Acting *as if*: Self-interested players act as if others will mirror their moves

Matthew Cashman[1], Drazen Prelec[1]*
[1]Massachusetts Institute of Technology

Word count: 4,959 excluding abstract, methods, references, and figure legends

Authors Note:

Matthew Cashman, Sloan School of Management, MIT; Drazen Prelec, Sloan School of Management, Department of Economics, Department of Brain and Cognitive Sciences, MIT

Author Contributions:

Matthew Cashman contributed to the design and implementation of experiments, to the analysis plan, and to drafting the manuscript. Drazen Prelec contributed to the design of the experiments and to drafting the manuscript.

**Please address correspondence to Matthew Cashman; matt@cashman.science*

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. We interpret this as reflecting differences in players' sense of impact on the collective outcome: Subjective impact is highest when other players have not yet moved. Three results provide further support: (1) The 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.

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 (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 for a meta-analysis). 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 decisions others make, but it is always better for the group if everyone cooperates.

There may, however, be circumstances in which even self-interested players who are only trying to maximize their own payouts end up cooperating. Such a phenomenon would point the way towards interventions that increase cooperation even among the self-interested in addition to illuminating how such decisions are made. It could be the case that self-interested players are cooperating because cooperation maximizes their individual payoff on the assumption that those moving after them are more likely to copy their move. In that case they behave *as if* they can influence others' moves without any communication. *Quasi-magical thinking* (QMT, (Shafir & Tversky, 1992) 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. We do not bring data to bear on the psychological phenomena underlying quasi-magical thinking here; rather, we investigate whether players in sequential games without observation are acting *as if.* We add to the existing literature by demonstrating positional order effects, ensuring players are maximizing payouts, considering the implications of players' predictions, and by testing the effects of breaking the "magical" causal connection between a player and those moving after him.

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 coordinated based on the order of play itself absent any information about others. Cooper et al. (1993) is the genesis of a modest body of work looking at positional order effects in coordination games such as Battle of the Sexes (Güth et al., 1998; Rapoport, 1997; Weber et al., 2004). Positional order effects are effects driven solely by common knowledge about order of play in games with no observation of others' moves.There are two equilibria in these games, one better for the first-mover and the other better for the second-mover. 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 (Budescu et al., 1995, 1997; Budescu & Au, 2002) games where players try to request as much as possible—but not too much in total—from a fixed pool of resources, and a related game called the Step-Level PGG, where players receive nothing at all from the public good if the total amount contributed is too low (Chen et al., 1996; Rapoport, 1997). 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 resource dilemmas or step-level PGGs have 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 group is on average, best off if everyone cooperates, but you are best off if you defect. The key difference between games with an element of coordination and those without is that in games with no reward for coordination there is no reason to condition your play on others' decisions, and therefore no obvious way to use order to inform play. However, there is evidence that suggests causal thinking about others is active even in situations without clear causal linkages rewarding coordination, and uncertainty about the state of the world seems to activate this sort of reasoning.

In an early study, Quattrone & 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 (38 undergraduates). They report that subjects who are performing a task that involves 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 subjects are changing the data they themselves produce in order to receive good news in apparent disregard of the causal relationship.

In related work, Shafir & Tversky (1992) explore nonconsequential reasoning—by which they mean reasoning that at least appears to either not produce estimates of the consequences of an action, or which ignores the consequences of that action. This class of decisions violate the sure thing principle, which states that if X is preferred to Y under all states of the world, then X should still be preferred to Y even if the state of the world is unknown. For instance, there are many people who would prefer to pay for a vacation to Hawaii in the event that they pass an exam *and* in the event that they fail, but who would also prefer *not* to buy in the case where the outcome of the exam is unknown (Tversky & Shafir, 1992). They refer to this pattern of events as "accept when win, accept when lose, reject when do not know" and refer to it as the "disjunction effect". In an experiment using the Prisoner's Dilemma (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 similar to the illusion of control (Langer, 1975; Stefan & David, 2013 for a review) and control heuristics (Thompson et al., 1998), which offer differing accounts of why the general phenomena might emerge.

Masel (2007) offers a formalization of quasi-magical thinking where players, upon observing additional information during the game, update their prior distributions in the usual fashion—one's own behavior being just another data point. Daley & Sadowski (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 importantly different flavors of uncertainty at play here: "closed fates" uncertainty about a counterpart's move when the move is not known to the player has been made and is therefore fixed, and "open fates" uncertainty about a counterpart's move that has yet to be made at all or which is presently being made (Morris et al., 1998). Miller and Gunasegaram (1990) demonstrated that, while events in the past are considered fixed, future events are treated as mutable. Moreover, future actions are perceived as more intentional and blameworthy than otherwise identical past actions (Burns et al., 2012).

Subsequent work on social dilemmas without coordination 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 QMT, 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, rematching subjects randomly after each round and doing so in small pools of students from the same university. Many also use Prisoners' Dilemmas, a two-person version of a PGG. The PGG enables ruling out any specific first-mover/leader effects (Eichenseer, 2023) or end-of-game effects (Figuières et al., 2012) that are distinct from effects driven merely by sequential order given PGGs accommodate an arbitrary number of players. In addition, many subjects are probably not even trying to maximize their own payoffs in these tasks—limiting what we can infer based on play. Budescu et al. (1997) report that 47% of their subjects are classified as "cooperative" (maximize joint own + other gains) and 2% as "altruistic" (maximizing others' gains); that is to say, half of their subjects 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 distribution of behavior as they come to the game, it is a substantial problem if we are making the assumption that players are attempting to maximizing profits.

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 that there is a decline in cooperation with increasing order ("order effect", earlier movers cooperate more than later movers) in a three-person PGG and that this effect is driven by respondents who are "Individualistic" on the Social Value Orientation (SVO, Murphy et al., 2011) scale. SVO is a measure of willingness to give up gains in order to benefit others. In the SVO battery, respondents play a series of incentivized games similar to Dictator games where they allocate funds between themselves and someone else. In these games respondents can choose to forego gains (or even pay costs) to help or hurt the other player. If respondents 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 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 order 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. Our chief interest is respondents who are trying to maximize their own financial rewards, but those who arrived at the experiment clearly self-interested 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 order effect observed in respondents who arrived at earlier studies already self-interested.

Study 4 deploys the technique from Study 3 to ask whether we would observe an order effect in the case where all players after the focal player have their contribution decisions delegated to a random process. If an order effect is present when random movers are behind the focal player, but absent when random movers are to move ahead of 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 is at play.

All studies are real-time games which are 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 game compensation, 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 respondents 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. When players are playing an SPGG, they are playing with the groupmates they chatted with in real time. All experiments have simultaneous-play PGG control conditions, and all players pass familiarization tasks and comprehension checks. Data from players who miss a single comprehension check is excluded from all analyses, but players who fail a comprehension check still complete the task. Because they are part of a group that is playing in real time, their moves are necessary to calculate payoffs. All experiments share the following three up-front comprehension and attention check questions:

> Q1: *Do any of the other players* ***know how much YOU decide to contribute?***
>
> Q2: *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?*
>
> Q3: *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. Later studies incorporate more extensive training and comprehension check regimes. All statistical tests were two-tailed.

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. Along with standard measures, we estimated participants' interpersonal utility tradeoffs using the SVO scale, which divides almost all[1] participants into two categories, Individualistic and Prosocial. Order effects should be more pronounced for participants who are primarily interested in their own payoff (Individualists) if participants are acting *as if*. In contrast, Prosocials should be less sensitive to order of play, as altruistic motivation should not be biased toward future players.

We do not meet the pre-registered threshold for an order effect in this study. This may be due to the study being under-powered to detect an effect. We do find evidence for the preregistered order effect when pooling pilot data with data collected post-preregistration. First-movers contribute more than later players, though we do not resolve a difference between second- and third-movers and therefore do not meet the conditions of the preregistration which specified backwards-difference coding. A linear regression of contribution on move order yields a significant negative slope, $\beta = -0.042$, 95% CI = [-0.081, -0.003], $F(1, 780) = 4.7$, $p = 0.03$. First-movers contribute more than second-movers ($p=.013$) and more than third-movers ($p=.031$). The difference between second- and third-movers' contributions is not significant.

We find support for the preregistered prediction that the order effect is concentrated among participants classified as Individualistic in the SVO task. Participants classified as Prosocial exhibit no significant differences in contribution

[1] Nearly all subjects were SVO classified as Individualistic or Prosocial; two respondents were classified as Altruistic, and two as Competitive. These respondents' data are excluded from analyses.

levels as function of order, while we do see a difference between the first-mover data and grouped second- and third-mover contributions (β = -0.142, 95% CI = [-0.247, -0.037], $F(1, 288) = 7.104$, $p = 0.008$) among Individualistic players. As with the aggregated data, we do not see the hypothesized difference between positions two and three among respondents SVO-classified as Individualistic.

In addition, we find some support for the pre-registered prediction that correlations between a player's own move and her predictions of future players' moves are stronger going forward in time vs. backwards. The interaction term in the pre-registered 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 Individualistic players only a significant equation is found, β = 17.251, 95% CI = [5.176, 29.325], $F(3, 576) = 97.587$, $p = 0.005$. There is no effect among Prosocial players, β = 0.106, 95% CI = [-12.953, 13.165], $F(3, 614) = 98.503$, $p = 0.9873$

2 Study 2: 5-Person Sequential Public Goods Game

Study 2 was a sequential 5-person PGG with no observation where respondents were classified based on an SVO task performed at the end of the study. In Study 2 we do not meet our pre-registered threshold to detect an order effect for the interaction between order and SVO category. A programming error meant that the time Player 1 and Player 2 had to make a decision was not correct, sometimes being shorter than for other players and sometimes close to zero. Once players with shortened time are excluded, the study is under-powered to detect this effect.

If we restrict our analysis to all players whose SVO degree measure was +/- 10° the predicted pattern appears despite the anomaly with positions 1 and 2. Perfect self-interested play is within +/- 7.82° SVO, and since slider input devices introduce some 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. The pre-registered analysis applied to these most clearly self-interested players yields a significant regression equation, β = -0.052, 95% CI = [-0.094, -0.01], $F(1, 154) = 5.992$, $p = 0.015$. The pre-registered prediction that the partial correlation between one's own contribution and prediction of others' contributions, controlling for population prediction, is stronger going forward is well-supported. For the forward direction we find n = 1092, Pearson's r = 0.39, 95% CI = [0.34, 0.44], $p < 0.001$, and for backwards in time n = 1304, Pearson's r = 0.28, 95% CI = [0.23, 0.33], $p < 0.001$. A bootstrapped analysis of the difference, 0.11, yields a 95% CI = [0.02, 0.20] and p=0.019. Study 1 and Study 2 were sufficient evidence of positional order effects among the clearly payoff-maximizing to justify moving on to Study 3.

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

Study 3 tests whether prompting all subjects to maximize personal payoffs generates a positional order effect similar to that observed among subjects who arrive at the experiment already trying to maximize their own payoffs. The main difference from Study 2 is that respondents did not perform the SVO filtering task. Instead, respondents 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 but is otherwise the same as 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 (order = 5). This condition was incorporated to control for the possibility of effects dependent on time spent waiting. While waiting, respondents are shown the task's standard wait screen which incorporates the option to play a simple game to keep respondents engaged with the task. Second, Study 3 incorporates an interactive practice game after the instructions and comprehension questions. This practice game asks respondents to calculate the correct answers to questions about payoffs for hypothetical players in a PGG. Respondents are paid for correct answers, and they can make multiple attempts at any given question, limited only by time. Third, participants in Study 3 move in lock-step with one another. Each page in the study takes an allotted amount of time no matter the respondent's behavior. This is to ensure that information cannot leak to other players in one's group via response times. For instance, Player 2 might notice that Player 1 made a decision rather quickly if Player 1 is allowed to advance from the Contribution page as quickly as she likes, since Player 1's advance triggers Player 2's decision period. Finally, Study 3 incorporates an improved up-front English fluency check that relies on a native speaker's ability to quickly complete idioms in order to ensure respondents 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 using data from respondents who pass comprehension checks detects the interaction effect ($\beta = -0.189$, 95% CI = [-0.294, -0.074], $F(3, 78) = 5.038$, $p = 0.006$ for the interaction). Respondents 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.

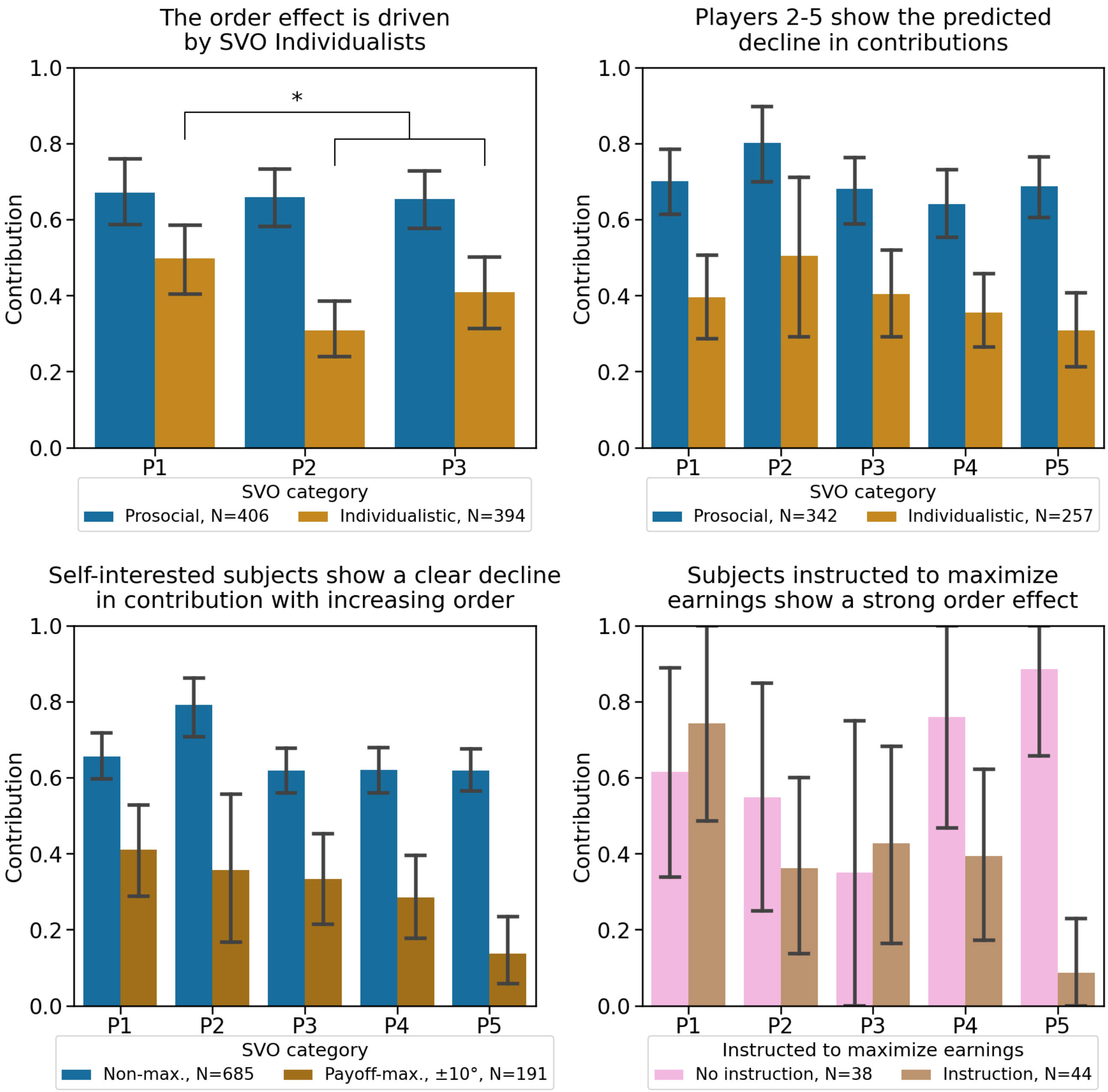

Figure 1: [Upper left] In Study 1, change in contribution with order is driven by subjects SVO-classified as Individualistic. [Upper right] SVO-individualist players show a decline in contribution with increasing order from Players 2-5. Study 2 suffered from technical problems with Players 1 and 2. [Lower left] Payoff-maximizing players show the predicted decline of contribution to the public good with increasing order in Study 2. [Lower right] Study 3 shows the hypothesized decline with order among those who were merely instructed to be greedy.

All subjects who passed comprehension checks, 95% CIs.

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 respondents are either told that every player *before* them has their contribution determined randomly ("Random Before"), or that every player moving *after* them has 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 graphical representations similar to that shown in *Figure 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, since there is no one before Player 1 and no one after Player 5.

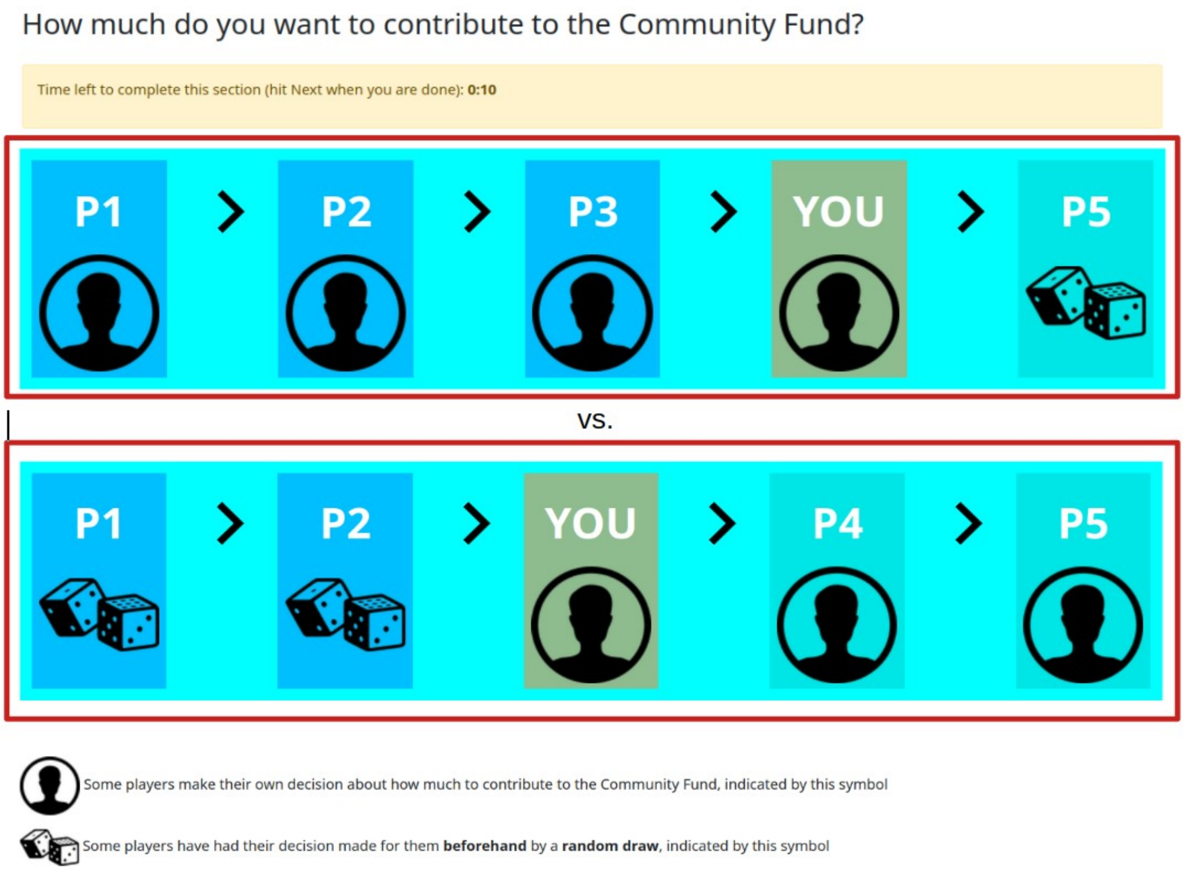

Figure 2: The stimuli on the Contribution page are shown in two conditions, in red boxes: above, Random After for Player 4 and below, Random Before for Player 3. Players see a graphical representation of their position relative to other players that clearly conveys which players are having their moves made by a random process. This is in addition to a previous screen that explains how some players are having their moves made for them by random processes.

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 ~ order * random_before + wealth, differing from previous analyses in that it controls for a measure of wealth, finds the effect. A significant regression equation was found, β = -0.079, 95% CI = [-0.134, -0.022], $F(4, 435) = 3.946$, $p = 0.0059$ for the interaction order * random_before. We also find a significant equation not controlling for wealth, β = -0.081, 95% CI = [-0.136, -0.024], $F(3, 436) = 4.276$, $p = 0.005$. Among players told the opposite, that everyone moving after them has their move made randomly, we observe no positional order effect. 75% of players in this experiment contribute either 100% or 0% of their endowment, and the effect size and direction are preserved in this subset, β = -0.096, 95% CI = [-0.163, -0.028], $F(4, 354) = 3.849$, $p = 0.006$ for the interaction. 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% of players who passed the initial checks), we observe a larger effect size (β = -0.101, 95% CI = [-0.158, -0.043], $F(4, 359) = 4.356$, $p = 0.0009$ for the interaction). This gives further reason to believe that the effect is concentrated among respondents who truly understand the game. We do not observe a difference between the two simultaneous-play control conditions, one with no delay and one with a delay similar to that which Player 5 experiences before moving, which rules out the effects being due to mere time in the experiment.

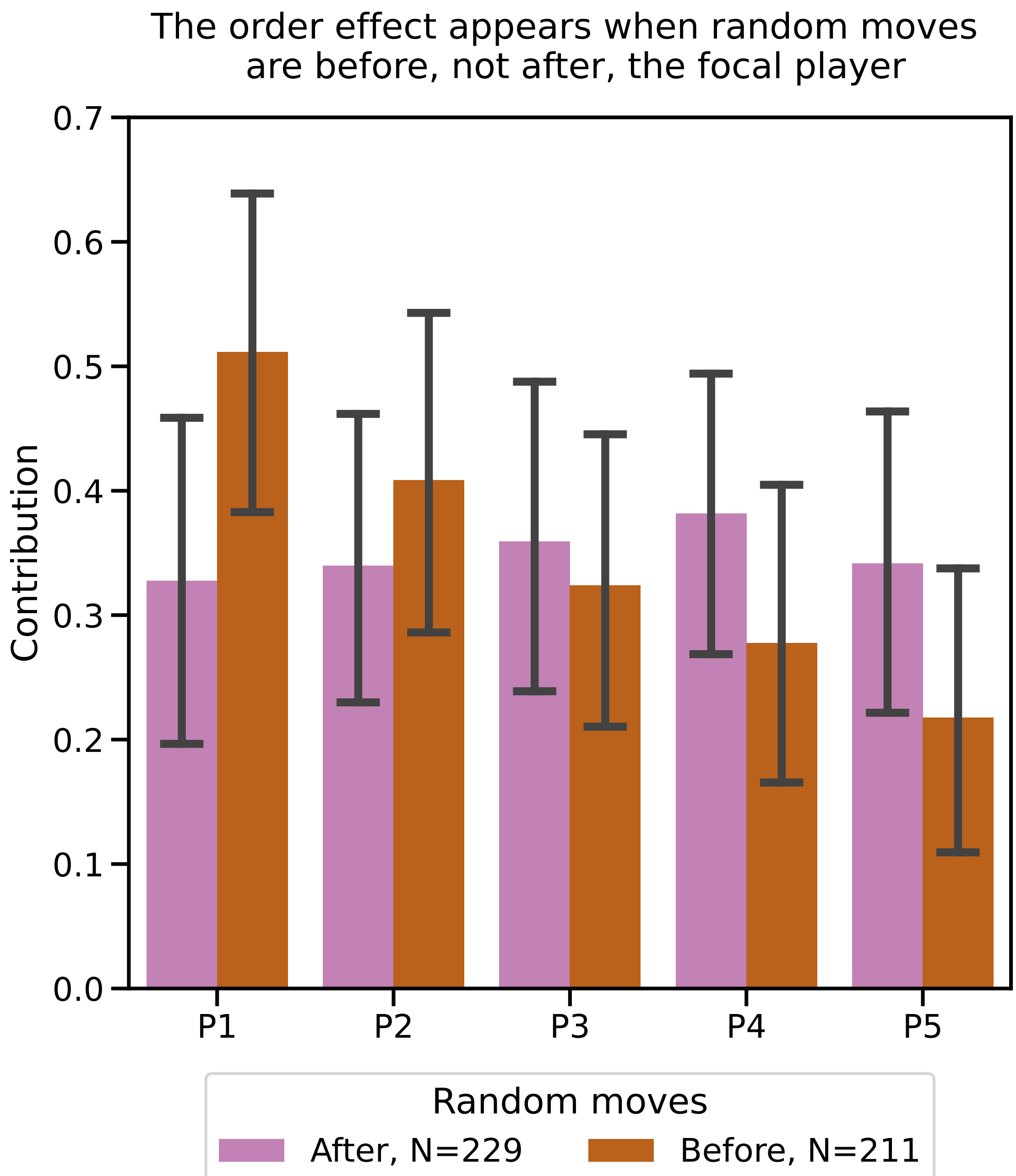

Figure 3: Study 4 shows a decline in contribution to the public good among players who are told that all players moving after them are making their own moves, and all players moving before them are having their moves made randomly. No effect is observed among players who are told that everyone moving after them has a move selected at random. Subjects who passed comprehension checks. 95% CIs.

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 (in this case the effect would propagate backwards in time), but *if* I cooperate, others will cooperate. It is also the case that players are willing to bet that people moving after them will make a move that is more similar to theirs than will the people moving before them, which would be expected in the case that players are acting *as if*. These results are consistent with quasi-magical thinking, and they make it clear that the distinction between open and closed fates is important for behavior. 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. 80% of subjects contribute either 0 or 100% of their endowment, and the positional order effect is stronger in this subset. A formalization of this model is included in the appendix.

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 focal player assumes some similarity between himself and the other players it may seem reasonable to look to his own behavior as a source of information 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). Social projection 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, and that the effect is conditional on (a) the signal being costly (since signals that are too easy to generate are not informative) and (b) some prior uncertainty about the characteristics (since being quite sure about these types means self-signals are uninformative in comparison to what is already known) (Bernheim & Thomadsen, 2005; Bodner & Prelec, 2003; Dhar & Wertenbroch, 2012; Mijovic-Prelec & Prelec, 2010). 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 Quattrone & Tversky (1984). However, like with social projection, the usual formulation of self-signaling does not naturally provide a direction in time for the effect. On the usual formalization it is possible to self-signal about open and closed fates, and so an explanation of why subjects only consider open fates would be required. This process of maximizing "news value" is also reminiscent of Evidential Decision Theory (Gibbard & Harper, 1978).

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? 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. The fact that this question occurs in the moral domain may imply that the moral phenomenon is a special case of a more general strategy that we observe 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.

Finally, whatever the mechanism, understanding a means by which 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: if I vote then it is more likely that other people will vote; 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 there are people deciding for themselves to contribute—or not—at a later time.

Methods

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

A convenience sample provided by Amazon Mechanical Turk (MTurk) was selected for all experiments because it is a reasonable approximation of American adults for our purposes. Studies 3 and 4 used a panel filtered by Cloud Research due to declines in the quality of MTurk samples. 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 covariates of interest. All experiments also involved extensive training and comprehension checks. Data from respondents who failed one or more pre-play comprehension check questions was excluded. All experiments are real-time online group tasks, where respondents 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.

1 Study 1

Participants. 1668 U.S.-based participants from Amazon Mechanical Turk completed the study. Median total pay per respondent (including bonuses for accurate predictions) is $3.16 (*SD* = 0.90), yielding an hourly rate of $18.46 per hour at 10 minutes' duration (SD = 8.38). Of 1668 respondents, 69% (1151) passed all of the comprehension check questions. Data from batches 1 and 8 were excluded due to technical problems resulting in server crashes during the experiment. Analysis is limited to the 60% (1002 total; 800 sequential) responses which passed comprehension checks and were not in batches 1 or 8. To estimate the sample size required, we performed a power analysis via simulation using pilot data.

Materials and procedure. Study 1 is a one-shot 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, 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, respondents 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.

Respondents 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? As with the Prisoner’s Dilemma, 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 supplemental online materials). Players in the simultaneous condition do not see any indication of sequence since they are moving simultaneously. Respondents then complete prediction questions, and then a Social Value Orientation (SVO) slider battery (Murphy et al., 2011; code based on Bakker, 2016/2019)[2]. The SVO battery measures preferences for how to allocate resources between oneself and others. The standard battery categorizes respondents 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 pre-registered analysis used to investigate the impact of order on contribution is a linear regression contribution ~ 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 pre-registered analysis for predictions of others' moves going forward is the prediction ~ own response interacted with a binary future vs. past variable.

2 Study 2

Participants. 1089 U.S.-based participants from Amazon Mechanical Turk completed the study. Median total pay per respondent (including bonuses for accurate predictions) is \$3.40 (*SD* = 0.53), yielding an hourly rate of \$11.18 per hour at 18 minutes' duration (SD = 3.83). Of 1089 respondents, 66% (720 total, 599 sequential)

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

passed all of the up-front comprehension check questions. Time on the decision-making page for players 1 and 2 was variable due to a programming error, and data from respondents who got less than the designed time was excluded. 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 identical to Study 1, with the exception that there are five players rather than three, that the up-front chat was 90 instead of 30 seconds, and that respondents complete the SVO slider battery before the PGG.

Analysis. The preregistered analysis used to investigate the impact of order on contribution is a simple OLS linear regression contribution ~ order (excluding simultaneous participants). Backwards-difference coding (as specified in Study 1) would 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.

3 Study 3

Participants. 86 U.S.-based participants from Amazon Mechanical Turk completed the study via Cloud Research. Median total pay per respondent (including bonuses for accurate predictions) is \$3.99 (*SD* = 1.25), yielding an hourly rate of \$15.99 per hour at 15.2 minutes duration (SD = 4.39). Of the 86 respondents who completed the task, 82 (95.0%) passed all of the up-front comprehension check questions.

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. Respondents 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 ~ order * instruct_or_no, with instruct_or_no being a binary indicator of whether or not respondents were instructed to be self-interested.

4 Study 4

Participants. 539 U.S.-based participants from Amazon Mechanical Turk via Cloud Research completed the study, with 440 in sequential conditions and 99 in simultaneous conditions. Median total pay per respondent (including bonuses for accurate predictions) is $4.24 (*SD* = 1.19), yielding an hourly rate of $16.13 per hour at 16.0 minutes duration (SD = 3.97). Of 539 respondents, 440 (82%) passed all of the up-front comprehension check questions. 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"). 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 ~ 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.

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

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). Self-signaling and diagnostic utility in everyday decision making. In I. Brocas & J. D. Carrillo (Eds.), *The Psychology of Economic Decisions*. Oxford University Press.

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

Chen, X.-P., Au, W. T., & Komorita, S. 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.

Croson, R. T. A. (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

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

Eichenseer, M. (2023). Leading-by-example in public goods experiments: What do we know? *The Leadership Quarterly*, 101695. https://doi.org/10.1016/j.leaqua.2023.101695

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

Gibbard, A., & Harper, W. L. (1978). Counterfactuals and two kinds of expected utility. In *Ifs: Conditionals, belief, decision, chance and time* (pp. 153–190). Springer.

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

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

Hristova, E., & Grinberg, M. (2010). *Testing Two Explanations for the Disjunction Effect in Prisoner's Dilemma Games: Complexity and Quasi-Magical Thinking*. Annual Meeting of the Cognitive Science Society.

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

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. H., & 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

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

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), Article 1. https://doi.org/10.1038/s41467-017-00826-8

Thompson, S. C., Armstrong, W., & Thomas, C. (1998). Illusions of control, underestimations, and accuracy: A control heuristic explanation. *Psychological Bulletin*, *123*(2), 143–161. https://doi.org/10.1037/0033-2909.123.2.143

Tversky, A., & Shafir, E. (1992). The Disjunction Effect in Choice under Uncertainty. *Psychological Science*, *3*(5), 305–310. https://doi.org/10/b2v8xx

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.

Zelmer, J. (2003). Linear Public Goods Experiments: A Meta-Analysis. *Experimental Economics*, *6*(3), 299–310. https://doi.org/10.1023/A:1026277420119

Appendix

1 Preregistrations

Please note that these pre-registrations reference a quadratic effect, which is unrelated to the order effect and will be the subject of a separate paper.

Study 1:
OSF preregistration: https://osf.io/3vsxk

Anonymous link to registration for review: https://osf.io/3vsxk/?view_only=bf35d2d3d39d48b68869c2cf78bf8e2b

Study 2:
OSF preregistration: https://osf.io/gw8nc

Anonymous link to registration for review: https://osf.io/gw8nc/?view_only=aa0c4825dac4469a82f0156b77390e3c

Study 3:
no preregistration

Study 4:
OSF preregistration: https://osf.io/3kepm

Anonymous link to registration for review: https://osf.io/3kepm/?view_only=614de27fdf4b40a0bad48847f32c879d

2 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, i*s 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:

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

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

3 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.

3.1 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, $\mathrm{i{=}1,\ \ldots,\ n}$. Let a_i denote the contribution of player i, $0 \leq a_i \leq 1$, and p_i the payoff to player i.

$$p_i = 1 - a_i + \frac{m}{n} \sum_{k=1}^{n} a_k \qquad \text{(Equation 2)}$$

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

$$u_i\left(a_1, \ldots, a_n\right) = \left(1 - s_i\right) p_i + s_i m a_i \qquad \text{(Equation 3)}$$

where p_i is determined by the game formula, eq. 1 . 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, $u_i = m a_i$. 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: (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.

3.2 Decision dependent expectations

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

$$a_i = 1 \Longleftrightarrow E\left[u_i \mid a_i = 1, s_i\right] > E\left[u_i \mid a_i = 0, s_i\right] (1) \qquad \text{(Equation 4)}$$

A player knows the value of their prosocial parameter and hence also knows the utility function in eq. 1. If he were just a spectator, not making a decision, his expectation of the contribution a_k of another, randomly selected player k would exhibit projection, along the lines of Bayesian updating. The simplest version of such updating is linear:

$$E\left[a_k \mid s_i\right] = b + cs_i(1) \qquad \text{(Equation 5)}$$

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

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

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

where $\mathrm{E}[a_k \mid s_i] = b + cs_i$ from eq. 4 is substituted in the final line.

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

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

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

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

The decision criterion in eq. 3 can be expressed as:

$$E[u_i a_i = 1, s_i] - E[u_i a_i = 0, s_i] = (1-s_i)E[p_i a_i = 1, s_i] + s_i m - (1-s_i)E[p_i a_i = 0, s_i]$$

$$= (1-s_i)(E[p_i a_i = 1, s_i] - E[p_i a_i = 0, s_i]) + s_i m$$

$$= (1-s_i)(-1 + mnE[k = 1^n a_k a_i = 1, s_i] - mnE[k = 1^n a_k a_i = 0, s_i]) + s_i m$$

where the first line follows from equation 2 and the third line from equation 3.

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

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

Substituting into the criterion,

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

(Equation 6)

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:

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

(Equation 7)

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 respondent 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 magical thinking 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 eq. 6) 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.