


Acting *as if* drives cooperation among the purely self-interestedMatthew P. Cashman*¹ and Drazen Prelec¹

¹Sloan School of Management, Massachusetts Institute of Technology, 100 Main Street,
Cambridge, MA 02142

Author NoteMatthew Cashman  <https://orcid.org/0000-0002-1041-6128>

We wish to thank Adam Bear and Danica Mijovic-Prelec for comments on the manuscript. This research was supported by the MIT Sloan School of Management.

*Correspondence concerning this article should be addressed to Matthew Cashman, Sloan School of Management, Massachusetts Institute of Technology, 100 Main Street, Cambridge, MA 02142. E-mail: matt@cashman.science.

Abstract

Collective action often requires the private contribution of effort or money in a setting where any single person has a tiny influence on the target outcome, as in collective compensation schemes, voting, or social causes. Theoretical accounts of cooperation include prosocial motivation, norms and reputation, reciprocity, kinship, and cognitive heuristics like team thinking, all of which explain why people end up *wanting* to help others. We provide evidence for a different psychological mechanism, one which explains cooperation even among the self-interested: acting as if future decision-makers will mirror one's own unobserved action. In one-shot Public Goods Games where players move sequentially but do not observe others' moves, contributions to the public good are highest among those moving first and decline progressively, with last-movers making the smallest contributions. This pattern is consistent with players acting as if those yet to move will copy the move they have made. Three further results provide support for this interpretation: (1) this positional order effect is generated by players who are acting to maximize their own payoffs (as assessed by a series of Dictator games), (2) instructing players to maximize their own payoff increases the effect, and (3) the effect is eliminated if the moves of future players, but not those of past players, are determined randomly. Acting *as if* is relevant to firms using collective compensation schemes, who should consider this on top of classical economic incentives, and to policymakers who wish to encourage working for the greater good.

Keywords: Cooperation, Sequential Games, Public Goods Game, Game Theory, Collective Action, Pay-for-Performance

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

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 & Muthukrishna, 2021; Rand & 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, see Appendix A). 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 it is 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 among the most self-interested in addition to illuminating how such decisions are made. *Quasi-magical thinking* (Shafir & 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, nor is it a theory of social decision-making. 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 two additional stipulations. First, that this thinking is biased towards future, as-yet unmade moves, and second, that these moves are to be made by other *agents*, in particular. Here, we explore the behavior of specifically self-interested players who cooperate because they believe cooperation maximizes their individual payoff. We theorize that they make the assumption that those moving after them (who have not yet made a decision) will make the same move they have, and choose their own move to maximize their own payoffs conditional on others doing as they have done. In this case, where there is no way to influence others, they behave as if they can in fact 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 do this by first demonstrating that positional order effects exist *only* among the most self-interested players and, second, by eliciting and removing the effect via 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 her. We conclude by discussing several possible reasons acting *as if* may have arisen, and elaborating on practical implications and future directions for research.

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 & 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 et al., 2018). As N, the number of people in a bonus grouping, increases, the impact of any

given 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 her 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.

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 (1944 / 2004) developed extensive form notation based purely on preliminaryity 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 of a sequential version in which no moves are observed (see Figure 1). By the mid 1980s, there was a small chorus raising the question of whether ignoring timing is a good idea (Kohlberg & 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), for instance, reports public goods, resource dilemma, and coordination games that are sensitive to order of play alone.

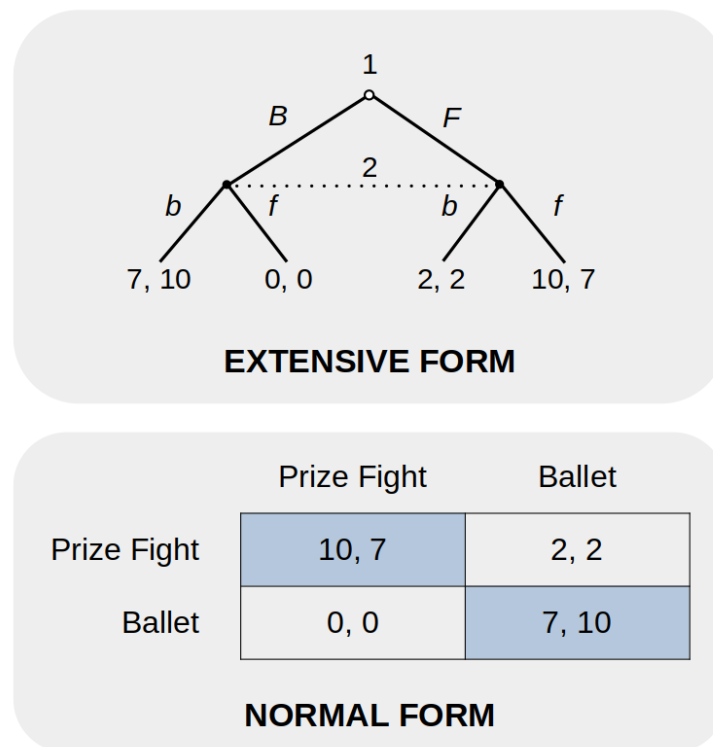


Figure 1

Battle of the Sexes game in Extensive and Normal Form. Note: The Extensive Form representation of the game captures the chronological ordering of moves, with Player 1 choosing B or F before Player 2 chooses b or f . The Normal Form representation of the same game disregards information about chronological ordering of moves, meaning a simultaneous-move version and a sequential-move version appear exactly the same.

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 et al., 1998; Rapoport, 1997; Weber et al., 2004), or they will “agree” that the first-mover should get the largest share of the gains in common-pool resource games (Budescu & Au, 2002; Budescu et al., 1995, 1997) and step-level PGGs (X.-P. Chen et al., 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, Quattrone and Tversky (1984) 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 has, of course, 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 and Tversky (1992) explore nonconsequential reasoning, reasoning about what to do 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 & Shafir, 1992). They call this pattern of events “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 & David, 2013); however, that work focuses on repeated tasks that do not generally involve other minds as games do. 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 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” and “open fates”. Closed fates uncertainty is uncertainty about a counterpart’s move when that move is not known to the player, but has already been made and is therefore fixed, whereas open fates uncertainty is uncertainty about a counterpart’s move when that move has yet to be made at all (or, perhaps, is presently being made) (Morris et al., 1998). Miller and Gunasegaram (1990) demonstrate 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 et al., 2012). Harris et al. (2011) document a surprising reversal: when an experimenter initiates a random process, adult participants prefer to guess the outcome *after* it has been determined. However, when the participants themselves start this process, they prefer to guess the outcome *before* it has

been determined. This suggests that causal thinking is naturally and sensibly directed towards the future. The distinction between closed fates and open fates is clearly relevant for behavior, but is little-explored in the context of decision-making.

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 & Grinberg, 2010; Morris et al., 1998; Shafir & 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 & Ehrhart, 2005; Figuières et al., 2012; Robinson et al., 2010; Steiger & 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 et al. (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 well-being. This proportion is reflected our own data as well. 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 standard assumption that players are trying to maximize profits.

Overview of 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 players tend to give less to the public good as their position in the sequence of players nears the end. In addition, we show this effect is driven by “individualistic” players who are interested in maximizing their own direct payoffs, as opposed to “prosocial” players whose behavior reflects other-regarding preferences. Prosocial players have reasons to cooperate that are entirely independent of order, so we would expect any order effects that are present to be much reduced. The work in Studies 1a - 2a 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-dictator game 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.¹ 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 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 Appendix B). Conversely, those who are individualistic (and therefore tend towards maximizing their own payoffs) might

¹ The standard battery produces a continuous SVO “angle” measure that allows for categorization 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) categories.

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 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—consistent with causal thinking about others being at play.

All studies are real-time, one-shot linear SPGGs with a multiplier of two (considering the PD as a two-person PGG with contribution of 0 or 100% of endowment). Participants contribute three main inputs: comprehension checks, game playing decisions, and predictions of the responses of other players. Apart from a base payment and proceeds from the game, correct answers to comprehension checks are directly incentivized and accurate predictions of others’ moves are incentivized based on the mean squared error between the actual and predicted value.

In all studies players participate in a brief text chat with their groupmates before learning about the task. The purpose of the chat is to assure participants that 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 except Study 1a have simultaneous-play PGG control conditions (see Appendix C), and all players pass

familiarization tasks and comprehension checks. All experiments except 1a² 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 is necessary that they move in order to finish the game. Later studies incorporate additional training and comprehension checks.

In total we tested 3686 participants distributed across five experiments. A convenience sample provided by Amazon Mechanical Turk (MTurk) was selected for Studies 1a, 1b and 1c because it was a reasonable approximation of American adults for our purposes. We selected CloudResearch’s filtered MTurk panel for Studies 2a and 2b because it provided among the highest-quality online panel data (Douglas et al., 2023; Hauser et al., 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.

Research Transparency Statement

We have made study materials, data, and analysis scripts available on osf.io. All studies except for Study 2a were preregistered on osf.io (see Appendix F for details), and all experiments we have conducted involving sequential games are reported here. All

² Study 1a’s comprehension check questions are similar, and are detailed in the relevant methods section, .

preregistrations include hypotheses, manipulations of the independent variables, measures for the dependent variables, desired sample size, exclusion criteria, and an analysis plan. All studies were approved by MIT's Committee on the Use of Humans as Experimental Subjects (COUHES) and comply with all relevant privacy and ethical regulations. We obtained electronic consent from all participants. The authors used artificial intelligence as part of an integrated development environment when writing the data analysis scripts, and the authors have no known competing interests to declare. All experiment software was written in the open-source oTree framework (D. L. Chen et al., 2016), and data analysis was conducted in python using statsmodels (Seabold & Perktold, 2010) for linear models. We will now detail each study.

Study 1a: Sequential Prisoner's Dilemma

Study 1a is a Sequential Prisoner's 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 as noted a Prisoner's Dilemma is formally equivalent to 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. It included a manipulation that involved mentalizing about the other player crossed with supplying information about behavior at the population level or not, as well as a standard sequential PD and matched control conditions.

Participants

2371 U.S.-based participants from Amazon Mechanical Turk completed the study. Of those, 45% (1075) passed the comprehension check questions. Among participants who passed the comprehension checks, mean total pay per participant (including bonuses for accurate predictions) is \$0.90 ($SD = 0.19$), yielding an hourly rate of \$10.20 at an average 6.0 minutes duration.

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)

Method

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)³. 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?

³ Since participants are grouped together for real-time experiments, 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.

5. What year is it?

Player 1 moves first, followed by Player 2. After making a 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 then exit the experiment and are paid.

Results

53% of players cooperated (SD=0.50) and the average payoff was \$0.58 (SD=0.18). First-movers cooperate more than second-movers. A logistic regression of Transfer decision on Order reaches significance (Odds Ratio = 0.751, 95% CI = [0.591, 0.955], $z = -2.333$, $p = 0.020$). 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 ($M_{teammate-population} = -2.25$), while those who transferred believe their teammate is more likely to transfer ($M_{teammate-population} = +2.76$) A linear regression shows a significant effect, $\beta = 5.01$, 95% CI=[3.211, 6.809], $F(1, 1073) = 29.874$, $p < 0.001$.

Discussion

Study 1a provides some evidence for the idea that first-movers would be more likely to cooperate than those moving afterwards despite the only difference between conditions being the knowledge of where one is in the sequence of moves. This is consistent with players acting as if others will do as they have done, and sparked the desire to investigate further. In addition, we see strong evidence that players view their own moves as indicative of the moves of others. The positional order effect is the main subject of the next study.

Study 1b: 3-Person Sequential Public Goods Game

Study 1b tests for a positional order effect alone, removing the additional manipulations found in Study 1a. We also sought to investigate whether any such effect is

driven by players who are trying to maximize their own payoffs, or by those who are keeping others' interests in mind in addition to their own.

Participants

1444 U.S.-based participants from Amazon Mechanical Turk completed the study. Of those 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, 782 were in the Sequential condition and 220 were in the Simultaneous condition. Mean total pay per participant (including bonuses for accurate predictions) was \$3.30 ($SD = 0.74$), yielding an hourly rate of \$19.94 ($SD = 7.62$) at 9.8 minutes average duration.

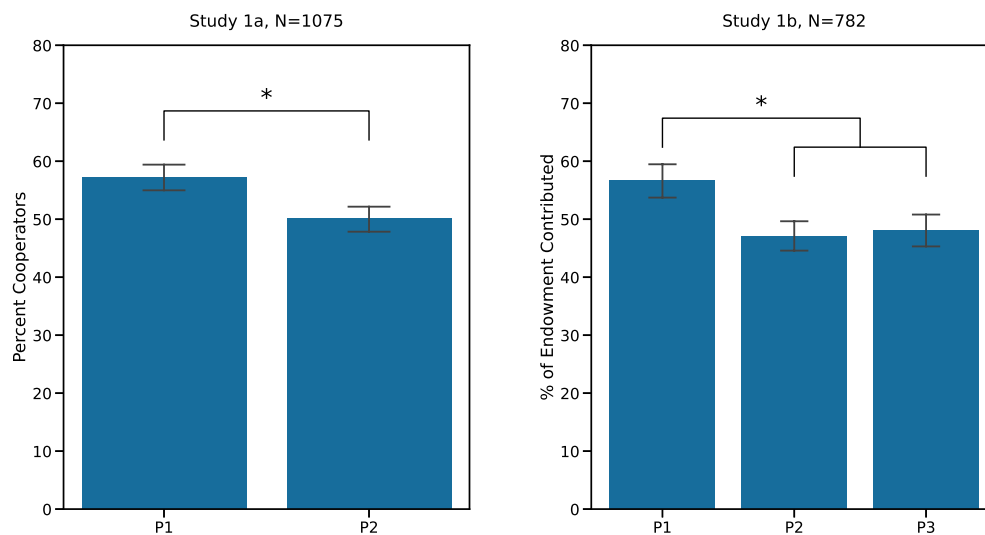


Figure 2

Studies 1a and 1b. Contributions to the public good fall with increasing order. Note: We observed a decline in contribution to the public good with increasing order in Studies 1a and 1b, which led to further investigation. Participants who passed comprehension checks, SEMs, * $p < 0.05$.

Method

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, once groups are formed, they are placed in a chat room for 30 seconds 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 afterwards 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 they are compensated based on the mean squared error between predictions and actual values, and then an SVO-dictator game slider battery (Murphy et al., 2011)⁴. Players complete some demographic questions, exit the experiment, and are

⁴ SVO is measured post-treatment, but we do not observe an effect of treatment on SVO, $N = 601$,

paid.

Results

Study 1b investigated how contribution to the public good varied with order of play, with the additional prediction that any order effects will be driven by individualistic players but not in prosocial players, as determined by an SVO battery. Almost all⁵ participants are classified as either individualistic or prosocial. The mean contribution for individualistic participants was 40.4% (SD=0.44) of the \$1.00 endowment, and for prosocials 66.1% (SD=0.41). 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. 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, $N = 782$, $\beta = -4.235$, 95% CI = [-8.139, -0.322], $p = 0.034$ ⁶. Participants classified as prosocial exhibit no significant differences in contribution levels as function of order, $N = 309$, $\beta = -1.449$, 95% CI = [-11.133, 8.487], $p = 0.774$, while we do see a difference among individualistic participants between first-mover data and grouped second- and third-mover contributions, $N = 290$, $\beta = -14.242$, 95% CI = [-24.729, -3.584], $p = 0.008$, see Figure 2. As hypothesized variation in contribution by order is driven by individualistic participants, but interestingly, third-mover contributions among individualistics were on average larger than those of second-movers (31.0%, SD=0.38 vs. 41.0%, SD=0.46), though the difference is not significant in a linear model, $N = 190$, $\beta = 9.994$, 95% CI = [-2.008,

$\beta = 1.027$, 95% CI = [-0.344, 2.414], $p = 0.145$. See Appendix E for the distribution.

⁵ One participant was classified as Altruistic, and one as Competitive. These participants' data are excluded from SVO analyses.

⁶ Contributions to the public good were distributed non-normally in all cases, with modes at 0%, 50%, and 100% of the endowment. When using OLS linear regressions in this and subsequent studies we bootstrap standard errors and p-values to make results robust to non-normal errors and so do not report F statistics.

22.044], $p = 0.101$.

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, $N = 1198$, $\beta = 0.074$, 95% CI = [-0.004, 0.151], $p = 0.067$ with cluster robust errors, but when applied to only individualistic players a significant equation is found, $N = 580$, $\beta = 0.172$, 95% CI = [0.059, 0.287], $p = 0.004$. There is no effect among prosocial players, $N = 618$, $\beta = 0.002$, 95% CI = [-0.116, 0.120], $p = 0.970$.

Discussion

Study 1b was designed to investigate pure order effects, and provided evidence that any such effects are driven by the actions of players who tend to maximize their own direct payouts. 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, with some variation. Further, 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, which is consistent with *as if* thinking among the self-interested. Study 1b provides evidence for a relationship between positional order effects and self-interest but is not definitive, so investigation with a more sensitive study was warranted. We continue with Study 1c, a refinement of this study.

Study 1c: 4-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 five from three, though only data from P2 - P5 are reported due to a technical error affecting P1.

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.

Method

Study 1c is a one-shot SPGG similar in structure to Study 1b but incorporating a number of improvements. It was designed for five players rather than the three of Study 1b, which allows for better resolution of order effects. The up-front chat was 60 instead of 30 seconds, allowing a little more time for a short conversation (30 seconds was usually only enough for greetings). Players also do not have a chat box while waiting for their group to fill; previously, chat time was waiting time + 30 seconds, which led to variable total chat times. Instead, in Study 1c players have a simple game on the wait screen. In addition, a player's place in the sequence is made much more salient. Players are shown their position in the sequence as they wait for others to move in addition to the simple game. Inputs are also changed to sliders with anchors that display the consequences (e.g., "KEEP FOR SELF: \$0.43 \leftrightarrow CONTRIBUTE TO FUND: \$0.57"), as opposed to simple numerical input boxes. These changes taken together were meant to provide a better experience for the participant.⁷

⁷ As with Study 1b, SVO is measured post-treatment with a battery of slider of dictator games, but we do not observe an effect of order on SVO angle, $N = 373$, $\beta = -0.495$, 95% CI = [-1.898, 0.900], $p = 0.491$.

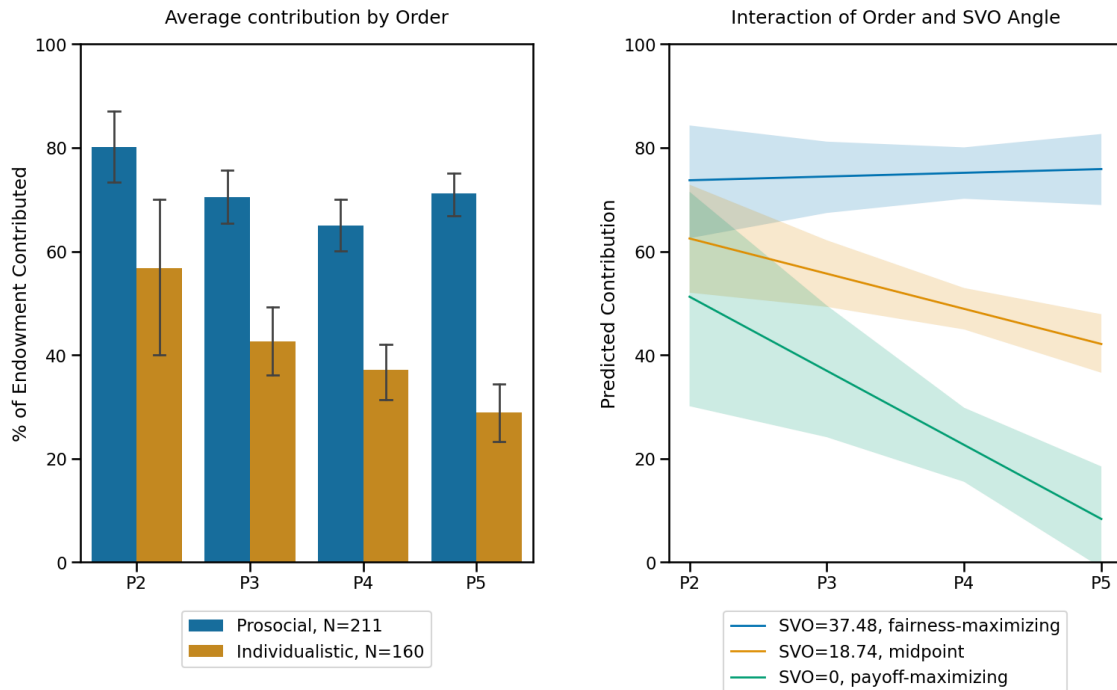


Figure 3

Study 1c. Positional order effects become stronger as participants become more self-interested. Note: [Left] We observe a decline in contributions with increasing order among individualistic, but not prosocial, participants. Participants who passed comprehension checks, SEMs. [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, 95% CIs.

Results

Study 1c provides further evidence that, among specifically self-interested players⁸, contribution to the public good declines as position in the order increases. A technical error meant that all first-movers and some later players were forced to make decisions too quickly, often timing out. We have excluded all affected players, which includes all first

⁸ Similar to Study 1b, nearly all participants were SVO classified as individualistic or prosocial; two were classified as altruistic, and these participants' data are excluded from categorical SVO analyses. See Appendix E for the distribution.

movers and a decreasing proportion of subsequent players (since response time allotted for a given player was a function of previous players' response times). All included players received the full allotted decision time. We believe the marginal benefit of having 5 vs. 4 players is minimal, so we exclude affected players and report results from 783 unaffected participants here, 484 of whom passed comprehension checks (373 in the Sequential treatment, and 111 in the Simultaneous treatment).

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, \$1.66 (SD=0.40) vs. prosocials \$1.48 (SD=0.38).

For Players 2-5 the preregistered linear regression (contribution \sim order) among SVO-individualistic players reaches significance: $N = 160$, $\beta = -7.928$, 95% CI = [-15.255, -0.331], $p = 0.040$. 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 \sim order*SVO_angle) we see $N = 373$, $\beta = 0.401$, 95% CI = [0.106, 0.685], $p = 0.009$ for the interaction. As expected, the effect of order on contribution becomes larger as SVO angle nears 0° (perfect self-interested play), see Figure 3.

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], $p = 0.002$.⁹

⁹ There are more backwards-facing predictions than forwards-facing because data from Player 1 was excluded, and Player 1 only produces forward-facing predictions.

Discussion

Study 1c further investigates the positional order effect among self-interested players that was hinted at by Studies 1a and 1b, and does so with a study that is better-suited to answering that specific question. 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.

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

Having good evidence 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 filling five-person groups with only self-interested people was not fast enough to be a viable strategy, so we explored alternatives. Study 2a tests whether merely *instructing* all comers to maximize personal payoffs generates a positional order effect similar to that observed among participants who arrive at the study already self-interested. 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.

Participants

197 U.S.-based participants from CloudResearch's filtered MTurk panel passed up-front bot checks and completed the study. Of those, 183 (93.0%) of those passed all of the up-front comprehension check questions and have their data included. The population was 37% female with an average age of 40. Mean total pay per participant (including bonuses for accurate predictions) was \$3.81 ($SD = 0.65$), yielding an hourly rate of \$14.25 at 16.6 minutes average duration.

Method

Study 2a served as a testbed for some important improvements to the 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.

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, $\text{contribution} \sim \text{order} * \text{instruct_or_no}$, detects the interaction effect, $N = 130$, $\beta = -15.745$, 95% CI = [-25.968, -5.303], $p = 0.005$.

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, which can be deployed when appropriate, rather than a stable feature of some subset of the population. Study 2a makes use of a small sample, and consequently there is substantial noise in estimates, but this study gave us enough confidence to deploy this technique in the next, larger experiment.

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

Study 2b incorporates the improvements from Study 2a and extends it in order to test the underlying psychological mechanism. It does this 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 she cannot see their moves.

Participants

747 U.S.-based participants from CloudResearch's filtered MTurk panel completed the study. Of those, 617 (83%) passed the up-front comprehension checks and have their data included in analyses (487 in the Sequential treatment, and 130 in the Simultaneous treatment). Among those, mean total pay per participant (including bonuses for accurate predictions) is \$4.75 ($SD = 0.67$), yielding a mean hourly rate of \$15.16 at 18.4 minutes average duration. To estimate the sample size required, we performed a power analysis via simulation using pilot data and data from previous experiments. Comprehension check pass rates and effect sizes from Cloud Research panels are larger relative to direct Mechanical Turk, meaning smaller sample sizes are required.

Method

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 Random Before and Random After, fully crossed with orders 1-5 and two simultaneous-play treatments. As in Study 2a, one simultaneous condition did not have any wait time, equivalent to moving first in the sequential game, while the other had a delay equivalent to the wait time 5th-movers experience in the sequential game. In the Random Before condition, players are told that everyone *before* them in the sequence has their decision about how much to contribute to the public good made by a random process, while in the Random After condition they are told that everyone *after* them has their decision made by a random process. Players are presented with a page that explains the setup, and are presented with symbols that make it clear which players' moves were randomly decided. They see the graphical representations in Figure 4 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 risks of using deception were warranted. Participants were debriefed at the end of the experiment.

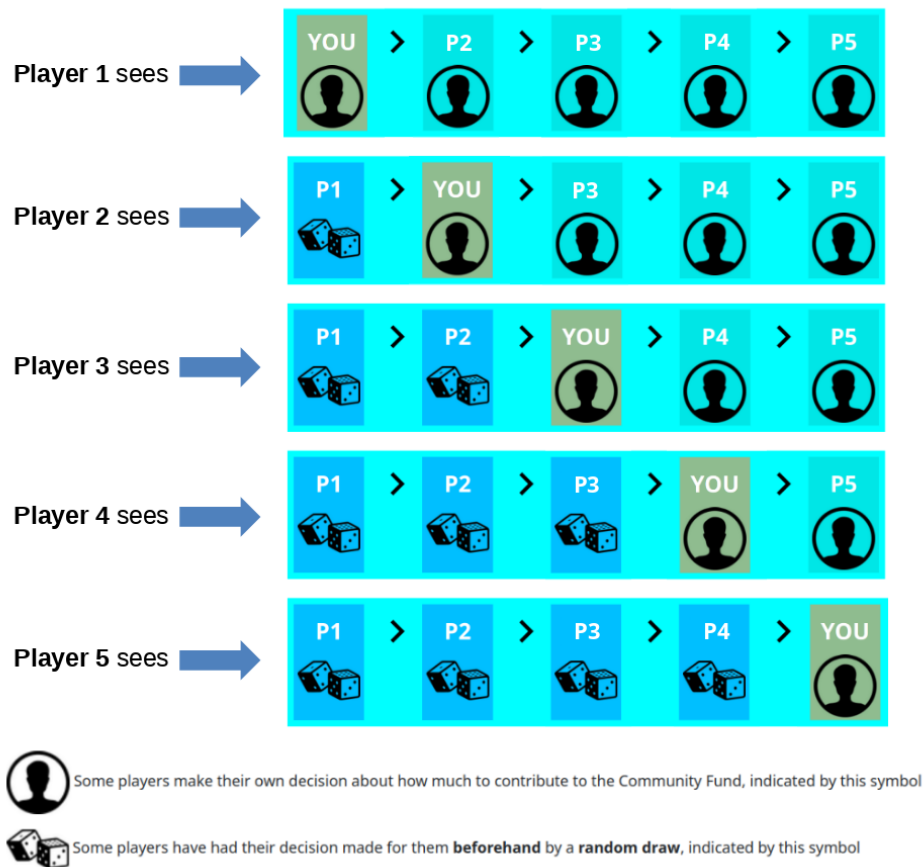


Figure 4

Study 2b. Stimuli for the Random Before condition. Note: 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.

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 35% of \$1.00 endowment (SD=0.43), and for Random After 34% (SD=0.42). Random Before and Random After participants left with the about same profit from the SPGG on average, \$1.40 (SD=0.41) for Random Before vs. \$1.44 (SD=0.39) for Random After.

Players who are told that everyone moving *before* them has her move determined randomly and everyone moving *after* them will decide on what move to make show a positional order effect. The preregistered linear regression contribution \sim order * random_before + wealth, differing from previous analyses in that it controls for a subjective measure of wealth, finds the effect, $N = 485$, $\beta = -8.005$, 95% CI = [-13.184, -2.769], $p = 0.004$; $\beta_{std} = -0.332$, 95% CI_{std} = [-0.547, -0.113] for the interaction (see Figure 5). 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. We see a marginally significant effect for wealth, $N = 485$, $\beta = -3.057$, 95% CI = [-6.225, 0.133], $p = 0.063$, indicating that as self-reported feelings of wealth increase, contributions to the public good decrease. The regression without wealth also finds the effect, $N = 487$, $\beta = -8.160$, 95% CI = [-13.277, -2.838], $p = 0.003$. When we restrict the main analysis to only those players who passed a second set of comprehension checks at the end of the experiment (80.1% of players who passed the initial checks), we observe a larger effect size when comparing standardized coefficients ($N = 403$, $\beta = -9.369$, 95% CI = [-14.802, -3.877], $p = 0.001$; $\beta_{std} = -0.401$, 95% CI_{std} = [-0.631, -0.167] for the preregistered analysis). This gives us reason to believe that the effect is concentrated among participants who understand the rules of the game best. 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. 75.5% of participants

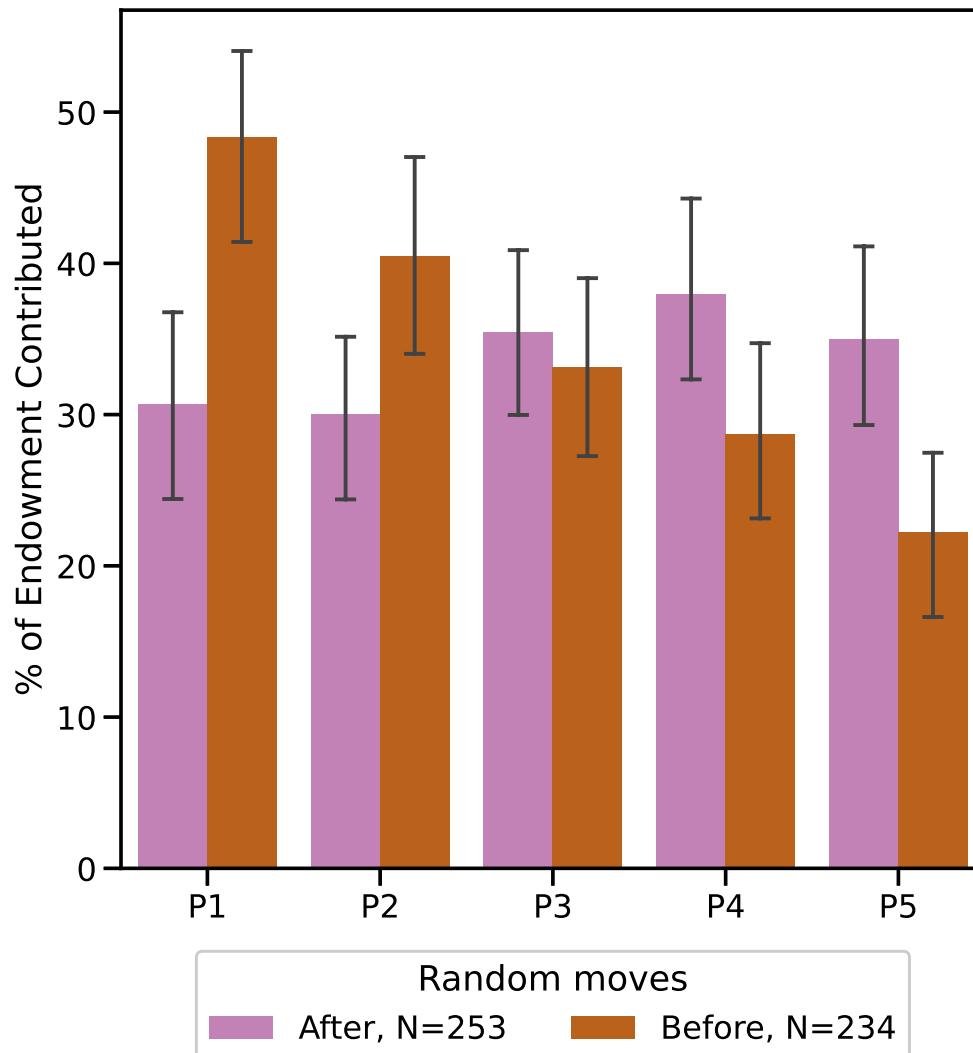


Figure 5

Study 2b. The order effect appears when random moves are before, not after, the focal player. Note: 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.

give either 0 or 100% of their endowment, and among these participants effect size increases relative to the overall analysis ($N = 364$, $\beta = -9.391$, 95% CI = [-15.937, -2.931],

$p = 0.004$; $\beta_{std} = -0.354$, 95% $CI_{std} = [-0.601, -0.111]$ for the preregistered analysis). See Appendix D for a comparison of these subsets. We do not observe a difference between the no delay (timing equivalent to P1, mean contribution = \$0.38) and long delay (80 seconds, timing equivalent to P5, mean contribution = \$0.47) simultaneous-move control conditions using a T test, $t(128) = -1.08$, $p = 0.282$. This shows 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 $\sim \text{contribution} * \text{future/past}$ yields $N = 932$, $\beta = 0.366$, 95% $CI = [0.254, 0.480]$, $p < 0.001$, and for Random After $N = 1004$, $\beta = -0.451$, 95% $CI = [-0.553, -0.348]$, $p < 0.001$ with cluster robust 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.

Discussion

Players who are trying to maximize their own payoffs adjust their contributions to a public good downwards as their place in a sequence of players approaches the end, 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 had their moves made randomly for them. It appears that the positional order effect in SPGGs requires that players believe there are other people who will freely make a choice involved in the game, and that those people have not yet made their decision. 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 54% of the endowment across all participants, prosocial and individualistic, whereas mean contribution among Random After participants here is 37%. Player 5 in the Random After condition has no random moves to contemplate at all, since she moves last, and therefore plays a standard 5-person SPGG—and contributes considerably less than non-instructed populations, confirming the efficacy of the manipulation.

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. Among players who are maximizing personal profit in the game, 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, and the fact that the effect is stronger in the direction of open fates (towards the future) 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. Critically, this is not *conditional* cooperation (Thöni & Volk, 2018) since there is nothing for a player to condition on in a one-shot sequential game with no observation, nor can reputation come into play in an anonymous online context. 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*—and for the same reason, an implied causal connection. 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, and choose their own move to maximize their own profits conditional on everyone yet to move doing as they have done. This implies a sharp step between 100% contribution and 0% contribution, which is observed in the data. 75.5% of participants contribute either 0 or 100% of their endowment, and the positional order effect is stronger in this subset. The effect is also stronger in the 80.1% of participants who pass both pre- and post-comprehension checks, implying the effect gets stronger as players better-understand the game (see Appendix D for analyses of these subsets). A formalization of this model is included in Appendix B.

Our theory predicts that some players, such as highly prosocial participants and individualistic participants who are at the beginning of a sequence, ought to 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 fully understand the game will pass comprehension checks as well (Andreoni, 1995; Houser & Kurzban, 2002), leading to noise in responses. Indeed, Kümmerli et al. (2010) report results from PGG experiments that explicitly incentivize full cooperation—and which never reach it. Some participants will also evince different levels of prosociality in the SPGG vs. in the SVO measure, leading to inconsistent behavior. Just so, Ackermann and Murphy (2019) report being able to explain more than 50% of variance in contribution decisions using repeated measures of social preferences and beliefs about others—but not much more than 50%. The sources of noise typical for PGGs and similar games suggest that we would see

attenuated effects relative to the pure strategy.

Rationality and adaptivity

Acting *as if* may fit the data we observe as a proximal psychological mechanism, 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 herself and other players her 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 & Thomadsen, 2005; Bodner & Prelec, 2003; Dhar & Wertenbroch, 2012; Mijovic-Prelec & 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 latent 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 & Harper, 1978), Subjective Expected Relative Similarity (Fischer & Savranovski, 2023), Superrationality (Hofstadter, 1983), and Newcomb’s problem (Lewis, 1979; Nozick, 1969; Wolpert & Benford, 2013). These theories all, to some degree, try to make sense of doing the thing that *correlates* with success, rather than the thing that *causes* success (causes—on some model of the world). It may be the case that defecting in a PGG is payoff-maximizing, all else equal, but for a payoff-maximizing player to defect for the right reasons requires a remarkable commitment to the correctness of one’s mental model of the game and the world immediately surrounding it. We often, even usually, lack such confidence in our understanding of a given situation. Even if we had the confidence, we might not have the computational resources to make use of that understanding. On this view, acting *as if* presents a psychological mechanism which can be useful given this type of uncertainty and paucity of resources, manifesting a decision where the “best” decision is out of reach. It is computationally cheap (everyone who hasn’t moved simply does what you do), and given a population who all act the same it will, in fact, be payoff-maximizing.

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 behavior may not, in fact, be specific to the moral domain; universalization could be a special case of the more general *as if* 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 towards the future, potentially making any of self-signaling, social projection, or universalization viable explanations for why acting *as if* might manifest.

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 also more similarity among players in a small-scale society, 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 that acting *as if* is payoff-maximizing 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

¹⁰ cognitive machinery that must be painstakingly inculcated, such as moral universals (e.g., "Stealing is wrong."). It remains for future work to tease apart these possibilities.

¹⁰ And still not universally-held, (Henrich et al., 2023)

Broader Implications, Future Directions, and Conclusion

Finally, whatever the adaptive story, understanding why even the most self-interested actors might decide to contribute to the public good is relevant to many managerial and policy decisions. This is particularly true in organizational contexts, where collective action is a requirement. For example, an employee might say to herself: many other people will be in my shoes in the future, so if I work late 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 the agency of future decision-makers. We might remind the self-interested of how other people like them will be deciding to contribute—or not—at a later time. Though our data show only a subset of the population is clearly trying to act in their own interests, this subset is the source of frustration for managers and policymakers since others are willing to act in the interests of the group to begin with. Interventions that affect specifically those who are *not* already working for the greater good are, indeed, an important goal for public policy and similar applications. But there is no need to apply any intervention to just those who are acting in their own interests in a given moment, which would require identifying them in the first place. In the case where we remind everyone that there will be other people in their shoes in the future, making the same decision, we would expect no effect on individuals who are already prosocial and acting *as if* among the self-interested. Combining a reminder that others will be making the same decision in the future with a steer towards more effective choices for those who are already trying to help others may be an effective strategy, and we leave empirical validation for future work. Similarly, this research is limited to one-shot sequential games, and the space of tasks where behavior is consistent with acting *as if* could be further explored. Future research should delineate the boundary conditions of acting *as if*, exploring uncertainty, higher stakes, different multipliers, and different tasks

altogether. The extent to which acting *as if* generalizes across different organization types, management structures, social groups, and cultures also remains to be investigated, and we would expect significant variation within and between individuals and larger groups.

Acting *as if* may be one reason why self-interested people cooperate, and encouraging it could pay off, but people who do this are nevertheless in error from a self-interested point of view. All else being equal, 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 people acting *as if* would flourish compared to one composed of people pursuing their self-interest according to normative decision theory. Given a world where collective action problems abound, from a team's performance within an organization to climate change, understanding the psychological foundations of cooperation among the self-interested could offer new approaches for addressing these challenges.

References

- Abele, S., & Ehrhart, K.-M. (2005). The timing effect in public good games. *Journal of Experimental Social Psychology, 41*(5), 470–481. <https://doi.org/10/bkgq9d>
- Ackermann, K. A., & Murphy, R. O. (2019). Explaining cooperative behavior in public goods games: How preferences and beliefs affect contribution levels. *Games, 10*(1), 15. <https://doi.org/10.3390/g10010015>
- Andreoni, J. (1995). Cooperation in public-goods experiments: Kindness or confusion? *American Economic Review, 85*(4), 891–904. Retrieved February 3, 2025, from https://econpapers.repec.org/article/aeaaecrev/v_3a85_3ay_3a1995_3ai_3a4_3ap_3a891-904.htm
- Bernheim, B. D., & Thomadsen, R. (2005). Memory and anticipation. *The Economic Journal, 115*(503), 271–304. <https://doi.org/10.1111/j.1468-0297.2005.00989.x>
- Bodner, R., & Prelec, D. (2003, February 13). Self-signaling and diagnostic utility in everyday decision making. In I. Brocas & J. D. Carrillo (Eds.), *The psychology of economic decisions* (pp. 105–124). Oxford University Press. <https://doi.org/10.1093/oso/9780199251063.003.0006>
- Budescu, D. V., & Au, W. T. (2002). A model of sequential effects in common pool resource dilemmas. *Journal of Behavioral Decision Making, 15*(1), 37–63. <https://doi.org/10.1002/bdm.402>
- Budescu, D. V., Au, W. T., & Chen, X.-P. (1997). Effects of protocol of play and social orientation on behavior in sequential resource dilemmas. *Organizational Behavior and Human Decision Processes, 69*(3), 179–193. <https://doi.org/10.1006/obhd.1997.2684>
- Budescu, D. V., Suleiman, R., & Rapoport, A. (1995). Positional order and group size effects in resource dilemmas with uncertain resources. *Organizational Behavior and Human Decision Processes, 61*(3), 225–238. <https://doi.org/10.1006/obhd.1995.1018>

- Burns, Z. C., Caruso, E. M., & Bartels, D. M. (2012). Predicting premeditation: Future behavior is seen as more intentional than past behavior. *Journal of Experimental Psychology: General*, *141*(2), 227–232. <https://doi.org/10/bz6ngt>
- Carpenter, J., Robbett, A., & Akbar, P. A. (2018). Profit sharing and peer reporting. *Management Science*, *64*(9), 4261–4276. <https://doi.org/10.1287/mnsc.2017.2831>
- Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97. <https://doi.org/10/bj42>
- Chen, X.-P., Au, W. T., & Komorita, S. (1996). Sequential choice in a step-level public goods dilemma: The effects of criticality and uncertainty. *Organizational Behavior and Human Decision Processes*, *65*(1), 37–47. <https://doi.org/10.1006/obhd.1996.0003>
- Cooper, R., DeJong, D. V., Forsythe, R., & Ross, T. W. (1993). Forward induction in the battle-of-the-sexes games. *The American Economic Review*, *83*(5), 1303–1316. Retrieved November 24, 2021, from <https://www.jstor.org/stable/2117562>
- Croson, R. T. (1999). The disjunction effect and reason-based choice in games. *Organizational Behavior and Human Decision Processes*, *80*(2), 118–133. <https://doi.org/10/fhrh8g>
- Daley, B., & Sadowski, P. (2017). Magical thinking: A representation result. *Theoretical Economics*, *12*(2), 909–956. <https://doi.org/10/f99g8r>
- Dawes, R. M. (1989). Statistical criteria for establishing a truly false consensus effect. *Journal of Experimental Social Psychology*, *25*(1), 1–17. [https://doi.org/10.1016/0022-1031\(89\)90036-X](https://doi.org/10.1016/0022-1031(89)90036-X)
- Delfgaauw, J., Dur, R., Onemu, O., & Sol, J. (2022). Team incentives, social cohesion, and performance: A natural field experiment. *Management Science*, *68*(1), 230–256. <https://doi.org/10.1287/mnsc.2020.3901>

- Dhar, R., & Wertenbroch, K. (2012). Self-signaling and the costs and benefits of temptation in consumer choice. *Journal of Marketing Research*, *49*(1), 15–25.
<https://doi.org/10/bbng3z>
- Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, prolific, CloudResearch, qualtrics, and SONA. *PLOS ONE*, *18*(3), e0279720. <https://doi.org/10.1371/journal.pone.0279720>
- Figuières, C., Masclet, D., & Willinger, M. (2012). Vanishing leadership and declining reciprocity in a sequential contributions experiment. *Economic Inquiry*, *50*(3), 567–584. <https://doi.org/10.1111/j.1465-7295.2011.00415.x>
- Fischer, I., & Savranovski, L. (2023). The effect of similarity perceptions on human cooperation and confrontation. *Scientific Reports*, *13*(1), 19849.
<https://doi.org/10.1038/s41598-023-46609-8>
- Gibbard, A., & Harper, W. L. (1978). Counterfactuals and two kinds of expected utility. In W. L. Harper, R. Stalnaker, & G. Pearce (Eds.), *Ifs: Conditionals, belief, decision, chance and time* (pp. 153–190). Springer Netherlands.
https://doi.org/10.1007/978-94-009-9117-0_8
- Güth, W., Huck, S., & Rapoport, A. (1998). The limitations of the positional order effect: Can it support silent threats and non-equilibrium behavior? *Journal of Economic Behavior & Organization*, *34*(2), 313–325.
[https://doi.org/10.1016/S0167-2681\(97\)00057-7](https://doi.org/10.1016/S0167-2681(97)00057-7)
- Harris, A. J., Rowley, M. G., Beck, S. R., Robinson, E. J., & McColgan, K. L. (2011). Agency affects adults', but not children's, guessing preferences in a game of chance. *Quarterly Journal of Experimental Psychology*, *64*(9), 1772–1787.
<https://doi.org/10.1080/17470218.2011.582126>
- Hauser, D. J., Moss, A. J., Rosenzweig, C., Jaffe, S. N., Robinson, J., & Litman, L. (2023). Evaluating CloudResearch's approved group as a solution for problematic data

- quality on MTurk. *Behavior Research Methods*, 55(8), 3953–3964.
<https://doi.org/10.3758/s13428-022-01999-x>
- Henrich, J., Blasi, D. E., Curtin, C. M., Davis, H. E., Hong, Z., Kelly, D., & Kroupin, I. (2023). A cultural species and its cognitive phenotypes: Implications for philosophy. *Review of Philosophy and Psychology*, 14(2), 349–386.
<https://doi.org/10.1007/s13164-021-00612-y>
- Henrich, J., & Muthukrishna, M. (2021). The origins and psychology of human cooperation. *Annual Review of Psychology*, 72(1), 207–240.
<https://doi.org/10.1146/annurev-psych-081920-042106>
- Hoch, S. J. (1987). Perceived consensus and predictive accuracy: The pros and cons of projection. *Journal of Personality and Social Psychology*, 53(2), 221–234.
<https://doi.org/10.1037/0022-3514.53.2.221>
- Hofstadter, D. R. (1983). Metamagical themas. *Scientific American*, 248(6), 14–29.
Retrieved August 31, 2023, from
<http://www.jstor.org.libproxy.mit.edu/stable/24968913>
- Houser, D., & Kurzban, R. (2002). Revisiting kindness and confusion in public goods experiments. *American Economic Review*, 92(4), 1062–1069.
<https://doi.org/10.1257/00028280260344605>
- Hristova, E., & Grinberg, M. (2010). Testing two explanations for the disjunction effect in prisoner’s dilemma games: Complexity and quasi-magical thinking. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 32.
<https://escholarship.org/uc/item/3t20s2q7>
- Knez, M., & Simester, D. (2001). Firm-wide incentives and mutual monitoring at continental airlines. *Journal of Labor Economics*, 19(4), 743–772.
<https://doi.org/10.1086/322820>
- Kohlberg, E., & Mertens, J.-F. (1986). On the strategic stability of equilibria. *Econometrica*, 54(5), 1003–1037. <https://doi.org/10.2307/1912320>

- Kreps, D. M. (1990). *Game theory and economic modelling*. Oxford University Press.
- Kümmerli, R., Burton-Chellew, M. N., Ross-Gillespie, A., & West, S. A. (2010). Resistance to extreme strategies, rather than prosocial preferences, can explain human cooperation in public goods games. *Proceedings of the National Academy of Sciences*, *107*(22), 10125–10130. <https://doi.org/10.1073/pnas.1000829107>
- Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, *32*(2), 311–328. <https://doi.org/10.1037/0022-3514.32.2.311>
- Levine, S., Kleiman-Weiner, M., Schulz, L., Tenenbaum, J., & Cushman, F. (2020). The logic of universalization guides moral judgment. *Proceedings of the National Academy of Sciences*, *117*(42), 26158–26169. <https://doi.org/10.1073/pnas.2014505117>
- Lewis, D. (1979). Prisoners' dilemma is a newcomb problem. *Philosophy & Public Affairs*, *8*(3), 235–240. <http://www.jstor.org/stable/2265034>
- Luce, D. R. (1992). Where does subjective expected utility fail descriptively? *Journal of Risk and Uncertainty*, *5*(1), 5–27. <https://doi.org/10.1007/BF00208784>
- Masel, J. (2007). A bayesian model of quasi-magical thinking can explain observed cooperation in the public good game. *Journal of Economic Behavior & Organization*, *64*(2), 216–231. <https://doi.org/10.1016/j.jebo.2005.07.003>
- Mijovic-Prelec, D., & Prelec, D. (2010). Self-deception as self-signalling: A model and experimental evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1538), 227–240. <https://doi.org/10/c2tqv2>
- Miller, D. T., & Gunasegaram, S. (1990). Temporal order and the perceived mutability of events: Implications for blame assignment. *Journal of Personality and Social Psychology*, *59*(6), 1111–1118. <https://doi.org/10.1037/0022-3514.59.6.1111>
- Morris, M. W., Sim, D. L., & Girotto, V. (1998). Distinguishing sources of cooperation in the one-round prisoner's dilemma: Evidence for cooperative decisions based on the

- illusion of control. *Journal of Experimental Social Psychology*, *34*(5), 494–512.
<https://doi.org/10/d4cs3w>
- Murphy, R. O., Ackermann, K. A., & Handgraaf, M. (2011). Measuring social value orientation. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.1804189>
- Nozick, R. (1969). Newcomb's problem and two principles of choice. In N. Rescher (Ed.), *Essays in honor of carl g. hempel* (pp. 114–146). Springer Netherlands.
https://doi.org/10.1007/978-94-017-1466-2_7
- Nyberg, A. J., Maltarich, M. A., Abdulsalam, D. “, Essman, S. M., & Cragun, O. (2018). Collective pay for performance: A cross-disciplinary review and meta-analysis. *Journal of Management*, *44*(6), 2433–2472.
<https://doi.org/10.1177/0149206318770732>
- Quattrone, G. A., & Tversky, A. (1984). Causal versus diagnostic contingencies: On self-deception and on the voter's illusion. *Journal of Personality and Social Psychology*, *46*(2), 237–248. <https://doi.org/10/dr4gxj>
- Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in Cognitive Sciences*, *17*(8), 413–425. <https://doi.org/10.1016/j.tics.2013.06.003>
- Rapoport, A. (1997). Order of play in strategically equivalent games in extensive form. *International Journal of Game Theory*, *26*(1), 113–136.
<https://doi.org/10.1007/BF01262516>
- Robinson, A. E., Sloman, S. A., Hagmayer, Y., & Hertzog, C. K. (2010). Causality in solving economic problems. *The Journal of Problem Solving*, *3*(1).
<https://doi.org/10/ggdjxn>
- Roemer, J. E. (2010). Kantian equilibrium. *The Scandinavian Journal of Economics*, *112*(1), 1–24. <https://doi.org/10.1111/j.1467-9442.2009.01592.x>
- Roemer, J. E. (2015). Kantian optimization: A microfoundation for cooperation. *Journal of Public Economics*, *127*, 45–57. <https://doi.org/10.1016/j.jpubeco.2014.03.011>

- Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. *9th Python in Science Conference*.
- Shafir, E., & Tversky, A. (1992). Thinking through uncertainty: Nonconsequential reasoning and choice. *Cognitive Psychology*, *24*(4), 449–474.
<https://doi.org/10/d6thrq>
- Stefan, S., & David, D. (2013). Recent developments in the experimental investigation of the illusion of control. a meta-analytic review: A meta-analysis of the illusion of control. *Journal of Applied Social Psychology*, *43*(2), 377–386.
<https://doi.org/10.1111/j.1559-1816.2013.01007.x>
- Steiger, E.-M., & Zultan, R. (2014). See no evil: Information chains and reciprocity. *Journal of Public Economics*, *109*, 1–12. <https://doi.org/10.1016/j.jpubeco.2013.10.006>
- Tarantola, T., Kumaran, D., Dayan, P., & De Martino, B. (2017). Prior preferences beneficially influence social and non-social learning. *Nature Communications*, *8*(1), 817. <https://doi.org/10.1038/s41467-017-00826-8>
- Thöni, C., & Volk, S. (2018). Conditional cooperation: Review and refinement. *Economics Letters*, *171*, 37–40. <https://doi.org/10.1016/j.econlet.2018.06.022>
- Tversky, A., & Shafir, E. (1992). The disjunction effect in choice under uncertainty. *Psychological Science*, *3*(5), 305–310.
<https://doi.org/10.1111/j.1467-9280.1992.tb00678.x>
- Von Neumann, J., & Morgenstern, O. (2004). *Theory of games and economic behavior* (60th anniversary ed). Princeton University Press. (Original work published 1944)
- Weber, R. A., Camerer, C. F., & Knez, M. (2004). Timing and virtual observability in ultimatum bargaining and “weak link” coordination games. *Experimental Economics*, *7*, 25–48.
- Wolpert, D. H., & Benford, G. (2013). The lesson of newcomb’s paradox. *Synthese*, *190*(9), 1637–1646. <https://doi.org/10.1007/s11229-011-9899-3>

Zelmer, J. (2003). Linear public goods experiments: A meta-analysis. *Experimental Economics*, 6(3), 299–310. <https://doi.org/10.1023/A:1026277420119>

Appendix 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:

$$p = \frac{mc}{n} + e(1 - a) \quad (\text{A1})$$

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.

Appendix B

Model

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

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, $i = 1, \dots, n$. Let a_i denote the contribution of player i , $0 \leq a_i \leq 1$, and p_i the payoff to player i .

$$p_i = 1 - a_i + \frac{m}{n} \sum_{k=1}^n a_k \quad (\text{B1})$$

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 :

$$u_i(a_1, \dots, a_n) = (1 - s_i)p_i + s_i m a_i \quad (\text{B2})$$

where p_i is determined by the game formula, (B1). 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 her contribution to the public good, $u_i = m a_i$. 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 can: (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.

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:

$$a_i = 1 \iff \mathbb{E}[u_i \mid a_i = 1, s_i] > \mathbb{E}[u_i \mid a_i = 0, s_i] \quad (\text{B3})$$

A player knows the value of their prosocial parameter and hence also knows the utility function in (B1). If she were just a spectator, not making a decision, her 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, with an intercept b and slope c :

$$\mathbb{E}[a_k \mid s_i] = b + cs_i \quad (\text{B4})$$

Prosocial players are more optimistic about the overall contribution level, all 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 i . We assume:

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

where $\mathbb{E}[a_k \mid s_i] = b + cs_i$ from (B4) is substituted in the final line.

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

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

There is perceived causality with respect to future players, proportional to the *as if* influence parameter d :

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

The decision criterion in (B3) can be expressed as:

$$\begin{aligned} \mathbb{E}[u_i | a_i = 1, s_i] - \mathbb{E}[u_i | a_i = 0, s_i] &= (1 - s_i)\mathbb{E}[p_i | a_i = 1, s_i] + s_i m - (1 - s_i)\mathbb{E}[p_i | a_i = 0, s_i] \\ &= (1 - s_i)(\mathbb{E}[p_i | a_i = 1, s_i] - \mathbb{E}[p_i | a_i = 0, s_i]) + s_i m \\ &= (1 - s_i) \left(-1 + \frac{m}{n} \mathbb{E} \left[\sum_{k=1}^n a_k | a_i = 1, s_i \right] - \frac{m}{n} \mathbb{E} \left[\sum_{k=1}^n a_k | a_i = 0, s_i \right] \right) + s_i m \end{aligned} \quad (\text{B5})$$

where the first line follows from (B2) and the third line from (B3).

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

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

Substituting into the criterion,

$$\mathbb{E}[u_i | a_i = 1, s_i] - \mathbb{E}[u_i | a_i = 0, s_i] = (1 - s_i) \left(-1 + \frac{m}{n} (1 + d(n - i)) \right) + s_i m. \quad (\text{B6})$$

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

$$\mathbb{E}[u_i | a_i = 1, s_i] - \mathbb{E}[u_i | a_i = 0, s_i] = 0 \iff d^*(i) = \frac{-m - smn + n}{m(n - i)} \quad (\text{B7})$$

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 (B7)) is steeper if s_i is smaller, that is, if players are more self-interested. To show this, we differentiate:

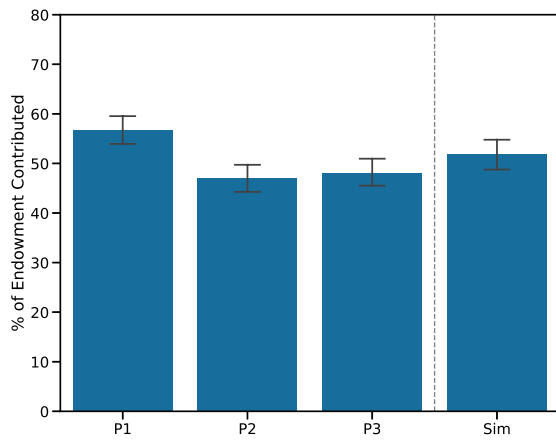
$$\frac{dd^*(i)}{di} = \frac{1}{(n-i)^2} \left(\frac{n-m}{m} - \frac{s_i}{(1-s_i)}n \right)$$

which is decreasing in s_i . This is the hypothesized interaction of order and prosociality.

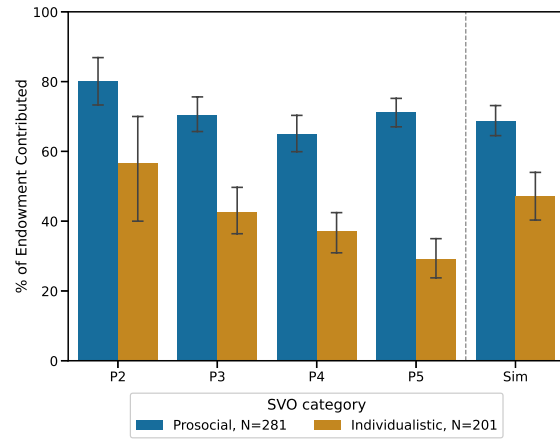
Less prosocial players will exhibit a stronger effect. Conversely, the positional order effect should disappear if s_i is sufficiently high.

Appendix C

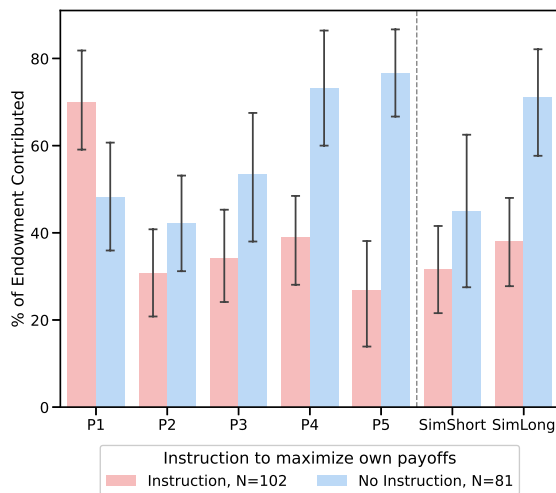
Simultaneous-move data



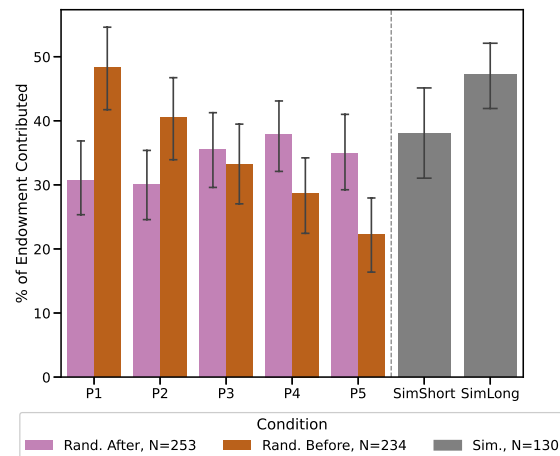
(a) Study 1b: results with simultaneous condition



(b) Study 1c: results with simultaneous condition



(c) Study 2a: results with simultaneous conditions



(d) Study 2b: results with simultaneous conditions

Figure C1

Data from simultaneous-move conditions that were present in Studies 1b, 1c, 2a, and 2b.

Studies 1b, 1c, 2a, and 2b included simultaneous-move conditions, conditions where participants were not assigned to a position in a sequence but rather moved all at once as

in a standard PGG. The nature of the random moves before vs. after manipulation in Study 2b meant it was not possible to treat simultaneous-move participants (since it is not possible to have moves be either before or after if everyone moves at once). In general, the contributions of self-interested simultaneous-movers are similar to those of self-interested players moving in the middle of a sequence, perhaps reflecting the intermediate state between “closed” and “open” fates represented by “right now”. Prosocial simultaneous-movers show little difference from prosocial sequential players at any position. Simultaneous-move contribution data is reported in Figure C1

Appendix D

Study 2b subsets

Overall data

Study 2b results for the total data set are as follows, for convenience:

Positional order effects

The preregistered linear regression contribution \sim order * random_before + wealth finds the effect,

$N = 485$, $\beta = -8.005$, 95% CI = [-13.184, -2.769], $p = 0.004$; $\beta_{std} = -0.332$, 95% CI_{std} = [-0.547, -0.113] for the interaction

We also find a significant equation not controlling for wealth,

$N = 487$, $\beta = -8.160$, 95% CI = [-13.277, -2.838], $p = 0.003$; $\beta_{std} = -0.338$, 95% CI_{std} = [-0.552, -0.118]

Predictions of others' moves by own move, forwards vs. backwards

The preregistered linear regression predicted_value \sim contribution * binary_position finds the effect with cluster robust errors,

Participants in the Random Before condition

$N = 932$, $\beta = 0.366$, 95% CI = [0.254, 0.480], $p < 0.001$

Participants in the Random After condition

$N = 1004$, $\beta = -0.451$, 95% CI = [-0.553, -0.348], $p < 0.001$

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 her 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 *as if* 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*.

Positional order effects

Among the 75.5% of participants give either 0 or 100% of their endowment, the preregistered linear regression $\text{contribution} \sim \text{order} * \text{random_before} + \text{wealth}$ finds the effect,

$N = 364$, $\beta = -9.391$, 95% CI = [-15.937, -2.931], $p = 0.004$; $\beta_{std} = -0.354$, 95% CI_{std} = [-0.601, -0.111] for the interaction

We also find a significant equation not controlling for wealth,

$N = 365$, $\beta = -9.588$, 95% CI = [-16.121, -3.087], $p = 0.003$; $\beta_{std} = -0.361$, 95% CI_{std} = [-0.608, -0.117]

Predictions of others' moves by own move, forwards vs. backwards

The preregistered linear regression $\text{predicted_value} \sim \text{contribution} * \text{binary_position}$ finds the effect with cluster robust errors,

Participants in the Random Before condition

$N = 704$, $\beta = 0.351$, 95% CI = [0.226, 0.474], $p < 0.001$

Participants in the Random After condition

$N = 748$, $\beta = -0.491$, 95% CI = [-0.608, -0.376], $p < 0.001$

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.

Positional order effects

Among participants who passed a second set of comprehension checks at the end of the experiment (80.1% of those who passed the initial checks), the preregistered linear regression contribution \sim order * random_before + wealth finds the effect,

$N = 403$, $\beta = -9.369$, 95% CI = [-14.802, -3.877], $p = 0.001$; $\beta_{std} = -0.401$, 95% CI_{std} = [-0.631, -0.167] for the interaction

We also find a significant equation not controlling for wealth,

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

$N = 403$, $\beta = -9.504$, 95% CI = [-14.944, -3.988], $p = 0.001$; $\beta_{std} = -0.407$, 95% CI_{std} = [-0.638, -0.170]

Predictions of others' moves by own move, forwards vs. backwards

The preregistered linear regression $\text{predicted_value} \sim \text{contribution} * \text{binary_position}$ finds the effect with cluster robust errors,

Participants in the Random Before condition

$N = 768$, $\beta = 0.442$, 95% CI = [0.321, 0.562], $p < 0.001$

Participants in the Random After condition

$N = 836$, $\beta = -0.467$, 95% CI = [-0.573, -0.359], $p < 0.001$

Appendix E

Social Value Orientation distributional data

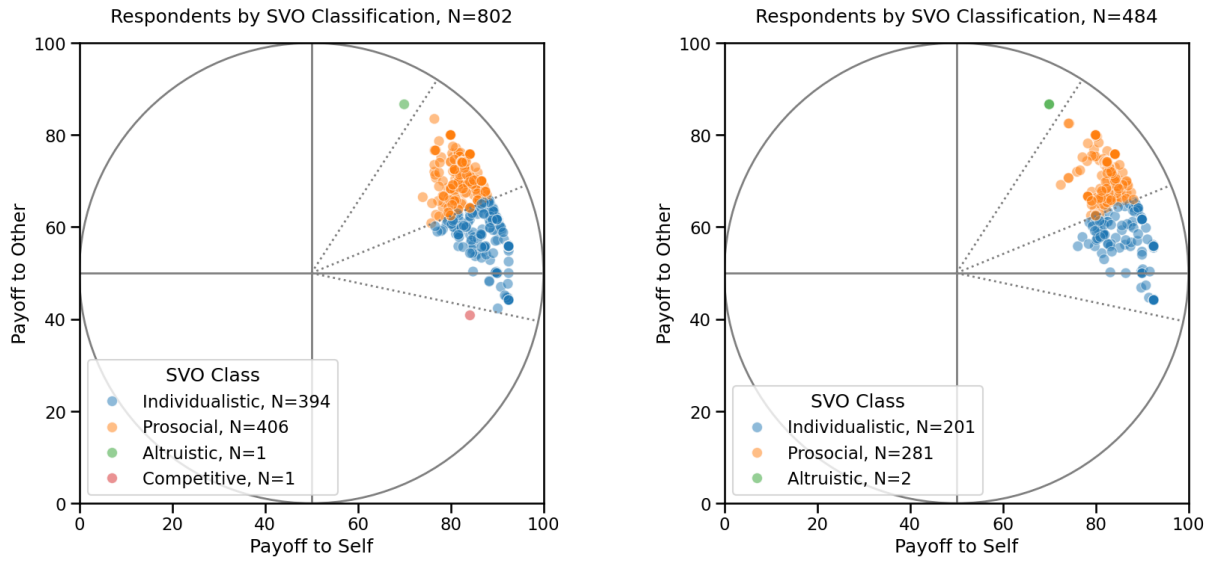


Figure E1

Social Value Orientation distributions for Studies 1b (left) and 1c (right).

Sequential-condition participants who passed comprehension checks.

Social Value Orientation distributional information is reported in Figure E1 for participants from Studies 1b and 1c. Participants filled out an SVO dictator game slider task at the end of the experiments.

Appendix F

Preregistrations

- Study 1a's preregistration can be found on osf.io. Study 1a deviated from the preregistration in that we stopped data collection half-way through, with 2,400 rather than 4,800 participants. Effect sizes for the included manipulations other than order were minimal and we did not want to use further resources.
- Study 1b's preregistration can be found on osf.io. 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 preregistration. 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. The preregistration should have indicated 773 participants, to bring us to the budgeted 1,000 total.
- Study 1c's preregistration can be found on osf.io. We preregistered 800 participants who supply usable data and ended up slightly short due to not having perfect control over how many participants finish, with 783. This took 1298 subjects total, rather than 1700, due to a higher-than-expected comprehension checks pass rate.
- Study 2a was not preregistered.
- Study 2b's preregistration can be found on osf.io. The preregistration specifies 500 participants in the sequential conditions, and we have data from 487. It incorporates hypotheses about quadratic effects, which are the subject of a separate paper.