Baader, Malte; Gächter, Simon; Lee, Kyeongtae; Sefton, Martin

Working Paper

Social Preferences and the Variability of Conditional Cooperation

IZA Discussion Papers, No. 15523

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Baader, Malte; Gächter, Simon; Lee, Kyeongtae; Sefton, Martin (2022) : Social Preferences and the Variability of Conditional Cooperation, IZA Discussion Papers, No. 15523, Institute of Labor Economics (IZA), Bonn

This Version is available at:
http://hdl.handle.net/10419/265744

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15523

Social Preferences and the Variability of Conditional Cooperation

Malte Baader
Simon Gächter
Kyeongtae Lee
Martin Sefton

AUGUST 2022

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15523

Social Preferences and the Variability of Conditional Cooperation

Malte Baader
University of Zurich

Simon Gächter
University of Nottingham, IZA and CESifo

Kyeongtae Lee
Bank of Korea

Martin Sefton
University of Nottingham

AUGUST 2022

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Social Preferences and the Variability of Conditional Cooperation[*]

We experimentally examine how the incentive to defect in a social dilemma affects conditional cooperation. In our first study we conduct online experiments in which subjects play eight Sequential Prisoner's Dilemma games with payoffs systematically varied across games. We find that few second movers are conditionally cooperative (i.e., cooperate if and only if the first mover cooperates) in all eight games. Instead, most second-movers change strategies between games. The rate of conditional cooperation is higher when the own gain from defecting is lower and when the loss imposed on the first mover from defecting is higher. This pattern is consistent with both social preference models and stochastic choice models. To explore which model explains our findings we employ a second study to jointly estimate noise and social preference parameters at the individual level. The majority of our subjects place significantly positive weight on others' payoffs, supporting the underlying role of social preferences in conditional cooperation.

JEL Classification: A13, C91

Keywords: Sequential Prisoner's Dilemma, conditional cooperation, social preferences

Corresponding author:
Martin Sefton
Centre for Decision Research and Experimental Economics (CeDEx)
University of Nottingham
University Park
Nottingham, NG7 2RD
United Kingdom
E-mail: martin.sefton@nottingham.ac.uk

[*] This work was supported by the European Research Council [grant numbers ERC-AdG 295707 COOPERATION and ERC-AdG 101020453 PRINCIPLES] and the Economic and Social Research Council [grant number ES/K002201/1]. Ethical approval for the experiments was obtained from the Nottingham School of Economics Research Ethics Committee. We are grateful to Robin Cubitt, Friederike Mengel, Brown Bag audiences at the Centre of Decision Research and Experimental Economics, and attendees at the UEA Behavioural Game Theory Workshop on Cooperation, Norwich 2022, for their comments and suggestions. S.G. acknowledges the hospitality of briq Bonn while working on this paper.

1. Introduction

Conditional cooperation is widely observed in social dilemmas. Whereas the pursuit of narrowly defined selfish interests would result in a lack of cooperation, many people are willing to forgo their selfish interests and cooperate, but only if others do so as well. This pattern of behavior is particularly clear in controlled experiments investigating contributions to public goods.[1] These experiments also reveal substantial heterogeneity: for example, in some of these studies (see Thöni and Volk (2018) for a review) some group members are classified as "free-riders" (i.e., defecting regardless of the behavior of others), others as "conditional cooperators" (i.e., cooperating if others do so), and still others as "unconditional cooperators" (i.e., cooperating independently of the behavior of others). Identifying these heterogeneities is crucial when trying to understand what makes individuals cooperate and which measures to take to further enhance cooperation.

Despite this commonly applied classification, not much is known, however, about whether it reflects stable personality traits whereby the participant would exhibit similar behavioral patterns in similar situations, or whether the classification applies only to the specific experimental setting and parameters. There is also surprisingly little evidence on how the specific material payoffs of the game affect conditional cooperation. In case that the degree of conditional cooperation varies with game parameters, it is fundamental to understand the mechanisms and account for them when studying group cooperation. To provide first insights into this dimension, in this study, we examine whether the behavioral pattern exhibited by a given participant, such as conditional cooperation, varies in response to changes in the material incentives.

Examining the within-subject variability of conditional cooperation across payoff variations is important for at least two reasons. First, it allows us to understand the nature of conditional cooperation: whether conditional cooperation reflects underlying social preferences, or whether conditional cooperation reflects a desire to reciprocate the cooperation of others in a way that is robust to changes in material incentives.[2] Social preference models,

[1] See, e.g., Bilancini, et al. (2022); Brandts and Schram (2001); Chaudhuri and Paichayontvijit (2006); Croson (2007); Cubitt, et al. (2017); Fischbacher, et al. (2001); Fischbacher and Gächter (2010); Furtner, et al. (2021); Gächter, et al. (2017b); Gächter, et al. (2022); Isler, et al. (2021); Keser and van Winden (2000); Kocher, et al. (2008). For reviews see Chaudhuri (2011); Fehr and Schurtenberger (2018); Gächter (2007); and Thöni and Volk (2018).

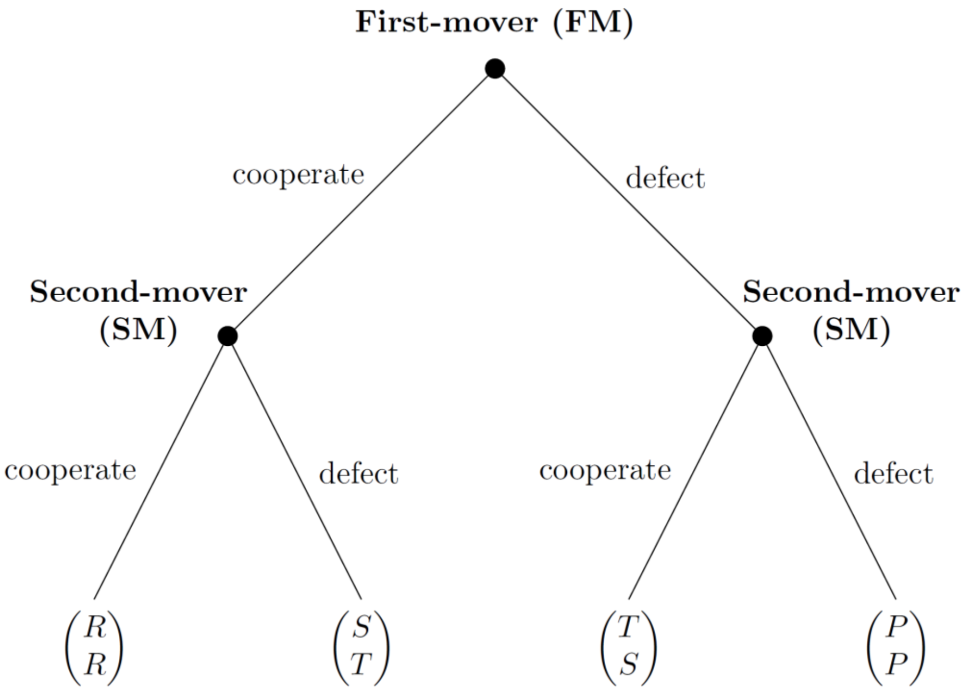
[2] As discussed by Bardsley and Sausgruber (2005), Fehr and Fischbacher (2004), Gächter, et al. (2017a), and Katuščák and Miklánek (2018), conformity to what is perceived as "socially appropriate" and willingness to sacrifice material payoffs in order to follow such norms could also be a candidate explanation for conditional cooperation.

which define preferences over one's own and other's material payoffs (e.g., Andreoni and Miller (2002); Bolton and Ockenfels (2000); Charness and Rabin (2002); Cox, et al. (2007); Cox, et al. (2008); Fehr and Schmidt (1999)), are capable of explaining conditional cooperation, but at the same time these models predict that it will be influenced by material incentives. In contrast, if conditional cooperation reflects a principled stand against free-riding, eschewing material gains to reciprocate the cooperation of others, then conditional cooperation is expected to be robust across payoff variations.

Second, the efficacy of interventions to promote cooperation depends on whether conditional cooperation is influenced by payoff variations. For example, leading by example would be an effective mechanism to achieve cooperative outcomes if followers are generally conditionally cooperative (e.g., Gächter, et al. (2012)). On the other hand, if conditional cooperation is sensitive to payoffs, then this implies that there are settings where leading by example is ineffective.

We study within-subject variability of conditional cooperation using two experimental designs. Our first experimental design is based on the sequential prisoner's dilemma shown in Figure 1.

FIGURE 1. The Sequential Prisoner’s Dilemma (SPD)



$Note : T > R > P > S; 2R > T + S$

In this game First-mover (FM) chooses either cooperate or defect, and, after observing this choice, Second-mover (SM) chooses either cooperate or defect. Combined earnings are maximized when both cooperate, resulting in each player receiving R. However, if FM cooperates, SM maximizes own earnings by defecting, in which case FM receives S and SM receives T. If FM defects, SM maximizes own earnings by also defecting, so that each player receives P. Thus, a selfish SM who maximizes own earnings should defect regardless of FM's choice.

In our experiments, subjects make decisions in the role of both FM and SM for eight different games with varying payoffs. We elicit SM strategies by asking how the subject would respond to defect and how they would respond to cooperate, with their actual decision being determined by their response to their opponent's actual FM choice. To rule out confounding factors, such as belief updating, no feedback on any of the individual games is provided until the end of the experiment. This simple experimental design allows us to examine whether conditional cooperation is stable across varying payoffs, and, if not, how conditional cooperation is affected by changes in incentives.

In our first study, we conducted two online experiments, the first using Amazon Mechanical Turk (AMT) and to explore the robustness of our findings, we replicated the study using a University of Nottingham (UoN) student subject pool. AMT workers are demographically more diverse and older than typical student subject pools. From the results of related research studying cooperation rates in simultaneous PDs (Gächter, et al. (2021)) and public goods games (Arechar, et al. (2018)) we expect AMT workers to exhibit a higher level of cooperativeness than the younger UoN student subject pool. Comparing AMT workers and UoN students thus provides us with an opportunity to gauge variability of conditional cooperation across subject pools.

For our AMT experiment we find that the proportion of conditional cooperators varies between 27% and 48% across games, while for our UoN experiment the proportion varies between 28% and 48%. Fractions of free rider strategies range from 40% to 55% in AMT and from 43% to 59% in UoN. Moreover, in both experiments more than 70% of subjects change their SM strategy at least once across games. An implication of this is that any classification of individuals as "conditional cooperators" or "free-riders" in one game should not be generalized to other games with different material payoffs: a conditional cooperator in one game may be a free-rider in another, and vice versa. Additionally, we also find that changes in behavior are systematically related to payoffs: SM are more likely to conditionally cooperate when they have less to gain from free-riding, and when free-riding has a larger negative impact on the FM's earnings.

This pattern is consistent with the predictions of several social preference models (e.g., Charness and Rabin (2002); Fehr and Schmidt (1999)), but it is also consistent with stochastic choice models where subject choices are determined by selfish preferences plus noise (e.g., Anderson, et al. (1998)). Thus, we developed a second experimental study to jointly estimate individual-level noise and social preference parameters. In this study, which we conducted as a laboratory experiment using our UoN student subject pool and a modified version of the SPD, subjects make choices either as FM or as SM in sixty-four games with varying payoffs,

As in Study 1, in Study 2 we find substantial heterogeneity across individuals. Some SMs consistently respond to cooperation by maximizing their own payoff (12%), while others consistently respond by cooperating (6%). However, most SMs vary their responses to cooperation across games in a way that is systematically, though not deterministically, related to payoffs. For 72% of SMs we estimate a random utility model incorporating social preferences and find that most of these (representing 66% of all SMs) have significantly positive social preference parameters that place a positive weight on their opponent's payoff. Thus, our results support the view that conditional cooperation reflects underlying social preferences.

The remainder of the paper is organized as follows. In Section 2, we place our contributions in the related literature that examines the variability of conditional cooperation and the relationship between social preferences and conditional cooperation. In Section 3 we describe the design and results of our online experiments, and in Section 4 we present the design and results of our lab experiment. In Section 5 we conclude.

2. Related literature and our contributions

A number of previous papers have examined the variability of conditional cooperation over time, by measuring conditional cooperation repeatedly but keeping payoff functions constant. The results are mixed. Brosig, et al. (2007) conducted similar SPDs to us three times within three months using the same subjects and random-matching and found that the rate of conditional cooperation diminished across repetitions. This finding is supported by Andreozzi, et al. (2020) that also found conditional cooperation diminished with repetition. Exploring public goods games, Muller, et al. (2008) elicited subjects' strategies across five repetitions. Although only 37% of subjects always chose the same strategy across all five games, previous choices were useful predictors of subsequent choices. For example, 69% of subjects who conditionally cooperated in any of the first four games also conditionally cooperated in the fifth game. Volk, et al. (2012) elicited subjects' strategies in a public goods game three times over the course of five months and observed that conditional cooperation was remarkably stable over time. Half of their subjects chose the same strategy in all three games, and 71% of these conditionally cooperated. In a closely related analysis, Gächter, et al. (2022) report stability rates of 66% and 59% in their provision and maintenance versions of public goods games played four months apart.[3]

Our approach differs from this previous literature because we examine robustness whilst varying the payoffs across the eight SPDs people play. We are not aware of any study examining how within-subject variation of payoffs affects conditional cooperation.[4] We are only aware of three studies that examine whether payoff variation affects conditional cooperation between subjects. Thöni and Volk (2018) found that the proportion of conditional cooperators is similar across 17 public goods experiments, which employ different parameters (i.e., MPCRs, group size). In contrast, Clark and Sefton (2001), using a between-subjects SPD experiment in which subjects played repeatedly against changing opponents with feedback on the outcomes of each play, found that doubling the temptation payoff, *T*, resulted in a significantly lower rate of conditional cooperation. Our studies differ from these in that we ask subjects to make decisions in multiple games with systematically varying payoffs and without feedback across games. This within-subject design allows us to examine how changes in payoffs affect conditional cooperation at the individual level.

Our paper also contributes to the literature of social preferences explaining decisions in experimental public goods games. One of the first papers in this literature is Blanco, et al. (2011). They measure parameters of disadvantageous and advantageous inequality aversion (Fehr and Schmidt (1999)) using ultimatum and modified dictator games and then have the same subjects play among others an SPD and a public good. They find that the elicited preference parameters predict decisions at the aggregate level but not so much at the individual

[3] Two further studies (Eichenseer and Moser (2020) and Mullett, et al. (2020)) examine the variability of conditional cooperation across different contexts by comparing behavior in a public goods game and a SPD. Both studies report that subjects who are conditionally cooperative in a SPD are also conditionally cooperative in a public goods game.

[4] Several studies examine how decisions in the simultaneous prisoner's dilemmas are influenced by payoff variations (e.g., Ahn, et al. (2001); Au, et al. (2012); Charness, et al. (2016); Engel and Zhurakhovska (2016); Mengel (2018); Ng and Au (2016); Schmidt, et al. (2001); Vlaev and Chater (2006)). See Gächter, et al. (2021) for a discussion of these papers and a systematic experimental analysis of the role of payoff parameters for cooperation in prisoner’s dilemma experiments.

level. Hedegaard, et al. (2021) elicit distributional preferences in a representative Danish sample and then use them to explain behavior in trust and public goods games. Our approach differs from these papers. Unlike them, we do not elicit preference parameters in some games to predict behavior in others. Instead, we estimate preference parameters using a series of SPDs that systematically vary the payoff parameters.

Our approach therefore directly relates to an experimental literature testing models of social preferences and estimating social preference parameters (see Cooper and Kagel (2016), for a review). Many experiments in this area are based on designs where individuals are randomly assigned to different treatments and tests of models are based on making treatment comparisons. It is typically the case that there are too few observations on individual subjects to estimate individual preference parameters, and so estimations are based on population regressions (e.g., Charness and Rabin (2002)). We take a fundamentally different approach by having subjects make many choices in a modified version of the SPD game with varying payoffs, enabling us to estimate preference parameters at the individual level. In this regard, our Study 2 experiment is most closely related to a literature initiated by Palfrey and Prisbrey (1997) who estimate altruism and warm glow parameters in public goods games, and Andreoni and Miller (2002), and Fisman, et al. (2007), who estimate individual preferences for giving by having subjects make choices in modified dictator games with varying endowments/prices of giving. A more recent related paper is Bruhin, et al. (2019), who estimate structural social preference models from binary choices, although their emphasis is on finite mixture models.[5]

3. Study 1: Measuring the variability of conditional cooperation

3.1. Experimental design

Our first experimental design is based on the sequential prisoner's dilemma game of Figure 1. Our main interest concerns how changes in payoffs affect conditional cooperation focusing on two factors. First, LOSS ≔ $(R - S)/R$ refers to FM's loss when SM responds to cooperation by defecting rather than cooperating. Second, GAIN ≔ $(T - R)/R$ refers to SM's gains from responding to cooperation by defecting rather than cooperating. We also manipulate the efficiency gains from cooperation, EFF ≔ $(R - P)/R$.

[5] Our approach is also similar to that used in a considerable literature on individual choice experiments where individual risk preferences are estimated from responses to a battery of lottery choices (see, for example, Hey and Orme (1994), Andersen, et al. (2008)).

Table 1 shows the payoff parameterization used in our experiment and the resulting values of EFF, LOSS and GAIN.[6] Payoffs were chosen to be strictly positive multiples of ten in order to avoid zero or non-rounded payoffs. All games satisfy the conditions $T > R > P > S$ and $2R > T + S$, so that mutual cooperation maximizes combined earnings, but, assuming players are selfish, own-earnings maximizers, the Nash equilibrium is always mutual defection. R (500) is constant across all games while there are two distinct values of P (200, 400). Thus, we study games with two different levels of efficiency. There are also two distinct values of T (600, 800) and four distinct values of S (20, 90, 40, 180). Note that with this parameterization we study a 2×2 variation in LOSS and GAIN for each level of efficiency.

TABLE 1. Payoff parameters for Sequential Prisoner's Dilemma Games

Game	*R*	*P*	*S*	*T*	EFF	LOSS	GAIN
G1	500	200	90	600	0.60	0.82	0.20
G2	500	200	20	600	0.60	0.96	0.20
G3	500	200	90	800	0.60	0.82	0.60
G4	500	200	20	800	0.60	0.96	0.60
G5	500	400	180	600	0.20	0.64	0.20
G6	500	400	40	600	0.20	0.92	0.20
G7	500	400	180	800	0.20	0.64	0.60
G8	500	400	40	800	0.20	0.92	0.60

Note: EFF = $(R - P)/R$; LOSS = $(R - S)/R$; GAIN = $(T - R)/R$.

3.2. Experimental procedures

We conducted our initial online interactive experiment in Spring 2019 using Amazon MTurk across five sessions with a total of 138 participants. We refer to this as our AMT experiment. To further examine the robustness of our first data collection, we replicated the AMT experiment with a different subject pool in Summer 2021, referred to as our UoN experiment, using students from the University of Nottingham who had signed up to a subject database for participating in experiments. For this experiment we used ORSEE (Greiner (2015)) to recruit subjects and conducted an additional three sessions with a total of 152 participants. Between both data sets we observe significant differences in demographics, such as age, gender and ethnic composition, allowing us to explore potential treatment effects for a range of individual

[6] This is the same parameterization used in Gächter, et al. (2021) for studying cooperation in simultaneous PDs.

characteristics (see Appendix A for the details). Both online experiments were programmed using LIONESS Lab (Giamattei, et al. (2020)), and the same program was used for both experiments, with only minimal changes to the instructions related to the subject pools (for the instructions, see Appendix B).

Each participant was paired with another subject after they had read the instructions and passed some control questions. Each pair then played all eight games of Table 1 with no feedback between games. For each game, subjects had to answer eight additional control questions about the payoffs before making decisions. These additional control questions were intended to ensure that subjects understood the implications of their decisions and recognized the payoff changes across games. Subjects then made decisions as FM as well as SM. Both decision tasks were presented on the same screen. For the FM decision, they simply chose whether to cooperate or to defect. For the SM decision, we asked subjects to decide in the following two situations: i) if FM cooperates, and ii) if FM defects. Therefore, we elicited SM strategies using the strategy method (Selten (1967)).[7] Rather than use the terms "cooperate" or "defect", we labeled options neutrally as A or B, with labeling randomly chosen at the pair level in each game. To control for potential order effects, we randomized the sequence of games and the order of tasks (i.e., placing the FM or SM decision at the top of the screen) at the pair level. Once subjects completed the tasks for all games, we asked them to complete a short post-experimental questionnaire eliciting basic demographic information.

We paired subjects with another participant on a real-time basis, and they made decisions in each game at the same time. That is, they could not proceed to the next game until both had completed their decisions for the current game.[8]

To elicit subjects' responses in an incentive-compatible way, we implemented the following payment scheme. At the end of the session, one of the eight games was randomly chosen at the pair-level for payment. If both subjects completed the entire experiment, they

[7] Regarding potential differences between responses elicited using the strategy method and those using a direct response method, previous studies found no statistical differences in subjects' responses between these two methods (see Brandts and Charness (2000), and Brandts and Charness (2011) for a review and Keser and Kliemt (2021) for a recent discussion). See also Fischbacher and Gächter (2010); Fischbacher, et al. (2012); Gächter, et al. (2017b); Isler, et al. (2021); Gächter, et al. (2022) who all find that combining strategies with beliefs explains contributions in direct response public goods games supporting the behavioral validity of the strategy method.

[8] To reduce the risk of decreased attention due to long waiting times as subjects waited for their opponent to decide, we took the following measures. Before participants entered the experiment, we told them to avoid distractions during the experiment. In addition, participants who were inactive for more than 30 seconds (i.e., no mouse movement or no keyboard input) got an alert voice message and a blinking text on their browser. If an inactive participant did not respond to the alert message for a further 30 seconds, such an inactive participant was removed from the experiment and the remaining person was able to continue the experiment.

were paid according to the outcome of this game as follows. One of the pair was randomly chosen to be FM, and the other was selected to be SM. Then, subjects were reminded of their decisions and informed about the outcome for this game. As mentioned above, for SM's decision we used their conditional response to FM's decision. If one of the pair had dropped out during the experiment, the computer randomly selected the payoff-relevant game for the remaining subject. Then the computer randomly selected one out of four monetary outcomes (i.e., T, $R, P,$ or S) of the chosen game for payment to the remaining subject. We explained this payment scheme clearly in the instructions. This payment procedure gives subjects a monetary incentive to take both FM decisions and SM decisions seriously in all games as any of these decisions can become payoff relevant.

In line with other online experiments, there was a non-negligible attrition rate: 32 out of 138 AMT subjects (23%) and 9 out of 152 UoN subjects (6%) dropped out during the experiment.[9] For subjects who completed the AMT experiment, the average age was 34.2 years (s.d. 10.2 years) and 37% were female, while for the UoN experiment the average age was 22.5 years (s.d. 4.6 years) and 58% were female. AMT subjects' earnings ranged from $1.20 to $9.00, averaging $4.59, while UoN subjects' earnings ranged from £1.20 to £9.00, averaging £4.53. On average, the experiment lasted about 30 minutes and subjects were informed of their payment immediately upon completion of the experiment and were paid within 24 hours.

3.3 Results

For our analysis, we only include the decisions of subjects who completed the experiment: thus, our data set consists of 848 AMT observations (106 subjects × eight games) and 1144 UoN observations (143 subjects × eight games). In line with our research question, our focus is on SM decisions as these provide a direct measure of conditional cooperation. These conditionally cooperative strategies, where SM responds to cooperation with cooperation and defection with defection, make up 38% of the strategies elicited in the AMT experiment, and 37% of the strategies elicited in the UoN experiment. Free-riding strategies (i.e., unconditional defection) make up 45% of the AMT strategies and 53% of the UoN strategies. There are relatively few unconditionally cooperative strategies (AMT: 12%; UoN: 7%) and even fewer strategies that respond to defection with cooperation and cooperation with defection (AMT: 5%: UoN: 3%)

[9] The dropout rate in our AMT sessions is similar to that of related interactive online experiments. For example, Arechar, et al. (2018) report a 20% dropout rate in their interactive four-player public goods game, and Gächter, et al. (2021) reports a 24% dropout rate in their interactive eight simultaneous prisoner's dilemma games.

Thus, as in other social dilemma experiments (e.g., Fischbacher, et al. (2001); Fallucchi, et al. (2019); Gächter, et al. (2022); Isler, et al. (2021); Miettinen, et al. (2020); Muller, et al. (2008); Thöni and Volk (2018)), conditional cooperation and free-riding make up the bulk of elicited strategies (87% in aggregate).

However, we find that most subjects cannot be unambiguously classified as a 'free-rider' or 'conditional cooperator' because they vary their strategy between games. Of AMT subjects, 13% always free-ride, 13% always conditionally cooperate, and the remainder change strategies at least once across games. For the UoN experiment, 19% always free-ride, 10% always conditionally cooperate, and the remainder change strategies at least once across games.

Table 2 reports the proportion of conditionally cooperative strategies, broken down by game, for the AMT and UoN experiments. Panel (a) gives the proportion in the AMT experiment depending on the levels of LOSS and GAIN for the high efficiency games, while panel (b) does the same for the UoN experiment. Results for low efficiency games are reported in panels (c) and (d) in a similar fashion. Table 2 also reports changes in proportions, and McNemar tests of significance of these, for pairwise comparisons between games where LOSS or GAIN varies and other payoffs are held constant.[10]

We find a consistent pattern in both experiments, where for all four possible combinations of efficiency and GAIN, conditional cooperation is higher in the high than the low LOSS game, and for all four possible combinations of efficiency and LOSS conditional cooperation is higher in the low than the high GAIN game. To test whether conditional cooperation varies significantly across games, we conduct pairwise comparisons using McNemar's test. In the AMT experiment, subjects are significantly less likely to conditionally cooperate when GAIN increases in the low efficiency games. In the UoN experiment the effect of GAIN is significant except when efficiency is low and LOSS is low, and the effect of LOSS is significant in the low efficiency game with low GAIN. For other cases the differences are not statistically significant.

[10] We conducted a similar analysis for the other strategies. For free-riding the results mirror those shown in Table 2 very closely: where conditional cooperation increases between games free-riding increases, and vice versa. The proportions of the other two strategies are small and do not vary across games in a systematic way (only 1 of 32 pairwise comparisons is significant at the 5% level). See Appendix A, Table A3-5 for details.

TABLE 2. Proportions of Conditionally Cooperative Strategies

(a) High Efficiency Games (EFF = 0.6) – AMT

		LOSS		
		Low (= 0.82)	*High* (= 0.96)	Δ (*H-L*)
	Low (= 0.20)	35.8%	42.5%	+6.7%
GAIN	*High* (= 0.60)	33.0%	38.7%	+5.7%
	Δ (*H-L*)	-2.8%	-3.8%	

(b) High Efficiency Games (EFF = 0.6) – UoN

		LOSS		
		Low (= 0.82)	*High* (= 0.96)	Δ (*H-L*)
	Low (= 0.20)	37.8%	44.1%	+6.3%
GAIN	*High* (= 0.60)	28.0%	32.9%	+4.9%
	Δ (*H-L*)	-9.8%**	-11.2%**	

(c) Low Efficiency Games (EFF = 0.2) – AMT

		LOSS		
		Low (= 0.64)	*High* (= 0.92)	Δ (*H-L*)
	Low (= 0.20)	44.3%	48.1%	+3.8%
GAIN	*High* (= 0.60)	27.4%	34.9%	+7.5%
	Δ (*H-L*)	-16.9%***	-13.2%**	

(d) Low Efficiency Games (EFF = 0.2) – UoN

		LOSS		
		Low (= 0.64)	*High* (= 0.92)	Δ (*H-L*)
	Low (= 0.20)	36.4%	48.3%	+11.9%***
GAIN	*High* (= 0.60)	32.2%	38.5%	+6.3%
	Δ (*H-L*)	-4.2%	-9.8%**	

Note: EFF = $(R - P)/R$, LOSS = $(R - S)/R$, GAIN = $(T - R)/R$. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$ indicate p-values based on McNemar's test.

To summarize how incentives affect conditional cooperation, controlling for demographic characteristics, we present the marginal effects from logit regressions for our AMT and UoN experiments in Table 3. The dependent variable is 1 if the subject conditionally cooperated and 0 otherwise. Our experiments manipulate LOSS and GAIN across two levels of EFF, and so we report how conditional cooperation varies with the indices of LOSS and GAIN separately for our two levels of efficiency. These regressions control for individual and task characteristics, round, and session effects. A full set of results is given in Table A3 in Appendix A.

TABLE 3. Determinants of Conditional Cooperation

	AMT	UoN
LOSS*EFF$_{high}$	0.175	0.289***
	(0.120)	(0.089)
LOSS*EFF$_{low}$	0.294**	0.332***
	(0.118)	(0.089)
GAIN*EFF$_{high}$	-0.150	-0.234***
	(0.092)	(0.072)
GAIN*EFF$_{low}$	-0.335***	-0.159**
	(0.089)	(0.076)
Observations	800	1,104

Notes: Average marginal effects from logit regression with robust standard errors clustered at the individual level. Dependent variable = 1 if SM conditionally cooperated, 0 otherwise. LOSS = $(R - S)/R$, GAIN = $(T - R)/R$, EFF$_{high}$: indicator variable for $(R - P)/R = 0.6$, EFF$_{low}$: indicator variable for $(R - P)/R = 0.2$. Regressions include controls for individual characteristics, task characteristics, round and session effects.
** $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$*

As in Table 2, we see consistent patterns in the sign of the effects of GAIN and LOSS on conditional cooperation. As GAIN increases and it becomes more profitable for a SM to free-ride, we observe lower levels of conditional cooperation, and as LOSS increases and free-riding has a larger negative impact on FM's earnings we observe more conditional cooperation. These effects are significant at the 5% level or lower in the UoN experiment, while for the AMT experiment the effects are significant at the 5% level or lower for the low EFF games, but not so for the high EFF games.

Including controls allows us to explore how conditional cooperation is affected by the heterogeneity in demographics across our two experiments. More detailed results are given in Appendix A Table A5, but here we note a significant effect of age and gender for the UoN sample, where older and female students are more likely to conditionally cooperate. In contrast to our expectation, we do not observe the same result on age in the AMT sample even though we have more age variation in the AMT data. We do, however, find significant heterogeneity in conditional cooperation between ethnicities in the AMT sample.

These results also control for task characteristics (whether cooperation was labelled A or B, and whether FM or SM decision fields were presented first), round effects, and session effects. In the UoN experiment neither round dummies ($\chi^2(7) = 9.75$, $p = 0.203$) nor session dummies ($\chi^2(2) = 2.13$, $p = 0.344$) are jointly significant. For AMT, however, there are significant round effects ($\chi^2(7) = 15.47$, $p = 0.030$), reflecting a higher rate of conditional cooperation in the first round compares to later rounds, and significant session effects ($\chi^2(4) = 10.38$, $p = 0.035$) effects. In neither experiment are the task characteristics significant.

In summary, across our online experiments we find that conditional cooperation varies across games. Most subjects change strategies across games, and this switching between strategies varies systematically with the distributional consequences of free-riding relative to conditionally cooperating. Subjects are more likely to conditionally cooperate when free-riding imposes larger losses on the FM, or when free-riding provides smaller gains for oneself. Our finding that strategies are sensitive to the cost imposed on the opponent as well as the gain to self suggests that a substantial proportion of subjects care not only about their own material payoffs but also about the other's material payoffs. Moreover, the way conditional cooperation varies with LOSS and GAIN is consistent with the predictions of several distributional preference models. For example, consider the Fehr and Schmidt (1999) model of inequality aversion or the “distributional preference” model by Charness and Rabin (2002). According to these models the SM maximizes utility by defecting in response to defection, while the optimal response to cooperation depends on how much weight the SM places on the disadvantaged FM’s payoff (Charness and Rabin’s ρ parameter) or the marginal disutility from earning more than the FM (Fehr and Schmidt’s β parameter). Applied to our game SM will conditionally cooperate if ρ (or β) > GAIN/(GAIN + LOSS), and free-ride otherwise. Thus, given a distribution of preference parameters in the population, more individuals in the population will conditionally cooperate when GAIN is lower, or LOSS is higher.

However, it should be noted that behavior is not perfectly aligned with these models: we observe some individuals sometimes unconditionally cooperate, some individuals sometimes switch from free riding to conditionally cooperating when GAIN increases, and so on. At best, our data is consistent with noisy versions of these models.[11] In fact, the systematic effects of GAIN and LOSS are also consistent with stochastic choice models in which subjects maximize selfish utility with error. For example, in Appendix C we present a quantal response equilibrium analysis and show that the QRE probability of conditional cooperation increases with LOSS and decreases with GAIN. Since our data does not allow us to distinguish which of these alternative models drives our results, we designed a further experiment to separately estimate the effects of social preferences and noise.

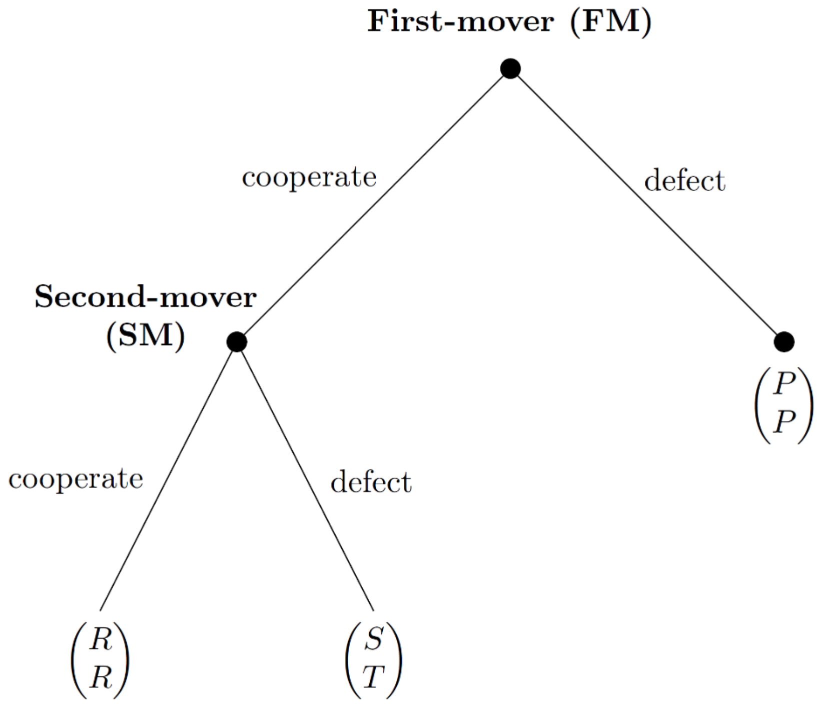
4. Study 2: Explaining the variability of conditional cooperation

Our second experimental design attempts to jointly estimate noise and preference parameters at an individual level. To obtain meaningful estimates it is necessary to have a subject play many games and so we modified our initial design in a number of ways. First, we used an in-person lab experiment to avoid problems of attrition and to enhance control. Second, we simplified the task by having subjects play in fixed roles. Third, we simplify the SM decision by having second movers only make a choice in response to cooperation – effectively, we hardwire defection as a response to defection (based on the results from our earlier design, where the response to defection is defect in over 85% of cases, we think not much is lost from this simplification). The game implemented in our lab experiment is shown in Figure 2.

Despite these changes, we kept most of the features identical to our first experiment. To elicit SM responses to cooperation we retained the contingent choice element of our earlier design. That is, FM and SM made choices at the same time and a SM choice was only relevant for payoffs if the FM cooperated. Thus, both players make binary choices in each game.

[11] There is considerable evidence that errors and confusion play a significant role for behavior in public goods games. See, for instance, Andreoni (1995), Bayer, et al. (2013), Burton-Chellew, et al. (2016); Ferraro and Vossler (2010); Houser and Kurzban (2002); Palfrey and Prisbrey (1997). Errors may also affect conditional cooperation to some extent (e.g., Fosgaard, et al. (2017); Gächter, et al. (2022)).

FIGURE 2. The Modified Sequential Prisoner's Dilemma Game



Additionally, as in our earlier design, we kept R (500) constant in all games and used the same four values of S (20, 40, 90, 180). We expanded the set of values of P (100, 200, 300, 400), and T (400, 600, 800, 1000), to obtain 64 games. These include the 8 games of our original design, and 22 more games satisfying the PD conditions $T > R > P > S, 2R > T + S$. In addition, there are 15 games where $T > R > P > S$, but $T + S > 2R$, so that the Nash equilibrium outcome is mutual defection, while combined earnings are maximized when FM cooperates and SM defects. In addition, there are 16 games where $R > T$, so that the SM maximizes own earnings by cooperating. For these games the Nash equilibrium outcome is mutual cooperation (and in one of these $S > P$ so the FM has a dominant strategy to cooperate). Finally, there are 3 more games where $S > P$ and so a FM has a dominant strategy to cooperate. (A complete list of games and parameters is provided in Appendix F, Table F1.) Thus, most of our games are dilemmas but the inclusion of other games means that a subject motivated to maximize own earnings cannot achieve this by using a simple heuristic of always defecting, and similarly a subject motivated to maximize combined earnings cannot use a simple heuristic of always cooperate. This feature of our experimental design provides us with an additional opportunity to examine the attentiveness of subjects and gain some insights into the rationalizability of choices.

4.2 Experimental procedures

We conducted our experiment in June 2022 in the CeDEx lab using University of Nottingham students. We conducted 13 sessions with a total of 194 participants (97 SMs). Subjects were recruited using ORSEE (Greiner, 2015) and the experiment was conducted with the software LIONESS Lab (Giamattei, et al. (2020)). 42% of the subjects were female and the average age was 22.1 yrs (s.d. 3.69 years). The experiment was pre-registered (AEARCTR-0009536).[12]

At the beginning of the session each participant was given a set of instructions, and these were read aloud by the experimenter. Subjects then answered control questions before beginning the decision-making part of the session. As in the online experiments, subjects were anonymously paired with another subject and then played all 64 games with no feedback between games.

In contrast to our online study, we asked subjects to make their choices on a graphical implementation of the decision tree as outlined in Figure 2 (see instructions in Appendix D). We again utilized neutral labels, where for each game, the FM was Person A and chose between options A1 and A2, while the SM was described as Person B and chose between options B1 and B2. In addition, we elicited beliefs about the other person's choice. As before, to control for potential order effects, we randomized the sequence of games at the pair level. Once subjects completed the tasks for all games, we asked them to complete a short post-experimental questionnaire eliciting basic demographic information.

At the end of the session two games were randomly chosen for each pair. One of these games was used to determine additional earnings based on game choices, based on an exchange rate of £0.02 per point. The other game was used to determine additional earnings based on beliefs. Subjects were rewarded in lottery tickets using a binarized scoring rule (Hossain and Okui (2013)), and these determined their chances of winning a prize of 200 points (i.e., £4). The instructions did not describe the precise binarized scoring rule to subjects. Instead, they were told that they maximized their chances of winning the prize by reporting their beliefs as accurately as possible. The instructions also offered to reveal the precise mechanism after the experiment to interested subjects (only one subject took up the offer). We adopted this procedure following Danz, et al. (2022) who show that despite a potential centrality bias using the binarized scoring rule, not outlining the details of the incentive mechanism results in most accurate belief elicitations.

[12] https://www.socialscienceregistry.org/trials/9536. We aimed for 200 participants but due to show-up problems we ended up with 194.

Subjects received a £5 show up fee and earnings ranged from £5.40 to £25.00, averaging £16. On average, the experiment lasted about 60 minutes, including the completion of a post-experimental questionnaire. Subjects were informed of their payment immediately upon completion of the experiment and were paid within 24 hours.

4.3 Econometric model

To jointly estimate the effects of noise and social preferences we use a random utility model incorporating social preferences. First, following Fehr and Schmidt (1999) and Charness and Rabin (2002) we assume SM's utility depends on both own-earnings and other's earnings as follows:

$$u_{SM}(\pi_{FM}, \pi_{FM}) = \rho\pi_{FM} + (1-\rho)\pi_{SM} = \pi_{SM} - \rho(\pi_{SM} - \pi_{FM})$$

The parameter ρ is the weight that SM places on FM's payoff when SM earns at least as much as FM in the Charness-Rabin model. It can also be interpreted as the marginal disutility from advantageous inequality in the Fehr-Schmidt model (their β parameter). In all 64 games SM earns at least as much as FM, and so we do not need to distinguish between the weights placed on the other's payoff when ahead and when behind, or between advantageous and disadvantageous inequality.

Second, we assume SM holds beliefs about FMs choice and assigns probability q to FM cooperating. Given these assumptions and our payoff parameterization, SM's expected utility from cooperating is

$$V_c = qR + (1-q)P$$

and the expected utility from defecting is

$$V_d = q(T - \rho(T - S)) + (1-q)P.$$

We assume SM follows a stochastic choice rule and cooperates if and only if

$$V_d - V_c < Z, \text{ where } Z \sim N(\mu, \sigma^2).$$

Thus, SM cooperates unless the expected utility gain from defecting exceeds a normally distributed noise term with mean μ. A positive parameter μ thus results in a bias toward cooperation that is independent of payoffs. That is, even if the choice rule is deterministic ($\sigma = 0$), and even if the expected payoff from defection exceeds the expected payoff from cooperation, SM may choose to cooperate if the expected utility difference is less than μ. The higher is μ, the more likely it is that SM will cooperate even when expected payoff maximization points toward defecting. Including μ in our econometric specification therefore accounts for potential individual bias and provides us with more robust estimates of the social preference parameter.

From this choice rule, letting $\Phi(\cdot)$ denote the standard normal distribution function, the probability of cooperating as a function of the payoffs, beliefs, preference, and noise parameters is:

$$\Pr\{SM\ cooperates\} = \Phi(\mu/\sigma + q(R-T)/\sigma + \rho q(T-S))/\sigma).$$

We then estimate μ, σ and the preference parameter ρ for each subject using maximum likelihood probit, providing us with individual estimates corresponding to our key dimensions of interest, social preferences (ρ) as well as noise (σ).

Prior to estimation we conducted Monte Carlo simulations to explore the properties of the estimators under alternative data generating processes (see Appendix E). The main take-aways from our simulations are, first, that sometimes the model fails to produce sensible estimates. Partly this reflects the familiar under-identification problem in discrete choice models, whereby maximum likelihood parameters cannot be estimated when the data is perfectly predicted by the regressors. This occurs if a subject cooperates in every decision, or defects in every decision. But it also occurs, for example, if choices follow deterministically from the social preference model so that a subject cooperates if and only if $\rho > (T-R)/(T-S)$ for some value of ρ. Another reason for failure to produce sensible estimates is that estimates of σ may be negative. Related to this, when estimates of σ are very close to zero, corresponding estimates of ρ can vary wildly. Thus, in our data analysis we treat individuals with $\hat{\sigma} < 0$ or $|\hat{\rho}| > 1$ as outliers.

Second, controlling for potential bias in choices, $\mu \neq 0$, by including a constant in the regression, and controlling for variability in beliefs across games by including beliefs in the regression, are important parts of our estimation strategy. When the data generating process includes a bias term, $\mu \neq 0$, estimates from a model without a constant are severely biased, whereas when the data generating process does not include a bias term, $\mu = 0$, estimating a model with a constant comes at a small price (mainly, adding a constant term increases the chance of running into identification problems). Similarly, when choices are based on expected utility differences and depend on q, estimating a model assuming $q = 1$ leads to severely biased estimates.

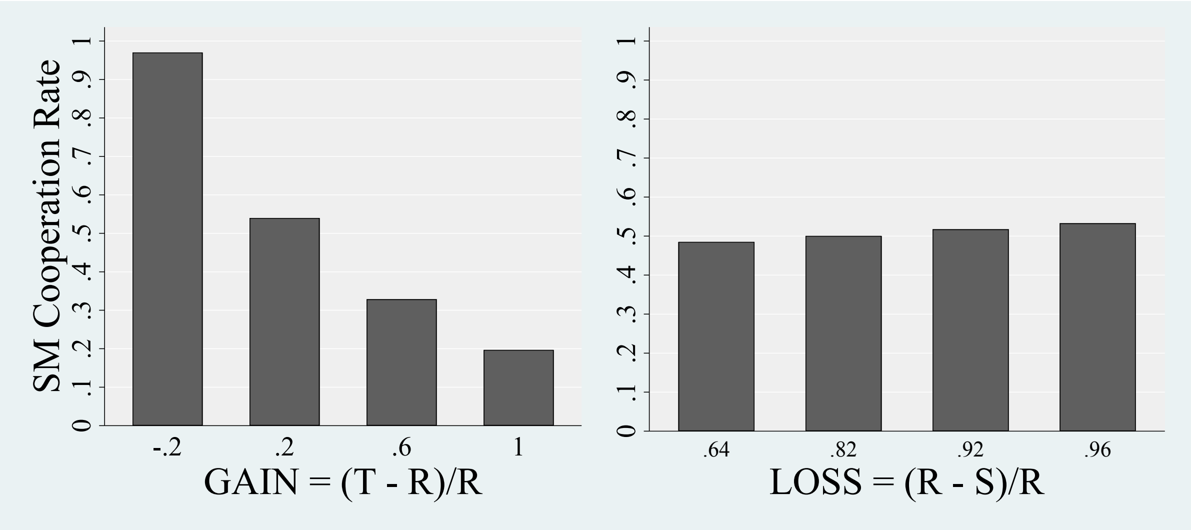
Third, standard errors can be very large, particularly when σ is large. An implication of this is that point estimates may be very imprecise estimates of underlying parameters. On the other hand, the finite-sample bias in the estimate of ρ is small, and so the average estimate of ρ across many individuals gives a useful estimate of the underlying mean parameter in the population.

4.4 Results

4.4.1 Descriptive statistics

Before estimating our model, we give some summary information on choices and beliefs.[13] Figure 3 shows how cooperative choices relate to GAIN and LOSS. As in our online experiment, the effect of GAIN is very clear: the higher is gain, the lower is the cooperation rate. Recall, that in our design SM's payoff from defecting is lower than the payoff from cooperating in 16 games (i.e., in the $T = 400$ games, which implies GAIN = -0.2)). As it turns out, SM almost always cooperates in these games. Note also that in these games the private gain from cooperating is quite small (100 points), and the same as the private gain from defecting when $T = 600$. The fact that the cooperation rate in the $T = 400$ games is almost 100% while the cooperation rate in the $T = 600$ games is substantially above zero is a first hint that cooperation reflects more than just selfish motives plus error.[14]

FIGURE 3. SM Cooperation Rate by GAIN and LOSS



[13] See Appendix F, Table F1, for a complete table of average beliefs and average choices for each of the 64 games.

[14] This is not conclusive evidence because it ignores the role of beliefs. SM may have a higher expectation that FM cooperates when $T = 400$ than when $T = 600$, and so the expected private gain from cooperating when $T = 400$ may be higher than the expected private gain from defecting when $T = 600$.

With respect to LOSS, as in our online experiment, there is a positive relationship, although much less pronounced than for the case of GAIN. These results are further confirmed in a logit regression model that examines how SM cooperation varies with GAIN and LOSS for the four levels of efficiency. The results are reported in Table 4 and replicate previous results from the online experiment: SM cooperation increases with LOSS and decreases with GAIN, though the effect size of GAIN is substantially greater (c.f. Table 3).

TABLE 4. Determinants of SM Cooperation (64 games)

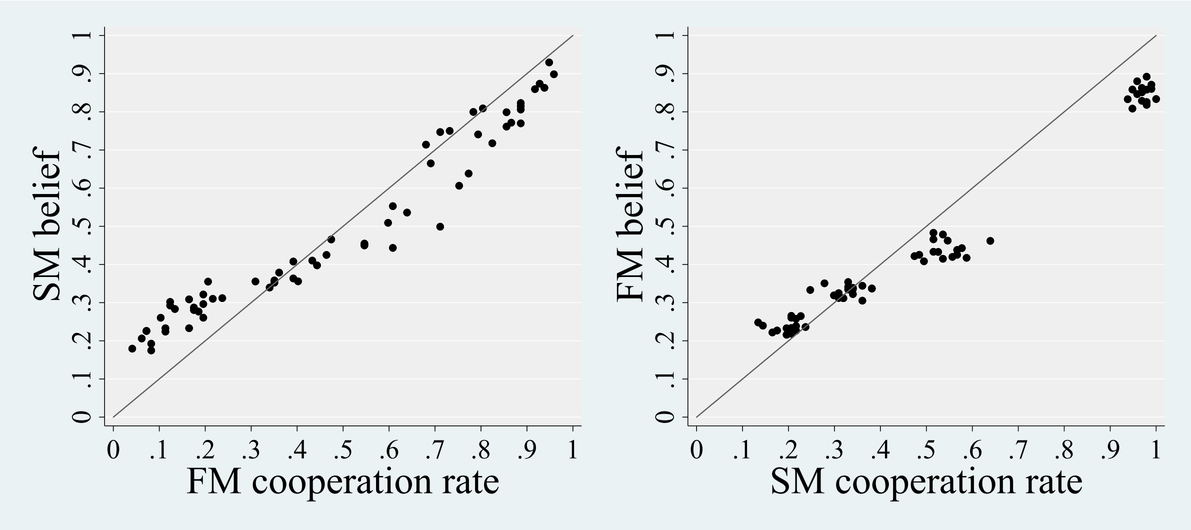
LOSS*EFF$_{=0.8}$	0.145*** (0.040)
LOSS*EFF$_{=0.6}$	0.152*** (0.041)
LOSS*EFF$_{=0.4}$	0.131*** (0.042)
LOSS*EFF$_{=0.2}$	0.146*** (0.041)
GAIN*EFF$_{=0.8}$	-0.561*** (0.020)
GAIN*EFF$_{=0.6}$	-0.560*** (0.018)
GAIN*EFF$_{=0.4}$	-0.512*** (0.020)
GAIN*EFF$_{=0.2}$	-0.556*** (0.020)
Observations	6,016

Notes: Average marginal effects from logit regression with robust standard errors clustered at the individual level. Dependent variable = 1 if SM conditionally cooperated, 0 otherwise. LOSS = (R – S)/R, GAIN = (T – R)/R, EFF = (R – P)/R. Controls: demographic variables, round and session effects. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$

Next, we examine how *beliefs* relate to choices in our 64 games. Figure 4 shows how beliefs about the other player's choice are related to the other player's actual choice. The left panel shows SM beliefs against the actual cooperation rates of FMs. The Spearman correlation is 0.984 (p < 0.001), suggesting that SM beliefs are quite well-calibrated, although there is a tendency for SM to somewhat over-estimate low FM cooperation rates and under-estimate FM high cooperation rates.

In the right panel of Figure 4 we present FM beliefs about SM choices. Again, there is a high correlation across the 64 games: the Spearman correlation coefficient is 0.949 (p < 0.001). For FMs, there is a clear clustering of beliefs, and it reflects differences in GAIN. Thus, for the $T = 400$ games FMs expect SMs to cooperate at a high rate (the average belief is 85%), although the actual SM cooperation rate is in fact even higher than this (97%). At $T = 1000$ on average FM expect SMs to cooperate 24% of the time (slightly over-estimating the cooperation rate of 20%).

FIGURE 4. Beliefs and Choices



4.4.2 Estimation results

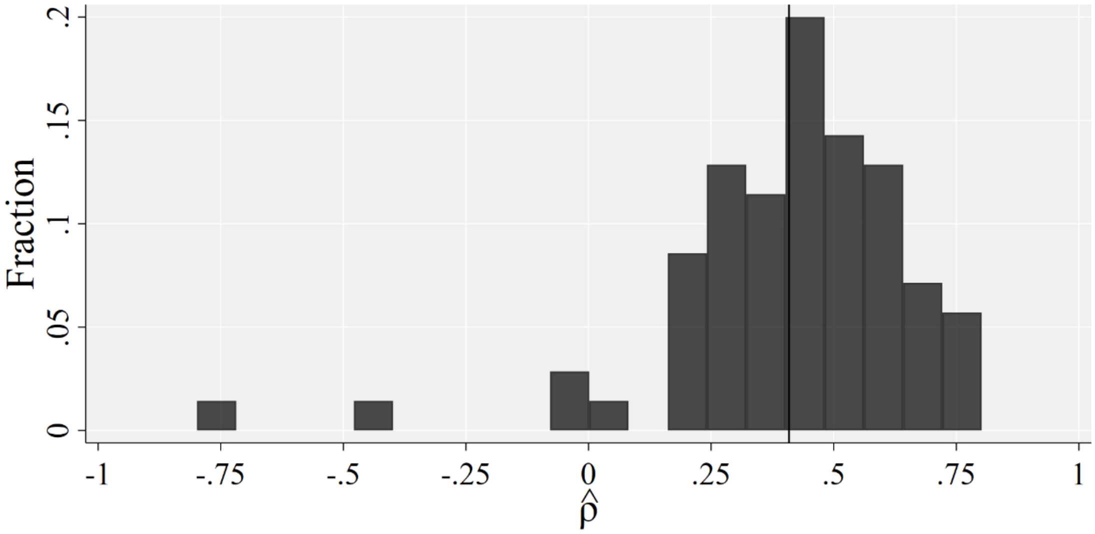
Next, we turn to the estimation of our social preferences model using maximum likelihood probit regressions. We are unable to estimate parameters for 26 of 97 SMs for whom a linear combination of regressors perfectly predict choices. For example, 6 SMs always chose to cooperate, and 12 SMs always chose to maximise own-earnings (i.e., defect when $T > R$ and cooperate when $T < R$). Of the remaining 71 SMs for whom we can estimate parameters, one is estimated with $|\hat{\rho}| > 1$, which we exclude as an outlier. The rest of our analysis of SMs is based on the remaining sub-sample of 70 SMs. A complete list of the individual estimates is provided in Table F2 in Appendix F.

Figure 5 presents a histogram of the 70 ρ parameter estimates. The average estimate is 0.41 (s.d. 0.25), and using this to estimate the mean ρ parameter in the population we conclude the mean parameter is statistically significantly different from zero (p < 0.001).[15] Of the 70 subjects, 64 have significantly positive estimates of ρ. Thus, the majority of our second movers have significantly positive social preference parameters and place positive weight on FM's

[15] To test this hypothesis, we use the average estimate in our sample, 0.41, as our estimate of the population mean, and for a standard error of this estimate we use $se(\sum\hat{\rho}/n) = (1/n)\sqrt{(\sum se(\hat{\rho}_i)^2)}$ = 0.1137.

payoff. This presents clear evidence in line with models of social preferences to explain variability in conditional cooperation under changing payoff parameters. It is interesting to compare these results with previous experiments that use modified dictator games to estimate the Fehr-Schmidt β parameter. In line with our results, Blanco et al. (2011) give an average estimate of β of 0.47 (s.d. 0.31), while Beranek, et al. (2015) (using UoN students) find an average estimate of β = 0.48 (s.d. 0.29).

FIGURE 5. Distribution of SM social preference parameter (ρ) estimates



For the noise parameters, the average estimate of μ is 5.36 (s.d. 289.76) and the average estimate of σ is 62.48 (s.d. 109.93). Note, the average bias estimate is very small, and we cannot reject the hypothesis that the mean bias in the population is zero (p = 0.978).[16] However, there is substantial variability in the sample: many subjects have a large estimated bias and 38 estimates are significantly negative (i.e., displaying a bias toward defection). Both of these additional pieces of evidence further support that social preferences are in fact the main driver of our previously established results.

For our sub-sample of 70 subjects Table 5 summarizes the predictive accuracy of our model and compares it with the predictive accuracy of two alternative models. For our first alternative, we simply predict cooperation using a probit model with a constant. That is, for each SM we predict cooperation (defection) in all games if that subject cooperates (defects) in most games). Note that this model must successfully predict at least 50% of an SM's choices.

[16] We again follow the same approach as before, using our average estimate, $\sum\hat{\mu}/n = 5.36$, as our estimate of the mean bias and for a standard error we use $se(\sum\hat{\mu}/n) = (1/n)\sqrt{(\sum se(\hat{\mu}_i)^2)} = 195.77$.

Across our sub-sample we find it predicts 67% of choices correctly. For a second alternative, we predict cooperation using a random utility model with ρ constrained to be zero (i.e., for each SM we predict decisions based on a constant and expected utility differences, where utility depends only on own-payoff). This increases predictive success considerably, and correctly predicts 80% of SM's choices, on average. Adding social preferences to the model increases the predictive success to 89%. We also report another widely used measure of predictive success, the pseudo R^2 measure. Since these measures of predictive success generally increase with the number of predictors in the model, we also report McFadden's adjusted Pseudo R^2, which penalizes for the number of predictors. Even using this measure, the model with social preferences improves predictive success relative to the other models.

TABLE 5. Measures of Predictive Success

Model	Average Hit Ratio	Average Pseudo R^2	Average adjusted Pseudo R^2
Constant	0.67	0	-0.03
Random utility with selfish preferences	0.80	0.31	0.25
Random utility with social preferences	0.89	0.56	0.47

4.4.3 Reciprocity

Our social preference model as outlined above is based on preferences defined over the distribution of payoffs. This means that when deciding how to respond to cooperation by FM, SM weighs up the utility of cooperation, which depends on R, and the utility from defection, which depends on S and T. The payoff parameter P does not directly enter SM's utility. However, it is possible that P does in fact matter for SM choices if subjects have reciprocal preferences. That is, SM considers that FM is being kind by cooperating (see, e.g., Falk, et al. (2003)), and so SM cooperates to reward this act. How kind FM is to SM could be measured in alternative ways that depend on P, the payoff that FM forgoes by choosing to cooperate. One can argue that $P-S$ is a relevant measure of kindness, as by cooperating FM forgoes P and risks getting S. Alternatively, one could argue that $R-P$ is a more relevant measure of kindness from cooperating as these are the cooperative gains being offered to SM.

To test whether some form of reciprocation is playing a role in our setting, we do not model reciprocal preferences explicitly, but rather we simply test whether the weight SM places on FMs payoff changes with P. That is, we suppose that SM's probability of cooperating is given by:

$$\Pr\{SM\ cooperates\} = \Phi\left(\mu/\sigma + q(R-T)/\sigma + q\sum_k \rho_k\, \mathbb{I}_{P=P_k}(T-S))/\sigma\right)$$

where $\mathbb{I}_{P=P_k}$ is the indicator function for the four possible values of P. We estimate separate ρ_k parameters for each value and test the hypothesis that all four are equal.

We can estimate this model for 68 subjects (70%). Of these, there are only four subjects for whom the weights significantly vary with P at the 10% level. Thus, in our specification, we can only find limited evidence of reciprocity as an explanation of conditional cooperation. It appears that most of the variation in cooperation we observe can be explained by heterogeneity in social preferences rather than due to additional reciprocal concerns.

4.4.4 First Mover choices

Although our main focus is on SM's response to cooperation, FM cooperation rates also vary substantially across the 64 games, ranging from as little as 4.1% up to 95.9% in another.[17] To further examine a potential explanation behind FM cooperation, we can also apply the random utility model with social preferences to FM choices. Let FM's utility be given by

$$u_{FM}(\pi_{FM}, \pi_{SM}) = \tau\pi_{SM} + (1-\tau)\pi_{FM} = \pi_{FM} + \tau(\pi_{SM} - \pi_{FM}).$$

Here, the parameter τ is the weight FM places on SM's payoff when SM earns more than FM (the Charness-Rabin σ parameter). (Recall, in our games SM always earns at least as much as FM.) It can also be interpreted as minus one times Fehr-Schmidt's disadvantageous inequality aversion parameter.

With this utility function, FM's utility from defecting is P, and since this determines the outcome with certainty FM's expected utility from defecting is

$$V_d = P.$$

FM's expected utility from cooperating is

$$V_c = qR + (1-q)(S + \tau(T-S))$$

where q is the probability FM assigns to SM cooperating. Using the random utility choice rule, FM cooperates if and only if

[17] Table F1 in Appendix F reports average FM cooperation rates for each of the 64 games

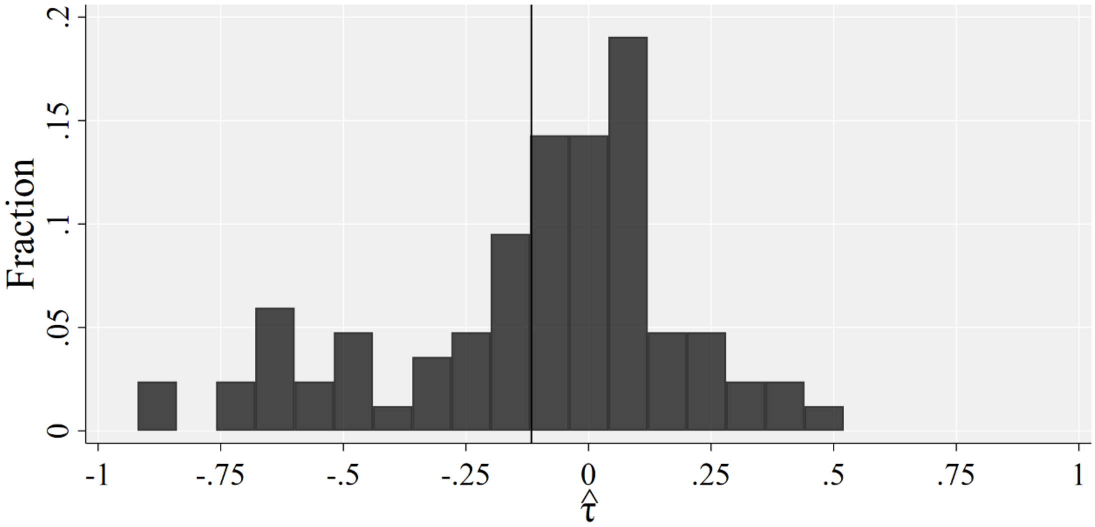
$$V_d - V_c < Z, \text{ where } Z \sim N(\mu, \sigma^2),$$

and from this it follows that FM cooperates with probability:

$$\Pr\{FM\ cooperates\} = \Phi(\mu/\sigma + (qR + (1-q)S - P)/\sigma + \tau(1-q)(T-S)/\sigma).$$

As before, we estimate the parameters μ, σ and τ using maximum likelihood probit regressions. We are able to estimate individual parameters for 84 of 97 FMs. The average estimate of τ is -0.11 (s.d. 0.30). Sixty-two of the estimates are insignificantly different from zero. Eleven are significantly negative and 13 are significantly positive. Despite this heterogeneity we find that the population mean of τ is significantly different from zero (p = 0.010) providing evidence for other-regarding preferences.[18] Figure 6 shows a histogram of the estimates.

FIGURE 6. Distribution of FM social preference parameter (τ) estimates



Most notably, we find that the weight FM places on SM's payoff tends to be lower than the weight SM places on FM's payoff. This is consistent with the assumption in Charness and Rabin (2002) that the weight placed on the other's payoff depends on whether a player is ahead or behind. Note also that a substantial proportion of FMs put a positive weight on SM's payoff, even when SM earns more – these subjects are not consistent with difference aversion models since only those subjects with $\tau < 0$ are averse to disadvantageous inequality in the sense of Fehr and Schmidt. Lastly, as in the case for SM, for the FM we also find large heterogeneity in

[18] In line with the SM analysis, we compute the standard error of our estimate as $se(\sum\hat{\tau}/n) = (1/n)\surd(\sum se(\hat{\tau}_i)^2) = 0.044$.

our estimated bias with a small and insignificant population mean of $\hat{\mu}$ = -0.18 (s.d. 126.84; p = 0.980).[19]

5. Discussion and conclusion

To our knowledge, our study is the first to empirically examine the within-subject variability of conditional cooperation when payoffs vary. To do this, we have subjects play eight one-shot sequential prisoner's dilemma games with varying payoff parameters. We find that conditional cooperation varies across games, and most subjects change strategies at least once across games. This switching between strategies varies systematically with the distributional consequences of free-riding relative to conditionally cooperating. Subjects conditionally cooperate more often when free-riding imposes larger losses on the FM, or when free-riding provides smaller gains for oneself. Further, in a second study, we jointly estimate social preference parameters and noise parameters at the individual level and find that the majority of our subjects place a significantly positive weight on others' payoff.

These findings provide two important implications. First, the within-subject variation of conditional cooperation with payoffs suggests that conditional cooperation should be viewed as an endogenous behavior arising from interaction between underlying motives and payoff variations, rather than a preference itself (Arifovic and Ledyard (2012)). A majority of subjects change their SM strategy when material payoffs change, and so classifications of individuals as conditional cooperators or free-riders should not be generalized to other games with different material payoffs. Moreover, identical payoffs could be internalized differently by individuals implying that across samples it is possible to observe different classifications in types even whilst maintaining the same game parameters. Lastly, it must be noted that other studies have elicited "conditional cooperation preferences" at the individual level and found them to be predictive of behavior in another, related, game (e.g. Eichenseer and Moser, 2020, Mullet et al., 2020). Our view is that these findings are not in conflict with ours, but we suggest a different interpretation whereby subjects classified as conditional cooperators can be viewed as subjects with sufficiently strong preference parameters, and it is these parameters that explain behavior in the related game.

Second, our findings support the underlying role of social preferences in conditional cooperation. We find that simple distributional preferences, where preferences depend on the

[19] Following our previous approach, we use the following to compute the standard error of our estimate: as $se(\sum\hat{\mu}/n) = (1/n)\surd(\sum se(\hat{\mu}_i)^2) = 10.70$.

distribution of payoffs, can explain a lot of conditional cooperation, and we find little support for reciprocity augmented models. This may reflect our particular focus on a sequential dilemma game in which only positive reciprocity can play a role, and the well-known evidence that positive reciprocity concerns are generally weaker than negative reciprocity (e.g., Abbink, et al. (2000); Offerman (2002)). We also find that weights placed on other's payoffs vary substantially between first movers and second movers. Since, in our games, second movers earn at least as much as first movers in any outcome, this finding is consistent with individuals placing less weight on other's payoff when others are ahead and more weight on other's payoffs when others are behind. These results are also qualitatively consistent with those reported in Bruhin, et al. (2019) in that they also find subjects place more weight on other's earnings when others are behind.

References

Abbink, K., Irlenbusch, B., & Renner, E. (2000). The moonlighting game - an experimental study on reciprocity and retribution. *Journal of Economic Behavior & Organization*, *42*, 265-277.

Ahn, T. K., Ostrom, E., Schmidt, D., Shupp, R., & Walker, J. M. (2001). Cooperation in pd games: Fear, greed, and history of play. *Public Choice*, *106*, 137-155.

Andersen, S., Harrison, G., Lau, M. I., & Rutström, E. E. (2008). Eliciting risk and time preferences. *Econometrica*, *76*, 583–618.

Anderson, S. P., Goeree, J. K., & Holt, C. A. (1998). A theoretical analysis of altruism and decision error in public goods games. *Journal of Public Economics*, *70*, 297-323.

Andreoni, J. (1995). Cooperation in public-goods experiments - kindness or confusion? *American Economic Review*, *85*, 891-904.

Andreoni, J., & Miller, J. (2002). Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica*, *70*, 737-753.

Andreozzi, L., Ploner, M., & Saral, A. S. (2020). The stability of conditional cooperation: Beliefs alone cannot explain the decline of cooperation in social dilemmas. *Scientific Reports*, *10*, 13610.

Arechar, A. A., Gächter, S., & Molleman, L. (2018). Conducting interactive experiments online. *Experimental Economics*, *21*, 99-131.

Arifovic, J., & Ledyard, J. (2012). Individual evolutionary learning, other-regarding preferences, and the voluntary contributions mechanism. *Journal of Public Economics*, *96*, 808-823.

Au, W. T., Lu, S., Leung, H., Yam, P., & Fung, J. M. Y. (2012). Risk and prisoner's dilemma: A reinterpretation of coombs' re-parameterization. *Journal of Behavioral Decision Making*, *25*, 476-490.

Bardsley, N., & Sausgruber, R. (2005). Conformity and reciprocity in public good provision. *Journal of Economic Psychology*, *26*, 664-681.

Bayer, R.-C., Renner, E., & Sausgruber, R. (2013). Confusion and learning in the voluntary contributions game. *Experimental Economics*, *16*, 478-496.

Beranek, B., Cubitt, R., & Gächter, S. (2015). Stated and revealed inequality aversion in three subject pools. *Journal of the Economic Science Association*, *1*, 43-58.

Bilancini, E., Boncinelli, L., & Celadin, T. (2022). Social value orientation and conditional cooperation in the online one-shot public goods game. *Journal of Economic Behavior & Organization*, *200*, 243-272.

Blanco, M., Engelmann, D., & Normann, H. T. (2011). A within-subject analysis of other-regarding preferences. *Games and Economic Behavior*, *72*, 321-338.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review.*, *90*, 166-93.

Brandts, J., & Charness, G. (2000). Hot vs. Cold: Sequential responses and preference stability in experimental games. *Experimental Economics*, *2*, 227-238.

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

Brandts, J., & Schram, A. (2001). Cooperation and noise in public goods experiments: Applying the contribution function approach. *Journal of Public Economics*, *79*, 399-427.

Brosig, J., Reichmann, T., & Weimann, J. (2007). Selfish in the end? An investigation of consistency and stability of individual behaviour *Faculty of Economics and Management Magdeburg Working Paper Series*.

Bruhin, A., Fehr, E., & Schunk, D. (2019). The many faces of human sociality: Uncovering the distribution and stability of social preferences. *Journal of the European Economic Association*, *17*, 1025-1069.

Burton-Chellew, M. N., El Mouden, C., & West, S. A. (2016). Conditional cooperation and confusion in public-goods experiments. *Proceedings of the National Academy of Sciences*, *113*, 1291-1296.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics*, *117*, 817-69.

Charness, G., Rigotti, L., & Rustichini, A. (2016). Social surplus determines cooperation rates in the one-shot prisoner's dilemma. *Games and Economic Behavior*, *100*, 113-124.

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

Chaudhuri, A., & Paichayontvijit, T. (2006). Conditional cooperation and voluntary contributions to a public good. *Economics Bulletin*, *3*, 1-15.

Clark, K., & Sefton, M. (2001). The sequential prisoner's dilemma: Evidence on reciprocation. *Economic Journal*, *111*, 51-68.

Cooper, D. J., & Kagel, J. H. (2016). Other-regarding preferences: A selective survey of experimental results. In *Handbook of experimental economics, volume 2*, ed. J. H. Kagel, & A. E. Roth. Princeton: Princeton University Press.

Cox, J. C., Friedman, D., & Gjerstad, S. (2007). A tractable model of reciprocity and fairness. *Games and Economic Behavior*, *59*, 17-45.

Cox, J. C., Friedman, D., & Sadiraj, V. (2008). Revealed altruism. *Econometrica*, *76*, 31-69.

Croson, R. (2007). Theories of commitment, altruism and reciprocity: Evidence from linear public goods games. *Economic Inquiry*, *45*, 199-216.

Cubitt, R., Gächter, S., & Quercia, S. (2017). Conditional cooperation and betrayal aversion. *Journal of Economic Behavior & Organization*, *141*, 110-121.

Danz, D., Vesterlund, L., & Wilson, A. J. (2022). Belief elicitation and behavioral incentive compatibility. *American Economic Review*, in press.

Eichenseer, M., & Moser, J. (2020). Conditional cooperation: Type stability across games. *Economics Letters*, *188*, 108941.

Engel, C., & Zhurakhovska, L. (2016). When is the risk of cooperation worth taking? The prisoner's dilemma as a game of multiple motives. *Applied Economics Letters*, *23*, 1157-1161.

Falk, A., Fehr, E., & Fischbacher, U. (2003). On the nature of fair behavior. *Economic Inquiry*, *41*, 20-26.

Fallucchi, F., Luccasen, R. A., & Turocy, T. L. (2019). Identifying discrete behavioural types: A re-analysis of public goods game contributions by hierarchical clustering. *Journal of the Economic Science Association*, *5*, 238-254.

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. *Quarterly Journal of Economics*, *114*, 817-68.

Fehr, E., & Schurtenberger, I. (2018). Normative foundations of human cooperation. *Nature Human Behaviour*, *2*, 458-468.

Ferraro, P. J., & Vossler, C. A. (2010). The source and significance of confusion in public goods experiments. *The BE Journal of Economic Analysis & Policy*, *10*, 1-42.

Fischbacher, U., & Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public good experiments. *American Economic Review.*, *100*, 541–556.

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., Kariv, S., & Markovits, D. (2007). Individual preferences for giving. *American Economic Review*, *97*, 1858-1876.

Fosgaard, T. R., Hansen, L. G., & Wengström, E. (2017). Framing and misperception in public good experiments. *The Scandinavian Journal of Economics*, *119*, 435-456.

Furtner, N. C., Kocher, M. G., Martinsson, P., Matzat, D., & Wollbrant, C. (2021). Gender and cooperative preferences. *Journal of Economic Behavior & Organization*, *181*, 39-48.

Gächter, S. (2007). Conditional cooperation: Behavioral regularities from the lab and the field and their policy implications. In *Psychology and economics: A promising new cross-disciplinary field (CESifo seminar series)*, ed. B. S. Frey, & A. Stutzer. Cambridge: The MIT Press.

Gächter, S., Gerhards, L., & Nosenzo, D. (2017a). The importance of peers for compliance with norms of fair sharing. *European Economic Review*, *97*, 72-86.

Gächter, S., Kölle, F., & Quercia, S. (2017b). Reciprocity and the tragedies of maintaining and providing the commons. *Nature Human Behaviour*, *1*, 650-656.

Gächter, S., Kölle, F., & Quercia, S. (2022). Preferences and perceptions in provision and maintenance public goods. *Games and Economic Behavior*, *135*, 338-355.

Gächter, S., Lee, K., Sefton, M., & O.Weber, T. (2021). Risk, temptation, and efficiency in the one-shot prisoner's dilemma. *CESifo Working Paper No.9449*.

Gächter, S., Nosenzo, D., Renner, E., & Sefton, M. (2012). Who makes a good leader? Cooperativeness, optimism, and leading-by-example. *Economic Inquiry*, *50*, 953-967.

Giamattei, M., Yahosseini, K. S., Gächter, S., & Molleman, L. (2020). Lioness lab: A free web-based platform for conducting interactive experiments online. *Journal of the Economic Science Association*, *6*, 95-111.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with orsee. *Journal of the Economic Science Association*, *1*, 114-125.

Hedegaard, M., Kerschbamer, R., Müller, D., & Tyran, J.-R. (2021). Distributional preferences explain individual behavior across games and time. *Games and Economic Behavior*, *128*, 231-255.

Hey, J. D., & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica*, *92*, 1291-1326.

Hossain, T., & Okui, R. (2013). The binarized scoring rule. *The Review of Economic Studies*, *80*, 984-1001.

Houser, D., & Kurzban, R. (2002). Revisiting kindness and confusion in public goods experiments. *American Economic Review*, *92*, 1062-1069.

Isler, O., Gächter, S., Maule, A. J., & Starmer, C. (2021). Contextualised strong reciprocity explains selfless cooperation despite selfish intuitions and weak social heuristics. *Scientific Reports*, *11*, 13868.

Katuščák, P., & Miklánek, T. (2018). What drives conditional cooperation in public goods games? . *CERGE-EI Working Paper Series No. 631*.

Keser, C., & Kliemt, H. (2021). The strategy method as an instrument for the exploration of limited rationality in oligopoly game behavior (Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes) (by Reinhard Selten) In *The art of experimental economics. Twenty top papers reviewed*, ed. G. Charness, & M. Pingle. London: Routledge.

Keser, C., & Van Winden, F. (2000). Conditional cooperation and voluntary contributions to public goods. *Scandinavian Journal of Economics*, *102*, 23-39.

Kocher, M. G., Cherry, T., Kroll, S., Netzer, R. J., & Sutter, M. (2008). Conditional cooperation on three continents. *Economics Letters*, *101*, 175-178.

Mengel, F. (2018). Risk and temptation: A meta-study on prisoner's dilemma games. *The Economic Journal*, *128*, 3182-3209.

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.

Muller, L., Sefton, M., Steinberg, R., & Vesterlund, L. (2008). Strategic behavior and learning in repeated voluntary-contribution experiments. *Journal of Economic Behavior & Organization*, *67*, 782-793.

Mullett, T. L., Mcdonald, R. L., & Brown, G. D. A. (2020). Cooperation in public goods games predicts behavior in incentive-matched binary dilemmas: Evidence for stable prosociality. *Economic Inquiry*, *58*, 67-85.

Ng, G. T. T., & Au, W. T. (2016). Expectation and cooperation in prisoner's dilemmas: The moderating role of game riskiness. *Psychonomic Bulletin & Review*, *23*, 353-360.

Offerman, T. (2002). Hurting hurts more than helping helps. *European Economic Review*, *46*, 1423-1437.

Palfrey, T. R., & Prisbrey, J. E. (1997). Anomalous behavior in public goods experiments: How much and why? *American Economic Review*, *87*, 829-846.

Schmidt, D., Shupp, R., Walker, J., Ahn, T. K., & Ostrom, E. (2001). Dilemma games: Game parameters and matching protocols. *Journal of Economic Behavior & Organization*, *46*, 357-377.

Selten, R. (1967). Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes. In *Beiträge zur experimentellen Wirtschaftsforschung*, ed. H. Sauermann. Tübingen: J.C.B. Mohr (Paul Siebeck).

Thöni, C., & Volk, S. (2018). Conditional cooperation: Review and refinement. *Economics Letters*, *171*, 37-40.

Vlaev, I., & Chater, N. (2006). Game relativity: How context influences strategic decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 131-149.

Volk, S., Thöni, C., & Ruigrok, W. (2012). Temporal stability and psychological foundations of cooperation preferences. *Journal of Economic Behavior & Organization*, *81*, 664-676.