liang, A. (2023). Chapter 1: Information Partitions and Knowledge, Lecture notes.
Miettinen, T., Kosfeld, M., Fehr, E., & Weibull, J. (2020). Revealed preferences in a sequential prisoners' dilemma: a horse-race between six utility functions. *Journal of Economic Behavior & Organization*, 173, 1–25.
Nosenzo, D., & Tufano, F. (2017). The effect of voluntary participation on cooperation. *Journal of Economic Behavior & Organization*, 142, 307–319.
Noussair, C. N., Plott, C. R., & Riezman, R. G. (1995). An experimental investigation of the patterns of international trade. *The American Economic Review*, 85 (3), 462–491.
Oechssler, J., Roider, A., & Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization*, 72, 147–152.
Palfrey, T. R., & Rosenthal, H. (1994). Repeated play, cooperation and coordination: an experimental study. *The Review of Economic Studies*, 61, 545–565.
Proto, E., Rustichini, A., & Sofianos, A. (2019). Intelligence, personality, and gains from cooperation in repeated interactions. *Journal of Political Economy*, 127, 1351–1390.
Proto, E., Rustichini, A., & Sofianos, A. (2022). Intelligence, Errors, and Cooperation in Repeated Interactions. *The Review of Economic Studies*, 89, 2723–2767.
Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, 83, 1281–1302.
Raven, J. (2000). The Raven's progressive matrices: change and stability over culture and time. *Cognitive Psychology*, 41, 1–48.
Reuben, E., & Suetens, S. (2012). Revisiting strategic versus non-strategic cooperation. *Experimental Economics*, 15, 24–43.
Sabater-Grande, G., & Georgantzis, N. (2002). Accounting for risk aversion in repeated prisoners' dilemma games: An experimental test. *Journal of Economic Behavior & Organization*, 48, 37–50.
Schupp, J., & Gerlitz, J.-Y. (2008). BFI-S: big five inventory-SOEP. In *Zusammenstellung Sozialwissenschaftlicher Items und skalen. ZIS Version*, 12, 7.
Selten, R. (1967). Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes. In H. Sauermann (Eds.), *Beiträge zur Experimentellen Wirtschaftsforschung*. pp.136–168.
Selten, R., & Stoecker, R. (1986). End behavior in sequences of finite Prisoner's Dilemma supergames a learning theory approach. *Journal of Economic Behavior & Organization*, 7, 47–70.
Sherstyuk, K., Tarui, N., & Saijo, T. (2013). Payment schemes in infinite-horizon experimental games. *Experimental Economics*, 16, 125–153.
Volk, S., Thöni, C., & Ruigrok, W. (2012). Temporal stability and psychological foundations of cooperation preferences. *Journal of Economic Behavior & Organization*, 81, 664–676.

Cite this article: Kölle, F., Quercia, S., & Tripodi, E. (2025). Social preferences under the shadow of the future. *Experimental Economics*, 28(1), 22–42. https://doi.org/10.1017/eec.2024.11

https://doi.org/10.1017/eec.2024.11 Published online by Cambridge University Press