Cooperation as if by magic: selfish people make prosocial choices by acting as if their move will influence others

Matthew Cashman[1*] and Drazen Prelec[1,2,3]

[1*]Sloan School of Management.
[2]Department of Economics.
[3]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 100 Main Street, Cambridge, 02142, MA, USA.

*Corresponding author(s). E-mail(s): cashman@mit.edu;
Contributing authors: dprelec@mit.edu;

Abstract

Theoretical accounts of cooperation include pro-social motivation, norms and reputation, and cognitive heuristics like team thinking. We provide experimental evidence for a different psychological mechanism, one that, notably, explains cooperation even among the self-interested: quasi-magical thinking. In one-shot Public Goods Games where players move sequentially but do not observe others' moves, we find that contributions to the public good are highest at the beginning and decline as order increases—specifically among payoff-maximizers. We interpret this as players acting as if their moves will influence others who have not yet moved even though they know there is no causal linkage. Three results provide support: (1) This positional order effect is generated by players who are acting in their own interests, (2) instructing players to maximize their own payoff increases the effect, and (3) the effect is eliminated if the moves of future players, but not of past players, are determined randomly.

Keywords: Cooperation, Game theory, Public Goods, Causality

1 Introduction

Social cooperation without external monitoring is widely regarded as fundamental to human culture, sustaining teamwork, mass political participation, and personal sacrifice for family, tribe, or nation. People often face opportunities to incur an individual cost in exchange for a collective benefit, and there is a rich literature exploring the whys and wherefores[1][2]. 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)[3]. The PGG is used as a model of human cooperation because this tradeoff between the benefits accruing to the group via cooperation and the benefits accruing to the individual via defection captures the essence of cooperation problems humans solve on a daily basis. In standard linear PGGs it is always better for an individual player to defect no matter what others do, but it is always better for the group if everyone cooperates.

There may, however, be circumstances in which even players who are just maximizing their own payouts end up cooperating. Such phenomena would suggest there are interventions that increase cooperation even among the most self-interested, in addition to illuminating how such decisions are made. Here we investigate self-interested players who cooperate because cooperation maximizes their individual payoff—on the assumption that those moving after them will make the same move they have. In the case where there is no way to influence others they behave *as if* they can influence others' moves without any communication. *Quasi-magical thinking* (QMT)[4] is precisely the view that people making decisions under uncertainty act as if they have control over the actions of others even when they know it is impossible. 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 without observation (SPGG). If players are acting *as if*, effects should be proportional to the number of people yet to move: the positional order effect. We are agnostic about the precise psychological mechanism behind this phenomenon, but will discuss some possible interpretations in the conclusion. Our principal goal is to establish the existence of the phenomenon—that self-interested players are acting *as if*. We do this by demonstrating that positional order effects exist *only* among the most self-interested players and by breaking the "magical" causal connection between a player and those moving after him.

1.1 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. However, games with an element of coordination have the interesting property that players can coordinate based on pure order of play without any observation. Cooper et al.[5] is the genesis of a modest body of work looking at positional order effects in coordination games such as Battle of the Sexes[6][7][8]. Positional order effects are driven solely by common knowledge about order of play in games with no observation of others' moves: people tend to "agree" to play the first-mover's preferred equilibrium without any communication at all. There is also a line of work investigating of order of play in common-pool resource games[9][10][11], games where players try to request as much as possible—but not too much in total—from a fixed pool of resources, and the related Step-Level PGG, where players receive nothing if the total amount contributed is too low[12][7]. In these games there is a strong incentive to reach the threshold, and criticality (the importance of one's own move for this goal) seems to engender cooperation, whereas uncertainty may reduce cooperation. While neither game has an obvious first-mover advantage, it appears players share a common understanding that those earlier in the sequence can get away with cooperating less and therefore the group can use order to coordinate.

Social dilemmas *without* an element of coordination, games like the PGG, pit what is good for you against what is good for everyone else. The key difference between games with an element of coordination and those without is that in the latter 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 causal thinking about others even in situations without rewards for coordination, and uncertainty about the state of the world seems to activate this sort of reasoning.

In an early study[13], Quattrone & Tversky report evidence for what they term "diagnostic" actions—actions that have no direct causal relationship to desirable outcomes, but which are indicative of them (38 undergraduates). 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 that is believed to be indicative of having a bad heart. The experience of holding one's arm in water of course has no bearing on heart type, but it does appear participants are changing the data they themselves produce in order to receive good news in apparent disregard of the causal relationship.

In related work, Shafir & Tversky[4] explore nonconsequential reasoning—reasoning that at least appears to either not produce estimates of the consequences of an action, or which ignores the consequences of that action. This class of decisions violate the Sure Thing Principle, which states that if X is preferred to Y under all states of the world, then X should still be preferred to Y even if the state of the world is unknown. For instance, there are many people who would prefer to pay for a vacation to Hawaii in the event that they pass an exam *and* in the event that they fail, but who would also prefer *not* to buy in the case where the outcome of the exam is unknown[14]. They refer to this pattern of events as "accept when win, accept when lose, reject when do not know" and refer to it as the "disjunction effect". In an experiment using the Prisoner's Dilemma (80 undergraduates), 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 QMT as a possible explanation for the disjunction effect. The idea is reminiscent of the illusion of control[15][16]; however, that work focuses on repeated tasks that do not generally involve other minds.

Masel[17] offers a formalization of quasi-magical thinking where players, upon observing additional information during the game, update their prior distributions in the usual fashion—one's own behavior being just another data point. Daley & Sadowski[18] develop a similar model of magical thinking that applies to players' preferences over actions rather than outcomes. However, neither formalization incorporates the arrow of time within a single game. There are two flavors of uncertainty at play here: "closed fates" uncertainty is about a counterpart's move when that move is not known to the player but has already been made and is therefore fixed, and "open fates" uncertainty which is about a counterpart's move that has yet to be made at all (or which is presently being made)[19]. Miller and Gunasegaram[20] demonstrated that, while events in the past are considered fixed, future events are treated as mutable. Moreover, future actions are perceived as more intentional and blameworthy than otherwise identical past actions[21].

Subsequent work on 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[22][23][19][4]. However, evidence for positional order effects in sequential PDs or PGGs, games without obvious benefits to coordination, is lacking. When considering QMT, Shafir & Tversky did not distinguish between open fates and closed fates and so could not have measured an order effect. Morris et al.[19] 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[24][25][26][27]. 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. Many also use Prisoners' Dilemmas, a two-person version of a PGG. PGGs with more players enable ruling out any specific first-mover/leader effects[28] or end-of-game effects[25] that are distinct from effects driven merely by sequential order. In addition, many participants are probably not even trying to maximize their own payoffs in these tasks—limiting what we can infer based on play. Budescu et al.[10] 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. While this may be fine if the goal is to explain the behavior of participants as they come to the game, it is a substantial problem when making the assumption that players are trying to maximize profits.

2 Results

The first three studies set the scene for Study 4, which directly tests the causal linkages involved in acting *as if*. Study 1 establishes a decline in cooperation with increasing order (the positional order effect) in a three-person PGG[1], and shows this effect is driven by participants who are "Individualistic" on the Social Value Orientation (SVO)[29] scale. SVO is a measure of willingness to give up gains in order to benefit others. In the SVO battery, participants make a series of incentivized decisions similar to Dictator games where they allocate funds between themselves and someone else. Participants can choose to forego gains (or even pay costs) to help or hurt the other player. If participants are acting *as if*, players who are Prosocial on the SVO measure (meaning they are willing to forego gains to help others) would be expected to show no positional order effect because they will never have a reason to defect: it is nearly always payoff-maximizing for a prosocial player to cooperate, whether at the beginning of a sequence or the end. Conversely, those who are Individualistic (and therefore tend towards maximizing their own payoffs) might show an effect since the number of "open fates" varies with order. Study 2 expands this to a five-person PGG to better clarify the effect and give insight into any first- and last-mover effects. In these studies we are chiefly interested in payoff-maximizing players, but they are sufficiently rare in the study population that forming five-person groups composed of them in real time proved difficult. For this reason, Study 3 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 4 deploys the technique from Study 3 to ask whether we still observe a positional order effect in the case where all players after a focal player have their contribution decisions delegated to a random process. If the effect is present when random movers are *before* the focal player, but absent when random movers are *after* the focal player, this would indicate the effect requires having real people who have not yet made a decision, but who will, moving after the focal player—implying causal thinking about others is at play.

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

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

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

[1] See C for a Sequential Prisoner's Dilemma that shows the positional order effect and which predates work with SPGGs

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 their decisions are necessary to finish the game. Later studies incorporate more extensive training and comprehension check regimes. Most statistical tests are one-sided given directional preregistered predictions.

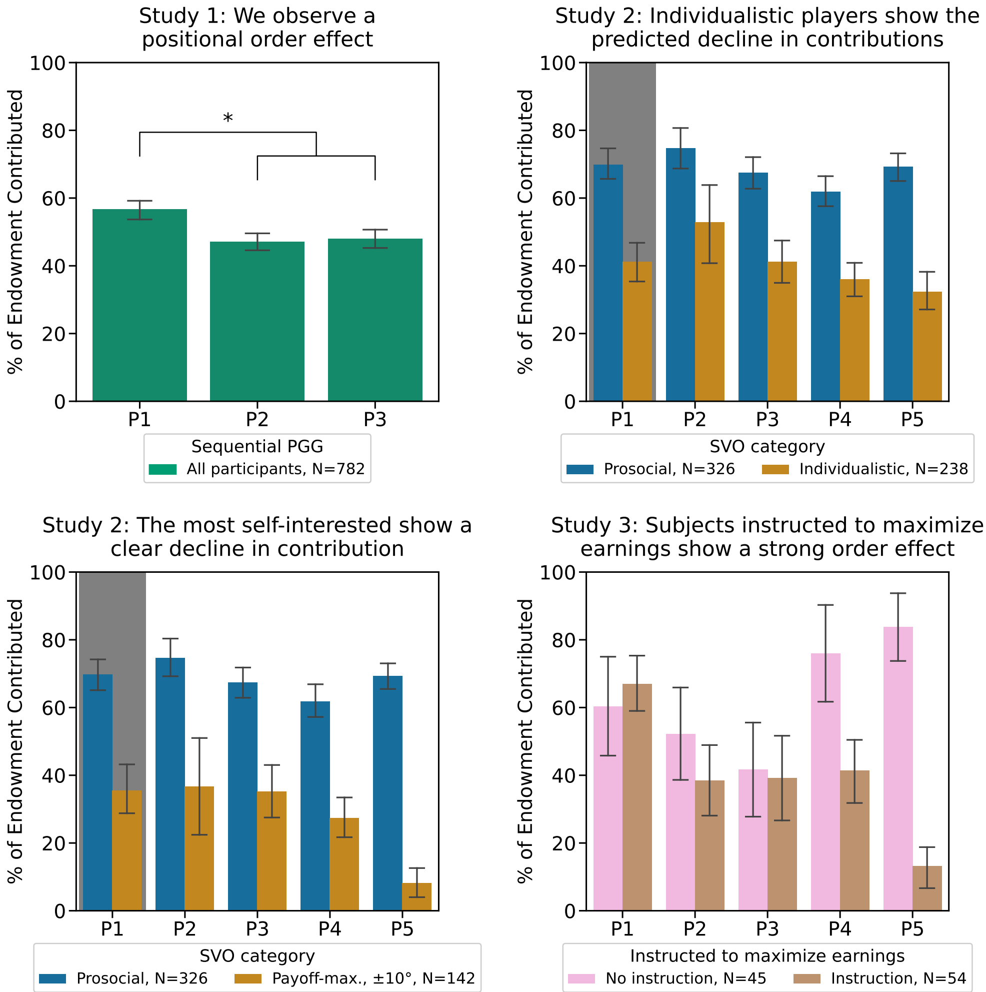

Fig. 1 [Upper left] In Study 1, change in contribution with order is driven by participants SVO-classified as Individualistic. [Upper right] SVO-individualist players show a decline in contribution with increasing order from Players 2-5. Study 2 suffered from technical problems that resulted in all P1 and some P2 not receiving enough decision time. Only P2 who received enough decision time are shown. [Lower left] Payoff-maximizing players in Study 2 show the predicted decline of contribution with increasing order despite technical problems. [Lower right] Study 3 shows the hypothesized decline with order among those who were merely instructed to be greedy. participants who passed comprehension checks, SEMs.

2.1 Study 1: 3-Person Sequential Public Goods Game

In Study 1, participants were randomized to position 1, 2, or 3 and played one round of a three-person SPGG with no observation. Our primary interest was how contribution to the public good varied with order of play. The SVO measure divides almost all[1] participants into two categories, Individualistic and Prosocial.

We do not meet the preregistered threshold for the positional order effect in this study. This may be due to the study being under-powered to detect an effect given the conditions of the preregistration, which specified backwards-difference coding. First-movers contribute more than later players, though we do not resolve a difference between second- and third-movers. A linear regression of contribution on move order yields a significant negative slope, $\beta = -4.244$, 95% CI = [-8.086, -0.401], $F(1, 780) = 4.7$, one-sided $p = 0.015$. First-movers contribute more than second-movers ($p = .007$) and more than third-movers ($p = .015$). The difference between second- and third-movers' contributions is not significant.

We find support for the preregistered prediction that the positional order effect is concentrated among participants classified as Individualistic in the SVO task. Participants classified as Prosocial exhibit no significant differences in contribution levels as function of order, while we do see a difference between the first-mover data and grouped second- and third-mover contributions ($\beta = -14.197$, 95% CI = [-24.681, -3.713], $F(1, 288) = 7.104$, one-sided $p = 0.004$) among Individualistic players. As with the aggregated data, we do not see the hypothesized difference between positions two and three among participants SVO-classified as Individualistic. A regression using the continuous SVO angle measure interacted with order (first or subsequent) does find significance, $\beta = 0.521$, 95% CI = [0.02, 1.021], $F(3, 597) = 31.14$, one-sided $p = 0.021$.

In addition, we find some support for the preregistered 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 does not find significance, but when applied to only Individualistic players a significant equation is found, $\beta = 0.173$, 95% CI = [0.023, 0.322], $F(3, 576) = 64.451$, one-sided $p = 0.012$ using cluster-robust errors at the participant level. There is no effect among Prosocial players, $\beta = 0.001$, 95% CI = [-0.13, 0.132], $F(3, 614) = 98.503$, one-sided $p = 0.494$.

2.2 Study 2: 5-Person Sequential Public Goods Game

Study 2 was a sequential 5-person PGG with no observation where participants were classified based on an SVO task[1] performed at the end of the study. A programming error affecting P1 and P2 meant that all first-movers were forced to respond too quickly, sometimes with no time at all, while only some second-movers were affected as the time they were allocated was a function of P1's response time. Including affected participants, for Study 2 we detect an effect for the interaction between order and SVO category, $\beta = -2.851$, 95% CI = [-6.259, 0.556], $F(1, 265) = 2.715$, one-sided $p =$

[1]Nearly all participants were SVO classified as Individualistic or Prosocial; one was classified as Altruistic, and one as Competitive. These participants' data are excluded from categorical SVO analyses.

0.050. When considering only players who were not affected by the error, Players 2-5, the preregistered regression also reaches significance: $\beta = -5.597$, 95% CI = [-12.215, 1.022], $F(1, 180) = 2.784$, one-sided $p = 0.048$. When interacting contribution with the continuous SVO angle measure instead of the categorical model we see $\beta = 0.36$, 95% CI = [0.091, 0.629], $F(3, 439) = 33.71$, one-sided $p = 0.004$ for P2-P5.

The SVO Individualistic category captures a fairly broad range of social preferences, and we wanted to examine the subset who are very clearly trying to maximize their own gains. Perfectly self-interested play is within ±7.82 SVO, and since slider input devices introduce some asymmetric trembling-hand noise ±10 should capture all players who are strictly attempting to maximize personal gains, as opposed to those who were merely classified as Individualistic. Among these most clearly self-interested participants the effect size in the preregistered analysis is notably increased, $\beta = -11.05$, 95% CI = [-19.356, -2.743], $F(1, 103) = 6.96$, one-sided $p = 0.005$, for P2-P5 unaffected by the error.

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 $n = 1254$, Pearson's $r = 0.34$, 95% CI = [0.29, 0.39], $p < 0.001$, and for backwards in time $n = 1314$, Pearson's $r = 0.27$, 95% CI = [0.21, 0.32], $p < 0.001$, z-score for the difference of 0.072 = 2.006, 95% CI = [0.002, 0.155], $p = 0.045$. Looking just at P2-5 with the correct decision time the forward correlation ($n = 533$) is $r = 0.43$, 95% CI = [0.36 0.49], backwards ($n = 1239$) is $r = 0.27$, 95% CI = [0.22 0.32]. The z-score for the difference of 0.154 is 3.394, 95% CI = [0.074, 0.271], $p = 0.001$. Study 1 and Study 2 were sufficient evidence of positional order effects among the clearly payoff-maximizing to justify moving on to Study 3.

2.3 Study 3: 5-Person Sequential Public Goods Game with induced self-interest

Because of the difficulty of filling experiments with subjects presenting as clearly self-interested, we wanted to test whether merely *prompting* all participants to maximize personal payoffs generates a positional order effect similar to that observed among Individualistic participants. The main difference from Study 2 is that we did not measure SVO in Study 3. Instead, participants were randomized to a condition with no prompt, or to a condition with the 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 3 incorporates four substantive improvements over Study 2. First, Study 3 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 control for the possibility of effects dependent on time spent waiting. While

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

waiting, participants are shown the task's standard wait screen which incorporates the option to play a simple game to keep participants engaged 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. Finally, Study 3 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.

We observe a positional order effect in this non-preregistered study. A linear regression of contribution on order interacted with a binary instructed/not instructed to maximize payoffs variable detects the interaction effect ($\beta = -16.543$, 95% CI = [-27.917, -5.169], $F(3, 95) = 5.309$, one-sided $p = 0.002$ for the interaction). participants receiving the prompt show a decline in contribution with increasing order. There is substantial noise in estimates of the means, but we felt this result provided enough confidence to justify deploying this technique in the next, larger experiment.

2.4 Study 4: 5-Person Sequential Public Goods Game with random moves

Study 4 incorporates the improvements from Study 3 and extends it by applying the instruction to act to maximize one's own payouts to all participants and at larger scale, but with two new conditions: all participants are either told that every player *before* them has had their contribution determined randomly ("Random Before"), or that every player moving *after* them has had their contribution determined randomly ("Random After"). This clarifies whether the positional order effect is driven by the fact that other *people*, specifically, will be moving after the focal player—even though he cannot see their moves. Players are presented with a page that explains the setup, and are presented with symbols that make clear which players' moves were randomly decided. They see the graphical representations in 2 on all pages from the point at which the concept of random moves is introduced until the end of the game. It may be noted that in this study Player 1 (in the Random Before condition) and Player 5 (in the Random After condition) play a standard sequential PGG in that they do not play with any players that have their contributions randomly determined at all.

We observe a decline in contribution with order only among those players who are told that everyone moving *before* them has his move determined randomly, while everyone moving *after* them is deciding on what move to make. The preregistered linear regression contribution $\sim$ order * random_before + wealth, differing from previous analyses in that it controls for a measure of wealth, finds the effect, $\beta = -6.402$, 95% CI = [-11.377, -1.426], $F(4, 501) = 3.604$, one-sided $p = 0.006$ for the interaction. Though participants have little time to talk and learn the rules of the game after their up-front text chat, there may be some worry about group-level effects. A mixed model that adds a group-level random effect shows the effect, $\beta = -6.567$, 95% CI = [-11.52,

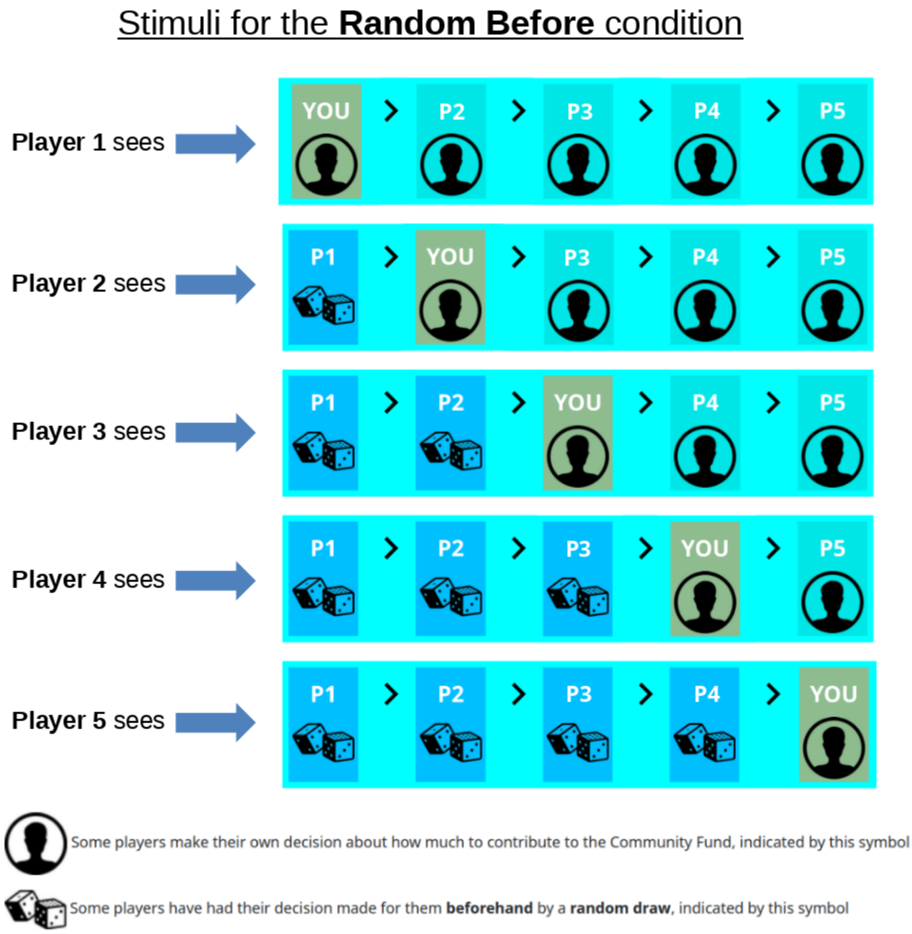

Fig. 2 Stimuli for the Random Before condition. Players see a graphical representation of their position relative to other players that clearly conveys which players are having their moves made by a random process. This is in addition to a previous screen that explains how some players are having their moves made for them by random processes. The Random After condition has the dice and human figures reversed.

-1.614], one-sided $p = 0.005$. We also find a significant equation not controlling for wealth, one-sided $p = 0.005$. When we restrict the main analysis to only those players who passed a second set of comprehension checks at the end of the experiment (76.6% of players who passed the initial checks), we observe a larger effect size ($\beta = -8.673$, 95% CI = [-14.136, -3.211], $F(4, 403) = 3.807$, one-sided $p = 0.001$ for the preregistered analysis; see E). This gives further reason to believe that the effect is concentrated among participants who truly understand the game. We do not observe a difference between the no delay (equivalent to P1, mean contribution = 41.1) and long delay (80 seconds, equivalent to P5, mean contribution = 48.3) simultaneous-move control conditions using a T-Test, t(129) = -0.90, p = 0.372. This rules out the effects being due to mere time in the experiment. On the model in B, self-interested participants who are acting *as if* should either contribute 0% of their endowment or 100% depending on where they are in the sequence. 67.3% of participants give either 0 or 100% of their endowment, and among these participants effect size increases ($\beta = -9.292$, 95% CI = [-16.179, -2.404], $F(4, 332) = 4.162$, one-sided $p = 0.004$ for the preregistered analysis, see D).

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 ∼ contribution*future/past yields $\beta = 0.374$, 95% CI = [0.224, 0.523], $F(3, 1000) = 74.307$, one-sided $p < 0.001$, and for Random After $\beta = -0.432$, 95% CI = [-0.57, -0.294], $F(3, 1012) = 77.198$, one-sided $p < 0.001$ with cluster-robust standard errors.

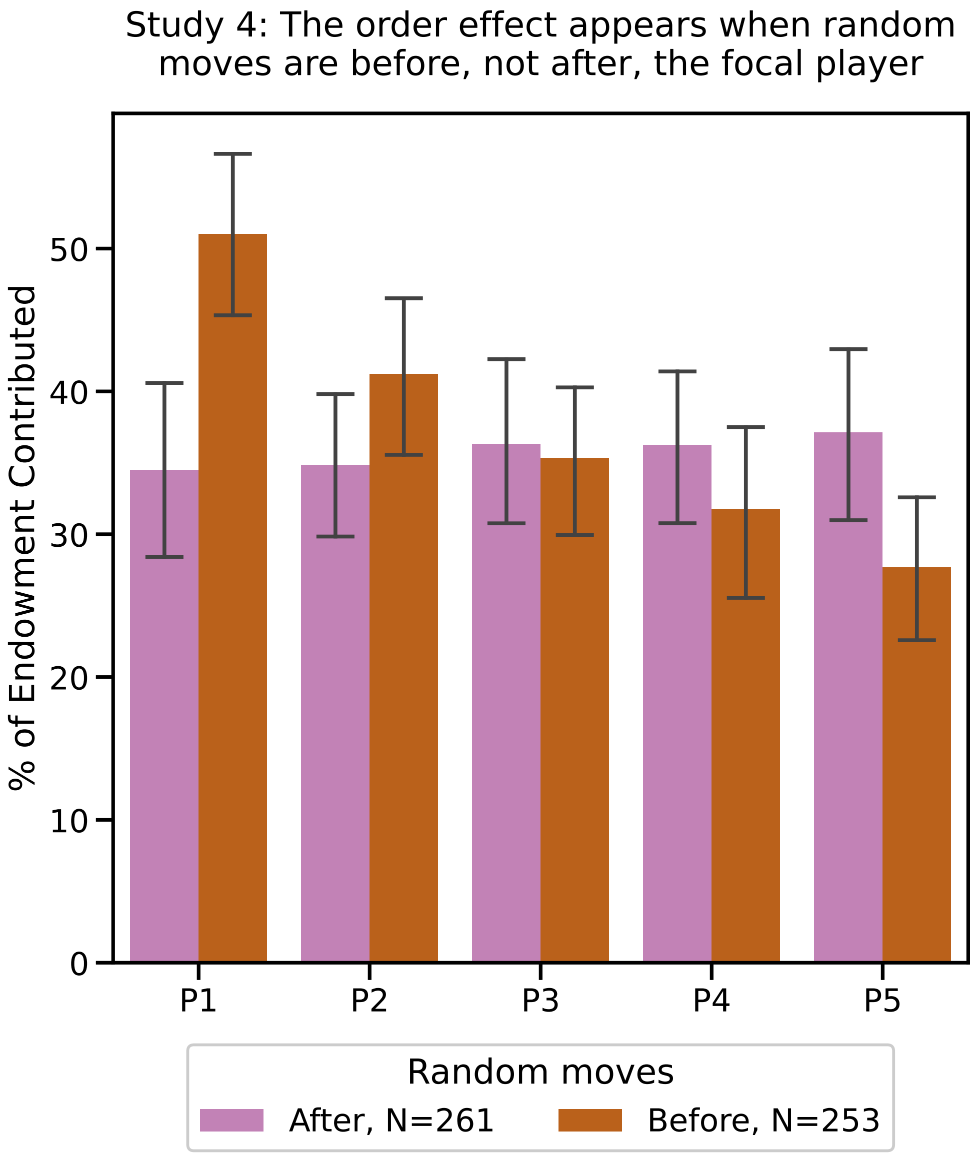

Fig. 3 Study 4 shows a decline in contribution to the public good among players who are told that all players moving after them are making their own moves, and all players moving before them are having their moves made randomly. No effect is observed among players who are told that everyone moving after them has a move selected at random. participants who passed comprehension checks. SEMs.

3 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. It appears that earlier movers tend to believe that contributing to the public good will maximize their payouts, and later movers believe that less contribution will maximize payouts—and so are more inclined to defect. The effect's absence when subsequent players have their moves made randomly suggests implicit causal thinking at play: It is not just *that* I cooperate that suggests others will cooperate, but *if* I cooperate, others will cooperate, so the effect is stronger in the direction of open fates. Participants are 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 be expected in the case that players are acting *as if* for the same reason. This behavior is consistent with quasi-magical thinking and makes it clear that the distinction between open and closed fates is important. We speculate that a simple model may capture something of the process generating this behavior specifically in self-interested agents: these agents understand the rules of the game and are trying to maximize their payouts—they just act as if everyone who has not yet moved will make the same move they do. This implies a sharp step between 100% contribution and 0% contribution, which is observed in the data. 67.3% of participants contribute either 0 or 100% of their endowment, and the positional order effect is stronger in this subset (see D). In addition, the effect is stronger in the 76.8% of participants who pass both pre- and post- comprehension checks, implying the effect is generated by people who understand the game. A formalization of this model is included in B.

Quasi-magical thinking may fit the data we observe, but it is not immediately obvious why this pattern of behavior is rational or adaptive. If, in the absence of other information, a player assumes some similarity between himself and other players his own behavior may 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[30][31][32] . 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.[33][34][35][36]. In the case of a PGG, self-interested players may be motivated to learn from their own behavior that others moving after them will also contribute, thereby raising their estimate of their payoffs. Adjusting your own estimate of your future profits upwards is pleasurable, so there is utility to be gained from that adjustment (diagnostic utility) in addition to the standard utility from the payout itself (outcome utility). Crucially, from the standpoint of both theory and empirical evidence, self-signaling does not require a perceived causal link between decisions and the underlying characteristic of interest; it can influence decisions even when their causal irrelevance is made obvious by experimental design as in [13]. 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[37].

A related body of work examines universalization as an explanatory model for many morally-relevant behaviors. The basic idea is that, at some level, people ask themselves "What if everyone did this?" in order to determine what they themselves should do. Roemer[38][39] 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.[40] 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. The fact that this question occurs in the moral domain may imply that the moral phenomenon is a special case of the more general strategy that we report evidence of here.

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

Finally, whatever the mechanism, understanding why even the most self-interested actors might decide to contribute to the public good is relevant to many practical policy questions. For example, an agent might say to herself: many other people will be in my shoes in the future, so if I vote then other people will too; if I conserve energy, then others will conserve as well; if I contribute to a public good, so will others—and this action is actually best for me independent of what's good for everyone else. Even purely self-interested individuals might feel that their investment of time or effort for a public cause will pay off, pointing to a class of interventions that highlight that other people like you will be deciding to contribute—or not—at a later time.

4 Methods

Ethics All studies reported here were approved by MIT's Committee on the Use of Humans as Experimental participants (COUHES) and comply with all relevant ethical regulations. We obtained electronic consent from all participants.

In total we tested 3615 participants distributed across four experiments. A convenience sample provided by Amazon Mechanical Turk (MTurk) was selected for Study

1 and Study 2 because it was a reasonable approximation of American adults for our purposes, but declines in the quality of MTurk data over the years these studies were conducted meant that we selected CloudResearch's filtered MTurk panel for Studies 3 and 4 because it provides among the highest-quality online panel data[41][42]. This work makes the point that these effects exist in human populations, and it is left for future work to examine how they vary across ages, sexes, SES, cultures, and other characteristics of interest. All experiment software was written in the oTree framework [43]. All experiments also involved training and comprehension checks. Data from participants who failed one or more pre-play comprehension check questions was excluded. All experiments are real-time online group tasks, where participants interact via text chat before learning the rules of the game in order to establish some sense that they are completing the task with actual people in real time. All studies except for Study 3 were preregistered on osf.io.

4.1 Study 1

Preregistration. The preregistration for Study 1 (https://osf.io/3vsxk) was registered on November 21, 2019. Study 1 deviated from the preregistration in that the preregistration specifies 1000 participants, while 775 were collected after the preregistration. When pooled with data from before the preregistration we reach 1002 participants. The budget was calculated assuming data from prior to the preregistration was included, and this should have been specified.

Participants. 1444 U.S.-based participants from Amazon Mechanical Turk completed the study. Mean total pay per participant (including bonuses for accurate predictions) is \$3.16 ($SD = 0.81$), yielding an hourly rate of \$18.46 per hour (SD = 8.06) at 10 minutes' duration. Of 1444 participants who passed up-front bot checks and finished the task, 69.0% (1002) passed all of the comprehension check questions and will have their data included. Data from batches 1 and 8 were excluded due to technical problems resulting in server crashes during the experiment. To estimate the sample size required, we performed a power analysis via simulation using pilot data. 200 participants produced contribution and prediction data, but failed to reach the SVO slider battery at the end of the task. This is primarily due to technical errors that arose in batches 2 and 13, resulting in crashes that stopped further progress in the task. These participants' data are included per the preregistration (since they answer the contribution question), but by necessity they do not feature in analyses that involve SVO.

Materials and procedure. Study 1 is a one-shot sequential sequential PGG with a multiplier of two. Three players can transfer any part of their individual \$1 endowment. The total transfer amount 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. The only difference in information among the players is knowledge of their position in the sequence. Each participant was assigned to one of four conditions: orders 1-3 and 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, the enter a wait room and form groups of three. Then they are placed in a chat room for 30 seconds after all players in their group have arrived to ensure participants believe the experiment is, in fact, a real game in real time with real people. After the chat, participants 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: (1) Do any of the other players know how much you decide to contribute? (2) No matter what the other players do, what earns you the most money? TRANSFERRING to the community fund or KEEPING your endowment? and (3) What year is it? Responses to the comprehension questions are only relevant to data analysis: players continue on whether or not they have answered correctly. Since players do not interact after the initial chat, players who fail the comprehension checks can have no further influence on those that pass. Players who fail comprehension checks remain in the game because the games are real games happening in real time, and so there moves are needed to calculate payouts without deception.

After having completed the comprehension questions, players 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 online materials). Players in the simultaneous condition do not see any indication of sequence since they are moving simultaneously. Participants then complete prediction questions, and then a Social Value Orientation (SVO) slider battery (Murphy et al., 2011; code based on Bakker, 2016/2019)[1]. The SVO battery measures preferences for how to allocate resources between oneself and others. The standard battery categorizes participants into Individualistic (concerned only with what is best for self), Competitive (maximize own outcomes as with Individualistic, but also minimize the outcomes for others), Prosocial (maximize outcomes for both self and other), and Altruistic (eager to give up own gains to help others). Players then exit the experiment and are paid.

Analysis. The preregistered analysis used to investigate the impact of order on contribution is a linear regression contribution $\sim$ order, with order treated as ordinal and backwards-difference coded. Backwards difference coding enforces a statistical significance test for each comparison, 1 vs. 2 and 2 vs. 3, enforcing a stepwise change from 1 to 2, 2 to 3, etc. The preregistered analysis for predictions of others’ moves going forward is the prediction $\sim$ own response interacted with a binary future vs. past variable.

4.2 Study 2

Preregistration. The preregistration for Study 2 (https://osf.io/gw8nc) was registered on April 6, 2021.

Participants. 1212 U.S.-based participants (43% female, average age 37) from Amazon Mechanical Turk completed the study. Mean total pay per participant (including bonuses for accurate predictions) is \$3.40 (SD = 0.52), yielding an hourly rate of

[1]SVO is measured post-treatment, but we do not observe an effect of treatment on SVO.

$10.87 per hour at 18.8 minutes' average duration. Of 1212 participants who passed up-front bot checks and finished the task, 752 (62%) passed all of the up-front comprehension check questions and will have their data included. Time on the decision-making page for players 1 and 2 was variable due to a programming error, and data from participants who got less than the designed time was excluded from analyses. To estimate the sample size required, we performed a power analysis via simulation using pilot data.

Materials and procedure. Study 2 is a one-shot sequential PGG very similar to Study 1, with the exception that there are five players rather than three and that the up-front chat was 90 instead of 30 seconds.

Analysis. The preregistered analysis used to investigate the impact of order on contribution is a simple OLS linear regression contribution $\sim$ order (excluding simultaneous participants) among SVO-Individualistic participants. Backwards-difference coding (as specified in Study 1) could have required unworkably large sample sizes per bootstrapped power analyses. The correlation between one's own contribution and those of groupmates being stronger going forward is investigated by calculating the partial correlation of prediction with own response controlling for a population-level prediction separately between forward- and backwards-predictions and then testing for the difference in correlation coefficients.

4.3 Study 3

Preregistration. This study was not preregistered.

Participants. 157 U.S.-based participants from Amazon Mechanical Turk via Cloud Research (37% female, average age 40) completed the study . Mean total pay per participant (including bonuses for accurate predictions) is $3.5 ($SD = 0.96$), yielding an hourly rate of $13.12 at 15.5 minutes average duration. Of the 157 participants who passed up-front bot checks and finished the task, 139 (89.0%) of those passed all of the up-front comprehension check questions and will have their data included.

Materials and procedure. Study 3 adds some features to the basic design from Study 2. In Study 3, 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.

Players are also randomized to a delayed simultaneous condition in addition to the simultaneous condition from previous studies, to control for effects that arise merely from waiting. participants randomized to the delayed simultaneous condition wait for 80 seconds on the standard wait page for the task (which contains a simple game they may play if they wish). In addition, players in Study 3 move in lock-step throughout the task. Instead of being able to advance on certain pages when they feel they are ready, players move in lock-step with a certain number of seconds allotted for each page (so subsequent players cannot infer anything from how quickly those previous to them have moved). Pages on which players make their contribution or make predictions do not force a player to stay for a certain amount of time, but rather let the player move on to a wait page when the decision has been made. The wait page soaks up any remaining time.

Analysis. There was no preregistration for Study 3 since it was meant to be a simple, fast test of whether or not instruction to be self-interested would produce a positional order effect. The analysis used is an OLS linear regression, contribution $\sim$ order * instruct_or_no, with instruct_or_no being a binary indicator of whether or not participants were instructed to be self-interested.

4.4 Study 4

Preregistration. The preregistration for Study 4 (https://osf.io/3kepm) was registered on November 8, 2022. The preregistration specifies a target of 500 participants; this was meant to refer to what would be required to detect the effect in the Sequential treatment (vs. Simultaneous), but this could have been clearer. We report data from 514 participants for the Sequential treatment P1 - P5.

Participants. 834 U.S.-based participants from Amazon Mechanical Turk via Cloud Research (45% female, average age 40) completed the study. Mean total pay per participant (including bonuses for accurate predictions) is \$4.10 ($SD = 1.42$), yielding an hourly rate of \$13.31 per hour at 18.5 minutes average duration. Of 834 participants who passed up-front bot checks and finished the task, 645 (77%) of those passed all of the up-front comprehension check questions, 514 of which were in the Sequential treatment. To estimate the sample size required, we performed a power analysis via simulation using pilot data and data from previous experiments.

Materials and procedure. Study 4 is a one-shot sequential PGG identical to Study 3, with the exception that players, instead of being randomized to get the instruction to maximize earnings or not, all players receive that instruction and instead they are randomized between two conditions, fully crossed with orders 1-5: 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 ("Random Before"), or players are told that everyone *after* them has their decision made by a random process ("Random After"). This study involved deception, as it was not true that everyone either before or after them was making their own decision or having their moves made randomly. As in Study 3, there are two simultaneous control conditions: one with a delay equivalent to the wait time 5th-movers experience in the sequential game, and one without which is equivalent to moving first.

Analysis. The preregistered analysis used to investigate the impact of order on contribution in Study 4 is a simple OLS linear regression that, in addition to what is used for Study 3, controls for self-reported wealth: contribution $\sim$ order + wealth among those who are told that players before them have their moves made randomly ("Random Before"). Wealth was added to the regression given the expectation, common across economics, that players' sensitivity to payoffs is modulated by the marginal change in their wealth or similar.

Supplementary information. Below are supplementary materials.

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:

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

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 pro-social motivation.

B.1 Prosocial preferences

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

$$p_i = 1 - a_i + \frac{m}{n} \sum_{k=1}^{n} a_k \tag{B2}$$

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:

$$u_i\left(a_1, \ldots, a_n\right) = \left(1 - s_i\right) p_i + s_i m a_i \tag{B3}$$

where p_i is determined by the game formula, B2 . A purely self-interested player ($s_i = 0$) will aim to maximize own payoff, $u_i = p_i$; a purely prosocial player ($s_i = 1$) will aim to maximize the impact of his contribution to the public good, $u_i = ma_i$. The prosocial motive, captured by the second term, thus reflects the impact of own contribution to the public good; other players' contributions enter the utility model only insofar they determine the first, self-interested utility term. In other words, players: (a) care how their action affects the payoffs of others, (b) care how other players' contribution affect their own payoff, but (c) do not care how other players' actions affect each others' payoffs.

B.2 Decision dependent expectations

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

$$a_i = 1 \iff \mathbb{E}\left[u_i \mid a_i = 1, s_i\right] > \mathbb{E}\left[u_i \mid a_i = 0, s_i\right] \tag{B4}$$

A player knows the value of their prosocial parameter and hence also knows the utility function in B2. If he were just a spectator, not making a decision, his expectation of the contribution of another, randomly selected player would exhibit projection, along the lines of Bayesian updating. The simplest version of such updating is linear:

$$\mathbb{E}\left[a_k \mid s_i\right] = b + cs_i \tag{B5}$$

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

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

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

where $\mathbb{E}\left[a_k \mid s_i\right] = b + cs_i$ from B5 is substituted in the final line.

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

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

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

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

The decision criterion in B4 can be expressed as:

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

where the first line follows from B3 and the third line from B4.

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

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

Substituting into the criterion,

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

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

$$\mathbb{E}\left[u_i \mid a_i=1,s_i\right]-\mathbb{E}\left[u_i \mid a_i=0,s_i\right]=0 \Longleftrightarrow d^*(i)=\frac{-m-smn+n}{m(n-i)} \tag{B7}$$

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 Study 0: Sequential Prisoner's Dilemma

Study 0 is a Sequential Prisoners' Dilemma (PD) study that predates studies 1-4 and produces the positional order effect. It is not reported in the main text because it is a Prisoner's Dilemma rather than a Public Goods Game, but it is very similar in structure to Studies 1-4. Study 0 contained a number of exploratory conditions designed to test theory of mind manipulations and the effect of having population-level information over which we collapse here.

Ethics All studies reported here were approved by MIT's Committee on the Use of Humans as Experimental participants (COUHES) and comply with all relevant ethical regulations. We obtained electronic consent from all participants.

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

Materials and procedure. The chat room and experimental platform was developed on the oTree framework [43]. Players arrive at the experiment web page, are consented, and then engage in a real effort task as attention and activity verification (transcribing nonsense sentences)[1]. The chat room then provides 30 seconds for exchanging a hello or brief message, confirming that their teammate is human rather than a computer algorithm. 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 C1:

Table C1 Study 0 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)

In Informed conditions, the payoff screen states that "About half (50%) of other players choose to TRANSFER, and half choose to KEEP,"[2] the screen also contains text boxes in the ToM and Irrelevant conditions. 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?

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

[2] Based on pilot data available at the time the experiment was run

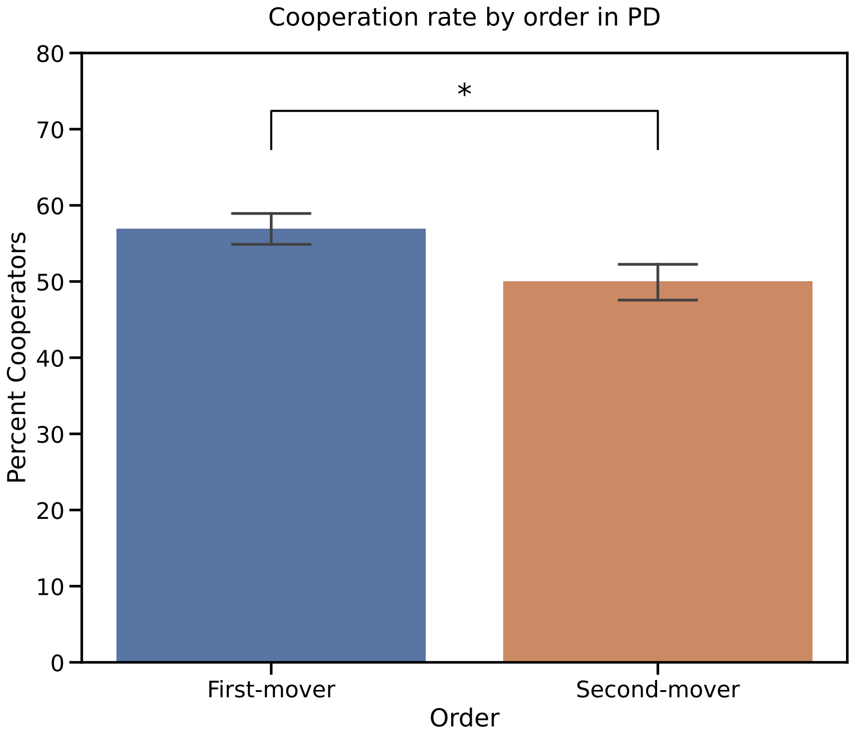

Fig. C1 Percent cooperation among first- versus second-movers in a sequential PD, 1075 of 2367 participants who passed comprehension checks, SEMs.

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?
5. What year is it?

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

C.0.1 Results

First-movers cooperate more than second-movers, which is consistent with our preregistered hypothesis.

A logistic regression of Transfer decision on Order reaches significance ($OddsRatio = 0.757$, $95\%CI = [0.596, 0.963]$, $z = -2.268$, $p = 0.023$). None of the other factors or interactions were near significance. We observed a strong non-preregistered impact of decision on perception of teammate's behavior relative to the population. Players who keep their endowment think that their teammate is less likely to transfer than the population at large ($M_{teammate-population} = -2.20$), while those who transferred believe their teammate is more likely to transfer ($M_{teammate-population} = +2.87$) A linear regression shows a significant effect, $\beta = 5.069$, $95\% CI = [3.272, 6.937]$, $F(1, 1071) = 30.4$, $p < 0.001$.

C.0.2 Discussion

The results support our preregistered hypothesis that first movers would be more likely to cooperate.

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

The formalization inB predicts that any given player who is both trying to maximize his own payoffs and who is acting *as if* will either give 100% of the endowment or 0%, with a sharp transition. The point at which the shift from 100% to 0% happens as order increases is a function of d, the magical influence parameter, when s_i, the player's prosociality, and m, the game's multiplier, are held constant. In Study 4 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 4 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*.

The preregistered linear regression contribution $\sim$ order * random_before + wealth, differing from previous analyses in that it controls for a measure of wealth, finds the effect.
$\beta = -9.292$, 95% CI = [-16.179, -2.404], $F(4, 332) = 4.162$, one-sided $p = 0.004$

We also find a significant equation not controlling for wealth
$\beta = -9.214$, 95% CI = [-16.061, -2.367], $F(3, 337) = 4.143$, one-sided $p = 0.004$

The short and long simultaneous conditions fail a t-test for difference in means. The mean of the long simultaneous condition is also higher than that for the short condition.
Short condition: M = 43.3, SD = 50.4, N = 30
Long condition: M = 49.2, SD = 50.4, N = 63
t(91) = -0.53, p = 0.601

Correlation between own move and predictions of others' moves

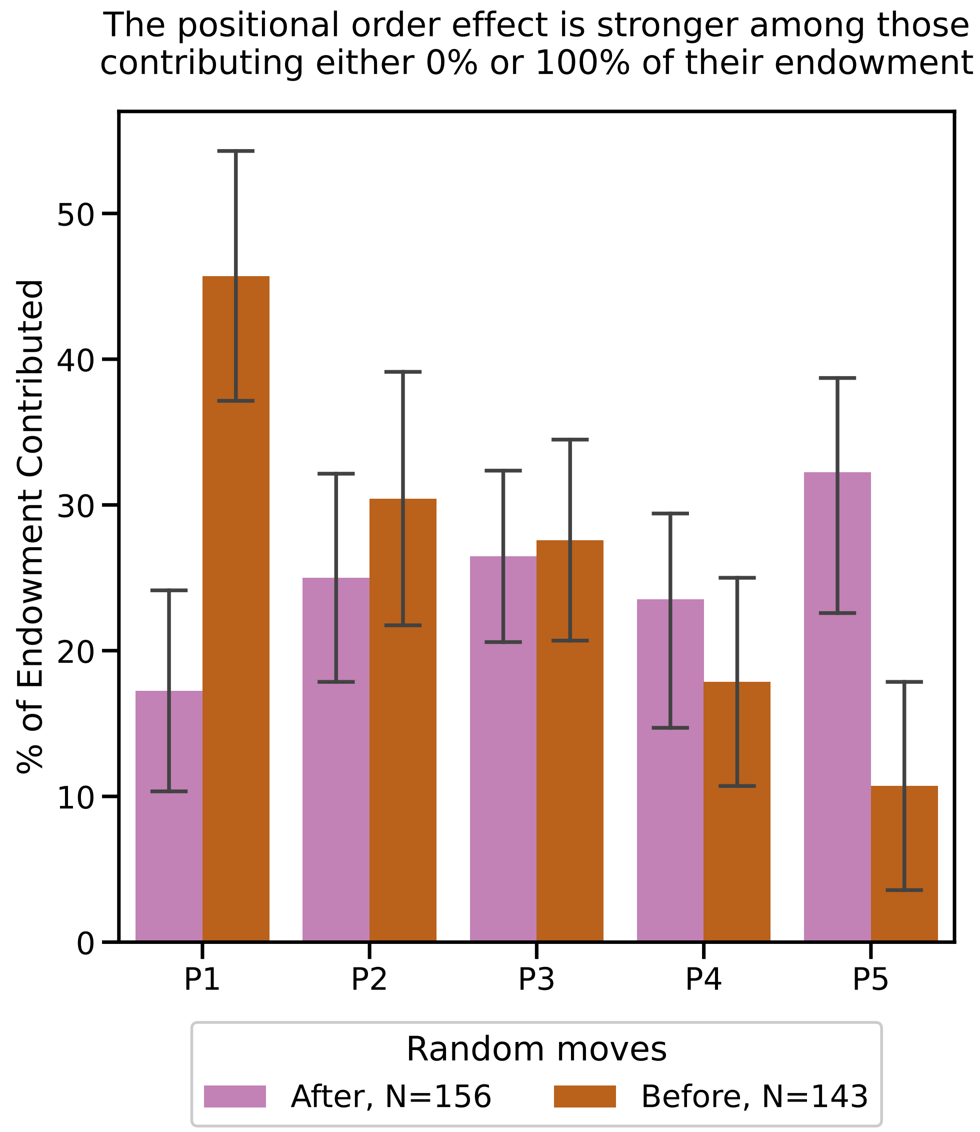

Fig. D2 Study 4 shows a decline in contribution to the public good among players who are told that all players moving after them are making their own moves, and all players moving before them are having their moves made randomly even when limiting responses to 100% of endowment and 0% of endowment. With this filter, effect size increases. SEMs.

Model 1a, all respondents passing comprehension checks
$\beta = -0.021$, 95% CI = [-0.094, 0.052], $F(3, 1516) = 247.521$, one-sided $p = 0.284$
Model 1b, random before people after
$\beta = 0.374$, 95% CI = [0.268, 0.479], $F(3, 644) = 139.237$, one-sided $p < 0.001$

With cluster-robust errors
$\beta = 0.374$, 95% CI = [0.21, 0.537], $F(3, 644) = 72.013$, one-sided $p < 0.001$
Model 1c, people before random after
$\beta = -0.476$, 95% CI = [-0.581, -0.371], $F(3, 684) = 125.649$, one-sided $p < 0.001$
With cluster-robust errors
$\beta = -0.476$, 95% CI = [-0.629, -0.323], $F(3, 684) = 68.42$, one-sided $p < 0.001$

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

Study 4 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 both understand the game and are tring to maximize their own personal payoffs.

The preregistered linear regression contribution ∼ order * random_before + wealth, differing from previous analyses in that it controls for a measure of wealth, finds the effect.
$\beta = -8.673$, 95% CI = [-14.136, -3.211], $F(4, 403) = 3.807$, one-sided $p = 0.001$
We also find a significant equation not controlling for wealth
$\beta = -8.682$, 95% CI = [-14.155, -3.209], $F(3, 404) = 4.217$, one-sided $p = 0.001$
The short and long simultaneous conditions fail a t-test for difference in means. The mean of the long simultaneous condition is also higher than that for the short condition.
Short condition: M = 32.7, SD = 43.0, N = 28
Long condition: M = 38.3, SD = 45.5, N = 58

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

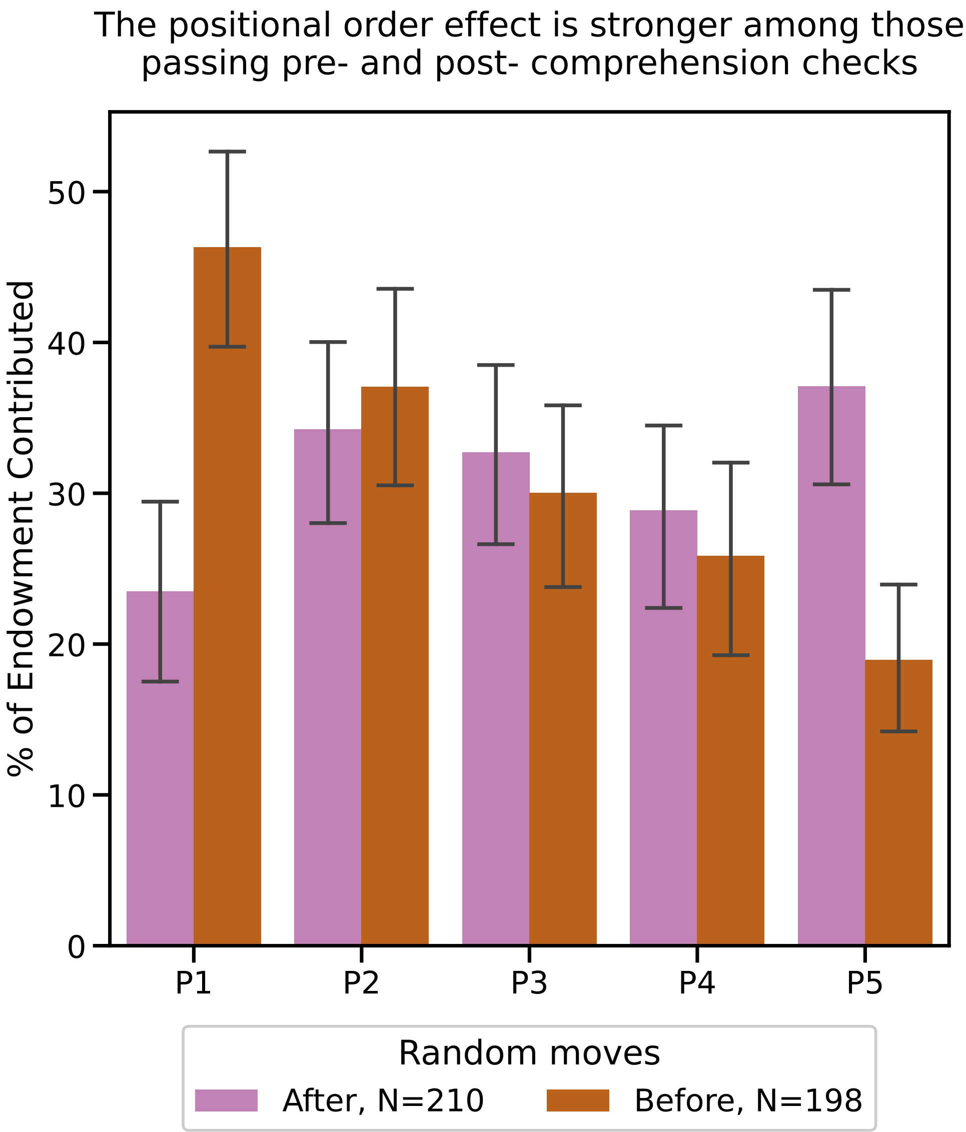

Fig. E3 Study 4 shows a decline in contribution to the public good among players who are told that all players moving after them are making their own moves, and all players moving before them are having their moves made randomly even when limiting responses those who pass both pre- and post-tast comprehension checks. With this filter, effect size increases. SEMs.

t(84) = -0.54, p = 0.588

Correlation between own move and predictions of others' moves

Model 1a, all respondents passing comprehension checks
$\beta = -0.007$, 95% CI = [-0.083, 0.07], $F(3, 1784) = 212.425$, one-sided $p = 0.431$
Model 1b, random before people after
$\beta = 0.45$, 95% CI = [0.338, 0.562], $F(3, 788) = 127.722$, one-sided $p < 0.001$
With cluster-robust errors
$\beta = 0.45$, 95% CI = [0.293, 0.607], $F(3, 788) = 73.402$, one-sided $p < 0.001$
Model 1c, people before random after
$\beta = -0.453$, 95% CI = [-0.56, -0.346], $F(3, 824) = 133.479$, one-sided $p < 0.001$
With cluster-robust errors
$\beta = -0.453$, 95% CI = [-0.597, -0.309], $F(3, 824) = 69.548$, one-sided $p < 0.001$

Appendix F Social Value Orientation distributional data

Social Value Orientation distributional information is reported for participants from Studies 1 and 2. participants filled out an SVO slider task at the end of the experiments.

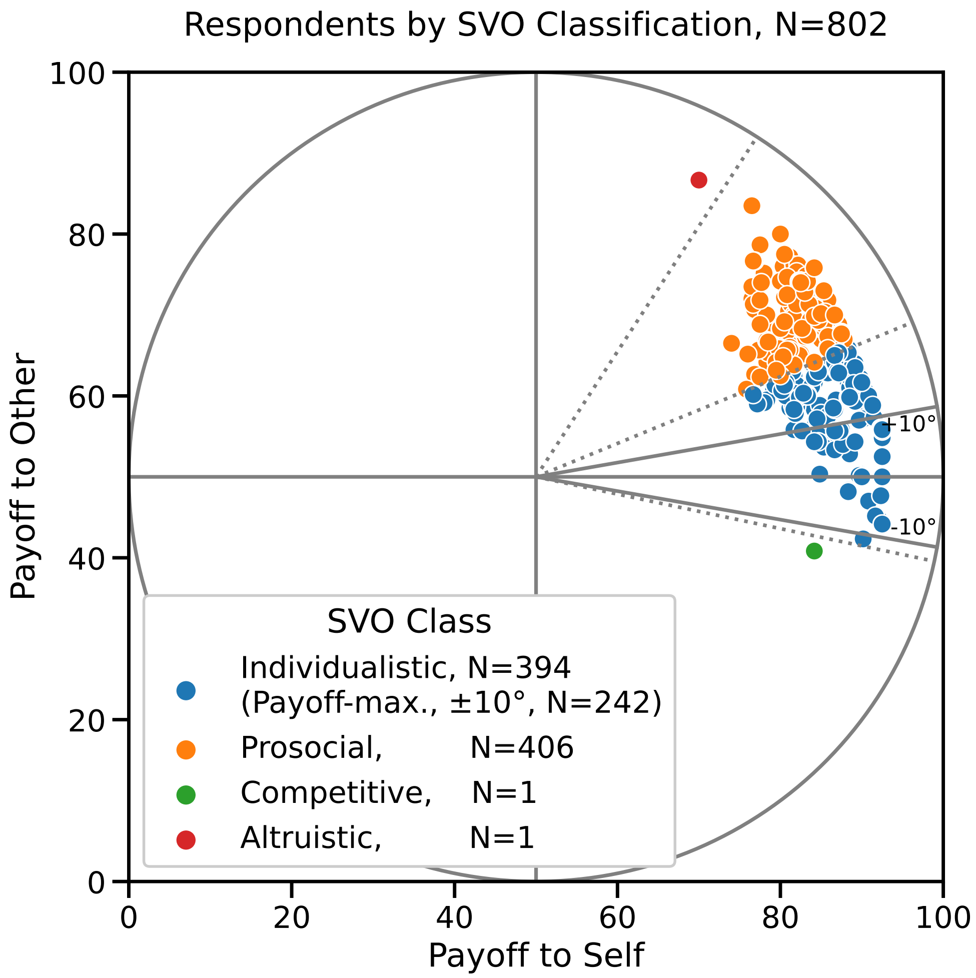

Fig. F4 Social Value Orientation distributional data for Study 1. Participants who passed comprehension checks.

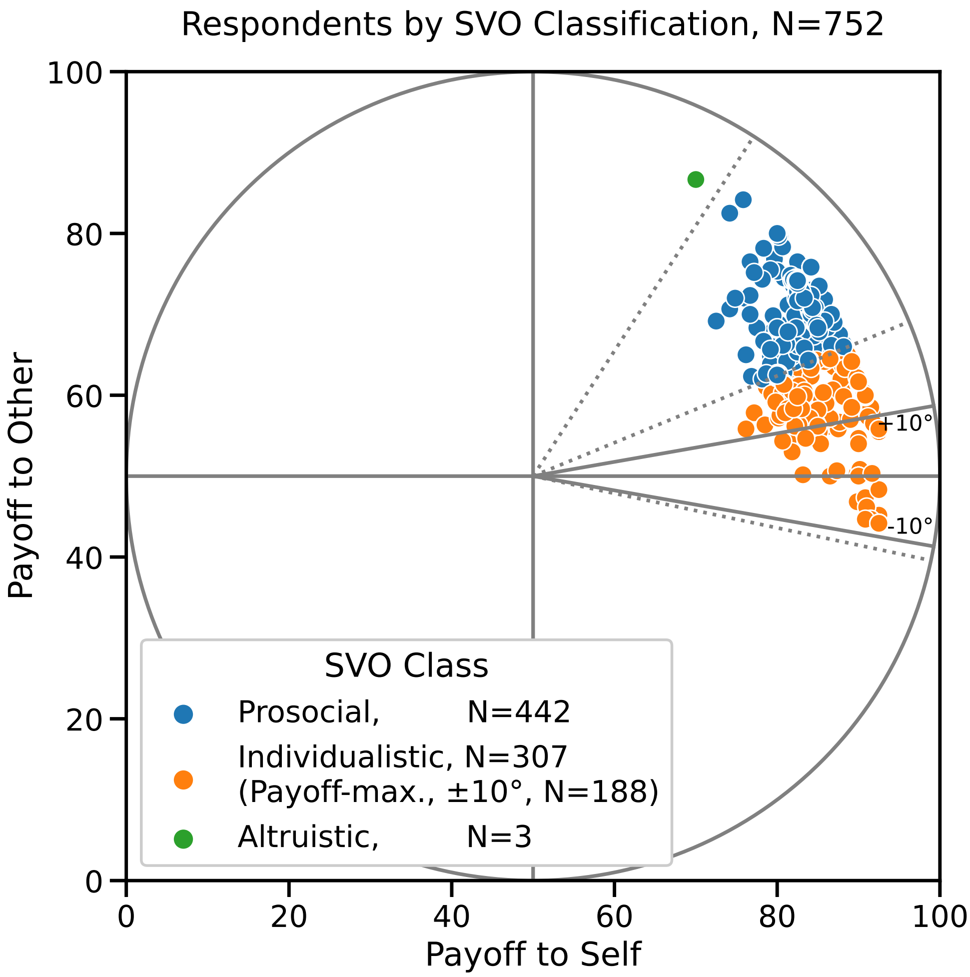

Fig. F5 Social Value Orientation distributional data for Study 2. Participants who passed comprehension checks.

References

[1] Henrich, J. & Muthukrishna, M. The Origins and Psychology of Human Cooperation. *Annual Review of Psychology* **72**, 207–240 (2021). URL https://www.annualreviews.org/doi/10.1146/annurev-psych-081920-042106.

[2] Rand, D. G. & Nowak, M. A. Human cooperation. *Trends in Cognitive Sciences* **17**, 413–425 (2013). URL https://linkinghub.elsevier.com/retrieve/pii/S1364661313001216.

[3] Zelmer, J. Linear Public Goods Experiments: A Meta-Analysis. *Experimental Economics* **6**, 299–310 (2003). URL http://link.springer.com/10.1023/A:1026277420119.

[4] Shafir, E. & Tversky, A. Thinking through uncertainty: Nonconsequential reasoning and choice. *Cognitive Psychology* **24**, 449–474 (1992). URL https://linkinghub.elsevier.com/retrieve/pii/001002859290015T. 00865.

[5] Cooper, R., DeJong, D. V., Forsythe, R. & Ross, T. W. Forward Induction in the Battle-of-the-Sexes Games. *The American Economic Review* **83**, 1303–1316 (1993). URL https://www.jstor.org/stable/2117562. Publisher: American Economic Association.

[6] Güth, W., Huck, S. & Rapoport, A. The limitations of the positional order effect: Can it support silent threats and non-equilibrium behavior? *Journal of Economic Behavior & Organization* **34**, 313–325 (1998). URL https://linkinghub.elsevier.com/retrieve/pii/S0167268197000577.

[7] Rapoport, A. Order of play in strategically equivalent games in extensive form. *International Journal of Game Theory* **26**, 113–136 (1997). URL http://link.springer.com/10.1007/BF01262516. 54 citations (Crossref) [2023-07-13].

[8] Weber, R. A., Camerer, C. F. & Knez, M. Timing and Virtual Observability in Ultimatum Bargaining and "Weak Link" Coordination Games. *Experimental Economics* **7**, 25–48 (2004).

[9] Budescu, D. V., Suleiman, R. & Rapoport, A. Positional Order and Group Size Effects in Resource Dilemmas with Uncertain Resources. *Organizational Behavior and Human Decision Processes* **61**, 225–238 (1995). URL https://linkinghub.elsevier.com/retrieve/pii/S0749597885710187. 59 citations (Crossref) [2023-08-31].

[10] Budescu, D. V., Au, W. T. & Chen, X.-P. Effects of Protocol of Play and Social Orientation on Behavior in Sequential Resource Dilemmas. *Organizational Behavior and Human Decision Processes* **69**, 179–193 (1997). URL https://www.sciencedirect.com/science/article/pii/S0749597897926844.

[11] Budescu, D. V. & Au, W. T. A model of sequential effects in common pool resource dilemmas. *Journal of Behavioral Decision Making* **15**, 37–63 (2002). URL https://onlinelibrary.wiley.com/doi/10.1002/bdm.402. 16 citations (Crossref) [2023-07-13].

[12] Chen, X.-P., Au, W. T. & Komorita, S. Sequential Choice in a Step-Level Public Goods Dilemma: The Effects of Criticality and Uncertainty. *Organizational Behavior and Human Decision Processes* **65**, 37–47 (1996). URL https://linkinghub.elsevier.com/retrieve/pii/S0749597896900035. 49 citations (Crossref) [2023-07-13].

[13] Quattrone, G. A. & Tversky, A. Causal versus diagnostic contingencies: On self-deception and on the voter's illusion. *Journal of Personality and Social Psychology* **46**, 237–248 (1984). 00479.

[14] Tversky, A. & Shafir, E. The Disjunction Effect in Choice under Uncertainty. *Psychological Science* **3**, 305–310 (1992). URL http://journals.sagepub.com/doi/10.1111/j.1467-9280.1992.tb00678.x.

[15] Langer, E. J. The illusion of control. *Journal of Personality and Social Psychology* **32**, 311–328 (1975). Place: US Publisher: American Psychological Association.

[16] Stefan, S. & David, D. 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**, 377–386 (2013). URL https://onlinelibrary.wiley.com/doi/10.1111/j.1559-1816.2013.01007.x.

[17] Masel, J. A Bayesian model of quasi-magical thinking can explain observed cooperation in the public good game. *Journal of Economic Behavior & Organization* **64**, 216–231 (2007). URL https://linkinghub.elsevier.com/retrieve/pii/S0167268106000977. 19 citations (Crossref) [2023-07-17].

[18] Daley, B. & Sadowski, P. Magical thinking: A representation result. *Theoretical Economics* **12**, 909–956 (2017). URL https://onlinelibrary.wiley.com/doi/10.3982/TE2099. 00008.

[19] Morris, M. W., Sim, D. L. & Girotto, V. 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**, 494–512 (1998). URL http://linkinghub.elsevier.com/retrieve/pii/S0022103198913610. 00090.

[20] Miller, D. T. & Gunasegaram, S. Temporal order and the perceived mutability of events: Implications for blame assignment. *Journal of Personality and Social Psychology* **59**, 1111–1118 (1990). URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.59.6.1111.

[21] Burns, Z. C., Caruso, E. M. & Bartels, D. M. Predicting premeditation: Future behavior is seen as more intentional than past behavior. *Journal of Experimental Psychology: General* **141**, 227–232 (2012). 00036.

[22] Croson, R. T. The Disjunction Effect and Reason-Based Choice in Games. *Organizational Behavior and Human Decision Processes* **80**, 118–133 (1999). URL https://linkinghub.elsevier.com/retrieve/pii/S0749597899928467. 00176.

[23] Hristova, E. & Grinberg, M. *Testing Two Explanations for the Disjunction Effect in Prisoner's Dilemma Games: Complexity and Quasi-Magical Thinking*, Vol. 32 (Cognitive Science Society, 2010). URL https://escholarship.org/uc/item/

3t20s2q7.

[24] Abele, S. & Ehrhart, K.-M. The timing effect in public good games. *Journal of Experimental Social Psychology* **41**, 470–481 (2005). URL https://linkinghub.elsevier.com/retrieve/pii/S0022103104001180. 00022.

[25] Figuières, C., Masclet, D. & Willinger, M. Vanishing Leadership and Declining Reciprocity in a Sequential Contributions Experiment. *Economic Inquiry* **50**, 567–584 (2012). URL https://onlinelibrary.wiley.com/doi/10.1111/j.1465-7295.2011.00415.x. 18 citations (Crossref) [2023-07-13].

[26] Robinson, A. E., Sloman, S. A., Hagmayer, Y. & Hertzog, C. K. Causality in Solving Economic Problems. *The Journal of Problem Solving* **3** (2010). URL https://docs.lib.purdue.edu/jps/vol3/iss1/6. 00006.

[27] Steiger, E.-M. & Zultan, R. See no evil: Information chains and reciprocity. *Journal of Public Economics* **109**, 1–12 (2014). URL https://linkinghub.elsevier.com/retrieve/pii/S0047272713002089. 13 citations (Crossref) [2023-07-13].

[28] Eichenseer, M. Leading-by-example in public goods experiments: What do we know? *The Leadership Quarterly* 101695 (2023). URL https://www.sciencedirect.com/science/article/pii/S1048984323000218.

[29] Murphy, R. O., Ackermann, K. A. & Handgraaf, M. Measuring Social Value Orientation. *SSRN Electronic Journal* (2011). URL http://www.ssrn.com/abstract=1804189.

[30] Dawes, R. M. Statistical criteria for establishing a truly false consensus effect. *Journal of Experimental Social Psychology* **25**, 1–17 (1989). URL https://linkinghub.elsevier.com/retrieve/pii/002210318990036X.

[31] Hoch, S. J. Perceived consensus and predictive accuracy: The pros and cons of projection. *Journal of Personality and Social Psychology* **53**, 221–234 (1987). URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.53.2.221.

[32] Tarantola, T., Kumaran, D., Dayan, P. & De Martino, B. Prior preferences beneficially influence social and non-social learning. *Nature Communications* **8**, 817 (2017). URL https://www.nature.com/articles/s41467-017-00826-8. Number: 1 Publisher: Nature Publishing Group.

[33] Bernheim, B. D. & Thomadsen, R. Memory and Anticipation. *The Economic Journal* **115**, 271–304 (2005). URL https://academic.oup.com/ej/article/115/503/271/5086053.

[34] Bodner, R. & Prelec, D. in *Self-signaling and diagnostic utility in everyday decision making* (eds Brocas, I. & Carrillo, J. D.) *The Psychology of Economic Decisions* (Oxford University Press, Oxford, [England] ; New York, 2003).

[35] Dhar, R. & Wertenbroch, K. Self-Signaling and the Costs and Benefits of Temptation in Consumer Choice. *Journal of Marketing Research* **49**, 15–25 (2012). URL http://journals.sagepub.com/doi/10.1509/jmr.10.0490. 00110.

[36] Mijovic-Prelec, D. & Prelec, D. Self-deception as self-signalling: a model and experimental evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 227–240 (2010). URL http://rstb.royalsocietypublishing.org/cgi/doi/10.1098/rstb.2009.0218. 00109.

[37] Gibbard, A. & Harper, W. L. in *Counterfactuals and Two Kinds of Expected Utility* (eds Harper, W. L., Stalnaker, R. & Pearce, G.) *Ifs: Conditionals, belief, decision, chance and time* 153–190 (Springer Netherlands, Dordrecht, 1978). URL http://link.springer.com/10.1007/978-94-009-9117-0_8.

[38] Roemer, J. E. Kantian Equilibrium. *The Scandinavian Journal of Economics* **112**, 1–24 (2010). URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9442.2009.01592.x. 77 citations (Crossref) [2023-07-17] _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9442.2009.01592.x.

[39] Roemer, J. E. Kantian optimization: A microfoundation for cooperation. *Journal of Public Economics* **127**, 45–57 (2015). URL https://www.sciencedirect.com/science/article/pii/S0047272714000607. 52 citations (Crossref) [2023-07-17].

[40] Levine, S., Kleiman-Weiner, M., Schulz, L., Tenenbaum, J. & Cushman, F. The logic of universalization guides moral judgment. *Proceedings of the National Academy of Sciences* **117**, 26158–26169 (2020). URL https://pnas.org/doi/full/10.1073/pnas.2014505117. 12 citations (Crossref) [2023-07-17].

[41] Hauser, D. J. *et al.* Evaluating CloudResearch's Approved Group as a solution for problematic data quality on MTurk. *Behavior Research Methods* **55**, 3953–3964 (2023). URL https://doi.org/10.3758/s13428-022-01999-x.

[42] Douglas, B. D., Ewell, P. J. & Brauer, M. Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLOS ONE* **18**, e0279720 (2023). URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0279720. Publisher: Public Library of Science.

[43] Chen, D. L., Schonger, M. & Wickens, C. oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance* **9**, 88–97 (2016). URL http://www.sciencedirect.com/science/article/pii/S2214635016000101. 00298.