Acting *as if* drives cooperation among the purely self-interested

Matthew Cashman[1]* Drazen Prelec[1]

[1]Sloan School of Management
Massachusetts Institute of Technology
{cashman, dprelec}@mit.edu

Abstract

Cooperation is puzzling when an individual acting alone has a small influence on the desired outcome, as is the case with collective pay-for-performance compensation schemes, voting, and participation in social causes—even more so when the individual in question is trying to maximize her own payoffs. We provide experimental evidence for a psychological mechanism that provides an explanation for cooperation specifically among those indifferent to the fates of others: acting *as if*. In one-shot Public Goods Games where players move one after another but do not observe others' moves, we find that contributions to the public good are highest at the beginning of the sequence, among those moving first, and decline as order increases such that last-movers make the lowest average contribution. This pattern is consistent with payoff-maximizing players acting as if those yet to move will make the same move they have even though they know there is no causal linkage. Four results provide support: (1) This positional order effect is generated by players who are acting in their own interests, (2) instructing players to maximize their own payoff increases the effect, (3) the effect is larger among those passing pre- and post-task comprehension checks, implying that those who best understand the task generate the effect, and (4) the effect is eliminated if the moves of future players, but not those of past players, are determined randomly. Firms designing pay-for-performance compensation schemes should consider this additional effect on top of classical economic incentives.

Keywords: Cooperation, Sequential Games, Public Goods Game, Game Theory

Contents

*Corresponding author

1 Introduction

Social cooperation without external monitoring is widely regarded as fundamental to human culture, sustaining teamwork, mass political participation, and personal sacrifice for family, tribe, or nation. People often face opportunities to incur an individual cost in exchange for a collective benefit, and there is a rich literature exploring the whys and wherefores (e.g., Henrich and Muthukrishna, 2021; Rand and Nowak, 2013). For example, a pedestrian can choose to throw litter into the gutter, or he can wait until he comes across a trash bin. A CEO might choose to move assets overseas in order to avoid taxes, or she might choose to avoid chicanery, keep assets domestically, and pay more in taxes—in the end, contributing to the public weal. Each choice involves a tradeoff between what is good for the agent and what is good for the group. This tradeoff is widely studied using Public Goods Games (PGGs, Zelmer, 2003). The PGG is used as a model of human cooperation because it captures this tension between the benefits accruing to a group via cooperation and the benefits accruing to an individual via defection, which is characteristic of many problems we solve on a daily basis. In standard linear PGGs it is always better for an individual to defect no matter what others do, but always better for the group if everyone cooperates.

Most accounts of cooperation in humans are stories about why people end up *wanting* to cooperate. There may, however, be circumstances in which even players who are indifferent to the fates of others end up cooperating. Such phenomena would suggest there are ways to increase cooperation even among the most self-interested in addition to illuminating how such decisions are made. *Quasi-magical thinking* (Shafir and Tversky, 1992) is precisely the view that people making decisions under uncertainty act as if their actions are causally linked to others' decisions, even when they know it is impossible. However, the theory does not include the arrow of time. Acting *as if* is the idea that people act ”as if” they can influence the actions of others, as in quasi-magical thinking, but with the additional stipulation that this thinking is biased towards the future, as-yet unmade moves of others. Here, we explore the behavior of specifically self-interested players who cooperate because they believe cooperation maximizes their individual payoff. We theorize that they do this on the assumption that those moving after them (who have not yet made a decision) will make the same move they have. In this case, where there is no way to influence others, they behave as if they can influence others' moves without any communication. We investigate this with a subtle variation on the classic one-shot PGG, changing it so that players within a single round move one after another but do not observe each others' moves: a sequential PGG (SPGG) without observation. If players are acting *as if*, cooperation should be proportional to the number of people *yet* to move: the positional order effect. Our goal is to establish the existence of acting *as if* among self-interested players, and we by do this by first demonstrating that positional order effects exist *only* among the most self-interested players and, second, we turn the effect on and off by manipulating the presence or absence of people making their own decisions in the future, thus making it possible or impossible to use the "quasi-magical" causal connection between a player and those moving after him. We conclude by discussing several possible underlying mechanisms and elaborating on practical implications.

This mechanism is of particular relevance when individuals are subject to collective reward or punishment, as is the case with "collective pay-for-performance" employee compensation packages (Knez and Simester, 2001; Nyberg et al., 2018). Mutual monitoring of effort is almost never straightforward, and employees therefore have incomplete information about the behavior of their peers (Delfgaauw et al., 2022). A critical variable is the size of the bonus group, sometimes referred to as "the 1/N problem" (Carpenter, Robbett, and Akbar, 2018). As N, the number of people in a bonus grouping, increases, the impact of any individual on the collective outcome declines. In the limiting case of company-wide profit sharing schemes, where one person is grouped with hundreds or thousands of others, the influence of any single employee on the collective outcome is essentially zero.

For example an employee working alone at her desk, burning the midnight oil must decide if spending additional time and effort at the margin working to further the firm's goals is worth it. All else being equal, in the case where she works hard and stays late and others also work hard and stay late, bonuses will be high and the work will feel worth it. In the case where others do not work hard (or only appear to work hard) and she does work hard, bonuses will be low and she will feel though she has been taken advantage of. How might she make the decision about whether to spend the next 10 minutes at her desk? In the absence of information about the behavior of others, acting *as if* can provide some insight. She can decide to act as if others will do as she has done: If she works hard and works late, that should raise her estimate of the likelihood that others will too—and therefore also raise her estimate of the marginal value of additional toil on behalf of the company's goals. This implies that firms should be more willing to implement collective pay-for-performance schemes than normative decision theory should predict.

1.1 Theoretical background: uncertainty, causality, and positional order effects in social dilemmas

Traditional game theory ignores the ordering of moves in time, focusing exclusively on what information is available when making a decision. von Neumann and Morgenstern (2004) developed extensive form notation based purely on preliminarity in information, ignoring the chronological ordering of moves (what they call "anteriority")—though they note that this only holds with perfect information. Because of this, the extensive form representation of a simultaneous game is identical to that for a sequential version in which no moves are observed. By the mid 1980s there was a small chorus raising the question of whether ignoring timing is a good idea (Kohlberg and Mertens, 1986; Kreps, 1990; Luce, 1992). Luce mentions the lack of a time variable in extensive form games (while real life inevitably involves one), and Kreps asks explicitly: "Can we find a pair of extensive form games that give rise to the same strategic form such that, when played by a reasonable participant population, there is a statistically significant difference in how the games are played?". The question has since been answered with a sure "yes" (Rapoport, 1997).

Games with an element of coordination have the interesting property that players can coordinate based on pure order of play without any additional information at all. For instance, people tend to "agree" to play the first-mover's preferred equilibrium without any communication at all in Battle of the Sexes (Cooper et al., 1993; Güth, Huck, and Rapoport, 1998; Rapoport, 1997; Weber, Camerer, and Knez, 2004), or they will "agree" that the first-mover should get the largest share of the gains in common-pool resource games (Budescu, Suleiman, and Rapoport, 1995; Budescu, Au, and X.-P. Chen, 1997; Budescu and Au, 2002) and step-level PGGs (X.-P. Chen, Au, and Komorita, 1996; Rapoport, 1997).

Social dilemmas *without* an element of coordination, games like the PGG, pit what is good for you against what is good for everyone else. In these games there is no reason to condition your play on others' decisions, and therefore no obvious reason for order to influence play. However, there is evidence that suggests people engage in causal thinking about others even in situations without rewards for coordination, and further there is reason to think that uncertainty about the state of the world activates this sort of reasoning.

In an early study (1984), Quattrone & Tversky report evidence for what they term "diagnostic" actions—actions that have no direct causal relationship to desirable outcomes, but which are indicative of them. They report that participants holding an arm in circulating ice water (a painful experience) are able to hold their arms in the water *longer* when they believe this is indicative of having a strong heart, and for shorter amounts of time when long durations are believed to be indicative of having a bad heart. The experience of holding one's arm in water of course has no bearing on heart type, but it does appear participants are changing the data they themselves produce in order to receive good news, in apparent disregard of the causal relationship.

In related work, Shafir & Tversky (1992) explore nonconsequential reasoning—reasoning that at least appears to either not produce estimates of the consequences of an action, or which ignores the consequences of that action. This class of decisions violate the Sure Thing Principle, which states that if X is preferred to Y under all states of the world, then X should still be preferred to Y even if the state of the world is unknown. For instance, there are many people who would prefer to pay for a vacation to Hawaii in the event that they pass an exam *and* in the event that they fail, but who would also prefer *not* to buy in the case where the outcome of the exam is unknown (Tversky and Shafir, 1992). They refer to this pattern of events as “accept when win, accept when lose, reject when do not know” and refer to it as the “disjunction effect”. In an experiment using the Prisoner’s Dilemma, they observe more cooperation in one-shot games when uncertainty about the other player’s move is highest: players cooperate more when they *do not know* the other player’s move than either when they know it is Defect or when they know it is Cooperate. The authors introduce the idea of quasi-magical thinking as a possible explanation for the disjunction effect. The idea is reminiscent of the illusion of control (Langer, 1975; Stefan and David, 2013); however, that work focuses on repeated tasks that do not generally involve other minds. Masel (2007) offers a formalization of quasi-magical thinking where players, upon observing additional information during the game, update their prior distributions in the usual fashion—one’s own behavior being just another data point. Daley & Sadowski (Daley and Sadowski, 2017) develop a similar model of magical thinking that applies to players’ preferences over actions rather than outcomes. However, neither formalization incorporates the arrow of time within a single game.

There are two flavors of uncertainty at play here: “closed fates” uncertainty is about a counterpart’s move when that move is not known to the player but has already been made and is therefore fixed, and “open fates” uncertainty which is about a counterpart’s move that has yet to be made at all (or which is presently being made) (Morris, Sim, and Girotto, 1998). Miller and Gunasegaram (1990) demonstrated that, while events in the past are considered fixed, future events are treated as mutable. Moreover, future actions are perceived as more intentional and blameworthy than otherwise identical past actions (Burns, Caruso, and Bartels, 2012). The distinction between ”closed fates” and ”open fates” is clearly relevant for behavior, but has been little-explored in the context of sequential games.

Subsequent work on sequential social dilemmas without observation is scant and mixed, but we can safely conclude that uncertainty matters. Uncertainty about the state of the world seems to push people towards more prosocial actions (Croson, 1999; Hristova and Grinberg, 2010; Morris, Sim, and Girotto, 1998; Shafir and Tversky, 1992). However, evidence for positional order effects in sequential PDs or PGGs, games without obvious benefits to coordination, is lacking. When considering quasi-magical thinking, Shafir & Tversky did not distinguish between open fates and closed fates and so could not have measured an order effect. Morris et al. (1998) report more cooperation in first-movers and larger effects in open fates vs. closed fates cases, but most studies incorporating sequential PDs or PGGs with no observation find no effect of order alone (Abele and Ehrhart, 2005; Figuières, Masclet, and Willinger, 2012; Robinson et al., 2010; Steiger and Zultan, 2014). These studies were, in general, not designed to investigate the effects of order of play alone and so tend to be under-powered to identify these effects. They also rematch participants randomly after each round and do so using small pools of students from the same university, not being quite as one-shot as could be hoped. In addition, many participants are probably not even trying to maximize their own direct payoffs in these tasks—limiting what we can infer based on play. Budescu, Au, and X.-P. Chen (1997) report that 47% of their participants are classified as “cooperative” (maximize joint own + other gains) and 2% as “altruistic” (maximizing others’ gains); that is to say, half of their participants are not playing the game to “win” by the usual standards of game theory, or they are optimizing for some joint outcome such as that of direct payoffs and social capital or emotional wellbeing. While this may be fine if the goal is to document the behavior of participants as they come to the game, it is a substantial problem when making the assumption that players are trying to maximize profits.

2 Studies

Below we present five empirical studies that investigate acting*as if* using sequential games. The first four studies develop the method and establish the positional order effect, where payoff-maximizing but not prosocial players tend to give less to the public good as their position in a sequence of players nears the end. This work sets the scene for Study 2b, which replicates the positional order effect and directly tests the causal linkages involved in acting *as if*. Study 1a shows a decline with increasing order in a Prisoner's Dilemma (a two-person PGG), Study 1b investigates the positional order effect in a three-person PGG, and shows this effect is driven by participants who are "Individualistic" on the SVO scale. SVO is a measure of willingness to give up gains in order to benefit others, and in the SVO battery participants make a series of incentivized decisions similar to Dictator games where they allocate funds between themselves and someone else. Participants can choose to forego gains (or even pay costs) to help or hurt the other player. If participants are acting *as if*, players who are Prosocial on the SVO measure (meaning they are willing to forego gains to help others) would be expected to show no positional order effect because they will never have a reason to defect: it is nearly always payoff-maximizing for a prosocial player to cooperate whether that player is at the beginning of a sequence or at the end (see formalization in appendices). Conversely, those who are Individualistic (and therefore tend towards maximizing their own payoffs) might show a decline in cooperation with increasing position in the sequence since the number of "open fates" left to influence decreases as order increases. Study 1c expands this to a four-person PGG. In these studies we are chiefly interested in payoff-maximizing players, but they are sufficiently rare in the study population that forming five-person groups composed of them in real time proved difficult. For this reason, Study 2a asks whether the mere instruction to maximize payouts also produces the positional order effect that was only observed among participants presenting as self-interested in earlier studies. Study 2b deploys the technique from Study 2a to ask whether we still observe a positional order effect in the case where all players after a focal player have their contribution decisions delegated to a random process. If the effect is present when random movers are *before* the focal player, but absent when random movers are *after* the focal player, this would indicate the effect requires having real people who have not yet made a decision, but who will, moving after the focal player—implying causal thinking about others is at play. We will now detail each study.

All studies are real-time, one-shot linear SPGGs with a multiplier of two. Participants contribute three main inputs: comprehension checks, game playing decisions, and predictions of the responses of other players. Apart from a base payment and game proceeds, correct answers to comprehension checks and accurate predictions are incentivized.

In all studies players participate in a brief text chat room with their groupmates before learning about the task. The purpose of the chat is to assure participants they are playing in real time with real people and, generally, to give the task more psychological reality than might be felt in an online task with no human interaction. The group they play the game with is the same group from the chat room. All experiments have simultaneous-play PGG control conditions, and all players pass familiarization tasks and comprehension checks. All experiments share the following three up-front comprehension and attention check questions:

1. *Do any of the other players* ***know how much YOU decide to contribute?***
2. *Jack and Jill are playing this game together. Jack decided to* ***TRANSFER*** *and Jill decided to* ***KEEP****. Who will make more money, Jack or Jill?*
3. *What year is it?*

Participants are given one chance to get each of these questions right, and a single wrong answer results in their data being excluded from analyses. Responses to the comprehension questions are only relevant to data analysis, however: players continue on whether or not they have answered correctly because it their moves are necessary in order to finish the game. Later studies incorporate additional training and comprehension check regimes. Materials, including data, analysis scripts, and stimuli, can be found at https://osf.io/mykzt/?view_only=248c1173bf664c78a8ffdcf637dddbd3.

In total we tested 6,100 participants distributed across four experiments. A convenience sample provided by Amazon Mechanical Turk (MTurk) was selected for Study 1b and Study 1c because it was a reasonable approximation of American adults for our purposes. Declines in the quality of MTurk data over the years these studies were conducted meant that we selected CloudResearch's filtered MTurk panel for Studies 2a and 2b because it provides among the highest-quality online panel data (Hauser et al., 2023; Douglas, Ewell, and Brauer, 2023). This work makes the point that these effects exist in human populations, and it is left for future work to examine how they vary across ages, sexes, SES, cultures, and other characteristics of interest. All experiment software was written in the oTree framework (D. L. Chen, Schonger, and Wickens, 2016). All studies except for Study 2a were preregistered on osf.io, and all studies were approved by MIT's Committee on the Use of Humans as Experimental participants (COUHES) and comply with all relevant ethical regulations. We obtained electronic consent from all participants.

2.1 Study 1a: Sequential Prisoner's Dilemma

Study 1a is a Sequential Prisoners' Dilemma (PD) study in which we first observed a positional order effect. It is a Prisoner's Dilemma rather than a Public Goods Game like the other studies in this series, but a Prisoner's Dilemma can be seen as a two-player Public Goods Game. It is similar in structure to Studies 1b and 1c. Study 1a contained a number of exploratory conditions not relevant to this line of inquiry, over which we collapse here. This study was preregistered, and the preregistration can be found at https://osf.io/dbcpv.

2.1.1 Participants

2367 U.S.-based participants from Amazon Mechanical Turk completed the study. Mean total pay per participant (including bonuses for accurate predictions) is \$0.71 (SD = 0.26), yielding an hourly rate of \$7.93 at an average 6.5 minutes' duration. Of 2367 participants who finished the task, 45% (1075) passed the comprehension check questions. Analysis is limited to these responses.

2.1.2 Materials and procedure

The chat room and experimental platform was developed on the oTree framework (D. L. Chen, Schonger, and Wickens, 2016). Players arrive at the experiment web page, are consented, and then engage in a real effort task as attention and activity verification (transcribing nonsense sentences)[1]. The chat room then provides 30 seconds for exchanging a hello or brief message, confirming that their teammate is indeed a real person. After leaving the chat room, the game is described as an allocation task where players choose to "keep" an initial endowment or "transfer" the endowment to the other player, with the transfer doubled before reaching the other player. The payoff matrix is given below in Table 1:

After reading the instructions players proceed to 5 comprehension tests:

1. Does the other player know what your move is?
2. If the other person TRANSFERS their money, what earns you the most money?
3. If the other person KEEPS their money, what earns you the most money?
4. If you choose to TRANSFER your money, do you make more money if the other person TRANSFERS or KEEPS?

[1] Since participants are grouped together for this real-time experiment, we must ensure that those who are being grouped are active directly before they are put into groups. If they are not, responsive players may be grouped with non-responsive players.

Table 1: Study 1a Sequential Prisoner's Dilemma payoff matrix

		Player 2	
		Transfer (cooperate)	Keep (defect)
Player 1	Transfer (cooperate)	(\$0.33, \$0.33)	(\$0.00, \$0.50)
	Keep (defect)	(\$0.50, \$0.00)	(\$0.16, \$0.16)

5. What year is it?

Player 2 waits while Player 1 moves, then makes a move on a screen similar to the one Player 1 saw. After indicating their move, both players predict "How likely is it that the other person in this game TRANSFERRED?" on a scale from 0-100. Players also answer a population version, "How likely is it that an average person who plays this game would TRANSFER?". Players exit the experiment and are paid.

2.1.3 Results

First-movers cooperate more than second-movers. A logistic regression of Transfer decision on Order reaches significance ($OddsRatio = 0.757$, $95\%CI = [0.596, 0.963]$, $z = -2.268$, $p = 0.023$). We also observed a strong non-preregistered impact of decision on perception of teammate's behavior relative to the population. Players who keep their endowment think that their teammate is less likely to transfer than the population at large ($M_{teammate-population} = -2.20$), while those who transferred believe their teammate is more likely to transfer ($M_{teammate-population} = +2.87$) A linear regression shows a significant effect, $\beta = 5.069$, $95\%CI = [3.272, 6.937]$, $F(1, 1071) = 30.4$, $p < 0.001$. 53% of players cooperated (SD=0.50) and the average payoff was \$0.58 (SD=0.18).

2.1.4 Discussion

Study 1a investigated a number of drivers of order effects not reported here, but the data taken together lends support to the idea that first-movers would be more likely to cooperate than those moving afterwards. In addition, we see strong evidence that players view their own moves as indicative of the moves of others. This led to further investigation of the positional order effect in the next study.

2.2 Study 1b: 3-Person Sequential Public Goods Game

Our primary goal with Study 1b was to test for a positional order effect, and we also sought to investigate whether any such effect is driven by players who are trying to maximize their own payoffs, or those who are keeping others' interests in mind in addition to their own. The preregistration for Study 1b can be found at https://osf.io/3vsxk.

2.2.1 Participants

1444 U.S.-based participants from Amazon Mechanical Turk completed the study. Of 1444 participants who passed up-front bot checks and finished the task, 69.0% (1002) passed all of the comprehension check questions. To estimate the sample size required, we performed a power analysis via simulation using pilot data. Among those 1002 participants, mean total pay per participant (including bonuses for accurate predictions) is \$3.31 ($SD = 0.74$), yielding an hourly rate of \$20.33 per hour (SD = 7.39) at 9.6 minutes' average duration.

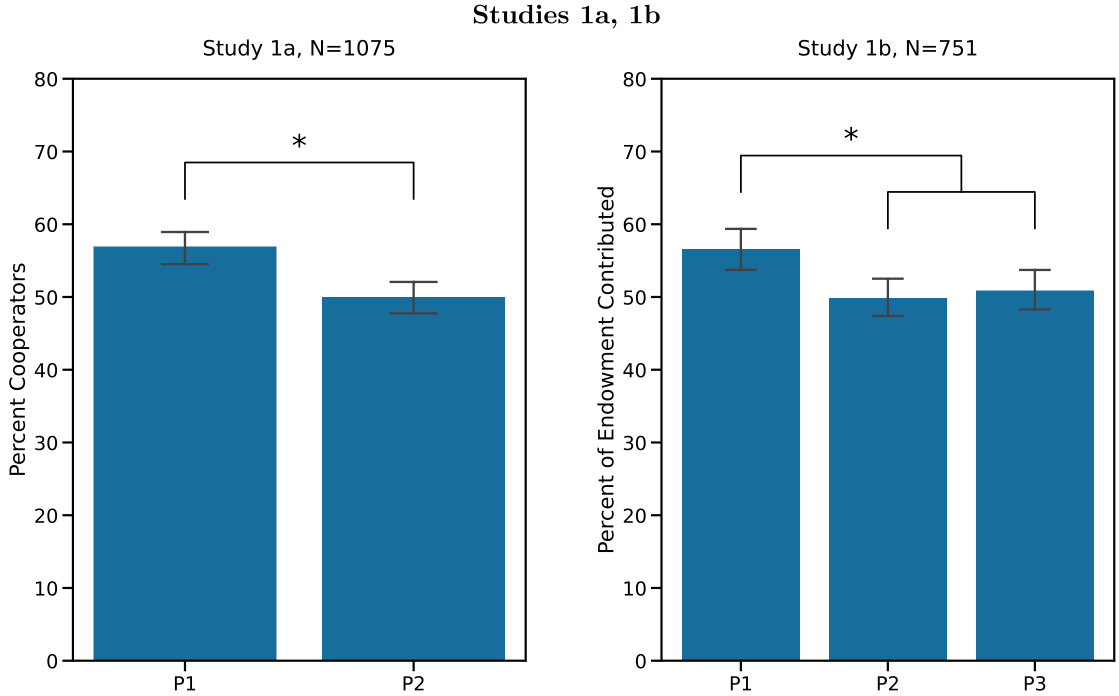


Figure 1: [Left] Data from Studes 1a and 1b suggest a decline in contribution to the public good with increasing order, and led to further investigation. Participants who passed comprehension checks, SEMs.

2.2.2 Materials and procedure

Study 1b is a one-shot SPGG with a multiplier of two. Three players can transfer any part of their individual $1 endowment to the public good. The total amount transferred from all participants is then doubled and distributed evenly among the players, irrespective of individual transfers. Order of play is determined randomly, with no communication among players during the game. The only difference in information among the players is knowledge of their position in the sequence. Each group was randomized to either the sequential game or a simultaneous-move condition. Players arrive at the experiment web page, complete a consent form, and then engage in a real effort task transcribing nonsense sentences in order to filter out bots. After this, they enter a wait room and form groups of three. Then they are placed in a chat room for 30 seconds after all players in their group have arrived to give participants the (correct) sense that the experiment is, in fact, a real game in real time with real people. After the chat, participants learn the game they will play. They are provided with an explanation of the rules of the game (which appear on every subsequent page for reference). The PGG is framed as a question of how much to contribute to a “Community Fund”. A player can “transfer” some or all of her endowment to the Community Fund, and she may “keep” some amount. Instructions include if-then statements about the consequences of certain moves to aid understanding.

Participants are then asked three comprehension and attention check questions, and then make their move. The contribution page includes a graphic at the top highlighting their place in the sequence of moves in red (see the stimuli in supplemental materials). Players in the simultaneous condition do not see any indication of sequence since they are moving simultaneously. Participants then complete incentivized prediction questions, where error is Brier-scored, and then a Social Value Orientation (SVO) slider battery

(Murphy, Ackermann, and Handgraaf, 2011) [2]. The SVO battery measures preferences for how to allocate resources between oneself and others. The standard battery categorizes participants into Individualistic (concerned only with what is best for self), Competitive (maximize own outcomes as with Individualistic, but also minimize the outcomes for others), Prosocial (maximize outcomes for both self and other), and Altruistic (eager to give up own gains to help others). Players complete some demographic questions, exit the experiment, and are paid.

2.2.3 Results

In Study 1b our primary interest was how contribution to the public good varied with order of play, with the additional prediction that order effects will appear in Individualistic players but not in Prosocial players, as determined by an SVO battery. The SVO measure divides almost all[3] participants into two categories: Individualistic and Prosocial. The mean contribution for Individualistic participants was 40.9% (SD=0.44) of the $1.00 endowment, and for Prosocials 68.1% (SD=0.40). Individualistic participants left with slightly more money from the SPGG itself, $1.63 (SD=0.44) vs. Prosocials $1.56 (SD=0.43).

We find some support for the positional order effect in this study. Contributions to the public good were distributed non-normally, with modes at 0%, 50%, and 100% of the endowment, so when using OLS linear regressions in this and subsequent studies we bootstrap standard errors and p-values. While the preregistered backwards-difference coded model contribution $\sim$ order does not find significance, we do see a decline in contribution without backwards-difference coding, $\beta = -4.224$, 95% CI = [-8.145, -0.420], $p = 0.031$, and when grouping data from P2 and P3 in the linear model contribution $\beta = -9.117$, 95% CI = [-15.639, -2.526], $p = 0.005$. Participants classified as Prosocial exhibit no significant differences in contribution levels as function of order, $\beta = -1.498$, 95% CI = [-10.925, 8.162], $p = 0.770$, while we do see a difference among Individualistic participants between first-mover data and grouped second- and third-mover contributions, $\beta = -14.133$, 95% CI = [-24.600, -3.352], $p = 0.009$.

In addition, we find support for the prediction that correlations between a player's own move and her predictions of other players' moves are stronger going forward in time vs. backwards. The interaction term in the preregistered regression of predictions of others' moves on the player's own contribution interacted with a binary future/past variable (predicted_value $\sim$ contribution * binary_position) does not find significance, $\beta = 0.075$, 95% CI = [-0.014, 0.164], $p = 0.097$, but when applied to only Individualistic players a significant equation is found, $\beta = 0.174$, 95% CI = [0.047, 0.302], $p = 0.009$. There is no effect among Prosocial players, $\beta = 0.002$, 95% CI = [-0.130, 0.136], $p = 0.989$.

2.2.4 Discussion

Taken together, we thought these results were suggestive of a relationship between positional order effects and self-interest but not definitive, so an investigation with a more sensitive study was warranted. Study 1b was designed to investigate pure order effects, and provided some evidence that these effects are driven by the actions of Individualistic players. Those who are most concerned with maximizing their own rewards tend to contribute less to a public good as their order in an SPGG increases, and more self-interested players are willing to bet that other players who have not yet moved will make moves more similar to their own relative to those who have already moved. We continue this investigation with Study 1c, a refinement of this study.

[2] SVO is measured post-treatment, but we do not observe an effect of treatment on SVO, $\beta = 1.033$, 95% CI = [-0.361, 2.409], $p = 0.147$

[3] One participant was classified as Altruistic, and one as Competitive. These participants' data are excluded from SVO analyses.

2.3 Study 1c: 5-Person Sequential Public Goods Game

Study 1c is an evolution of Study 1b that is aimed at investigating the role of a player's self-interest in producing positional order effects. It incorporates several refinements to Study 1b's design in an effort to make the experience more intuitive for participants. It also raises the number of players to 5 from 3, though only data from four are reported. The preregistration for Study 1c can be found at https://osf.io/gw8nc.

2.3.1 Participants

1298 U.S.-based participants from Amazon Mechanical Turk passed up-front bot checks and finished the task. Of those, 788 (62%) passed all of the up-front comprehension check questions and have their data included. Among these, 44% were female, average age was 37, and mean total pay per participant (including bonuses for accurate predictions) was \$3.52 ($SD = 0.51$), yielding an hourly rate of \$11.39 at 18.6 minutes' average duration. To estimate the sample size required, we performed a power analysis via simulation using pilot data.

2.3.2 Materials and procedure

Study 1c is a one-shot SPGG similar in structure to Study 1b. It incorporates a number of improvements: it was designed for five players rather than three, the up-front chat was 60 instead of 30 seconds, and inputs were sliders with anchors that displayed the consequences (e.g., "KEEP FOR SELF: \$0.43 ↔ CONTRIBUTE TO FUND: \$0.57"). As with Study 1b, SVO is measured post-treatment, but we do not observe an effect of order on SVO angle, $\beta = -0.478$, 95% CI = [-1.898, 0.881], $p = 0.472$.

2.3.3 Results

Study 1c provides further evidence that, among specifically self-interested players[4], contribution to the public good declines as position in the order increases. A timing error meant that all first-movers and some later players were forced to respond too quickly, with response time allotted being a function of previous players' response time. We believe the marginal benefit of having 5 vs. 4 players is minimal given that this study is mostly a replication of Study 1b, and so exclude affected players and report results from 484 unaffected participants here.

The mean contribution for Individualistic participants was 38% (SD=0.43) of the \$1.00 endowment, and for Prosocials 70% (SD=0.37). Individualistic participants left with slightly more money from the SPGG itself, \$1.70 (SD=0.43) vs. Prosocials \$1.47 (SD=0.41).

For Players 2-5 the preregistered linear regression (contribution ∼ order) among SVO-Individualistic players reaches significance: $\beta = -7.849$, 95% CI = [-14.957, -0.542], $p = 0.034$. We specified a categorical SVO measure in our preregistration, but the continuous SVO angle measure is strictly better in that it avoids throwing away information. When interacting contribution with the SVO angle measure (contribution ∼ order*SVO_angle) we see $\beta = 0.403$, 95% CI = [0.112, 0.685], $p = 0.004$ for the interaction. As expected, the effect of order on contribution becomes larger as SVO angle nears 0°(perfect self-interested play).

The preregistered prediction that the partial correlation between one's own contribution and prediction of others' contributions is stronger going forward in time (towards the open fates of those who have not yet moved) is well-supported. Among all participants, for the forward direction we find $n = 421$ Pearson's $r = 0.41$ 95% CI = [0.33 0.49] $p < 0.001$, and for backwards in time $n = 1071$ Pearson's $r = 0.25$ 95% CI = [0.2 0.31] $p < 0.001$, z-score for the difference of 0.159 = 3.11, 95% CI = [0.067, 0.285], $p = 0.002$.[5]

[4] Similar to Study 1b, nearly all participants were SVO classified as Individualistic or Prosocial; three were classified as Altruistic, and these participants' data are excluded from categorical SVO analyses.

[5] There are more backwards-facing predictions than forwards-facing because data from Player 1 was excluded, and Player 1 only produces forward-facing predictions.

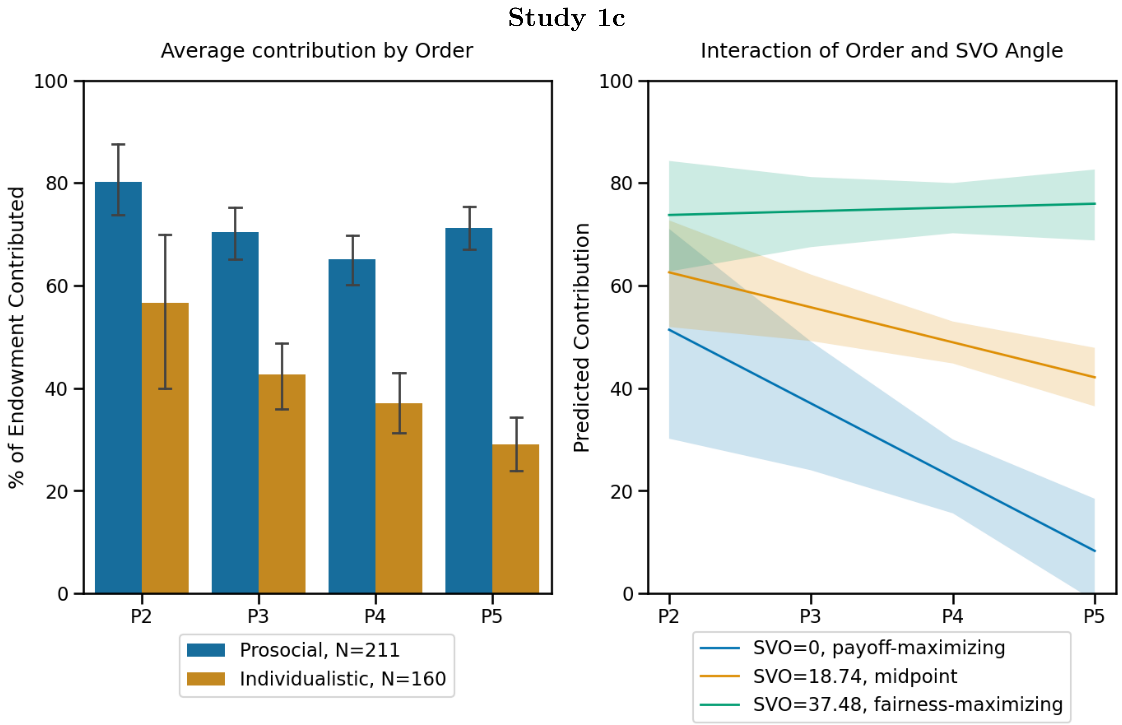


Figure 2: [Left] Study 1c shows a decline in contributions with increasing order that gets stronger as participants become more self-interested, as measured by SVO battery. [Right] Analysis with the more sensitive SVO angle measure, as opposed to categorical SVO classes, yields variation in the effect of order on contribution with SVO angle. Participants who passed comprehension checks, SEMs.

2.3.4 Discussion

Study 1c replicates the positional order effect among self-interested players that was suggested by Studies 1a and 1b. The evidence from Studies 1a-1c, taken together, was sufficient to make us confident in moving on to a second series of studies that probe the mechanism behind the positional order effect among self-interested participants.

2.4 Study 2a: 5-Person Sequential Public Goods Game with induced self-interest

Having established that positional order effects are driven by players who are trying to maximize their own payouts, we moved on to investigating the mechanism with Studies 2a and 2b. We first attempted to pre-test participants with the SVO measure in order to filter out participants who were not trying to maximize their own profits. Our real-time interaction paradigm meant that 5-person groups did not fill fast enough for that to be a viable strategy, so we explored alternatives. Study 2a tests whether merely *asking* all comers to maximize personal payoffs generates a positional order effect similar to that observed among Individualistic participants. This study was not preregistered; It was intended as a quick test of new software and the prompt to be greedy. Its primary purpose was to gauge the effectiveness of this prompt and inform confidence in deploying it at larger scale in Study 2b.

2.4.1 Participants

157 U.S.-based participants from Amazon Mechanical Turk via Cloud Research passed up-front bot checks and completed the study. Of those, 139 (89.0%) of those passed all of the up-front comprehension check questions and have their data included, a marked improvement over earlier panels direct from Mechanical Turk. The population was 37% female at an average age of 40. Mean total pay per participant (including bonuses for accurate predictions) was \$3.79 ($SD = 0.66$), yielding an hourly rate of \$14.05 at 15.5 minutes average duration.

2.4.2 Materials and procedure

Study 2a served as a testbed for some important improvements to the basic design from Study 1c. In Study 2a, SVO is not measured. Instead, players are randomized to an "Instruction" and a "No instruction" condition. In the Instruction condition, players see a prompt:

> Please try to play this game **however you think will make you the most money.** We understand that sometimes you want to help other people, but for the purposes of this experiment we want you to try to make as much money as possible.

In addition to the prompt, Study 2a incorporates four substantive improvements over Study 1c. First, Study 2a adds an additional simultaneous-play control condition that implements a delay of 80 seconds. These participants will wait about as long as sequential-condition players who are moving last (P5). This condition was incorporated to allow us to rule out effects dependent on time spent waiting. While waiting, participants are shown the task's standard wait screen which incorporates the option to play a simple game to encourage engagement with the task. Second, we incorporate an interactive practice game after the instructions and comprehension questions. This practice game asks participants to calculate the correct answers to questions about payoffs for hypothetical players in a PGG. Participants are paid for correct answers and they can make multiple attempts at any given question, limited only by time. Third, participants move in lock-step with one another. Each page in the study takes an allotted amount of time no matter the participant's behavior to ensure that information cannot leak to others via response times (e.g., P3 could move quickly relative to P2, and P4 could note the difference). Pages on which players make their contribution or prediction decisions do not force a player to stay for a certain amount of time, but rather let the player move on to a wait page that soaks up any remaining time. Finally, Study 2a incorporates an improved up-front English fluency filter that relies on a native speaker's ability to quickly complete idioms in order to ensure participants are real people who speak English fluently.

2.4.3 Results

We observe a positional order effect given the instruction to maximize payouts. A linear regression of contribution on order interacted with a binary instructed/not instructed to maximize payoffs variable, contribution $\sim$ order * instruct_or_no, detects the interaction effect, $\beta = -16.379$, 95% CI = [-25.985, -6.339], $p = 0.001$.

2.4.4 Discussion

Participants receiving the instruction to maximize payouts in Study 2a show a decline in contribution with increasing order. The fact that all comers to the experiment can be induced to produce the effect suggests it is a strategy held generally, and not just by some subset of the population. Study 2a makes use of a small sample, and consequently there is substantial noise in estimates, but this small study gave us enough confidence to deploying this technique in the next, larger experiment.

2.5 Study 2b: 5-Person Sequential Public Goods Game with random moves

Study 2b incorporates the improvements from Study 2a and extends it by applying the instruction to act to maximize one's own payouts to all participants and at larger scale, but with two new conditions: all participants are either told that every player *before* them has had their contribution determined randomly ("Random Before"), or that every player moving *after* them has had their contribution determined randomly ("Random After"). This clarifies whether the positional order effect is driven by the fact that other *people*, specifically, will be moving after the focal player—even though he cannot see their moves. The preregistration for Study 2b can be found at https://osf.io/3kepm. It incorporates hypotheses about quadratic effects, which are the subject of a separate paper.

2.5.1 Participants

834 U.S.-based participants from Amazon Mechanical Turk via Cloud Research completed the study. Of those, 645 (77%) passed the up-front comprehension checks and have their data included in analyses. Among those, mean total pay per participant (including bonuses for accurate predictions) is \$4.73 (SD = 0.64), yielding a mean hourly rate of \$15.10 at 18.4 minutes average duration. 514 of the 645 participants were in the Sequential treatment. To estimate the sample size required, we performed a power analysis via simulation using pilot data and data from previous experiments. Effect sizes from higher-quality Cloud Research panels are much larger relative to direct Mechanical Turk, meaning smaller sample sizes are required.

2.5.2 Materials and procedure

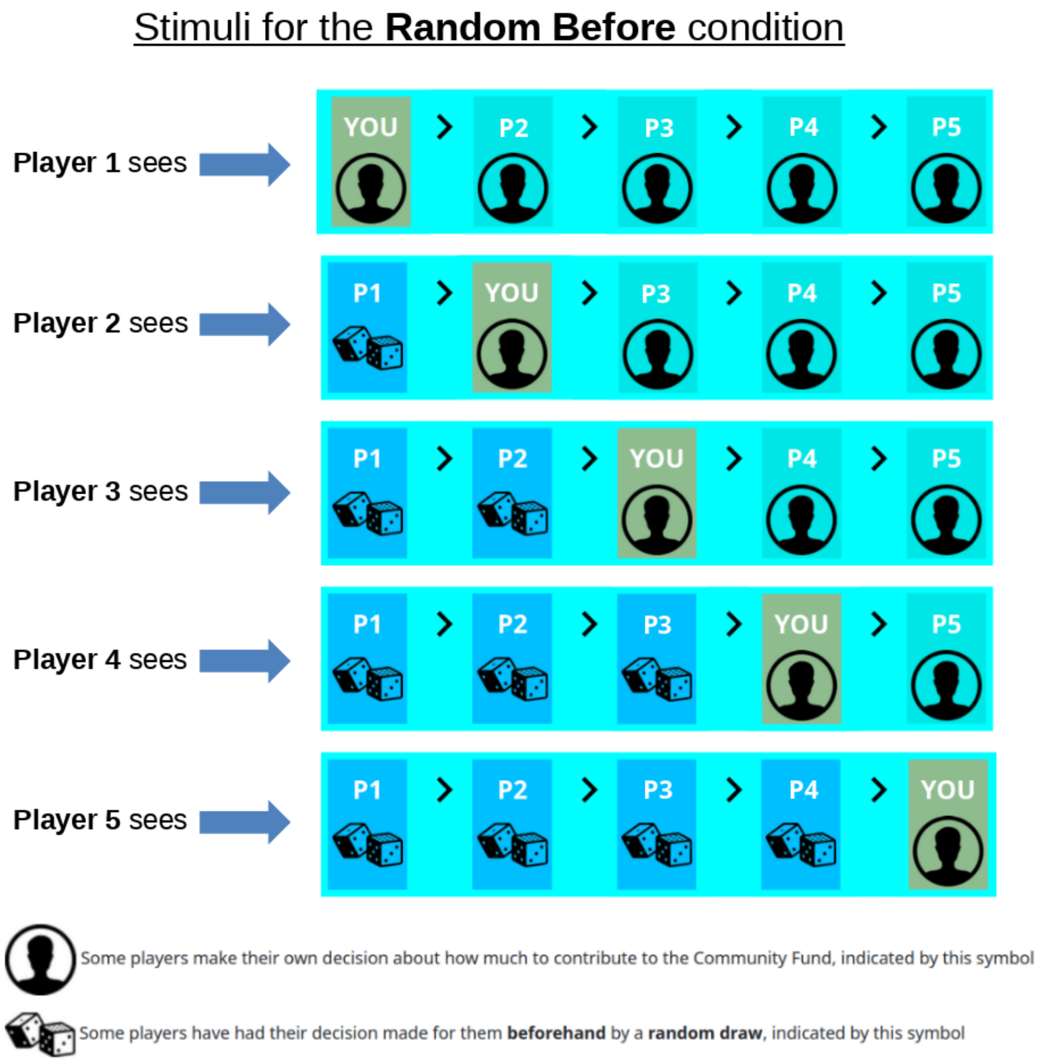
Study 2b is a one-shot sequential PGG identical to Study 2a, with the exception that all players receive the instruction to maximize earnings and they are randomized between two conditions, fully crossed with orders 1-5: players are told that either everyone *before* them in the sequence has their decision about how much to contribute to the public good made by a random process ("Random Before"), or that everyone *after* them has their decision made by a random process ("Random After"). Players are presented with a page that explains the setup, and are presented with symbols that make clear which players' moves were randomly decided. They see the graphical representations in 2.5.3 on all pages from the point at which the concept of random moves is introduced until the end of the game.

We used deception in this study. It was not true that everyone either before or after a given player was making their own decision or having their moves made randomly. Rather, each player in each five-person game made his or her own moves, and was merely *told* that the others in the game either made their own decisions or had them made randomly. Performing this study without deception would have meant four players who produce no data per five-person game, thus requiring five times as many participants at five times the cost. We determined this was unworkable, and that the costs of using deception were warranted. Participants were debriefed at the end of the experiment. As in Study 2a, there are two simultaneous control conditions: one with a delay equivalent to the wait time 5th-movers experience in the sequential game, and one without which is equivalent to moving first.

2.5.3 Results

Study 2b shows a clear effect of direction in time: we observe the positional order effect among players who are in the "Random Before" condition, but not among those in the "Random After" condition. The mean contribution for Random Before participants was 38% of \$1.00 endowment (SD=0.40), and for Random After 36% (SD=0.40). Random Before and Random After participants left with the same profit from the SPGG on average, \$1.38 (SD=0.39) for Random Before vs. \$1.43 (SD=0.38) for Random After.

We observe a decline in contribution with increasing order only among those players who are told that everyone moving *before* them has his move determined randomly, while everyone moving *after* them will



Stimuli for the **Random Before** condition

Some players make their own decision about how much to contribute to the Community Fund, indicated by this symbol

Some players have had their decision made for them **beforehand** by a **random draw**, indicated by this symbol

Figure 3: Players see a graphical representation of their position relative to other players that clearly conveys which players are having their moves made by a random process. This is in addition to a previous screen that explains how some players are having their moves made for them by random processes. The Random After condition has the dice and human figures reversed.

decide on what move to make. The preregistered linear regression contribution $\sim$ order * random_before + wealth, differing from previous analyses in that it controls for a measure of wealth, finds the effect, $\beta = -6.405$, 95% CI = [-11.284, -1.419], $p = 0.011$ for the interaction. Wealth was added to the regression given the expectation, common in economics, that players' sensitivity to payoffs is modulated by the marginal change in their wealth. The regression without wealth also finds the effect, $\beta = -6.332$, 95% CI = [-11.135, -1.369], $p = 0.012$. When we restrict the main analysis to only those players who passed a second set

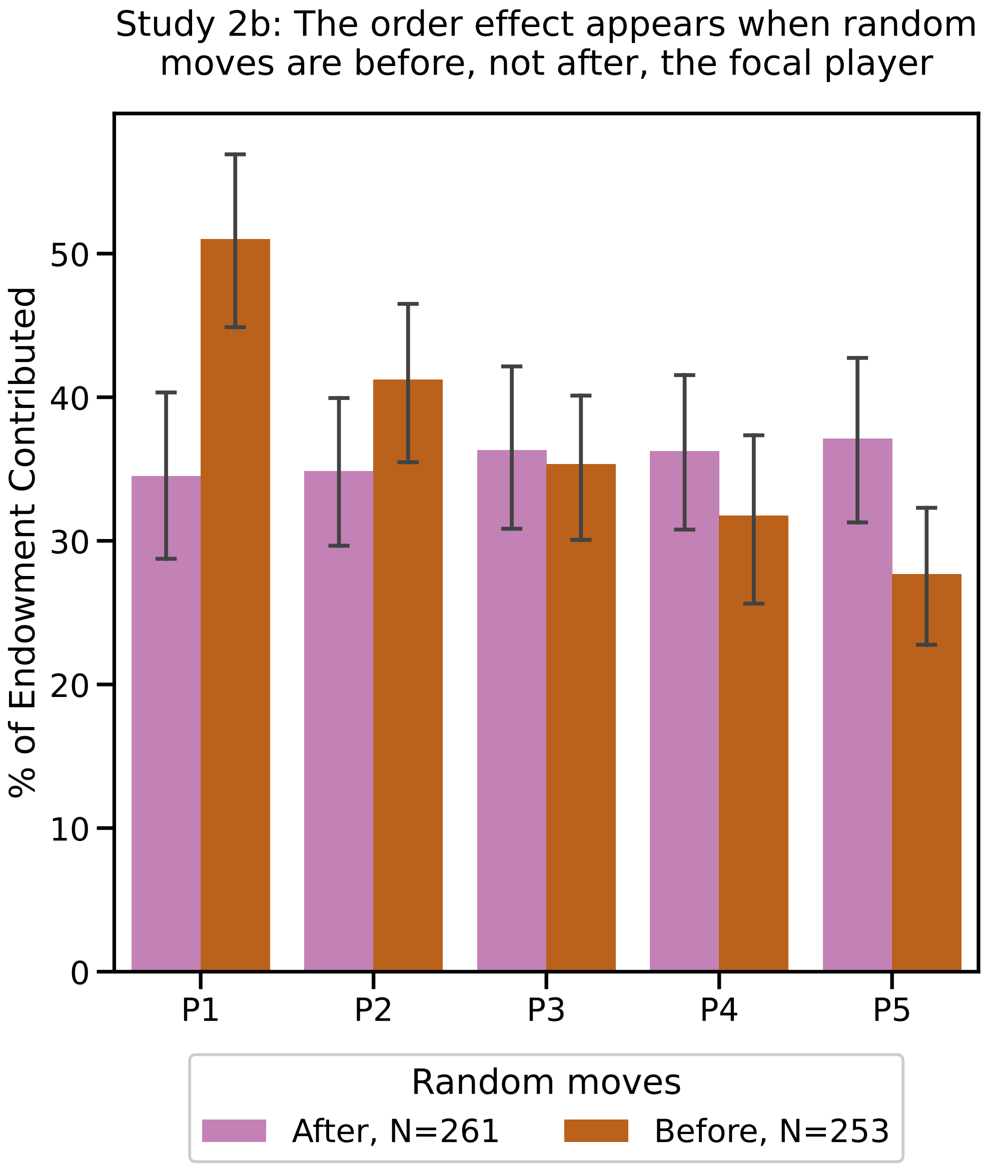


Figure 4: Study 2b shows a decline in contribution to the public good among players who are told that all players moving *after* them are making their own moves, and all players moving *before* them are having their moves made randomly. No effect is observed among players who are told that everyone moving after them has a move selected at random. Participants who passed comprehension checks. SEMs.

of comprehension checks at the end of the experiment (76.6% of players who passed the initial checks), we observe a larger effect size ($\beta = -8.624$, 95% CI = [-14.001, -3.202], $p = 0.002$ for the preregistered analysis; see Appendix E for detail). This gives further reason to believe that the effect is concentrated among participants who understand the rules of the game best. We did not preregister an exclusion for participants who time out on the contribution decision page, instead preferring to let participants allow the page to time out if they prefer to. However, it is possible to exclude participants who both never enter a contribution decision by moving the input slider (thus defaulting), and who time out; doing so, the effect size is increased, $\beta = -7.750$, 95% CI = [-13.393, -1.955], $p = 0.007$ for the interaction. On the formalization included in Appendix B, self-interested participants who are acting *as if* should either contribute 0% of their endowment or 100% depending on where they are in the sequence. 67.3% of participants give either 0 or 100% of their endowment, and among these participants effect size increases ($\beta = -9.467$, 95% CI = [-16.215, -2.540], $p = 0.009$ for the preregistered analysis, see Appendix D for detail). We do not observe a difference between the no delay (equivalent to P1, mean contribution = $0.41) and long delay (80 seconds, equivalent to P5, mean contribution = $0.48) simultaneous-move control conditions using a T-Test, t(129) = -0.90, p = 0.372. This suggests the effects are not due to mere time in the experiment; indeed, players in the long wait condition contributed slightly more than those in the short wait condition.

In addition, a player's own move is strongly predictive of her expectations of others' moves–but only in the direction people making their own moves are to be found. For the Random Before condition, the OLS model prediction $\sim$ contribution*future/past yields $\beta = 0.374$, 95% CI = [0.224, 0.523], $F(3, 1000) = 74.307$, $p < 0.001$, and for Random After $\beta = -0.432$, 95% CI = [-0.57, -0.294], $F(3, 1012) = 77.198$, $p < 0.001$ with cluster-robust standard errors. The coefficient for the interaction is positive in the Random Before condition, and negative in the Random After condition, indicating players are willing to bet their own move is predictive only of moves other people choose themselves.

2.5.4 Discussion

Players adjust their contributions to a public good downwards the nearer they are to moving last in a sequence, but only if there are people making their own decisions moving *after* them. We do not see any effect in players who are told all others moving *before* them have their moves made randomly for them. It appears that the positional order effect in SPGGs requires that players believe there are other minds who will freely make a choice, but have not yet, involved in the game. Further, the extent to which a player contributes to the public good is proportional to the number of others yet to freely make their own move. Players in the Random After condition all contribute approximately as much as Player 3 in the Random Before condition, as well, implying that acting *as if* raises contributions above baseline in the first half of the sequence and lowers them in the second half, with the middle player, Player 3 of 5, making approximately the same decision in both conditions. We also note that the mean contribution in the "Random After" condition, which shows no positional order effect, is considerably lower than what would be expected from a population that has not been instructed to maximize their own payoffs. Mean contribution in the populations sampled in Studies 1a - 1c was about 60% of the endowment across all participants, Prosocial and Individualistic, whereas mean contribution among "Random After" participants here is about 35%. Player 5 in the "Random After" condition has no random moves to contemplate at all, since he moves last, and therefore plays a standard 5-person SPGG–and contributes considerably less than non-instructed populations.

3 General Discussion

Reward-maximizing players in sequential PGGs without observation display a positional order effect: they cooperate more when they believe some players are yet to move, and this effect increases with the number of such uncommitted people moving after them. Four experiments support this view. For players who are maximizing personal profit in the game, then earlier movers tend to believe that contributing to the public good will maximize their payouts, and later movers tend to believe that defecting will maximize payouts. The effect's absence when specifically *subsequent* players have their moves made randomly is consistent with implicit causal thinking: It is not just *that* I cooperate that suggests others will cooperate, but *if* I cooperate, others will cooperate; the effect is stronger in the direction of open fates, towards the future. Participants are also willing to bet that people moving after them will make a move that is more similar to theirs relative to those moving before them, which is to be expected in the case that players are acting *as if*—for the same reason. This behavior makes it clear that the distinction between open and closed fates is important. We speculate that a simple model may capture something of the process generating this behavior specifically in self-interested agents: these agents understand the rules of the game and are trying to maximize their payouts—they just act as if everyone who has not yet moved will make the same move they do. This implies a sharp step between 100% contribution and 0% contribution, which is observed in the data. 67.3% of participants contribute either 0 or 100% of their endowment, and the positional order effect is stronger in this subset (see Appendix D for detail). The effect is also stronger in the 76.8% of participants who pass both pre- and post- comprehension checks, implying the effect is a function of players who understand the game. A formalization of this model is included in Appendix B.

Our theory suggests that some players, such as highly Prosocial participants and Indivdualistic participants who are at the beginning of a sequence, will give 100% of their endowment to the public good. However, prosocial participants are not at ceiling for contributions to the public good, nor are Individualistics who are at the beginning of the sequence. Part of this effect is due to noise, which will necessarily have an asymmetric effect (since it is not possible to contribute more than 100% of an endowment), lowering the ceiling for contributions. Some number of participants who do not understand the game well will pass comprehension checks, and some participants will evince different levels of prosociality in the SPGG vs. an SVO measure (Andreoni, 1995; Houser and Kurzban, 2002).

3.1 Underlying processes

Acting *as if* may fit the data we observe, but it is not immediately obvious why this pattern of behavior is rational or adaptive. If in the absence of other information a player assumes some similarity between himself and other players his own behavior may, in fact, be informative about others (as in self-signaling or social projection). Projection from personal decisions to collective behavior can be rational in the sense that it can be consistent with Bayes' rule (Dawes, 1989; Hoch, 1987; Tarantola et al., 2017). This could explain the sensitivity to other players making their own decisions (or not), but would not explain why the arrow of time ("closed fates" vs. "open fates") is important. Self-signaling via social projection could also explain cooperation among these self-interested agents. In a self-signaling account, individuals regard their own decisions as informative about their unknown "deep" characteristics, such as morality, affection, dedication or willpower. Self-signaling implies that individuals will favor decisions that generate good news (a positive self-signal) about these characteristics (Bernheim and Thomadsen, 2005; Bodner and Drazen Prelec, 2003; Dhar and Wertenbroch, 2012; Mijovic-Prelec and D. Prelec, 2010). In the case of an SPGG, the "good news" self-interested players may be motivated to learn is that the other people they are playing with will contribute to the public good, thereby raising their payoffs. Adjusting your estimate of your future profits upwards is pleasurable, so there is utility to be gained from that adjustment (diagnostic utility) in addition to the standard utility from the payout itself (outcome utility). Crucially, from the standpoint of both theory and empirical evidence, self-signaling does not require a perceived causal link between decisions and the underlying characteristic of interest; it can influence decisions even when their causal irrelevance is made obvious by experimental design as in Quattrone and Tversky (1984). However, as with social projection, the usual formulation of self-signaling does not naturally provide a direction in time for the effect. It is possible to self-signal about open and closed fates, and so an explanation of why participants only consider open fates would be required. This process of maximizing "news value" is also reminiscent of Evidential Decision Theory (Gibbard and Harper, 1978), Subjective Expected Relative Similarity (Fischer and Savranevski, 2023) and Superrationality (Hofstadter, 1983).

A related body of work examines universalization as an explanatory model for moral judgment. The basic idea is that, at some level, people ask themselves "What if everyone did this?" in order to determine what is right and wrong. Roemer (2010; 2015) develops the idea of a “Kantian equilibrium”, where each player asks: “if I deviate from my action and everyone else were to deviate in the same way, would I prefer the consequences of the new action profile versus not deviating at all?”, and Levine et al. (2020) present a computational model of universalization in moral judgment, along with evidence from vignette studies and, significantly, refine the motivating question to, “What if everyone felt free to do that?”, which adds a sense of temporal direction—towards open fates. This may not, in fact, be specific to the moral domain but rather a specific instance of the general mechanism we report here.

Self-signaling, social projection, and universalization could each lead to people acting as if their actions can influence other people without communicating, i.e., as if by magic—even when they correctly believe it to be impossible. However, maybe even magic has limits: it can be circumscribed by logic and commonsense metaphysics. In particular, past actions of other people may be unknown, but are not reversible. In contrast, future actions of other people are both unknown and potentially open to influence. These facts point to deeply-held priors that direct thoughts like these towards the future, potentially making any of self-signaling, social projection, or universalization viable underlying psychological mechanisms for what manifests as acting *as if* when taken in combination with these priors. It remains for future work to tease apart these psychological possibilities.

The question of why the phenomenon might manifest in the first place is a live one as well. It is possible that acting *as if* is adaptive in that it is directly payoff-maximizing in the case of small-scale societies, where baseline priors about information leakage are high (so even when one is alone one might feel watched, and for good reason—it might be true), there is more similarity among players, and a very high probability of close social and genetic connection to any other player as well. In small-scale societies the payoffs to cooperation are much closer to hand than in modern societies with high degrees of anonymity, and we may retain the basic instinct even having evolved many additional cooperation-stabilizing mechanisms. Another possibility is that acting *as if* reflects adaptation to an environment where, even if you can't quite see how defecting will harm you, it's probable there is some way it will. In this case, to ask oneself ”What if everyone else were to do this?” is to deploy a simulation of the game that informs the player about the likely *unobserved* payoff structure of the world. If acting *as if* is indeed adaptive, it may be a more primitive psychological shortcut to cooperative behavior that does not require more recently-evolved [6] cognitive machinery that must be painstakingly inculcated, such as moral universals (e.g., ”Stealing is wrong.”).

3.2 Broader Implications and Conclusion

Finally, whatever the mechanism, understanding why even the most self-interested actors might decide to contribute to the public good is relevant to many managerial and policy decisions. For example, an agent might say to herself: many other people will be in my shoes in the future, so if I vote then other people will too; if I conserve energy, then others will conserve as well; if I contribute to a public good, so will others—and this action is actually best for me independent of what's good for everyone else. Even the self-interested might feel that their investment of time or effort will pay off, pointing to a class of interventions that highlight that other people like you will be deciding to contribute—or not—at a later time. Though only a subset of the population may be acting in their own interests at any given time, they are likely to be the source of frustration for those trying to increase cooperation and prosocial behavior. Interventions that affect specifically those who are not already working for the greater good are, indeed, the goal for public policy and similar applications. Acting *as if* may be one reason why self-interested people act against their own interests, and encouraging it could pay off. These people are nevertheless in error from a self-interested point of view; contributing does not actually pay off either in our experiments or in similar situations in the real world. At the collective level, however, a team, company or community composed of individuals thinking like this may flourish compared to one composed of individuals pursuing their self-interest according

[6] And still not universally-held, (Henrich, Blasi, et al., 2023)

to normative decision theory.

Acknowledgements We wish to thank Adam Bear and Danica Mijovic-Prelec for comments on the manuscript and the MIT Neuroeconomics Lab for funding.

References

Abele, Susanne and Karl-Martin Ehrhart (Sept. 2005). "The timing effect in public good games". en. In: *Journal of Experimental Social Psychology* 41.5. 00022, pp. 470–481. ISSN: 00221031. DOI: 10/bkgq9d. URL: https://linkinghub.elsevier.com/retrieve/pii/S0022103104001180 (visited on 11/27/2019).

Andreoni, James (1995). "Cooperation in Public-Goods Experiments: Kindness or Confusion?" In: *American Economic Review* 85.4. Publisher: American Economic Association, pp. 891–904. URL: https://econpapers.repec.org/article/aeaaecrev/v_3a85_3ay_3a1995_3ai_3a4_3ap_3a891-904.htm (visited on 02/03/2025).

Bernheim, B. Douglas and Raphael Thomadsen (Apr. 2005). "Memory and Anticipation". en. In: *The Economic Journal* 115.503, pp. 271–304. ISSN: 0013-0133, 1468-0297. DOI: 10.1111/j.1468-0297.2005.00989.x. URL: https://academic.oup.com/ej/article/115/503/271/5086053 (visited on 06/08/2022).

Bodner, Ronit and Drazen Prelec (Feb. 2003). "Self-Signaling and Diagnostic Utility in Everyday Decision Making". en. In: *The Psychology of Economic Decisions*. Ed. by Isabelle Brocas and Juan D Carrillo. Oxford University PressOxford, pp. 105–124. ISBN: 978-0-19-925106-3. DOI: 10.1093/oso/9780199251063.003.0006. URL: https://academic.oup.com/book/51973/chapter/420800208 (visited on 01/03/2025).

Budescu, David V. and Wing Tung Au (Jan. 2002). "A model of sequential effects in common pool resource dilemmas". en. In: *Journal of Behavioral Decision Making* 15.1. 16 citations (Crossref) [2023-07-13], pp. 37–63. ISSN: 0894-3257, 1099-0771. DOI: 10.1002/bdm.402. URL: https://onlinelibrary.wiley.com/doi/10.1002/bdm.402 (visited on 11/24/2021).

Budescu, David V., Wing Tung Au, and Xiao-Ping Chen (Mar. 1997). "Effects of Protocol of Play and Social Orientation on Behavior in Sequential Resource Dilemmas". en. In: *Organizational Behavior and Human Decision Processes* 69.3, pp. 179–193. ISSN: 0749-5978. DOI: 10.1006/obhd.1997.2684. URL: https://www.sciencedirect.com/science/article/pii/S0749597897926844 (visited on 07/14/2023).

Budescu, David V., Ramzi Suleiman, and Amnon Rapoport (Mar. 1995). "Positional Order and Group Size Effects in Resource Dilemmas with Uncertain Resources". en. In: *Organizational Behavior and Human Decision Processes* 61.3. 59 citations (Crossref) [2023-08-31], pp. 225–238. ISSN: 07495978. DOI: 10.1006/obhd.1995.1018. URL: https://linkinghub.elsevier.com/retrieve/pii/S0749597885710187 (visited on 11/24/2021).

Burns, Zachary C., Eugene M. Caruso, and Daniel M. Bartels (2012). "Predicting premeditation: Future behavior is seen as more intentional than past behavior". In: *Journal of Experimental Psychology: General* 141.2. 00036, pp. 227–232. ISSN: 1939-2222(Electronic),0096-3445(Print). DOI: 10/bz6ngt.

Carpenter, Jeffrey, Andrea Robbett, and Prottoy A. Akbar (Sept. 2018). "Profit Sharing and Peer Reporting". en. In: *Management Science* 64.9, pp. 4261–4276. ISSN: 0025-1909, 1526-5501. DOI: 10.1287/mnsc.2017.2831. URL: http://pubsonline.informs.org/doi/10.1287/mnsc.2017.2831 (visited on 06/08/2022).

Chen, Daniel L., Martin Schonger, and Chris Wickens (Mar. 2016). "oTree—An open-source platform for laboratory, online, and field experiments". en. In: *Journal of Behavioral and Experimental Finance* 9. 00298, pp. 88–97. ISSN: 2214-6350. DOI: 10/bj42. URL: http://www.sciencedirect.com/science/article/pii/S2214635016000101 (visited on 01/11/2020).

Chen, Xiao-Ping, Wing Tung Au, and S.S. Komorita (Jan. 1996). “Sequential Choice in a Step-Level Public Goods Dilemma: The Effects of Criticality and Uncertainty”. en. In: *Organizational Behavior and Human Decision Processes* 65.1. 49 citations (Crossref) [2023-07-13], pp. 37–47. ISSN: 07495978. DOI: 10.1006/obhd.1996.0003. URL: https://linkinghub.elsevier.com/retrieve/pii/S0749597896900035 (visited on 11/24/2021).

Cooper, Russell et al. (1993). “Forward Induction in the Battle-of-the-Sexes Games”. In: *The American Economic Review* 83.5. Publisher: American Economic Association, pp. 1303–1316. ISSN: 0002-8282. URL: https://www.jstor.org/stable/2117562 (visited on 11/24/2021).

Croson, Rachel T.A. (Nov. 1999). “The Disjunction Effect and Reason-Based Choice in Games”. en. In: *Organizational Behavior and Human Decision Processes* 80.2. 00176, pp. 118–133. ISSN: 07495978. DOI: 10/fhrh8g. URL: https://linkinghub.elsevier.com/retrieve/pii/S0749597899928467 (visited on 09/28/2019).

Daley, Brendan and Philipp Sadowski (May 2017). “Magical thinking: A representation result”. en. In: *Theoretical Economics* 12.2. 00008, pp. 909–956. ISSN: 1555-7561. DOI: 10/f99g8r. URL: https://onlinelibrary.wiley.com/doi/10.3982/TE2099 (visited on 09/28/2019).

Dawes, Robyn M (Jan. 1989). “Statistical criteria for establishing a truly false consensus effect”. en. In: *Journal of Experimental Social Psychology* 25.1, pp. 1–17. ISSN: 00221031. DOI: 10.1016/0022-1031(89)90036-X. URL: https://linkinghub.elsevier.com/retrieve/pii/002210318990036X (visited on 08/06/2020).

Delfgaauw, Josse et al. (Jan. 2022). “Team Incentives, Social Cohesion, and Performance: A Natural Field Experiment”. en. In: *Management Science* 68.1, pp. 230–256. ISSN: 0025-1909, 1526-5501. DOI: 10.1287/mnsc.2020.3901. URL: http://pubsonline.informs.org/doi/10.1287/mnsc.2020.3901 (visited on 06/09/2022).

Dhar, Ravi and Klaus Wertenbroch (Feb. 2012). “Self-Signaling and the Costs and Benefits of Temptation in Consumer Choice”. en. In: *Journal of Marketing Research* 49.1. 00110, pp. 15–25. ISSN: 0022-2437, 1547-7193. DOI: 10/bbng3z. URL: http://journals.sagepub.com/doi/10.1509/jmr.10.0490 (visited on 01/11/2020).

Douglas, Benjamin D., Patrick J. Ewell, and Markus Brauer (Mar. 2023). “Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA”. en. In: *PLOS ONE* 18.3. Publisher: Public Library of Science, e0279720. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0279720. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0279720 (visited on 12/07/2023).

Figuières, Charles, David Masclet, and Marc Willinger (July 2012). “Vanishing Leadership and Declining Reciprocity in a Sequential Contributions Experiment”. en. In: *Economic Inquiry* 50.3. 18 citations (Crossref) [2023-07-13], pp. 567–584. ISSN: 00952583. DOI: 10.1111/j.1465-7295.2011.00415.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1465-7295.2011.00415.x (visited on 11/24/2021).

Fischer, Ilan and Lior Savranevski (Nov. 2023). “The effect of similarity perceptions on human cooperation and confrontation”. en. In: *Scientific Reports* 13.1. Publisher: Nature Publishing Group, p. 19849. ISSN: 2045-2322. DOI: 10.1038/s41598-023-46609-8. URL: https://www.nature.com/articles/s41598-023-46609-8 (visited on 06/21/2024).

Gibbard, Allan and William L. Harper (1978). “Counterfactuals and Two Kinds of Expected Utility”. In: *Ifs: Conditionals, belief, decision, chance and time.* Ed. by William L. Harper, Robert Stalnaker, and Glenn Pearce. Dordrecht: Springer Netherlands, pp. 153–190. ISBN: 978-90-277-1220-2. DOI: 10.1007/978-94-009-9117-0_8. URL: http://link.springer.com/10.1007/978-94-009-9117-0_8 (visited on 12/21/2023).

Güth, Werner, Steffen Huck, and Amnon Rapoport (Feb. 1998). “The limitations of the positional order effect: Can it support silent threats and non-equilibrium behavior?” en. In: *Journal of Economic Behavior &*

Organization 34.2, pp. 313–325. ISSN: 01672681. DOI: 10.1016/S0167-2681(97)00057-7. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167268197000577 (visited on 11/24/2021).

Hauser, David J. et al. (Dec. 2023). "Evaluating CloudResearch's Approved Group as a solution for problematic data quality on MTurk". en. In: *Behavior Research Methods* 55.8, pp. 3953–3964. ISSN: 1554-3528. DOI: 10.3758/s13428-022-01999-x. URL: https://doi.org/10.3758/s13428-022-01999-x (visited on 12/07/2023).

Henrich, Joseph, Damián E. Blasi, et al. (June 2023). "A Cultural Species and its Cognitive Phenotypes: Implications for Philosophy". en. In: *Review of Philosophy and Psychology* 14.2, pp. 349–386. ISSN: 1878-5166. DOI: 10.1007/s13164-021-00612-y. URL: https://doi.org/10.1007/s13164-021-00612-y (visited on 07/03/2024).

Henrich, Joseph and Michael Muthukrishna (Jan. 2021). "The Origins and Psychology of Human Cooperation". en. In: *Annual Review of Psychology* 72.1, pp. 207–240. ISSN: 0066-4308, 1545-2085. DOI: 10.1146/annurev-psych-081920-042106. URL: https://www.annualreviews.org/doi/10.1146/annurev-psych-081920-042106 (visited on 02/27/2023).

Hoch, Stephen J. (1987). "Perceived consensus and predictive accuracy: The pros and cons of projection." en. In: *Journal of Personality and Social Psychology* 53.2, pp. 221–234. ISSN: 1939-1315, 0022-3514. DOI: 10.1037/0022-3514.53.2.221. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.53.2.221 (visited on 06/09/2022).

Hofstadter, Douglas R. (1983). "Metamagical Themas". In: *Scientific American* 248.6. Publisher: Scientific American, a division of Nature America, Inc., pp. 14–29. ISSN: 00368733, 19467087. URL: http://www.jstor.org.libproxy.mit.edu/stable/24968913 (visited on 08/31/2023).

Houser, Daniel and Robert Kurzban (Aug. 2002). "Revisiting Kindness and Confusion in Public Goods Experiments". en. In: *American Economic Review* 92.4, pp. 1062–1069. ISSN: 0002-8282. DOI: 10.1257/00028280260344605. URL: https://pubs.aeaweb.org/doi/10.1257/00028280260344605 (visited on 02/03/2025).

Hristova, Evgenia and Maurice Grinberg (2010). "Testing Two Explanations for the Disjunction Effect in Prisoner's Dilemma Games: Complexity and Quasi-Magical Thinking". en. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 32. Cognitive Science Society. URL: https://escholarship.org/uc/item/3t20s2q7.

Knez, Marc and Duncan Simester (Oct. 2001). "Firm-Wide Incentives and Mutual Monitoring at Continental Airlines". en. In: *Journal of Labor Economics* 19.4, pp. 743–772. ISSN: 0734-306X, 1537-5307. DOI: 10.1086/322820. URL: https://www.journals.uchicago.edu/doi/10.1086/322820 (visited on 06/09/2022).

Kohlberg, Elon and Jean-Francois Mertens (1986). "On the Strategic Stability of Equilibria". In: *Econometrica* 54.5. Publisher: [Wiley, Econometric Society], pp. 1003–1037. ISSN: 0012-9682. DOI: 10.2307/1912320. URL: https://www.jstor.org/stable/1912320 (visited on 07/18/2023).

Kreps, David M. (1990). *Game Theory and Economic Modelling*. en. Google-Books-ID: qMoTDAAAQBAJ. Oxford University Press. ISBN: 978-0-19-828381-2.

Langer, Ellen J. (1975). "The illusion of control". In: *Journal of Personality and Social Psychology* 32.2. Place: US Publisher: American Psychological Association, pp. 311–328. ISSN: 1939-1315. DOI: 10.1037/0022-3514.32.2.311.

Levine, Sydney et al. (Oct. 2020). "The logic of universalization guides moral judgment". en. In: *Proceedings of the National Academy of Sciences* 117.42. 12 citations (Crossref) [2023-07-17], pp. 26158–26169. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.2014505117. URL: https://pnas.org/doi/full/10.1073/pnas.2014505117 (visited on 07/04/2023).

Luce, Duncan R. (Feb. 1992). "Where does subjective expected utility fail descriptively?" en. In: *Journal of Risk and Uncertainty* 5.1, pp. 5–27. ISSN: 1573-0476. DOI: 10.1007/BF00208784. URL: https://doi.org/10.1007/BF00208784 (visited on 07/18/2023).

Masel, Joanna (Oct. 2007). "A Bayesian model of quasi-magical thinking can explain observed cooperation in the public good game". en. In: *Journal of Economic Behavior & Organization* 64.2. 19 citations (Crossref) [2023-07-17], pp. 216–231. ISSN: 01672681. DOI: 10.1016/j.jebo.2005.07.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167268106000977 (visited on 07/04/2023).

Mijovic-Prelec, D. and D. Prelec (Jan. 2010). "Self-deception as self-signalling: a model and experimental evidence". en. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 365.1538. 00109, pp. 227–240. ISSN: 0962-8436, 1471-2970. DOI: 10/c2tqv2. URL: http://rstb.royalsocietypublishing.org/cgi/doi/10.1098/rstb.2009.0218 (visited on 11/28/2018).

Miller, Dale T. and Saku Gunasegaram (Dec. 1990). "Temporal order and the perceived mutability of events: Implications for blame assignment." en. In: *Journal of Personality and Social Psychology* 59.6, pp. 1111–1118. ISSN: 1939-1315, 0022-3514. DOI: 10.1037/0022-3514.59.6.1111. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.59.6.1111 (visited on 07/19/2023).

Morris, Michael W., Damien L.H. Sim, and Vittorio Girotto (Sept. 1998). "Distinguishing Sources of Cooperation in the One-Round Prisoner's Dilemma: Evidence for Cooperative Decisions Based on the Illusion of Control". en. In: *Journal of Experimental Social Psychology* 34.5. 00090, pp. 494–512. ISSN: 00221031. DOI: 10/d4cs3w. URL: http://linkinghub.elsevier.com/retrieve/pii/S0022103198913610 (visited on 03/05/2019).

Murphy, Ryan O., Kurt A. Ackermann, and Michel Handgraaf (2011). "Measuring Social Value Orientation". en. In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.1804189. URL: http://www.ssrn.com/abstract=1804189 (visited on 06/09/2022).

Nyberg, Anthony J. et al. (July 2018). "Collective Pay for Performance: A Cross-Disciplinary Review and Meta-Analysis". en. In: *Journal of Management* 44.6, pp. 2433–2472. ISSN: 0149-2063, 1557-1211. DOI: 10.1177/0149206318770732. URL: http://journals.sagepub.com/doi/10.1177/0149206318770732 (visited on 06/08/2022).

Quattrone, George A. and Amos Tversky (1984). "Causal versus diagnostic contingencies: On self-deception and on the voter's illusion". In: *Journal of Personality and Social Psychology* 46.2. 00479, pp. 237–248. ISSN: 1939-1315(Electronic),0022-3514(Print). DOI: 10/dr4gxj.

Rand, David G. and Martin A. Nowak (Aug. 2013). "Human cooperation". en. In: *Trends in Cognitive Sciences* 17.8, pp. 413–425. ISSN: 13646613. DOI: 10.1016/j.tics.2013.06.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S1364661313001216 (visited on 02/27/2023).

Rapoport, Amnon (Mar. 1997). "Order of play in strategically equivalent games in extensive form". en. In: *International Journal of Game Theory* 26.1. 54 citations (Crossref) [2023-07-13], pp. 113–136. ISSN: 0020-7276, 1432-1270. DOI: 10.1007/BF01262516. URL: http://link.springer.com/10.1007/BF01262516 (visited on 11/24/2021).

Robinson, A. Emanuel et al. (Oct. 2010). "Causality in Solving Economic Problems". en. In: *The Journal of Problem Solving* 3.1. 00006. ISSN: 1932-6246. DOI: 10/ggdjxn. URL: https://docs.lib.purdue.edu/jps/vol3/iss1/6 (visited on 11/27/2019).

Roemer, John E. (2010). "Kantian Equilibrium". en. In: *The Scandinavian Journal of Economics* 112.1. 77 citations (Crossref) [2023-07-17] _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9442.2009.01592.x, pp. 1–24. ISSN: 1467-9442. DOI: 10.1111/j.1467-9442.2009.01592.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9442.2009.01592.x (visited on 07/14/2023).

— (July 2015). "Kantian optimization: A microfoundation for cooperation". en. In: *Journal of Public Economics*. The Nordic Model 127. 52 citations (Crossref) [2023-07-17], pp. 45–57. ISSN: 0047-2727. DOI: 10.1016/j.jpubeco.2014.03.011. URL: https://www.sciencedirect.com/science/article/pii/S0047272714000607 (visited on 07/14/2023).

Shafir, Eldar and Amos Tversky (Oct. 1992). "Thinking through uncertainty: Nonconsequential reasoning and choice". en. In: *Cognitive Psychology* 24.4. 00865, pp. 449–474. ISSN: 00100285. DOI: 10/d6thrq. URL: https://linkinghub.elsevier.com/retrieve/pii/001002859290015T (visited on 01/14/2020).

Stefan, Simona and Daniel David (Feb. 2013). "Recent developments in the experimental investigation of the illusion of control. A meta-analytic review: A meta-analysis of the illusion of control". en. In: *Journal of Applied Social Psychology* 43.2, pp. 377–386. ISSN: 00219029. DOI: 10.1111/j.1559-1816.2013.01007.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1559-1816.2013.01007.x (visited on 03/14/2023).

Steiger, Eva-Maria and Ro'i Zultan (Jan. 2014). "See no evil: Information chains and reciprocity". en. In: *Journal of Public Economics* 109. 13 citations (Crossref) [2023-07-13], pp. 1–12. ISSN: 00472727. DOI: 10.1016/j.jpubeco.2013.10.006. URL: https://linkinghub.elsevier.com/retrieve/pii/S0047272713002089 (visited on 03/30/2023).

Tarantola, Tor et al. (Oct. 2017). "Prior preferences beneficially influence social and non-social learning". en. In: *Nature Communications* 8.1. Number: 1 Publisher: Nature Publishing Group, p. 817. ISSN: 2041-1723. DOI: 10.1038/s41467-017-00826-8. URL: https://www.nature.com/articles/s41467-017-00826-8 (visited on 07/19/2023).

Tversky, Amos and Eldar Shafir (Sept. 1992). "The Disjunction Effect in Choice under Uncertainty". en. In: *Psychological Science* 3.5, pp. 305–310. ISSN: 0956-7976, 1467-9280. DOI: 10.1111/j.1467-9280.1992.tb00678.x. URL: http://journals.sagepub.com/doi/10.1111/j.1467-9280.1992.tb00678.x (visited on 09/22/2023).

Von Neumann, John and Oskar Morgenstern (2004). *Theory of games and economic behavior*. 60th anniversary ed. OCLC: ocm56443320. Princeton, N.J. ; Woodstock: Princeton University Press. ISBN: 978-0-691-11993-9.

Weber, Roberto A, Colin F Camerer, and Marc Knez (2004). "Timing and Virtual Observability in Ultimatum Bargaining and "Weak Link" Coordination Games". en. In: *Experimental Economics* 7, pp. 25–48.

Zelmer, Jennifer (2003). "Linear Public Goods Experiments: A Meta-Analysis". In: *Experimental Economics* 6.3, pp. 299–310. ISSN: 13864157. DOI: 10.1023/A:1026277420119. URL: http://link.springer.com/10.1023/A:1026277420119 (visited on 02/28/2023).

Appendices

Appendices

A The Public Goods Game

In a standard PGG, n players are each given an endowment e, and are asked to decide what proportion of their endowments to contribute to the public good, from nothing to all of it. A given player's contribution to the public good is represented by a. The total amount from all the players that is contributed to the public good, c, is then multiplied by a multiplier m (which must be less than the number of players), and this amount is distributed evenly among all the players—even those who chose to contribute nothing. An individual player's payoff function in a standard simultaneous-move PGG is as follows:

$$p = \frac{mc}{n} + e(1 - a) \tag{1}$$

Consequently, whenever the multiplier m is less than the number of players n, the group as a whole does better if everyone contributes their entire endowment (cooperates), but each individual player is better off if he or she contributes nothing (defects). Put another way, the total amount of money in the group is maximized if everyone cooperates, but any individual player always makes more by defecting—independent of anyone else's moves. Because other players do not know your move, they cannot change their own moves in reaction to it. If a group plays the game only once, it is impossible to build reputations, enact retribution, or to reward others for their actions.

B Model

Here we provide a more precise statement of a model that generates the hypothesized interaction between the positional order effect and pro-social motivation.

B.1 Prosocial preferences

Consider a sequential PGG with n players endowed with 1 payoff unit each, and multiplier m, with $1 < m < n$. Players are indexed by their order of play in the sequence,$i = 1, ..., n$. Let a_i denote the contribution of player i, $i, 0 \leq a_i \leq 1$, and the payoff to player .

$$p_i = 1 - a_i + \frac{m}{n} \sum_{k=1}^{n} a_k \tag{2}$$

Prosocial preferences are modeled through a prosocial parameter s_i where $s_i = 0$ indicates pure self-interest and $s_i = 1$ pure prosocial motivation. In keeping with the experimental setup, we assume that players do not learn the specific contributions of other players. The utility of player i is therefore a function of the two variables the player does or will know, namely contribution a_i and payoff p_i:

$$u_i (a_1, \ldots, a_n) = (1 - s_i) p_i + s_i m a_i \tag{3}$$

where p_i is determined by the game formula, 2 . A purely self-interested player ($s_i = 0$) will aim to maximize own payoff, $u_i = p_i$; a purely prosocial player ($s_i = 1$) will aim to maximize the impact of his contribution to the public good, $u_i = ma_i$. The prosocial motive, captured by the second term, thus reflects the impact of own contribution to the public good; other players' contributions enter the utility model only insofar they determine the first, self-interested utility term. In other words, players: (a) care how their action affects the payoffs of others, (b) care how other players' contribution affect their own payoff, but (c) do not care how other players' actions affect each others' payoffs.

B.2 Decision dependent expectations

We assume that players compare expected utilities conditional on contributing ($a_i = 1$) or not contributing ($a_i = 0$), and choose whichever expected utility is higher (we ignore here fractional contributions). The decision criterion is therefore the difference between the two expected utilities:

$$a_i = 1 \Longleftrightarrow \mathbb{E}\left[u_i \mid a_i = 1, s_i\right] > \mathbb{E}\left[u_i \mid a_i = 0, s_i\right] \tag{4}$$

A player knows the value of their prosocial parameter and hence also knows the utility function in 2. If he were just a spectator, not making a decision, his expectation of the contribution of another, randomly selected player would exhibit projection, along the lines of Bayesian updating. The simplest version of such updating is linear:

$$\mathbb{E}\left[a_k \mid s_i\right] = b + cs_i \tag{5}$$

Prosocial players are more optimistic about the overall contribution level, other things equal.

The critical assumption we now make is that expectations of future players' contributions are additionally influenced by a player's own action, while expectations of prior players' contributions are not influenced. Let $a_{k<i}$ denote the contribution of any player moving before player i, and $a_{k>i}$ the contribution of any player moving after player. We assume:

$$\begin{aligned}
\mathbb{E}\left[a_{k<i} \mid a_i, s_i\right] &= b + cs_i \\
\mathbb{E}\left[a_{k>i} \mid a_i, s_i\right] &= b + cs_i + d\left(a_i - \mathbb{E}\left[a_k \mid s_i\right]\right) \\
&= (b-d) + (c-d)s_i + da_i
\end{aligned}$$

where $\mathbb{E}\left[a_k \mid s_i\right] = b + cs_i$ from 5 is substituted in the final line.

There is no perceived causality with respect to previous players, since expectations are the same irrespective of contribution:

$$\mathbb{E}\left[a_{k<i} \mid 1, s_i\right] - \mathbb{E}\left[a_{k<i} \mid 0, s_i\right] = 0$$

There is perceived causality with respect to future players, proportional to the "magical influence" parameter d:

$$\mathbb{E}\left[a_{k>i} \mid 1, s_i\right] - \mathbb{E}\left[a_{k>i} \mid 0, s_i\right] = d$$

The decision criterion in 4 can be expressed as:

$$\begin{aligned}
\mathbb{E}\left[u_i | a_i = 1, s_i\right] - \mathbb{E}\left[u_i | a_i = 0, s_i\right] &= (1-s_i)\mathbb{E}\left[p_i | a_i = 1, s_i\right] + s_i m - (1-s_i)\mathbb{E}\left[p_i | a_i = 0, s_i\right] \\
&= (1-s_i)\left(\mathbb{E}\left[p_i | a_i = 1, s_i\right] - \mathbb{E}\left[p_i | a_i = 0, s_i\right]\right) + s_i m \\
&= (1-s_i)\left(-1 + \frac{m}{n}\mathbb{E}\left[\sum_{k=1}^{n} a_k | a_i = 1, s_i\right] - \frac{m}{n}\mathbb{E}\left[\sum_{k=1}^{n} a_k | a_i = 0, s_i\right]\right) + s_i m
\end{aligned}$$

where the first line follows from 3 and the third line from 4.

Assuming that expectations about contributions of previous players are not affected by own contribution, the difference in expected total contribution resolves as:

$$\begin{aligned}\mathbb{E}\left[\sum_{k=1}^{n} a_k \mid a_i = 1, s_i\right] - \mathbb{E}\left[\sum_{k=1}^{n} a_k \mid a_i = 0, s_i\right] &= 1 + \mathbb{E}\left[\sum_{k=i+1}^{n} a_k \mid a_i = 1, s_i\right] - \mathbb{E}\left[\sum_{k=i+1}^{n} a_k \mid a_i = 0, s_i\right] \\ &= 1 + d(n-i)\end{aligned}$$

Substituting into the criterion,

$$\begin{aligned}&\mathbb{E}\left[u_i | a_i = 1, s_i\right] - \mathbb{E}\left[u_i | a_i = 0, s_i\right] \\ &\qquad = (1 - s_i)\left(-1 + \frac{m}{n}\left(1 + d(n-i)\right)\right) + s_i m. \qquad (6)\end{aligned}$$

For any particular value of s_i, the minimum magical influence parameter $d^*(i)$ that leads to $a_i = 1$, i.e., full contribution to the Public Good, is computed as:

$$\begin{aligned}&\mathbb{E}\left[u_i \mid a_i = 1, s_i\right] - \mathbb{E}\left[u_i \mid a_i = 0, s_i\right] = 0 \\ &\qquad \iff d^*(i) = \frac{-m - smn + n}{m(n-i)} \qquad (7)\end{aligned}$$

Note that $d^*(i)$ is increasing in i (if the expression is positive) and decreasing in s_i. The increase in i is the positional order effect: Players later in the sequence require a higher value of $d^*(i)$ in order to contribute. Assuming that d is an exogenous parameter with some distribution in the participant sample, fewer players will clear the cutoff and contribute if they are later in the sequence. The decrease in s_i simply indicates that prosocial players require less acting *as if* in order to contribute.

The second implication of the model is that the slope of this function with respect to i (the term in the brackets in 7) is steeper if s_i is smaller, that is, if players are more self-interested. To show this, we differentiate:

$$\frac{dd^*(i)}{di} = \frac{1}{(n-i)^2}\left(\frac{n-m}{m} - \frac{s_i}{(1-s_i)}n\right)$$

which is decreasing in s_i. This is the hypothesized interaction of order and prosociality. Less prosocial players will exhibit a stronger effect. Conversely, the positional order effect should disappear if s_i is sufficiently high.

C No timeouts: When considering only participants who actively make a contribution decision, effect sizes increase

We did not preregister an exclusion for subjects who let the contribution decision page time out, preferring instead to allow this behavior. Subjects may input a decision using the slider and then contemplate it, or merely prefer to wait for the next page to automatically appear. However, it may be reasonable to exclude subjects who both time out and never touch the input slider. The slider does not have a visible default state, so it is hard to believe subjects who never touched it were merely acquiescing to the default.

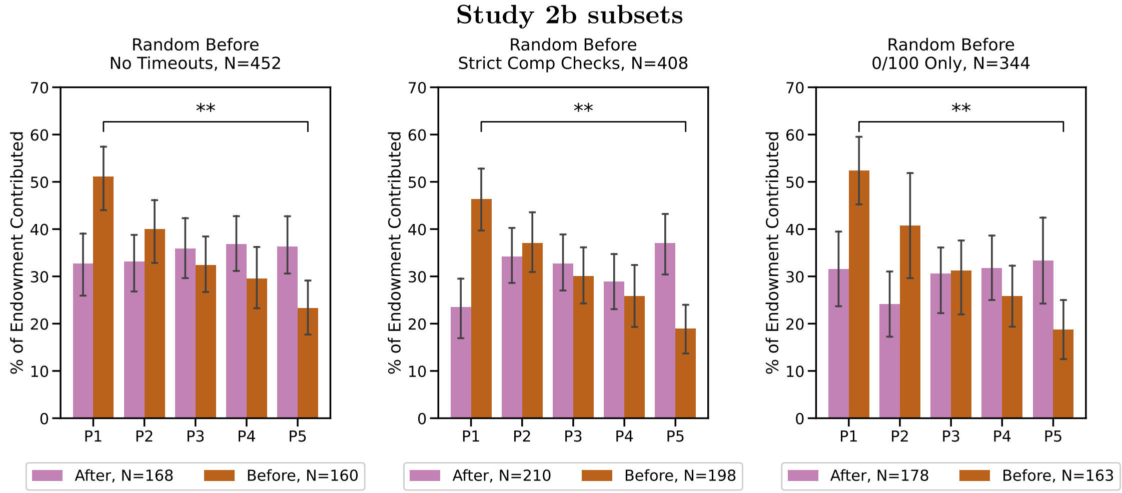


Figure 5: Study 2b shows a decline in contribution to the public good among players who are told that all players moving after them are making their own moves, and all players moving before them are having their moves made randomly. This holds when limiting responses to those who actively choose a contribution and advance the contribution page, those who pass both pre- and post-comprehension checks, and those who give either 100% or 0% of their endowment. In all these cases, effect size increases. SEMs.

C.1 Positional order effects

The preregistered linear regression contribution $\sim$ order * random_before + wealth finds the effect,
$\beta = -7.750$, 95% CI = [-13.393, -1.955], $p = 0.007$
We also find a significant equation not controlling for wealth,
$\beta = -7.787$, 95% CI = [-13.328, -2.126], $p = 0.007$

C.2 Correlation between own move and predictions of others' moves

Participants in the Random Before condition
$\beta = 0.382$, 95% CI = [0.278, 0.487], $p < 0.001$
Participants in the Random After condition
$\beta = -0.434$, 95% CI = [-0.536, -0.333], $p < 0.001$

D 0s and 1s: When considering only participants contributing all or nothing, effect sizes increase

The formalization in Appendix B predicts that any given player who is both trying to maximize his own payoffs and who is acting *as if* in accordance with the model will either give 100% of the endowment or 0% to the public good, with a sharp transition. The point at which the shift from 100% to 0% happens as order increases is a function of d, the magical influence parameter, when s_i, the player's prosociality, and m, the game's multiplier, are held constant. In Study 2b participants were instructed to maximize their own payoffs, and m is constant. Results from players who either give 0% or 100% of their endowment in Study 2b show increased effect sizes.

It may be the case that there is a weaker effect going backwards in time, towards players who have already made their moves. While our formalization only looks forward, our theoretical commitments merely see open fates as more compelling targets for acting *as if*.

D.1 Positional order effects

The preregistered linear regression contribution ∼ order * random_before + wealth finds the effect,
$\beta = -9.467$, 95% CI = [-16.215, -2.540], $p = 0.009$
We also find a significant equation not controlling for wealth,
$\beta = -9.383$, 95% CI = [-15.991, -2.664], $p = 0.008$

D.2 Correlation between own move and predictions of others' moves

Participants in the Random Before condition
$\beta = 0.370$, 95% CI = [0.264, 0.480], $p < 0.001$
Participants in the Random After condition
$\beta = -0.464$, 95% CI = [-0.574, -0.354], $p < 0.001$

E Strict comprehension checks: When considering only participants who pass both pre- and post- comprehension checks, effect sizes increase

Study 2b implemented several comprehension checks after the main task:

1. Could other players in the game see what choices you made? For instance, did other players know how much you chose to contribute?
   (a) NO, Other players could NOT see the choices I made in the game
   (b) YES, other players could see the choices I made in the game
2. Would you have more money right now if you had decided to contribute less to the Community Fund?[1]
   (a) NO, I would not have more money right now if I had decided to contribute less
   (b) YES, I would have more money right now if I had decided to contribute less
3. Is there any way the decisions you made while playing the game could have influenced what other players chose to do?
   (a) NO, my decisions could not influence what other players chose to do
   (b) YES, my decisions could influence what other players chose to do

The fact that effect sizes increase when using a stricter comprehension check regime gives further support to the claim that the positional order effect is generated by people who best understand the game and who are trying to maximize their own personal payoffs.

[1] This question is only applicable to participants who contributed something to the public good.

E.1 Positional order effects

The preregistered linear regression contribution ∼ order * random_before + wealth finds the effect,
$\beta = -8.624$, 95% CI = [-14.001, -3.202], $p = 0.002$
We also find a significant equation not controlling for wealth,
$\beta = -8.623$, 95% CI = [-13.995, -3.165], $p = 0.002$

E.2 Correlation between own move and predictions of others' moves

Participants in the Random Before condition
$\beta = 0.452$, 95% CI = [0.338, 0.564], $p < 0.001$
Participants in the Random After condition
$\beta = -0.453$, 95% CI = [-0.562, -0.347], $p < 0.001$

F Social Value Orientation distributional data

Social Value Orientation distributional information is reported in figure F and F for participants from Studies 1b and 1c. Participants filled out an SVO slider task at the end of the experiments.

G Stimuli

- Stimuli_Ex1b.pdf: Stimuli from Study 1b.
- Stimuli_Ex1c.pdf: Stimuli from Study 1c.
- Stimuli_Ex2a.pdf: Stimuli from Study 2a.
- Stimuli_Ex2b.pdf: Stimuli from Study 2b.

H Preregistrations

- Study 1b's preregistration can be found at https://osf.io/dbcpv. Study 1b deviated from the preregistration in that the actual design run was a 2x4, not 2x2x4.
- Study 1b's preregistration can be found at https://osf.io/3vsxk. Study 1b deviated from the preregistration in that the preregistration specifies data from 1,000 participants, while we actually collected data from 775 participants after the reregistration. When pooled with data from before the preregistration we reach 1002 participants. The budget for this study was planned for 1,000 participants total, rather than for 1,000 after the preregistration.
- Study 1c's preregistration can be found at https://osf.io/gw8nc. We preregistered 800 participants and ended up slightly short due to not having perfect control over how many participants finish.
- Study 2a was not preregistered.
- Study 2b's preregistration can be found at https://osf.io/3kepm. The preregistration specifies 500 participants in the sequential conditions.

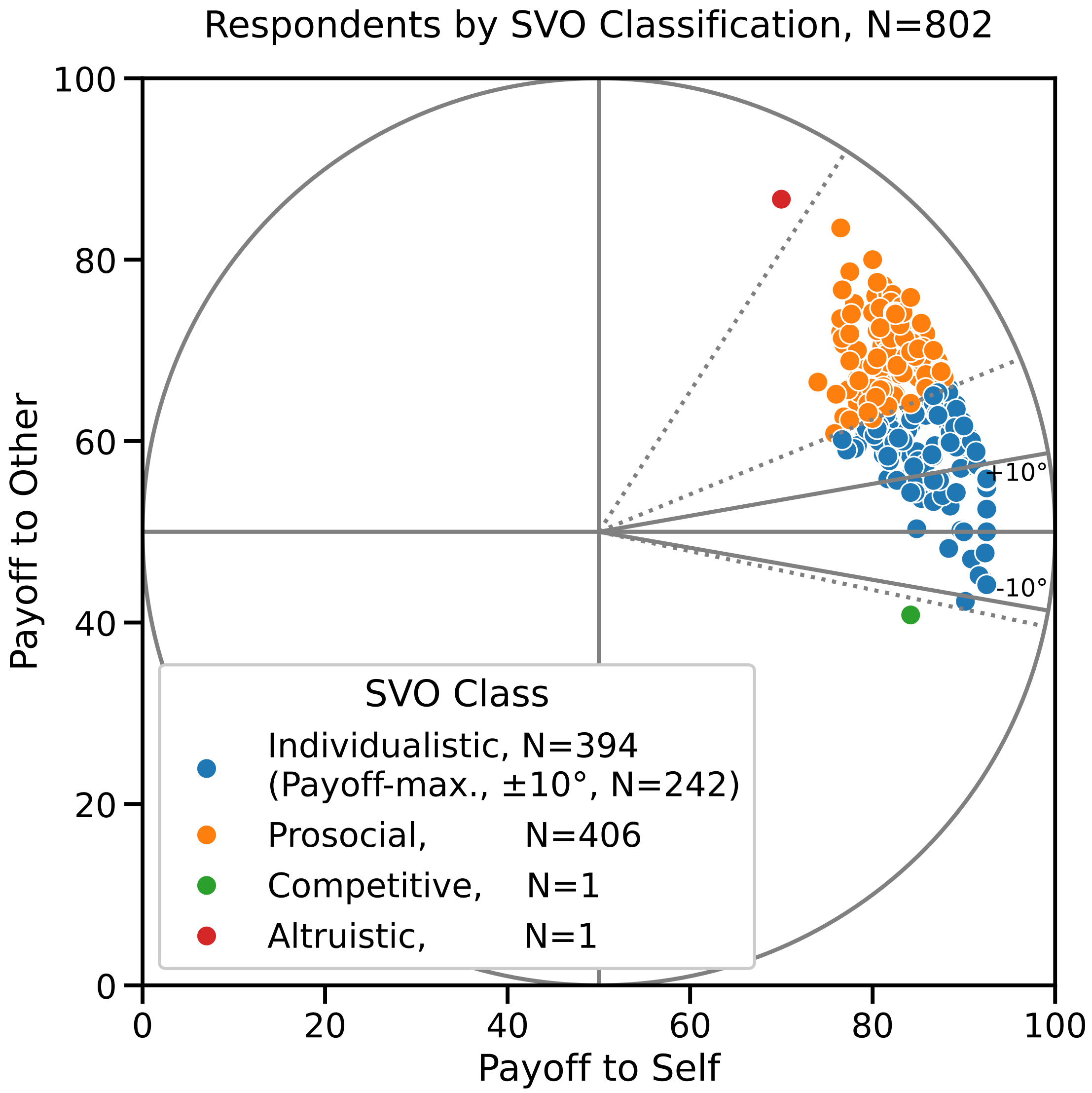


Figure 6: Social Value Orientation distributional data for Study 1b. Participants who passed comprehension checks.

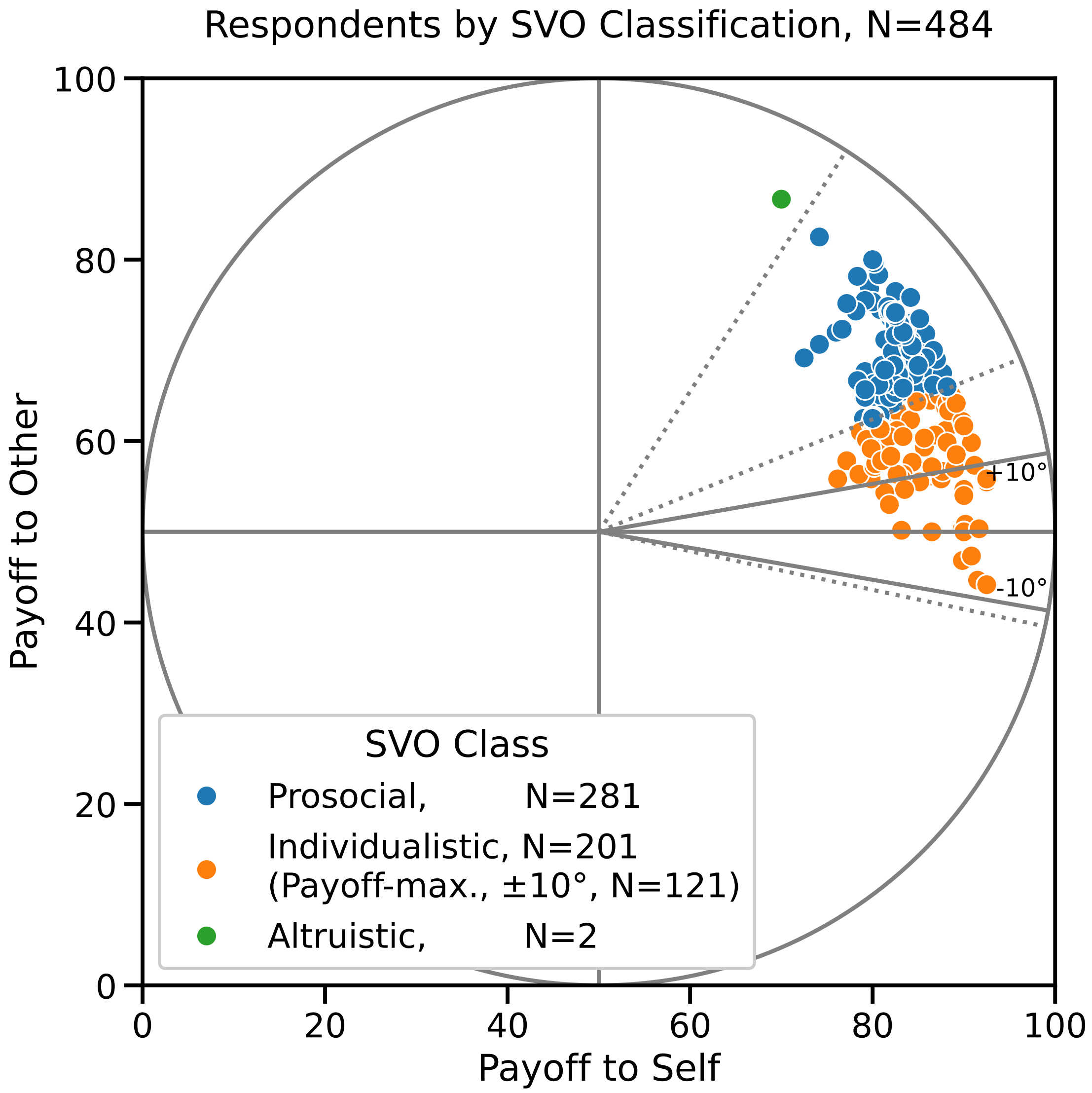


Figure 7: Social Value Orientation distributional data for Study 1c. Participants who passed comprehension checks.