Order of play matters for cooperation among selfish agents even when moves are unobserved

Matthew Cashman[1*], Drazen Prelec[1]

[1]Massachusetts Institute of Technology

Word count: 4,440 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 online 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

Theories of cooperation often bear on why people choose to cooperate, but what if there were circumstances in which people who are trying to be selfish end up cooperating anyway? We investigate this possibility with real-time, one-shot Public Goods Games where players move sequentially and do not observe others' moves. We find that cooperation among players trying to maximize their own payoffs increases with the number of people yet to move. Selfish players may be maximizing their payouts on the assumption that everyone moving after them will move as they did. In the absence of other information, it may be that subjects feel that their unobserved actions are informative about the behavior of others who are yet to move. A mechanism that induces cooperation among overtly selfish agents is of interest more broadly, with possible applications to policy.

Main

Social cooperation without external monitoring is widely regarded as a 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 (Rand & Nowak, 2013; Henrich & Muthukrishna, 2021). 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 tension between the benefits accruing to the group via cooperation and the benefits accruing to the individual via defection captures the essence of the types of cooperation problems humans solve on a daily basis.

There may, however, be circumstances in which even selfish players—players for whom there is no tradeoff, who are just trying to maximize their own payouts—end up acting prosocially. We investigate this possibility with a subtle variation on the classic PGG, a sequential game. If players are forced to move one after another, with no knowledge of each others' moves, and we observe an increase in cooperation among players who move earlier in the sequence (an increase borne of the mere knowledge that others will be acting in the same game after you), that is interesting for two reasons. First, knowing how to induce more cooperation is useful—and doubly so in the particular case of people who are trying to do best by themselves and who are indifferent to the costs to others. Second, it gives us a window into how these decisions are being made. It could be the case that selfish players are cooperating because they are calculating an expected payoff on the assumption that everyone moving subsequent to them will make the same move they have, providing valuable insight into possible mechanisms by which this cooperation emerges.

The games used here are "sequential" in that players move one after another, and players' moves are unobserved in that there is no information flow between players: any given player knows nothing about the decisions players who have already moved have made. For example, a five-person PGG is sequential with unobserved moves when the five players move one after another, but each player knows nothing about what the other players have done. Traditional game theory suggests that the order in which players in the same game are moving is irrelevant as long as they don't know anything about what moves others make, and therefore the distinction between events that have happened and have not happened—even if they are unknown—is lost. Because of this, the formal expression of the simultaneous version of a game is identical to that for a sequential version, where players move one after another.

There are several bodies of work that investigate the effects of agents acting one after another with observability. There is a rich literature investigating team effects, in which individual agents acting as part of a team optimize for the team's success, rather than for their own best interests, under certain conditions ((See Colman & Gold, 2018 for a review) Similarly, there is substantial work investigating leader- and follower- effects in games which have some element of sequential moves (e.g. Gächter et al., 2010), as well as sequence effects in economic games with observed moves, both one-shot and repeated (e.g. Figuières et al., 2012). Critically, the theoretical accounts from these literatures rely on the flow of information about earlier players' moves from early players to later players, and they rely on earlier players knowing that will happen. Evidence for order effects in the absence of any information flow at all may, then, be consequential for these theories.

Relevant to games without information flow, there is a related body of work on the illusion of control: the idea that there are circumstances under which people over-estimate the amount of control they have over a situation. This literature asserts that people are motivated to believe they have more control than they do over a situation, especially when the lack of control should be logical or observable. The feeling of control when there is none may be due to social motivations or to the desire to preserve self-esteem (Stefan & David, 2013 for a review). This provides evidence from other domains for the type of behavior that may lead to effects in sequential games with no information flow.

The literature on order effects in economic games *without* observation, games similar to those investigated here, is modest and inconclusive. Morris et al. (1998) report more cooperation among first-movers (66%) than among second-movers (53%) in a Prisoners' Dilemma, but do not replicate the effect in a subsequent larger experiment. Abele & Erhart (2005) describe two Public Goods Game experiments with undergraduates (N=86 and N=192), and Figuières et al. describe a voluntary contribution game (2012, N=32); both report no order effect for sequential conditions with no information flow. Budescu and Au (2002) describe more defection in earlier movers in a common pool resource game, and Weber et al. (2004) describe order effects in coordination games (but see Li, 2007).

Shafir and Tversky (1992) found that informing Prisoner's Dilemma players about their teammate's move reduced cooperation even if that teammate had chosen to cooperate. The increase in cooperation under uncertainty may reflect a feeling of influence over teammate's unknown action, which would be consistent with some sort of magical thinking. However, their study did not measure projection, nor indicate to participants whether the other player had already moved. Related, Chen and Zhong (2020) report that uncertainty about a dice game results in more moral behavior, which may be relevant to observed behavior in sequential games if uncertainty is fundamentally about how many people have yet to move, vs. the total number of other players. In a sequential game with no information flow there is no change in uncertainty with order in a formal sense. It may be interesting to note that, while any given player knows nothing about his groupmates' moves, those moves (once made) are known in a more general sense. Having been determined, they are at minimum known to others.

In a standard PGG *n* players are each given an endowment *e*, and asked to decide what proportion of the endowment to contribute to the public good, *a*, from nothing to all of it. 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 distributed *evenly* among *all* the players—even those who chose to contribute less or 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 enact retribution or reward others for their actions.

We will call the player of interest in a sequence of players the "focal player". If all players after the focal player were to make the same move as the focal player, we would expect to see a decline in cooperation with increasing position in the sequence among players trying to maximize their payouts.

As the focal player gets closer to the end of the sequence, cooperation will go down. We expect to see the most cooperation from Player 1 of 5, who will tend to cooperate more than Player 2, who will cooperate more than Player 3, etc. A simple model that captures this pattern incorporates *s*, the position of the focal player in the sequence, *c*, the average proportion of the endowment contributed by everyone *before* the focal player, and *d*, a discount applied to all contributions made after the focal player's move, which is simply a proportion by which players after the focal player copy the focal player's move, moving away from a zero contribution. We can express this as follows:

$$p = \frac{m(ce(s-1) + dae(n-s) + ae)}{n} + e(1-a) \qquad \text{(Equation 2)}$$

Solving for *s* where payoffs are equal given $a = 1$ or $a = 0$ yields:

$$s = \frac{m - n + mnd}{md} \qquad \text{(Equation 3)}$$

This characterizes the point in the sequence where the focal player's payoffs are not sensitive to her contribution. Before this point in the sequence the most profitable move for the focal player is to cooperate, and after this point it makes sense to defect. We would observe a decline in contribution on average among selfish players if there is heterogeneity between subjects in *d*, the extent to which focal players believe that subsequent players will make the same move as they did (the remaining variables, *m* and *n*, are exogenous). This is because the point in the sequence at which any given player would switch to defect varies with *d*. In a 5-person PGG, at $s = 1$ (the focal player is first to move), most players cooperate and give their entire endowment to the public good. At $s = 3$ (the focal player is third to move), some players will cooperate and some will defect, depending on *d*, the extent to which they believe subsequent players will make the same move. At $s = 5$ in a 5-person game, most players defect.

In four studies we test whether the temporal order of unobserved moves influences decisions. The first is a three-person Public Goods Game, and the remaining studies use a five-person Public Goods Game. Study 1 asks whether there is an order effect at all, and how that varies with Social Value Orientation (a measure of willingness to give up gains in order to benefit others). Study 2 expands this to five people to better rule out any first-mover and last-mover effects. People who arrived at the experiment clearly selfish are sufficiently rare in the study population that forming five-person groups in real time proved difficult, so Study 3 asks whether merely instructing respondents to maximize their own payouts is effective. Study 4 deploys the technique from Study 3 and asks whether it is necessary to have people moving after you who are making their own decisions (vs. having their decisions made for them randomly) in order to observe an order effect. In all studies participants contribute three 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.

Results

All studies are linear PGGs with a multiplier of two, and share several characteristics. First, before learning what game they are to play, players participate in a chat room with their groupmates in order to serve as a rough and ready Turing test. This also gives the task more psychological reality than might otherwise be felt in an online task with no human interaction. Second, all games are real-time interactions between real players. When players are playing a PGG, they are playing with the groupmates they chatted with in real time. Third, all pre-play attention checks, comprehension questions, gameplay decisions, and predictions are incentivized. Participants are paid more for correct answers. Fourth, all experiments involve simultaneous-play PGG control conditions. Fifth, all players pass familiarization tasks and comprehension checks. Data from players who miss a single comprehension check is excluded from the main analyses. 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 that data being excluded. Later studies incorporate more extensive training and comprehension check regimes.

Study 1: 3-Person Sequential Public Goods Game

Participants played one round of a three-person PGG game. 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 Social Value Orientation (SVO) scale (Murphy et al., 2011), which divides most participants into two categories, Individualist and Prosocial. Individualists are working to maximize their own outcomes and are indifferent to others' outcomes, while Prosocials try to maximize their own outcomes but do take others' outcomes into account. Order effects should be more pronounced for participants who are primarily interested in their own payoff (Individualists). In contrast, Prosocials should be less sensitive to order of play, as altruistic motivation should not be biased toward future players.

We find good evidence for the preregistered order effect pooling with pilot data, with first-movers contributing 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 preregistraton. 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.[1]

We also find support for the preregistered prediction that the order effect is concentrated among participants classified as Individualistic in the SVO task. (Figure 3). 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 ($\beta = -0.142$, 95% CI = [-0.248, -0.035], $F(1, 288) = 7.104$, $p = 0.008$).[2] As with the aggregated data, we do not see the hypothesized difference between positions two and three among subjects SVO-classified as Individualistic.

1 See the supplemental online materials for a discussion of the effects variation in times spent waiting within an order

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

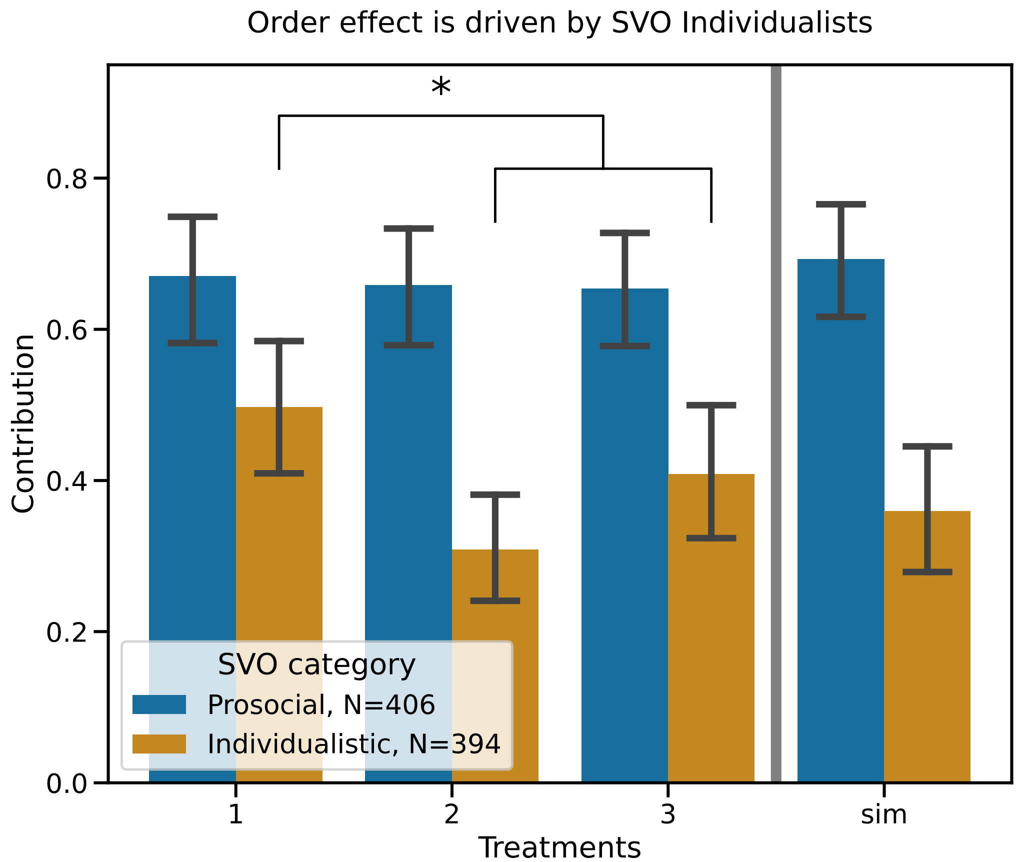

Figure 1: Change in contribution with order is driven by subjects SVO-classified as Individualistic. Subjects who passed comprehension checks. 95% CIs.

Study 2: 5-Person Sequential Public Goods Game

A five-person PGG allows for more insight into the order effect, especially given the potential for effects due to being either first in a sequence of any length ("leader effects", e.g. Eichenseer, 2019)or last. We report results from a sequential 5-person game where subjects were classified based on an SVO task performed up-front. In Study 2 we do not meet our pre-registered threshold to detect an order effect, β = 0.011, 95% CI = [-0.032, 0.055], $F(3, 595) = 33.285$, $p = 0.6206$ for the interaction. A programming error meant that time Player 1 and Player 2 had to make a decision was not correct, sometimes being shorter than for other players. We do observe the predicted effect in Players 2-5.

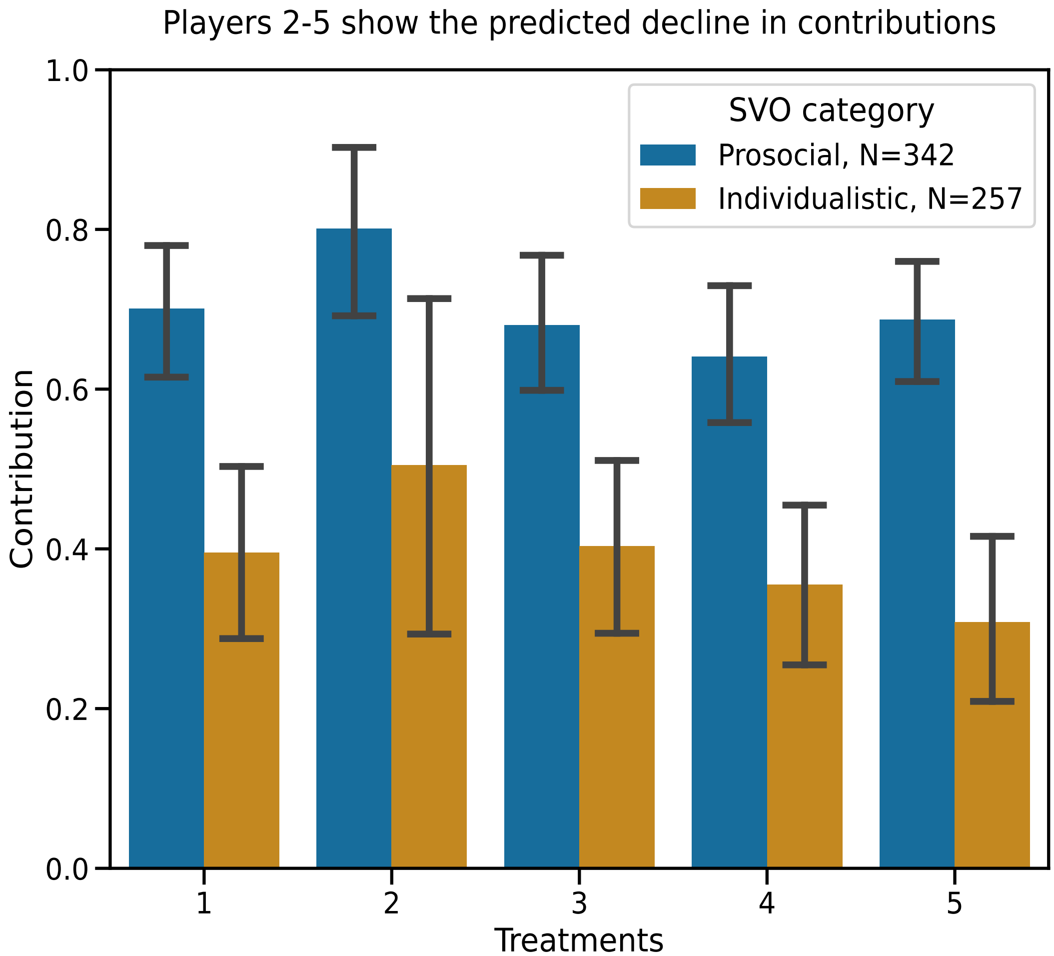

Figure 2: SVO-individualist players show a decline in contribution with increasing order from Player 4, and an anomalous result for Player 1. Subjects who passed comprehension checks. 95% CIs.

We noted that the effect became apparent among all positions if analysis is to subjects who were close to 0 degrees on the SVO scale, i.e. those who were most clearly maximizing their own returns, as opposed to those who were merely classified as "Individualistic". If we restrict our analysis to all players whose SVO degree measure was +/- 10 degrees (players who are maximizing their own returns or very close to it), the predicted pattern appears.

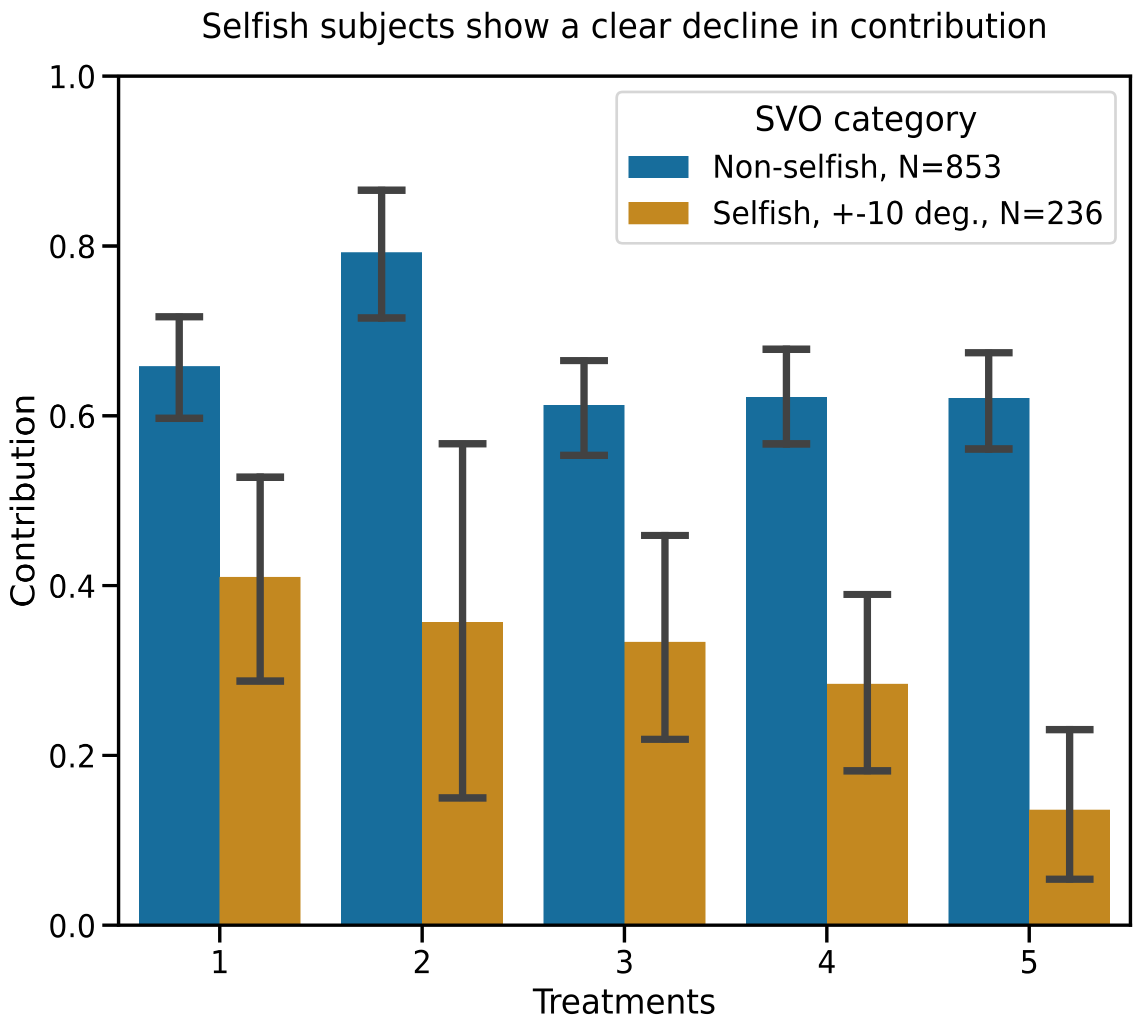

Figure 3: Players whose SVO degree measure is +/-10 degrees show the predicted decline of contribution to the public good with increasing order. All subjects. 95% CIs.

The experiment was not sufficiently powered to detect an effect among the most selfish +/-10 degrees SVO population who passed comprehension checks, but a linear regression of contribution on order interacted with a binary selfish / not selfish variable using data from all subjects does detect the effect, $\beta = -0.042$, 95% CI = [-0.084, -0.0], $F(3, 879) = 45.661$, $p = 0.043$. No-delay and delayed simultaneous conditions show no effect. The preregistered tests for the partial correlation between predictions of other group members' moves and the focal player's moves being stronger going forward shows no effect.

Study 3: 5-Person Sequential PGG with induced self-interest

The observation from Study 2 that the most selfish subjects were those exhibiting the largest order effect led to the design of Study 3. Filling real-time 5-person games with enough selfish subjects proved impractical due to the rarity of subjects who score +/- 10 degrees on the SVO battery, so Study 3 was meant to efficiently examine if a mere prompt to act in one's own interests would allow us to replicate the pattern observed in subjects who arrived at earlier studies already selfish. The task is similar to Study 2, but has significant improvements. The main difference is that subjects did not perform the SVO filtering task. Instead, subjects 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.*

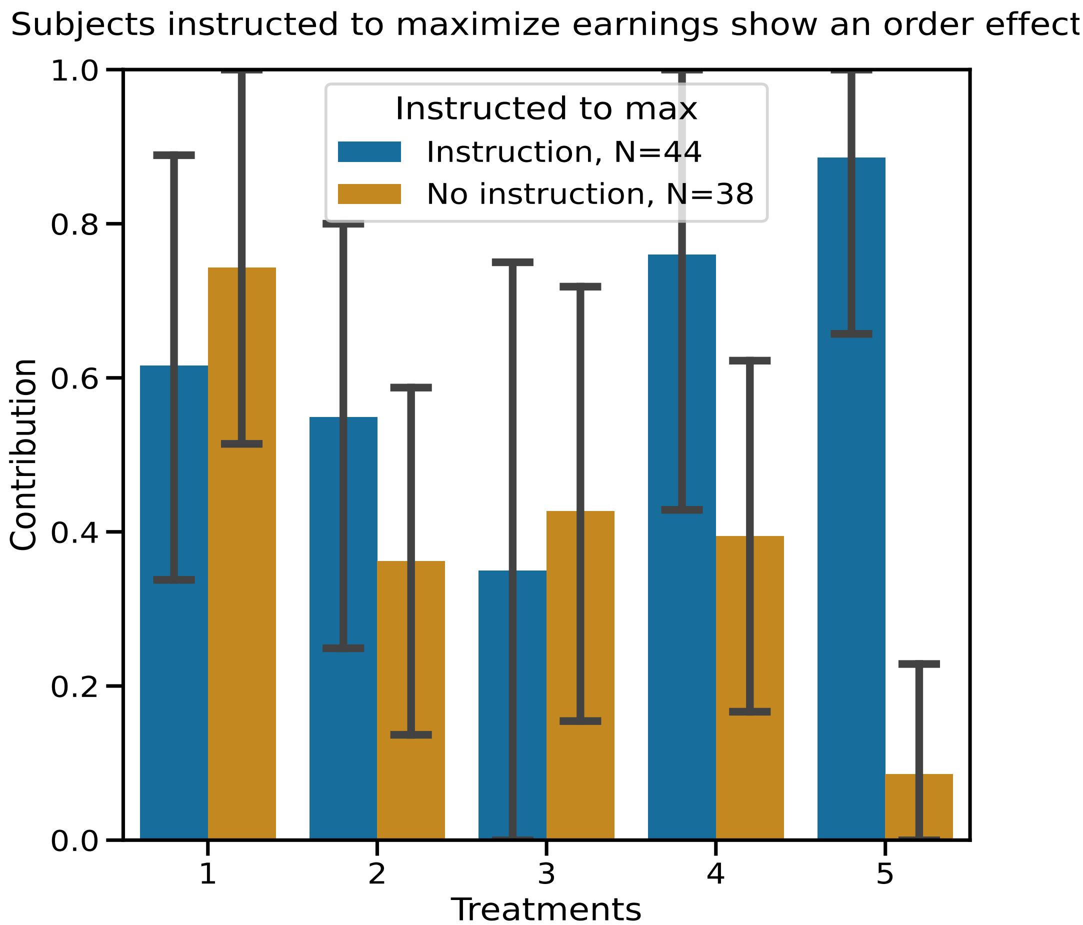

Figure 4: Study 3 shows the hypothesized decline with order among those who were instructed to be greedy. Subjects who passed comprehension checks. 95% CIs.

In addition to the prompt, Study 3 incorporates three additional substantive improvements. First, Study 3 incorporates an additional simultaneous-play control condition that incorporates 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 in order to control for the possibility of wait time effects. While waiting, subjects are shown the task's standard wait screen which incorporates the option to play a simple game to keep subjects engaged with the task. Second, Study 3 incorporates an interactive practice game after the instructions and comprehension questions. This practice game asks subjects to calculate the correct answers to questions about payoffs for hypothetical players in a PGG. Subjects 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 subject's behavior. This is to ensure that information cannot leak via response times. For instance, Player 2 might notice that Player 1 moved rather quickly if Player 1 is allowed to advance from the Contribution page as quickly as she likes.

We observe the order effect in this non-preregistered study. A linear regression of contribution on order interacted with a binary instructed to be greedy / not instructed to be greedy variable plus wealth using data from subjects who pass comprehension checks detects the interaction effect (β = -0.189, 95% CI = [-0.294, -0.074], $F(3, 78) = 5.038$, $p = 0.006$ for the interaction). Subjects receiving the prompt show a decline in contribution with increasing order.

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

Having collected some evidence that a prompt to maximize one's own payouts works as well as arriving at the experiment already wanting to, Study 4 extends Study 3 by applying the instruction to act to maximize one's own payouts to all participants and at larger scale, but with two additional conditions: all subjects are either told that every player before them has his or her contribution determined randomly ("Random Before"), or that every player moving after them has his or her contribution determined randomly ("Random After"). This allows some insight into whether the order effect is somehow 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 which players' moves were randomly decided clear. For instance, if the focal player is in position 3, players see graphical representations similar to that shown in *Figure 5* on all pages from the point at which the concept of random moves is introduced until the end of the game. Subjects in Study 4 continue to move in lock-step, preventing the flow of information to other players via response times.

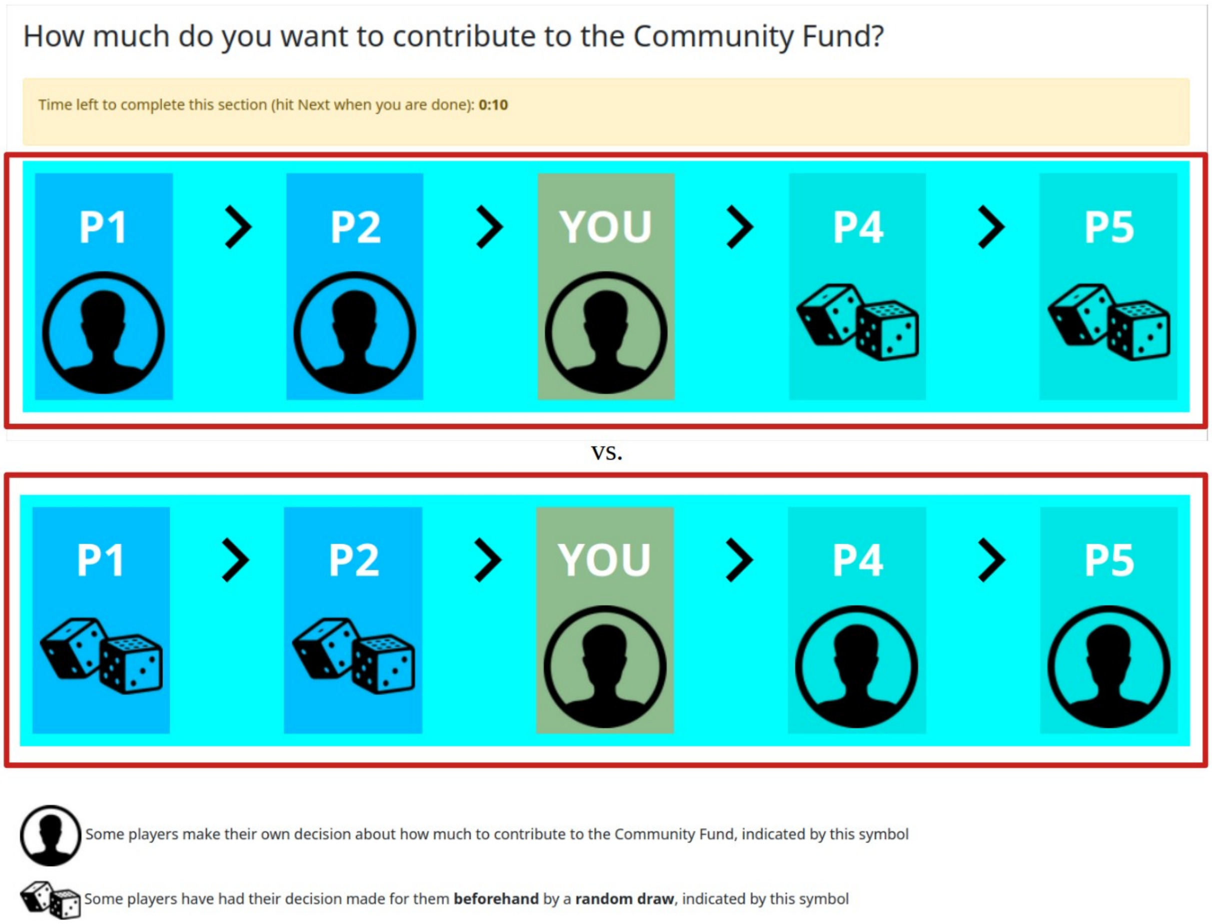

Figure 5: The graphical representation given to Player 3 on the Contribution page is shown in two conditions, in red boxes: Random After and Random Before. Players see a graphical representation of their position relative to other players that clearly conveys which players are having their moves made 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.

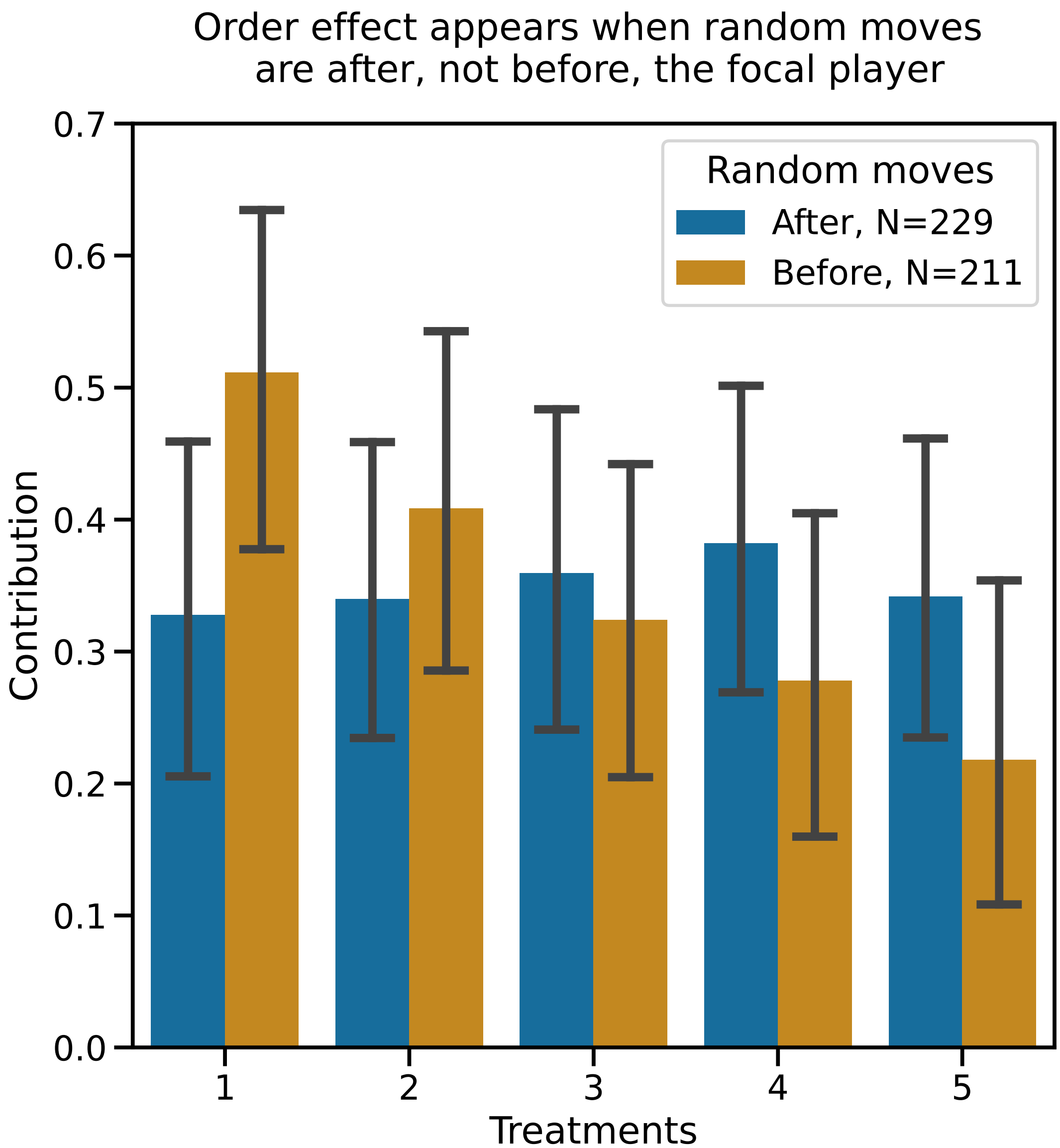

Figure 6: Study 4 shows a decline in contribution to the public good among players who are told that all other players moving after them are making their own moves, and all players moving before them are having their moves made randomly. No effect is observed among players who are told that everyone moving after them has a move selected at random. Subjects who passed comprehension checks. 95% CIs.

In Study 4 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 as usual . 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. Among players told the opposite, that everyone moving after them has their move made randomly, we observe no 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, $\beta = -0.096$, 95% CI = [-0.163, -0.028], $F(4, 354) = 3.849$, $p = 0.006$ for the interaction, lending credence to the idea that heterogeneity in the point at which the optimal move switches from cooperate to defect is driving the order effect. When we restrict the analysis to only those players who passed a second set of comprehension checks at the end of the experiment (80% of players), we observe a larger effect ($\beta = -0.101$, 95% CI = [-0.158, -0.043], $F(4, 359) = 4.356$, $p = 0.0009$ for the interaction). 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.

Discussion

Reward-maximizing players in sequential PGGs cooperate more when they believe there are people moving after them. Four experiments support this view. Substantial training, practice games, and comprehension checks provide evidence that participants understand the game. Furthermore, when we filter based on *ex post* comprehension checks in addition to the preregistered up-front checks, effect size in Study 4 increases. The fact that we observe this effect in participants who understand the game and who are trying to maximize their own rewards narrows the space of possible mechanisms: it appears that earlier movers tend to believe that more cooperation will maximize their payouts, and later movers are more inclined to defect. The order 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. This makes it clear that the distinction between events that have happened and those that have not yet happened is important for behavior in this case. We speculate that a simple model may capture something of the process generating this behavior specifically in selfish agents: these agents understand the rules of the game and are trying to maximize their payouts—they just act as if their move is informative about all subsequent players' moves in a sequential game, and make the move that maximizes payouts if everyone who has not yet moved were to make the same move.

It may be that there is a source of information about what others might do in a standard PGG. In the total absence of other information, it is possible that players look to their own behavior in an attempt to learn about what others will do via social projection (Dawes, 1989; Hoch, 1987). If 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. If this is the case, there may be mechanisms by which people who are selfish—who are trying to maximize their own payoffs independent of what is good for others—end up cooperating anyway. Projection from personal decisions to collective behavior rational in the sense that it can be consistent with Bayes' rule.

Self-signaling is another mechanism that may explain cooperation among selfish 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). Agents who are self-signaling are motivated to produce signals that give them good news. In the case of a PGG, selfish players may be motivated to learn from their own behavior that others 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 (Quattrone & Tversky, 1984). Projection from personal decisions to collective behavior, as in social projection, is consistent with Bayes' rule. With decisions, however, there is a causal component to projection. By freely choosing an action the individual also chooses the signal about collective behavior that the action delivers. Causal power over one's expectations about others' prosocial behavior may be motivationally, if not logically, equivalent to a feeling of power over their actual behavior.

It also may be the case that some participants are acting as if they have control even though they know they do not. These behaviors reflect what might be called magical thinking, where agents act as if they have control over events they know they cannot influence (Daley & Sadowski, 2017 for an overview). This account, in contrast to the preceding two, is more clearly irrational at some psychological level. All three of these potential mechanisms could be present in the same subject population.

Literature investigating the effects of order of agents' actions on behavior, such as that on teams, first- and last-mover effects, and that on sequential games with information flow may need to be revisited. Interpreted in the light of the results we report here, it may be the case that some of the findings in these bodies of literature are a function of sequence alone.

Finally, whatever the mechanism, understanding a means by which selfish actors might decide to contribute to the public good is relevant to many practical policy questions. For example, applying our findings 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 others. This could explain why even a purely selfish individual might feel that their investment of time or effort for a public cause will pay off, provided there are several others deciding 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 subjects.

Code availability: All statistical analysis was done in python. Satistical analysis scripts are available as part of the supplementary information and on osf.io.

Data availability: Anonymized data sets are available as part of the supplementary information and on osf.io.

A convenience sample provided by Amazon Mechanical Turk 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. 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 subjects who failed one or more pre-play comprehension check questions was excluded. All experiments are real-time online group tasks, where subjects 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.

Study 1

Participants. 1668 U.S.-based participants from Amazon Mechanical Turk completed the study. Median total pay per subject (including bonuses for accurate predictions) is \$3.16 (*SD* = 0.90), yielding an hourly rate o \$18.46 per hour at 10 minutes' duration (SD = 8.38). Of 1668 subjects, 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 1002 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, are consented, and then engage in a real effort task transcribing nonsense sentences. 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, subjects are provided with an explanation of the rules of the game (which appear on every subsequent page for reference). The Public Goods Game 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.

Subjects 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 PD, 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.

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. Subjects then complete prediction questions, and then a Social Value Orientation (SVO) slider battery (Murphy et al. 2011, code based on Bakker 2019)[3]. The SVO battery measures preferences for how to allocate resources between oneself and others. Respondents are often categorized 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 ~ 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.

[3] SVO is measured post-treatment, but we do not observe an effect of treatment on SVO. See supplementary online materials for a discussion of the effect of the treatment on SVO.

Study 2

Participants. 1089 U.S.-based participants from Amazon Mechanical Turk completed the study. Median total pay per subject (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 subjects, 66% (720) 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. Amazon Mechanical Turk was selected because it is a reasonable approximation of a representative sample of American adults for our purposes. 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 public goods game 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 subjects 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.

Study 3

Participants. 86 U.S.-based participants from Amazon Mechanical Turk completed the study via Cloud Research. Median total pay per subject (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 subjects 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 a "Selfish" and a "Non-Selfish" condition. In the Selfish 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 arrive merely from waiting. Subjects randomized to the delayed simultaneous condition wait for 80s 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, subjects have and that players move in lock-step with a certain number of second allotted for each page (so players cannot infer anything from how quickly those previous to them have moved).

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 selfish would produce an 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 selfish.

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 subject (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 subjects, 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 public goods game identical to Study 3, with the exception that players 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”), players are told that everyone after them has their decision made by a random process (“Random After”). In addition, as in Study 3, there are two simultaneous control conditions: one with a delay, and one without.

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

REGENERATE REFERENCES BEFORE FINAL DRAFT

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

Chen, Y., & Zhong, S. (2020). Uncertainty-Driven Moral Behavior. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3737959

Colman, A. M., & Gold, N. (2018). Team reasoning: Solving the puzzle of coordination. *Psychonomic Bulletin & Review*, *25*(5), 1770–1783. https://doi.org/10.3758/s13423-017-1399-0

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. (2019). Leading by Example in Public Goods Experiments: What Do We Know? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3441638

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

Gächter, S., Nosenzo, D., Renner, E., & Sefton, M. (2010). Sequential vs. simultaneous contributions to public goods: Experimental evidence. *Journal of Public Economics*, *94*(7–8), 515–522. https://doi.org/10.1016/j.jpubeco.2010.03.002

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

Li, T. (2007). Are there timing effects in coordination game experiments? *Economics Bulletin*, *3*(13), 1–9.

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

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

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. (2011). Measuring Social Value Orientation. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1804189

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

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

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