Conditional cooperation and confusion in public-goods experiments

Maxwell N. Burton-Chellew\textsuperscript{a,b,c}, Claire El Mouden\textsuperscript{a,c}, and Stuart A. West\textsuperscript{a,b,1}

\textsuperscript{a}Department of Zoology, University of Oxford, Oxford OX1 3PS, United Kingdom; \textsuperscript{b}Ca'valla Research Centre for Evolution and Human Sciences, Magdalen College, Oxford OX1 4AU, United Kingdom; and \textsuperscript{c}Sociology Group, Nuffield College, Oxford OX1 1NF, United Kingdom

Edited by Raghavendra Gadagkar, Indian Institute of Science, Bangalore, India, and approved December 9, 2015 (received for review May 18, 2015)

Economic experiments are often used to study if humans altruistically value the welfare of others. A canonical result from public-good games is that humans vary in how they value the welfare of others, dividing into fair-minded conditional cooperators, who match the cooperation of others, and selfish noncooperators. However, an alternative explanation for the data are that individuals vary in their understanding of how to maximize income, with misunderstanding leading to the appearance of cooperation. We show that (i) individuals divide into the same behavioral types when playing with computers, whom they cannot be concerned with the welfare of; (ii) behavior across games with computers and humans is correlated and can be explained by variation in understanding of how to maximize income; (iii) misunderstanding correlates with higher levels of cooperation; and (iv) standard control questions do not guarantee understanding. These results cast doubt on certain experimental methods and demonstrate that a common assumption in behavioral economics experiments, that choices reveal motivations, will not necessarily hold.

It is an accepted paradigm that humans can be divided into fair-minded cooperators that act for the good of the group and selfish “free riders” that exploit the altruism of others (1–16). This conclusion comes from the results of economic experiments, where people in small groups are given some money to play games with. Individuals can either keep the money for themselves or contribute to some cooperative project. The experimenter then typically shares the contributions out equally, but only after multiplying them in such a way that ensures contributions are beneficial to the whole group but personally costly to the contributor. The canonical result from these public-goods games is that most people can be classified into one of two types, with about 50\% being conditional cooperators, who approximately match the contributions of their groupmates, and about 25\% being free riders who sacrifice nothing (1, 2). The remaining players either contribute a relatively constant amount, regardless of what their groupmates do (unconditional cooperators), or they exhibit some other more complex behavioral pattern (1–6).

This division of people into distinct social types has been the accepted basis for new fields of research investigating the cultural, genetic, and neuronal bases of this variation (3–6, 17–20). Some studies have suggested these differences can be exploited by policies to make societies behave in a more public spirited way (21–25). The idea here is that traditional economic policies were erroneous because they only appealed to material self-interest (23). Instead policies could encourage greater cooperation by taking into account how different social types interact (8, 10) and appealing to people’s sense of fairness, especially in populations with more conditional cooperators (5, 25, 26).

This division of people into distinct social types relies on the assumption that an individual’s decisions in public-goods games can be used to accurately measure their social preferences. Specifically, that greater contributions to the cooperative project in the game reflect a greater valuing of the welfare of others, termed “prosociality.” However, this assumption is problematic because there are other possible explanations for the variation in behavior, such as variation in the extent to which individuals understand the game. For example, individuals might cooperate or cooperate conditionally, if they mistakenly think this will make them more money. There are many reasons why individuals might misunderstand the game, including responses to suggestive cues in the experimental setting or the game superficially reminding them of everyday scenarios where cooperation is favored (27–31).

If the variation in levels of cooperation during experiments were mainly due to variation in understanding, then the accepted division into behavioral types would be an artifact of how economic experiments are conducted, rather than any underlying difference in social preferences. Consequently, any research or governmental policies based on the division would be based on false premises. The relative importance of these alternative explanations for variation in game behavior is controversial; whereas some have argued that confused players are responsible for around 50\% of the observed cooperation in public-goods games (32, 33), others have argued that confused players only make up 6–10\% of the population (1, 2). We therefore tested between two competing explanations: differences in social preferences or differences in understanding.

We examined differences in behavior and understanding in three ways (SI Methods). First, we tested whether the same social types arise when individuals know they are playing public-goods games with computers and not other people. Any variation in behavior in this game could not be explained by social preferences and so would pose a problem for the accepted explanation. Second, we then made these individuals play with each other, to directly test how behavior is influenced by whether contributions benefit others or not. Third, we examined whether players understood the essential social dilemma of the game, by asking them whether the income-maximizing decision does or does not depend on what others do. This design allows us to determine whether variation in behavior correlates with understanding and
Results and Discussion

We set up a public-goods game in the same way as those that have been previously used to measure if there are distinct social types (1–6), using the same instructions, control questions, and parameter settings (1, 2). We placed individuals into groups of four players, where each player is given 20 monetary units (MUs) that they can either keep for themselves or partially/fullly contribute to a group project. We then multiplied all contributions to the group project by 1.6 before sharing them out equally among all four members. Therefore, each player lost 0.6 MU from each 1.0 MU they contributed to the public good, whereas their groupmates each gained 0.4 MU. Consequently, the strategy that maximizes individual financial gain is to contribute nothing (0 MU). Importantly, the return on any MU contributed is not altered by the contributions of others, and therefore the strategy to maximize financial gain is not altered depending on how others are playing. We first explained the public-goods game to all players, both on screen and on a piece of paper they kept throughout the experiment, before using the same control questions as used by previous studies ([SI Appendix]).

Cooperating with Computers. We explained to our players, after the initial instructions and control questions, that they would first be playing in a group with three computer players that would be playing randomly and that no other people would benefit from their contributions. All players had to click a button with the words, “I understand I am only playing the computer” before they could proceed ([SI Methods]). We also followed previous studies in using what is termed the strategy method to classify individuals according to how they vary their behavior depending on the possible behavior of their groupmates (1–6). In the strategy method, players have to make contributions for each and every possible mean integer contribution of their three group mates. In our version, the appropriate amount is then contributed from their account after the contributions of the computer “players” have been generated. To prevent any learning, we did not provide our players with any information on their earnings from this game.

We found that when playing with computers, individuals can be divided into the same behavioral types that have previously been observed when playing with humans (34) (Fig. 1). Specifically, we found that 21% (n = 15) are noncooperators (free riders) who contribute 0 MU, irrespective of the computer contribution, and 50% (n = 36) are conditional cooperators, who contribute more when the computer contributes more (1–6). These conditional cooperators are adjusting their behavior in response to the computer’s contribution, even though they have been told that their contributions will not benefit others and despite the fact that the income-maximizing strategy does not depend on how much the computer contributes. The remaining 29% (21) of players exhibited some other pattern ([SI Results]).

This distribution of different behavioral types playing with computers is strikingly similar to that previously observed when individuals are playing with other humans (χ² test comparing our distribution to an amalgam of the distributions reported in refs. 1 and 2: χ²(3) = 5.2, P = 0.156; Table 1, Table S1, and [SI Results]). Consequently, a variation in the regard for the welfare of others, or a social preference, is not required to explain why individuals vary in their behavior. The data from games with computers suggest that the standard methodology of public-good games using the strategy method may not provide a reliable measure of underlying social preferences.

It could be argued that games with computers are uninformative of human psychology because they put players in unnatural situations. However, this argument could just as easily be applied to many other economic games. For example, is it any more natural to ask players to respond to the decisions of others when there is no strategic reason to do so (1–6), or to punish anonymous individuals that they will never interact with again (35), or to report the number they privately roll on a dice to determine their payoff (36)? Laboratory studies are both advantageous, in that they allow precise control of the available incentives, and problematic, because they can remove important cues for natural behavior and because humans are not adapted for the laboratory (21, 37–46).

It could also be argued that theories of social preferences make no prediction for how people will play with computers and that therefore such treatments provide no relevant data for such theories (13). The key point here is how individuals behave in a single scenario (1), but to experimentally test how behavior compares across different scenarios (39, 47), because theories of social preferences do imply differences between situations when individuals know that others will benefit or not. Consequently, after playing with computers, we had our players play with humans, so that we could directly test how their behavior is affected by the knowledge that others will benefit from contributions (47).

Play with Computers Predicts Play with Humans. We next compared how well the above strategy method predicted play in unconditional games where players simultaneously and privately decide their contributions, as was done in refs. 2 and 48. We then had our players play one series of six such unconditional games with humans. We provided no feedback between decisions so that these six decisions essentially represented a single “one-shot” decision with no opportunities to influence or respond to the decisions of other player. The instructions made four clear references to playing with people and required the players to click an on-screen button with the words “I understand I am now playing with real people” before they could proceed ([SI Appendix]). We did not counterbalance the order of our treatments because we wanted to first classify our players on their ability to maximize their personal income before allowing them to play with humans.
and to provide a logical progression to our treatments (49, 50). In all cases, communication was forbidden, and we provided no feedback on earnings or the behavior of groupmates. This design prevents signaling, reciprocity, and learning and therefore minimizes any order effects (51–53).

We found that the behavioral types from the strategy method significantly predicted the level of cooperation in the subsequent unconditional games, both with computers [generalized linear model (GLM), contribution \( \sim \) type: \( F_{3,68} = 7.7, P < 0.001, R^2_{adj} \) from a linear model = 0.22] and with humans [linear model (LM), mean-contribution over six rounds \( \sim \) type: \( F_{3,68} = 6.9, P < 0.001, R^2_{adj} \) from a linear model = 0.12; Fig. 2A and Fig. S1]. Furthermore, controlling for individuals, there was no significant difference in the mean unconditional contributions between games with computers or humans (paired \( t \) test \( t_{(71)} = 0.7, P = 0.471; \) Fig. 2A and Table S2). These results show that individuals cooperate to the same degree, in the public-goods game, irrespective of whether they are playing computers or humans (correlation = 0.78, \( P < 0.001 \)). These conclusions are based on the classification scheme of ref. 2 but hold if we use our classifications from Fig. 1 (SI Results).

We also found that how individuals conditioned their behavior on their beliefs about the behavior of their groupmates did not differ in response to whether they were playing with computers or humans. In unconditional public-goods games, individuals appear to still conditionally cooperate, by correlating their contributions with their stated beliefs about their groupmates (2, 54). Therefore, at the same time players made their contribution decision, we asked them what they expected their groupmates would do. Specifically, what the mean contribution of their three groupmates would be. This way we could investigate if our players conditioned their contributions on the basis of their expectations.

In contrast to some previous studies, we did not financially reward individuals who better estimated the behavior of their groupmates (2, 55). The reason for this is that such incentives increase the length and complexity of the instructions and have been shown to influence the level of cooperation (55). Furthermore, the hypothesis of conditional cooperation stipulates that players are motivated to form accurate beliefs about their groupmates to match them, such that “beliefs have a causal effect on contributions” (54, p. 414). Our nonincentivized elicitation of beliefs is therefore merely asking putative conditional cooperators to record their already formed beliefs.

As in previous studies (2), we found that our players’ contributions were positively correlated with the amount that they expected their human group mates to contribute [generalized linear mixed model (GLMM) on six rounds of data: \( F_{1,405} = 152.9, P < 0.001, \beta = 0.210 \pm 0.017; \) Fig. S2]. This result demonstrates that financial rewards (incentives) for better estimates of group mates’ behavior are not required to recreate the standard pattern of behavior. However, we also found the same positive relationship between contributions and expectations when playing with computers (GLM: \( F_{1,20} = 17.0, P < 0.001, \beta = 0.173 \pm 0.046; \) Fig. S2). Analyzing all of the data together, the relationship between contributions and expectations did not differ significantly depending on whether groupmates were computers or humans (GLMM interaction: \( F_{1,486} = 2.5, P = 0.116, \) difference in \( \beta = 0.054 \pm 0.034; \) SI Results).

Overall, our results show that individuals behave in the same way, irrespective of whether they are playing computers or humans, even when controlling for beliefs (Figs. S2 and S3). Therefore, the previously observed differences in human behavior do not need to be explained by variation in the extent to which individuals care about fairness or the welfare of others.

### Table 1. Distribution of behavioral types does not differ between games with computers and humans

<table>
<thead>
<tr>
<th>Type</th>
<th>Ref. 1</th>
<th>Ref. 2</th>
<th>This study</th>
<th>Significance*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conditional</td>
<td>22 (50%)</td>
<td>77 (55%)</td>
<td>36 (50%)</td>
<td>0.676</td>
</tr>
<tr>
<td>Humped</td>
<td>6 (14%)</td>
<td>17 (12%)</td>
<td>7 (10%)</td>
<td>0.667</td>
</tr>
<tr>
<td>Unclassifiable</td>
<td>3 (7%)</td>
<td>14 (10%)</td>
<td>14 (19%)</td>
<td>0.033</td>
</tr>
<tr>
<td>Free-rider</td>
<td>13 (30%)</td>
<td>32 (23%)</td>
<td>15 (21%)</td>
<td>0.624</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>44</td>
<td>140</td>
<td>72</td>
<td>0.156</td>
</tr>
</tbody>
</table>

*Fisher’s exact test per type, \( \chi^2 \) test for totals.

†The three negative cooperators are classified as humped and the five unconditional cooperators as unclassifiable (SI Results).

**Fig. 2.** Play with computers predicts play with humans and conditional cooperators misunderstand the game. (A) The mean contribution (±95% CI) to the public-good grouped by behavioral type (Fig. 1). For all types, the mean levels of cooperation were not significantly different when playing with computers (dark gray) vs. when playing with humans (light gray). (B) The percentage (±95% modified Wald method CI) of players, separated by type, failing our beliefs test, which asked if players knew that the payoff maximizing decision did not depend on what groupmates contributed. Conditional cooperators were more likely to fail the beliefs test than noncooperators and were just as likely to fail as unclassified players, who were previously argued to be the only confused players (2).
Conditional Cooperators Misunderstand the Game. We hypothesized that variation in behavior largely reflects variation in understanding of the game. Specifically, that conditional cooperators tend to believe that the income-maximizing strategy depends on what others contribute, whereas noncooperators tend to realize that it does not. We tested this hypothesis by asking each player: “In the game, if a player wants to maximize his or her earnings in any one particular round, does the amount they should contribute depend on what the other people in their group contribute?” We allowed players to answer either: yes/sometimes/no/unsure. We found that only 21 (29%) of our 72 players passed this beliefs test, correctly answering (no) that the income-maximizing strategy does not depend on what others contribute in a one-shot game: with 33 (46%) answering yes that the contributions of others do matter; 11 (15%) answering that the contributions of others sometimes matter; and the remaining 7 (10%) answering that they were unsure.

As predicted by our hypothesis, we found that there was a significant correlation between beliefs about the game and behavior in the strategy method with computers (Fig. 2B, Tables S3 and S4, and SI Results). Specifically, individuals that correctly answered no tended to be noncooperative free riders and individuals that answered otherwise tended to be conditional or humped cooperators (GLM: $\chi^2 = 12.9, P = 0.002$; Fig. 2B). Conditional cooperators were more likely to answer incorrectly than noncooperators. Whereas 30 of the 36 (83%) conditional cooperators were incorrect, only 5 of the 15 (33%) noncooperators were incorrect [Fisher’s exact test (FET): $P < 0.001$; Fig. 2B]. Refs. 1 and 2 suggested that their 6–10% of unclassified players may have been confused players, yet Fig. 2B shows that conditional and humped cooperators are just as likely to answer incorrectly (36 of 43, 84%) as unclassified players (10 of 14, 71%, FET: $P = 0.436$).

As we did not incentivize responses to the above question, it might be argued that our players were not sufficiently motivated to answer correctly. However, there is no reason to believe that a lack of motivation can explain the significant correlation between type and response. Furthermore, if the incentives of the game with computers did not make players income maximizers, there is no reason to suppose that equally incentivizing this question would have motivated them to answer correctly.

It might also be argued that people playing with computers cannot help behaving as if they were playing with humans. However, this interpretation would: (i) be inconsistent with other studies showing that people discriminate behaviorally, neurologically, and psychologically between humans and computers when playing simpler games (19, 56–58); (ii) not explain why behavior significantly correlated with understanding (Fig. 2B and Tables S3 and S4); (iii) contradict the key assumption for theories of social preferences that players respond to the costs and benefits of the choices offered to them (59); and (iv) suggest that behavior reflects the payoffs of encounters in the real world, rather than the payoffs of the laboratory game (30, 38–46). Such ingraining of behavior would suggest a major problem for the way in which economic games have been used to measure social preferences (38, 41, 42, 60). In particular, behavior would reflect every expectation from the real world (39, 40), such as reputation concerns or the possibility of reciprocity, rather than the setup of the game and the true consequences of choices (43, 44). Although this could be useful for measuring cultural differences in how such games are perceived (29, 61), it would make the logic of measuring individual social preferences problematic (60). However, if players are bringing in their outside behavior, this could explain three results: (i) why many people have mistaken beliefs about the income-maximizing strategy; (ii) why players improve their income maximization with experience of economic games (53); and (iii) why people play games differently depending on how they are named or described (61).

Standard Control Questions Fail to Control for Understanding. Previous studies have required that their players correctly answer a series of control questions before allowing them to play (1–6). We followed a previous study by describing four scenarios and asked the players what the resultant incomes would be (2) (SI Appendix). For example, if all players contribute 20 MU, then how much would each player receive? Previous studies have assumed that ensuring all players have given correct answers to these four questions allow one to “safely assume that the players understood the game” (2, p. 543); however, these same studies have still classified 6–10% of their players as confused (1, 2).

We tested the assumption that correct answers indicate understanding. We did this by letting our players answer the questions
freely and then examining if correct answers to the 10 control questions ensured that individuals correctly identified the income-maximizing strategy, either in games with computers, or in our control question. We found that only 16 (22%) of our 72 players correctly answered all 10 control questions (SI Results). However, of these 16 players, only 6 (38%) got the income-maximizing strategy correct in both of the games with computers. In fact these 16 players that answered all questions correctly were not less likely to be conditional or lumped cooperators (8 of 16, 50%) than those that failed the standard control questions (36 of 56, 64%, FET: \( P = 0.744 \); Fig. 3A, 3E). Furthermore, only five (31%) replied correctly to our question about the game being interdependent or not, which was not significantly more than the 16 of the 56 (29%) players that failed the standard control questions (FET: \( P = 1.000 \); Fig. 3B).

Tellingly, even when we only consider the 16 individuals that answered all 10 standard control questions correctly, the responses to our beliefs test still predicted who cooperates or not with computers. Specifically, of the 16 above, all 5 of those that also passed our beliefs test were noncooperators vs. just 2 of the 11 who failed our beliefs test (FET: \( P < 0.005 \)). Although these sample sizes are small, we found the same qualitative results in a similar, but larger (\( n = 216 \)) study that did not contain the strategy method but did contain the same control questions and one-shot games with the computer and humans (SI Results). Therefore, answering the standard control questions correctly, contrary to the assumptions in previous studies, does not guarantee understanding (1-5).

Comprehenders Are Not Cooperators. It is possible that even if a large proportion of players misunderstand the game, those that do understand the game are still likely to be significantly altruistic. Some previous studies have concluded that around 50% of contributions are due to confusion (32, 33), leaving open the possibility that a substantial number of people who do understand the game still choose to cooperate. We investigated this possibility by examining the behavior of three different types of players, which could each be argued to have understood the game. Specifically, we examined the individuals that: (i) contributed 0 MU in both the strategy method and in the one-shot game with the computer (\( n = 13/72, 18% \)); (ii) answered all of the standard control questions correctly (\( n = 16, 22% \)); and (iii) that passed on our beliefs test (\( n = 21, 29% \)).

First, overall, players that maximized their income when playing with computers did not contribute significantly more than 0 MU when playing with humans (paired \( t \) test: \( t_{12} = 1.957, P = 0.074 \)). Individually, none of these players gave significantly more to humans than to computers (Table S5). These results show that players that successfully maximize their earnings when playing with computers do not contribute significantly more when told their contributions will benefit others. Second, players that answered all of the standard control question correctly showed no prosocial bias toward humans: not giving significantly more to humans (5.4 MU) than they did to computers (4.8 MU) (paired-samples \( t \) test: \( t_{15} = 0.6, P = 0.529 \)). Third, players that correctly answered our control question also showed no prosocial bias toward humans: not giving significantly more to humans (6.9 MU) than they did to computers (5.1 MU) (paired-samples \( t \) test: \( t_{21} = 1.7, P = 0.098 \)). Therefore, we find no evidence that there is a subpopulation of players that understand the game and have prosocial motives toward human players (SI Results).

Measuring Motivations. Finally, we investigated the motivations of all our players with a simple postgame questionnaire and found little evidence of prosociality. We asked our players, “What was your most important motivation in the games with real people? Please select the answer that best describes your motivations” and gave them a choice of five options (SI Results). Perhaps surprisingly, considering that there was no cost to players wishing to appear prosocial, 50% of players (\( n = 36 \) of 72) specified they had been motivated by making themselves the maximum money possible. Alternative options were making the most money for everyone (\( n = 13, 18% \)), for the group (\( n = 11, 15% \)), for others (\( n = 3, 4% \)), or making more than others (\( n = 1, 1% \)). The remaining players (\( n = 8, 11% \)) chose “other.” It is commonly assumed that the incentives in economic experiments make players more honest and thus appear less prosocial than they would in nonincentivized questionnaires. However, we found that the number declaring that they had been motivated by maximizing personal income (\( n = 36, 50% \)) was significantly more than the number of noncooperators in the strategy method (\( n = 15, 21% \), FET: \( P < 0.001 \)) or in the games with humans (\( n = 13, 18% \), FET: \( P < 0.001 \)).

When we compared motivations among different types of players, we found that nearly half (47%, 20/43 players) of the conditional and lumped cooperators declared they were motivated by self-interest. This proportion was not significantly different to the proportion of noncooperators declaring a self-interested motivation (73%, 11/15 players, FET: \( P = 0.131 \); Table S6).

Economic Games and Social Preferences. To conclude, our results strongly suggest that the previous division of humans into altruistic cooperators and selfish free riders was misleading. We showed that the strategy method reveals the same division even when individuals are playing with computers, and nobody benefits from their cooperation. Instead, the variation in behavior, even in the strategy method, can be explained by variation in understanding rather than variation in social preferences. For example, individuals previously categorized as fair-minded conditional cooperators tend to be individuals who misunderstand the nature of the game and think that, even in one-shot games, the income-maximizing decision depends on others.

There are a number of reasons why individuals might incorrectly think that the way to best maximize their income depends on the behavior of others (SI Discussion). First, the strategy method places an emphasis on the behavior of others, possibly suggesting their behavior is important, as it would be in a threshold public-goods game, for example (62). Second, the wording of instructions to players, with words such as invest, could suggest the game is risky, and hence dependent on how others invest (34). Third, the best strategy for many everyday situations may depend on what others are doing, and the game reminds them of such scenarios (30, 40). Points 2 and 3 could apply to a broad range of scenarios and not just games that used the strategy method. For example, cooperation was significantly reduced in another public-goods game experiment when players were explicitly informed that they “lose money on contributing” (63). We are not arguing that misunderstandings explain all aspects of behavior in economic games; rather, the possibility for misunderstanding needs to be considered when developing null hypotheses (47).

More generally, our results confirm that when attempting to measure social behaviors, even with the strategy method, it is not sufficient to merely record decisions with behavioral consequences and then infer social preferences (1-6). One also needs to manipulate these consequences to test whether this affects the behavior. Here when we removed any social effects from the consequences of players’ decisions, by having them knowingly play with computer-identified groupmates, their behavior is unchanged in both the strategy method and in the unconditional games. These results suggest that other existing paradigms from the fields of behavioral economics might be built on incorrect conclusions from experimental studies. The question is, which aspects of human sociality are these games actually measuring (38, 60, 64)? Numerous studies have made the implicit assumption that behavior in economic experiments perfectly corresponds to the underlying behavioral preferences or intentions of individuals (1, 2, 35, 59). We showed, in public-goods games, that when a competing hypothesis is considered, which does not make an assumption of perfect play and understanding, it is better able to
explain the data. A major task for the future is to develop and test competing hypotheses that do not assume perfect understanding and perfect play in other economic games.

Methods

Experiments were conducted using z-Tree (65) at the Centre for Experimental Social Sciences (CESS), Nuffield College, University of Oxford. Participants were recruited using ORSEE (66), from the general participant pool with the sole specification that they had not before participated in a public-goods experiment. The CESS laboratory has a policy of “no deception,” all experiments must pass the CESS ethical review board, and CESS obtains informed consent from all players. We provided all players with the same instruction and control questions (SI Appendix), which were compiled verbatim as much as possible from the online appendix of ref. 2. We provided the instructions both on screen and also on paper handouts that they could keep during the experiment (SI Methods).

ACKNOWLEDGMENTS

We thank Raghavendra Gadagkar; three anonymous reviewers; Miguel dos Santos for comments; the Centre for Experimental Social Sciences and the European Research Council, the Calleva Research Centre, and the John Fell Fund Oxford for funding; and our players.

Supporting Information

Burton-Chellew et al. 10.1073/pnas.1509740113

SI Methods

Experimental Methods. For an overview of our experimental design and treatment orders, please see SI Appendix. We included 24 players in each of our three sessions. All decisions were anonymous, and our players did not know who else was in their group. The mean self-reported age of our players was 39.7 ± 1.9 y (SEM), ranging from 18 to 74 y, and 38/34 reported that they were female/male.

We first gave our participants the instructions (SI Appendix) and standard control questions (SI Appendix), both copied verbatim as much as possible from ref. 2. We then told the players that before they would play with people, they would first be playing in a special case with computer players. The full text is available in SI Appendix, but the key text is “Before you begin, you are going to play this game in a special case. In this special case, you will be in a group of just you and the COMPUTER; “The computer will pick the decisions of the other 3 players. The computer will pick their decisions randomly and separately (so each computer player will make its own random decision)”; and “You are the only real person in the group, and only you will receive any money.” In contrast to a previous study (34), our initial game instructions did not mention playing with computer players, to prevent the possibility of our players being distracted during the instructions on the game’s payoffs or during the control questions by thoughts of computer players. All players had to click on an on-screen button with the words, “I understand I am only playing the computer” before they could proceed.

After explaining that they would be playing with computers and that nobody else could benefit from their contributions, we then explained that they would be playing the strategy method, using the same instructions copied verbatim as much as possible from ref. 2 (SI Appendix). Again, we only told our players about the strategy method after the initial instructions and control questions, and after we told them they would be playing with computers (instructions copied and as much as possible from ref. 2). This way their strategies were not merely recording the decisions of players that had already made up their mind of how to strategically respond to human players. Instead, they represented behavior in response to a choice that had no social consequences and therefore should not reflect any social preferences or concerns other than how to maximize personal income (“You are the only person in the group, and only you will receive any money”: SI Appendix). To prevent any learning, we did not provide our players with any information on their earnings from this game.

After the strategy method, which we used to categorize our participants as has been done previously (1–6), we then had our participants play a typical public-goods game, again with computers. In the typical public-goods game, players have to decide simultaneously on a single contribution with no knowledge of each other’s actions (SI Appendix). This way we could compare behavior in the strategy method with behavior in the typical public-goods game setting where conditional responses to the behavior of other players are not possible. We then had our participants continue to play the typical public-goods game, but now with people, for six rounds (SI Appendix). We provided no feedback during these six rounds; therefore, they were essentially a one-shot game that prevented signaling, reciprocity, and learning and therefore minimized any order effects (51–53), “You WILL NOT receive any information about the decisions of the other players, nor about your earnings in these rounds, nor will anyone else at any time except for the experimenter after the experiment.” (SI Appendix).

The first two treatments, with computers, are diagnostic tests of understanding and follow on naturally from the control questions used by others (1, 2). More specifically, our measure was of who did and did not maximize their earnings when there were no social concerns and provided a baseline measure of income maximization behavior. We could then compare our measure of income maximization with behavior in the subsequent public-goods game with people, where there were social concerns, rather than simply attributing all failures of income maximization to social preferences as has been done before (1, 2, 35, 59).

We did not counterbalance the order of our treatments because we wanted to first classify our participants on their ability to maximize their personal income before allowing our participants to play with humans. Overall our treatment order provided a logical progression (49, 50), and there was no reason to suspect treatment order effects as in all cases we provided no feedback and therefore any treatment-order effects should be minimal. We are also not arguing for any differences in behavior between treatments that could be correlated with or confounded by experience or time. Furthermore, we can compare our results in our opening treatments of games with computers to the highly stylized results of prior published research on public-goods games. As we show below (SI Results, Table 1, and Table S1), our distribution of social types is statistically equivalent to the prior research. In addition, as we also show below (SI Results) our players’ behavior in the unconditional game with computers is well in line with previous research. Our players contributed 39% of their endowment, which is very similar to the results reported in other studies with humans (2).

Statistical Methods. Below is a glossary of the terms we use in the results.

- FET: Fisher’s exact test for analyzing two-way contingency tables of categorical data.
- LM: Linear model. Used when modeling continuous data (e.g., mean contributions across six rounds of play) with one data point per individual.
- GLM: Generalized linear model with binomial-logit link. Used when modeling binary or proportional data (e.g., correct/incorrect answer; or contributions from 0 to 20 MU) that came from a single observation per individual (e.g., contributions to the computers in the one-shot game). Parameter estimation used the Hybrid method, with the Pearson χ² scale parameter, and a model-based estimator of the covariance matrix. Effect statistics used the Likelihood ratio χ² and profile-likelihood confidence intervals.
- GLMML: Generalized linear mixed model with binomial-logit link. Used when modeling binary or proportional data that came from repeated measures of the same individuals. We specified a first-order auto-regressive covariance structure for the residuals to control for repeated measures of the same individuals across rounds. Degrees of freedom were varied across tests according to the Satterthwaite approximation. Significance tests used model based covariances.

SI Results

The Distribution of Behavioral Types and Altruism Toward Computers. Refs. 1 and 2 were the first to classify players on the basis of their pattern of contributions in the strategy method. They separated
their players into four categories, defining three social types, termed conditional cooperators, humped/triangle cooperators, and free riders. Anyone they could not classify into these categories was described as either (1) or unclassifiable/confused (2). We classified our behavioral types according to this classification scheme but with the addition of two more types. Below we outline the criteria for all of the above categories and also how we categorized our players.

Conditional cooperators (n = 36; 50%) are defined by the method of ref. 1 that classifies conditional cooperators as those that have a contribution schedule that has a significant and positive Spearman’s rank correlation with the mean contribution of the three groupmates. In total, 36 of our 72 participants met these criteria at the P < 0.01 level (players 2, 3, 4, 5, 6, 7, 8, 9, 10, 13, 14, 15, 16, 18, 23, 27, 29, 31, 32, 33, 35, 36, 39, 45, 46, 48, 49, 51, 56, 57, 58, 64, 67, 69, 70, and 72; SI Appendix). Two other participants (players 21 and 40; SI Appendix) also had significant correlations at the conventional P < 0.05 level (P = 0.013 and 0.020 respectively) but we did not classify them as conditional cooperators; instead, they were humped and unclassified (see below), respectively.

Humped cooperators (n = 4; 6%) (also termed triangle cooperators, e.g., in ref. 2) are defined as those who both “increase their contributions with the contribution of others up to a point” and then “decrease their own contributions the more others contribute” (2, p. 542). We found that four of our players displayed such a pattern (players 21, 30, 42, and 62; SI Appendix), first increasing and then decreasing, on average, after an idiosyncratic threshold point, with a peak contribution when their computerized groupmates contributed 16, 10, 7, and 7 MU, respectively.

Negative cooperators (n = 3, 4%) are shown by a further three of our players (players 26, 50, and 65; SI Appendix) who exhibited a pattern that was not evident in ref. 1 or described in ref. 2, but is similar to both the above two patterns. We termed these participants as negative cooperators, because they showed significantly conditional cooperation (significant Spearman’s rank correlation), but with a negative correlation. When comparing the statistical distributions of our study with previously published studies, we needed to have the same number of categories, so we decided to categorize these negative cooperators as humped cooperators. This recategorization was because all of them can be described by the same rule as humped cooperators, increasing contributions and then decreasing contributions the more others contribute after a point (2, p. 542), if one considers their idiosyncratic turning point to be 0 MU.

Unconditional cooperators (n = 5; 7%) were any player that exhibits a constant positive contribution regardless of what their groupmates contribute. Ref. 1 classified their one such player as other. It is not clear if ref. 2 classified such players as confused/ unclassifiable or did not find any such participants. We found five (7%) such players (players 19, 24, 55, 60, and 63; SI Appendix).

Noncooperators (n = 15; 21%) are quite simply those that contribute 0 MU regardless of what their groupmates contribute and thus are in one sense similar to the above unconditional cooperators. These players in other studies are typically termed free riders, as they metaphorically enjoy the benefits of public goods such as a publicly funded transport service, without contributing to the costs. This is the payoff-maximizing strategy in the typical public-goods game and thus should be exhibited by all participants, regardless of social preference, when playing with computers, if they understand the game. We found 15 such players (players 1, 11, 20, 22, 25, 28, 37, 43, 44, 47, 52, 59, 61, 66, and 71; SI Appendix). We also had nine players (12.5%) that we chose not to categorize (unclassified; players 12, 17, 34, 38, 40, 41, 53, 54, and 68; SI Appendix).

As explained in the main text, the distribution of our different behavioral types when playing with computers is strikingly similar to that previously observed when individuals are playing with other humans. In comparison with refs. 1 and 2, we find no difference between all three distributions (χ² test: χ²(6) = 6.4, P = 0.384; no difference between our results and the first study (1) χ²(6) = 4.2, P = 0.248); or the latter study (2) χ²(6) = 3.8, P = 0.288), or compared with a combination of both the first and the latter study (χ²(3) = 5.2, P = 0.156).

We primarily compared our distribution to those reported by refs. 1 and 2 as these were the pioneering studies that developed the method we replicated. However, there are also other published studies measuring the frequency of conditional cooperators, primarily for the purpose of cross-cultural comparison. We therefore also compared our distribution of behavioral types to a collection of prior published distributions generated using the strategy method of refs. 1 and 2.

Including this study, we found, in total, eight studies detailing a total of 14 distributions from various locations using a total of 708 players. Two of the studies were with computerized groupmates and 12 with human groupmates (Table S1). In the 12 studies with humans, the frequency of conditional cooperators ranged from 42% to 63% for 11 of them, with one study reporting 81%, and ours was 50%. The frequency of noncooperators (free riders) ranged from 2% to 36%, and ours was 21%. The overall mean frequencies for each type across all human and computerized studies are reported in Table S1.

To compare the reported distributions for the 12 human studies with the two computerized studies, we used a χ² test on the combined absolute distributions of all human studies compared with both computer studies, and we found no significant difference (χ² test: χ²(3) = 2.6, P = 0.458; Table S1). For completeness, we also tested the distribution of each type separately, using FET, and found no significant differences for any type (conditional cooperators: P = 0.470; humped cooperators: P = 0.600; unclassified: P = 0.602; free riders: P = 0.168) (Table S1).

Play with Computers Predicts Play with Humans. We found that our players again cooperated with computers, contributing on average 39% of their endowment (7.7 MU; median = 8 MU, mode = 0 MU), a result in line with findings in other studies with humans (2). They then cooperated likewise with humans (mean = 41% of endowment, 8.1 MU, median = 8 MU, mode = 0 MU). As reported in the main text, we also found that the behavioral type in the strategy method with computers significantly predicted the contributions in an unconditional game with both computers and humans. The results in the main text are based on the classification scheme of ref. 2, which has the four categories presented in Fig. 2 (noncooperator, conditional cooperator, humped cooperator, and unclassified). The results are qualitatively the same when we use our classification scheme presented in Fig. 1, which has six categories (the same as ref. 2 plus negative cooperators and unconditional cooperators). The six-way classification scheme significantly predicted the one-shot contribution with computers (computers: GLM: F₆,₆₆ = 4.9, P = 0.001) and all six rounds with humans (the complete absence of any feedback on the behavior or earnings of the players converted the six games with humans into essentially one-shot encounters), regardless if we ran a LM of each player’s mean contribution over six rounds (F₆,₆₆ = 4.3, P = 0.002) or a GLMM of all six contributions, controlling for individual (F₆,₉₅ = 5.5, P < 0.001). The six-way classification scheme also significantly predicted contributions for four of the six human rounds analyzed separately (R1: F₆,₆₆ = 2.6, P = 0.032; R2: F₆,₆₆ = 3.3, P = 0.010; R3: F₆,₆₆ = 4.3, P = 0.002; R4: F₆,₆₆ = 1.8, P = 0.134; R5: F₆,₆₆ = 1.6, P = 0.172; R6: F₆,₆₆ = 5.7, P < 0.001; Fig. S1). There was also no significant interaction between behavioral type and groupmate type (computer or humans), suggesting that all types treated humans and computers alike (GLMM testing...
for an interaction effect on contributions between behavioral type and nature of groupmates: $F_{2,173} = 0.470, P = 0.799$.

Furthermore, at the individual level, contributions made in one-shot games with humans were correlated with and showed no significant difference to contributions in the one-shot game with computers. This significant similarity was true for the first round with humans (mean difference = 0.57 MU, correlation = 0.73, $P < 0.001$; paired $t$-test $t_{(71)} = 0.96, P = 0.339$), for each and every round, and for the mean contribution across all six rounds (mean difference = 0.38 MU, correlation = 0.78, $P < 0.001$; paired $t$-test $t_{(71)} = 0.73, P = 0.471$; Table S2).

As described in the main text, we also compared whether individuals differed in how they adjusted their behavior conditionally, depending on whether they were playing computers or humans. At the same time that they made their contribution decision, we asked all individuals what they expected the mean contribution of their groupmates would be. We did this to test whether (i) players’ contributions in the unconditional rounds are a function of their expectations about their groupmates’ mean contributions; (ii) their functions differed when playing with computers or humans; and (iii) their functions in the unconditional rounds corresponded to and were consistent with their behavior in the strategy method.

First, as in previous studies, and reported in the main text, we found that contributions were positively correlated with the amount that they expected their human group mates to contribute, confirming that one does not necessarily need to incentivize the reporting of expectations (GLMM on six rounds of data, contribution $~$ expectation: $F_{1,468} = 152.9, P < 0.001, \beta = 0.210 \pm 0.017$; Fig. S2). This correlation was true overall and separately for each of the six rounds with humans (GLM round 1: $F_{1,70} = 59.1, P < 0.001, \beta = 0.32 \pm 0.054, R^2_{\text{adj}} = 0.49$; GLM round 2: $F_{1,70} = 51.4, P < 0.001, \beta = 0.26 \pm 0.044, R^2_{\text{adj}} = 0.50$; GLM round 3: $F_{1,71} = 37.1, P < 0.001, \beta = 0.23 \pm 0.047, R^2_{\text{adj}} = 0.41$; GLM round 4: $F_{1,70} = 41.6, P < 0.001, \beta = 0.21 \pm 0.038, R^2_{\text{adj}} = 0.46$; GLM round 5: $F_{1,70} = 50.8, P < 0.001, \beta = 0.22 \pm 0.039, R^2_{\text{adj}} = 0.50$; GLM round 6: $F_{1,70} = 49.3, P < 0.001, \beta = 0.23 \pm 0.040, R^2_{\text{adj}} = 0.46$; note $R^2_{\text{adj}}$ are from LMs rather than GLMs that control for the binomial error structure).

Second, as we reported in the main text, we also found qualitatively the same relationship between contributions and expectations when playing with computers (GLM: $F_{1,70} = 17.0, P = 0.001, \beta = 0.17 \pm 0.046, R^2_{\text{adj}}$ from a linear model = 0.22; Fig. S2), and analyzing all of the data together shows that quantitatively this correlation did not significantly depend on whether groupmates were computers or humans (GLMM: interaction between nature of groupmates and expectations, $F_{1,468} = 2.5, P = 0.116, \text{difference in } \beta = 0.05 \pm 0.034$; Fig. S2). The above results were robust to various forms of analysis. The overall result remained qualitatively the same if we changed from model-based covariances to using robust covariances instead ($F_{1,96} = 1.1, P = 0.296$). The interaction also remained nonsignificant if we compared play with computers with each round of play with human groupmates separately (GLMM, computer vs. human round 1: $F_{1,80} = 1.0, P = 0.330$, difference in $\beta = 0.039 \pm 0.040$; computer vs. human round 2: $F_{1,83} = 0.4, P = 0.520$, difference in $\beta = 0.025 \pm 0.039$; computer vs. human round 3: $F_{1,92} = 0.0, P = 1.000$, difference in $\beta = 0.000 \pm 0.044$; computer vs. human round 4: $F_{1,94} = 0.1, P = 0.720$, difference in $\beta = 0.016 \pm 0.045$; computer vs. human round 5: $F_{1,95} = 0.1, P = 0.885$, difference in $\beta = 0.006 \pm 0.042$; computer vs. human round 6: $F_{1,96} = 0.1, P = 0.918$, difference in $\beta = 0.004 \pm 0.040$).

We also repeated the above analyses, comparing contributions as a function of expectations between the round with computers and the six rounds with humans, but separately for noncooperators ($n = 15$) and conditional cooperators ($n = 36$). Neither conditional cooperators nor noncooperators differ in how they do or do not condition their contributions on their expectations depending on the nature of their groupmates (GLMM: interaction between expectations and groupmates on contributions, conditional cooperators: $F_{1,243} = 0.001, P = 0.973$; noncooperators: $F_{1,100} = 3.418, P = 0.067$).

For completeness, we also compared the extent to which individuals contributed more or less than they expected their groupmates to contribute and compared whether this differed depending on playing with humans or computers. For instance, did those that contributed a little less than they expected of their human groupmates (selfish biased conditional cooperation; refs. 1 and 2) do the same when playing computers? This allowed us to ask if behavior differed when playing with computers or humans while controlling for expectations; for example, maybe players cooperated with computers but became more generous with humans? For the sake of simplicity, we merely compare the first, or final, round of play with humans with the round of play with computers, using a paired-samples $t$-test to control for individual.

On average, individuals contributed 1.1 (first round) or 1.5 MU (final round) less than they expected of their human groupmates and 2.3 MU less than they expected of the computer. These decisions were not significantly different in magnitude (paired $t$ tests, first round: $t_{(71)} = 1.9, P = 0.062$; final round: $t_{(71)} = 0.9, P = 0.375$). Repeating the above process for the behavioral types reveals the same results for the conditional cooperators (paired $t$ tests, first round: $t_{(35)} = -0.1, P = 0.936$; final round: $t_{(35)} = 0.4, P = 0.657$); bumped cooperators (paired $t$ tests, first round: $t_{(10)} = 0.3, P = 0.746$; final round: $t_{(6)} = -0.6, P = 0.601$); and the unclassified players (paired $t$ tests, first round: $t_{(14)} = -1.2, P = 0.258$; final round: $t_{(13)} = 0.2, P = 0.813$). Noncooperators, in contrast, do show a significant difference, contributing 7.8 MU less than their expectations of the computer, but only 3.0 (first round) or 3.5 MU (final round) less than their expectations of humans (paired $t$ tests, first round: $t_{(14)} = -2.6, P = 0.020$; final round: $t_{(14)} = -1.9, P = 0.08$). However, this is not because they contributed more to humans (they tended to contribute 0 MU regardless) but because they had higher expectations of their computerized groupmates (mean $= 9.1$ MU) than their human groupmates (mean $= 5.3$ MU; GLMM: $F_{1,99} = 9.5, P = 0.003$, leading to a larger discrepancy.

Third, we then investigated whether their contributions in the unconditional rounds corresponded to and were consistent with their behavior in the strategy method. Methodologically, if a particular individual contributed more to human groupmates than they did in the strategy method for the corresponding expectation, then we scored them as having a positive discrepancy and vice versa (Fig. S3). For the sake of simplicity, we only analyzed behavior from the first round of play with humans.

When comparing individual contributions, as a function of expectations, with behavior in the strategy method, nearly 50% of the 72 participants were perfectly consistent, contributing the identical amount in the one-shot encounter as they did in the strategy method for the corresponding expectation. This was true regardless if their groupmates were computers (34/72, 47%) or humans (33/72, 46%; FET: $P = 1.000$) and also suggests affective forecasting was not a serious problem (Fig. S3).

We found that our participants, when playing with either computers or humans, were equally likely to be consistent with their prior decision in the strategy method. We also found that those that deviated, also deviated equally, in both absolute terms (mean absolute discrepancy; vs. computers: $\pm 2.46$ MU and vs. humans: $\pm 3.29$ MU; paired $t$ test: $t_{(71)} = 0.83333, P = 0.074$) and in net terms (mean net discrepancy; vs. computers: $\pm 0.9028$ MU, and vs. humans: $\pm 0.9028$, paired $t$ test: $t_{(71)} = 0.0, P = 1.0$). These identical net discrepancies ($\pm 0.9028$) were not significantly different from 0 MU, either when playing with computers (LM: intercept, $F_{1,71} = 0.835, P = 0.083$) or, more importantly, when playing with humans (LM: intercept, $F_{1,71} = 2.014, P = 0.160$; Fig. S3).

In summary, although around 50% of players are not perfectly consistent when transferring from the strategy method to the
one-shot game, their shift in behavior does not depend on their group-mates being computers or humans, and they show no prosocial shift, being equally likely to become either more or less favorable to their human groupmates (mean net discrepancy of 0.90 MU not significantly different to 0 MU; Fig. S3).

**Conditional Cooperators Misunderstand the Game.** After we had our participants first play the strategy method with computers, then a one-shot game with computers, and then six one-shot games with humans, all without feedback to prevent any learning about the game’s payoffs or the propensities of other players, we then asked them a question to test whether they understood the game. Specifically, the question tested whether they knew that the payoff-maximizing strategy was independent of what other players contributed, or did they incorrectly believe that the game’s payoffs were interdependent. The question read “In the game, if a player wants to maximize his or her earnings in any one particular round, does the amount they should contribute depend on what the other people in their group contribute?” We allowed them to answer either yes/sometimes/no/unsure. Note as this was a multiple-choice format, some players may have answered correctly merely by chance. The full breakdown of results is available in Table S3, with responses separated by behavioral type as measured in the preceding strategy method with computerized groupmates.

As the classification scheme for different types can be made more or less complex, we have analyzed the responses using several different classification schemes, using logistic regression (GLM with binary-logit link) to test whether players responded correctly (chose no) or not (chose yes/sometimes/unsure). In all cases, the behavioral types correlated with the answers to our question on understanding (Table S4). Our most complex classification scheme (expansive) is represented in Fig. 1. It contains six behavioral types; the four from ref. 1, plus our two extra, which are negative cooperators and unconditional cooperators. This scheme significantly predicts if a player will answer our question about the nonstrategic nature of the game (GLM: $\chi^2(5) = 15.3, P = 0.009$). The next most complex uses the same criteria as ref. 1 but also includes negative cooperators (FO1 + negative), a phenomenon not occurring in ref. 1 and not documented in ref. 2. Again this scheme significantly predicts correct answers or not (GLM: $\chi^2(6) = 14.8, P = 0.005$). The next most complex classification scheme (FO1) is simply that of ref. 1. Here we had to choose which category is most suitable for the negative cooperators; as mentioned above, we chose to classify them as humped cooperators because humped cooperators also show a negative correlation, after an idiosyncratic point, between contributions and the contributions of others. Again this scheme significantly predicts understanding (GLM: $\chi^2(3) = 12.9, P = 0.005$).

More simply, we have a three-way classification scheme, which combines conditional and humped cooperators into one category compared with noncooperators and unclassified. Again this scheme significantly predicts understanding of the game (three-way, GLM: $\chi^2(2) = 12.9, P = 0.002$). Finally, we have two simple, binary classification schemes. First, do players cooperate (contribute) or not? The division of cooperators and noncooperators significantly predicts understanding (GLM: $\chi^2(1) = 11.9, P = 0.001$). As does our final scheme, which simply classifies players as making constant contributions (noncooperators and unconditional cooperators) or inconstant contributions (conditional, humped cooperators, and unclassified). The constant players may or may not cooperate but show no form of conditionality in any sense and thus act as though they believe the game’s payoffs are not interdependent. Again, this classification scheme significantly predicts the correct response to the question testing their beliefs and understanding about the game’s payoffs (GLM: $\chi^2(1) = 12.1, P = 0.001$).

Finally, evidence that many players were uncertain in how to play the game comes from the lack of stability in contributions exhibited by our players in the preceding six rounds of play with human groupmates. Such instability suggests a degree of bet hedging by many players. In these rounds, which had no information feedback and thus no learning about the game or one’s groupmates could occur, only 21 (29%) of the 72 players remained constant, and 13 of these were noncooperators (contributed 0 MU). On average, the 72 players changed their contributions by $\pm 3.2$ MU from the previous round (mean absolute changes for rounds 2, 3, 4, 5, and 6, respectively: $2.4, 3.3, 3.8, 3.3,$ and $3.2$ MU). On average, the mean range in contributions (maximum − minimum contribution) was $7.3$ MU, with $10$ (14%) players exhibiting the full range possible ($0−20$ MU).

**Standard Control Questions Fail to Control for Understanding.** The control questions are available in SI Appendix. We have sufficient experience of running public-goods game experiments to confirm the difficulty in getting all players to answer questions correctly in a timely manner. An experimenter nearly always has to get involved, either in person or on-screen, and try to help without helping too much, as is clear from the following quotes; “After completion of the questionnaire, the questions were publicly solved. Any remaining questions were answered in private.” (3, p. 176); “Subjects who showed problems in understanding the task were given assistance.” (4, p. 89); “Once all participants had completed the exercises the experimenter solved them in public. Any remaining questions that the subjects had were then answered in private.” (6, p. 151); “if any participant repeatedly failed to answer correctly, the experimenter provided an oral explanation.” (50, p. 200) (bold emphases added). An anonymous reviewer also claimed that in ref. 2: “Before the start of the experiment subjects had to read the instructions and to answer control questions and were instructed again when necessary”.

Clearly such processes can influence the participants and allow participants through that have not independently answered all questions. For this reason, we allowed our players to answer as they wished, without any further instruction. This allowed us to identify where each participant went wrong, to identify sources of misunderstanding, and to quantify, for the first time to our knowledge, how well people respond to such questions.

It is not always clear if a player answered the control questions correctly. In the strictest sense, 16 players (22%) answered all 10 questions, pertaining to the four scenarios, perfectly. The breakdown per scenario was as follows: scenario 1 (two questions), 47 of 72 (65%) players answered correctly; scenario 2 (two questions), 39 (54%) players answered correctly; scenario 3 (three questions), 23 (32%) players answered correctly; and scenario 4 (three questions), 37 (51%) players answered correctly. These correct response rates differ significantly ($\chi^2$ test: $\chi^2 = 16.6, P < 0.001$). However, some participants answered in ways that could be interpreted as correct based on certain ambiguities in the questions. For example, in scenario 1, nobody contributes, so nobody makes a profit on his or her initial endowment of 20 MU. Some players incorrectly entered 0 MU, perhaps meaning a gain of 0 MU relative to the endowment, although the question asked them to calculate the “total income (in MU).” Six players appeared to correctly calculate the returns from the public project, but either forgot or thought that one does not add the saved, noncontributed, MU. To be generous, one could also categorize these players as passing the test, to give a total of 22 players (31%) that answered the questions correctly.

As a robustness check of the results in the main text, we analyzed our data from another, larger, study conducted in the same laboratory. This study had 216 players and was very similar in that the players had no identical instructions and control questions that were copied verbatim ref. 35 instead of ref. 2. Ref. 35 uses eight control questions that are an identical subset covering the same four scenarios of ref. 2 but with only two subsections instead of three for scenarios 3 and 4. These players then also played a one-shot unconditional game with the computer and the six-stage one-shot unconditional game with humans and answered our control question.
on whether the game is interdependent or not (but in the post-experiment questionnaire rather than in the experiment). The main difference was that they did not complete a strategy method game.

Of these 216 players, 78 (36%) correctly answered all eight questions, and of these, only 41 (53%) maximized their income (contributed 0 MU) when playing with computers. These 78 players contributed no more on average when playing with humans (5.4 MU) than they did when playing the computer (4.6 MU, paired t test: \( t_{(77)} = 1.0, P = 0.311 \)). Regarding our control question on whether the game is interdependent or not, we found that only 29 (37%) of the 78 correctly answered that the game is not interdependent. However, 21 of these 29 (72%) maximized their income when playing with the computer, significantly more than compared with the 20 of the 49 (41%) that answered our question incorrectly (FET: \( P = 0.0097 \)). Thus, our results are robust when analyzing a larger sample than presented in the main text. This larger sample confirms that answering the standard control questions does not guarantee that players understand the game and that our control question significantly helps explain behavior even among those that answered all of the standard control questions correctly.

**Comprehenders Are Not Cooperators.** We conducted paired \( t \) tests between individual contributions to the computer and contributions to humans for those that could be deemed to understand the game (comprehenders). We examined the behavior of three types of players that could all be argued to be comprehenders. First, all of the individuals that contributed 0 MU in both the strategy method and in the one-shot game with the computer (\( n = 13 \) of 72, 18%) and thus could be argued to best comprehend the game as they demonstrated they knew the income-maximizing strategy. Of course some of these could be false positives as some players may not fully understand the game but still contribute 0 MU for various reasons. Second, the individuals that answered all of the standard control questions correctly and thus have previously been assumed to understand the game (\( n = 16, 22\% \)) (1–6). Third, the individuals that passed our beliefs test and arguably understood the essential dilemma of the game (\( n = 21, 29\% \)). Again some of these may be false positives, for two reasons: (i) as the test was a multiple-choice question meaning some could guess the correct answer; and (ii) some players may correctly think that the decisions of one’s groupmates are irrelevant to one’s attempts to maximize earnings but incorrectly believe that to contribute is always profitable.

In all three cases, we find that comprehenders are no more cooperative with humans than they are with computers. First, the 13 players that maximized their income when playing computers did not contribute significantly more than 0 MU in any individual round with humans (Table S5). The mean contributions by round were as follows: 1.6; 0.6; 1.0; 2.1; 1.9; and 0.0 MU for the six rounds, respectively. None of these were significantly different from 0 MU (paired \( t \) tests: \( R_1, t_{(12)} = 1.3, P = 0.224; R_2, t_{(12)} = 1.0, P = 0.337; R_3, t_{(12)} = 1.3, P = 0.227; R_4, t_{(12)} = 1.5, P = 0.168; R_5, t_{(12)} = 1.4, P = 0.175; R_6, t_{(12)} = 0.0, P = 1.0 \)), nor across all six rounds combined (paired \( t \) test: \( t_{(12)} = 1.957, P = 0.074 \)).

Second, the 16 players that answered all of the standard control questions correctly showed no prosocial bias toward humans, giving just as much to computers (4.8 MU) as they did to humans in each and every round. They did not contribute significantly more than 4.8 MU in any individual round with humans. The mean contributions by round were as follows: 5.3; 5.7; 4.6; 6.4; 6.2; and 4.3 MU for the six rounds, respectively. None of these were significantly different from 4.8 MU (paired \( t \) tests: \( R_1, t_{(15)} = 0.4, P = 0.708; R_2, t_{(15)} = 0.7, P = 0.482; R_3, t_{(15)} = 0.2, P = 0.815; R_4, t_{(15)} = 0.8, P = 0.450; R_5, t_{(15)} = 1.0, P = 0.338; R_6, t_{(15)} = 0.6, P = 0.562 \)) or across all six rounds combined (paired \( t \) test: \( t_{(15)} = 0.6, P = 0.529 \)).

Third, the 21 players that answered our control question correctly also showed no prosocial bias toward humans. They did not contribute significantly more than the 5.1 MU they gave to the computers in the first round in five of the six subsequent individual rounds with humans. The mean contributions by round were as follows: 7.0; 6.6; 6.7; 7.1; 7.6; 6.4; and 6.9 MU for the six rounds, respectively. Only one of these was significantly different from 5.1 MU (paired \( t \) tests: \( R_5, t_{(20)} = 2.1, P = 0.048 \)), and this would no longer be the case if we were to control for multiple testing (e.g., with a Bonferroni correction) (Table S6). Not one of the other rounds were significantly different (\( R_1, t_{(20)} = 1.4, P = 0.186; R_2, t_{(20)} = 1.4, P = 0.183; R_3, t_{(20)} = 1.5, P = 0.139; R_4, t_{(20)} = 1.4, P = 0.188; R_6, t_{(20)} = 1.3, P = 0.224 \)) or across all six rounds combined (paired \( t \) test: \( t_{(20)} = 1.7, P = 0.098 \)).

**Measuring Motivations.** We assumed that players playing with computers are trying to maximize their income as much as players are ever motivated to do so. We have then inferred that because behavior is unchanged when shifting from playing with computers to humans that players are equally motivated to try and maximize personal income when playing with humans. One may be tempted to ask, why not simply ask the players what their motivations were? Therefore, in a postgame questionnaire, conducted while the payments to our players were prepared, we asked our players about their motivations during their games with humans.

As this was a hypothetical questionnaire, players could make their responses appear prosocial without paying a cost to do so, contrary to typical economic decisions in experiments measuring social preferences. Therefore, responses to this question could be argued to provide a lower bound of self-interested motivations among our players. We asked, “What was your most important motivation in the games with real people? Please select the answer that best describes your motivations?” (numbers chose in brackets).

- Making myself the maximum money possible (36)
- Making other people the maximum money possible (3)
- Making everyone the maximum money possible (13)
- Making the group the maximum money possible (11)
- Making myself more money than other people (1)
- Other (8)

We consider options 1 and 5 to reveal a self-interested motivation (\( n = 37 \)), whereas options 2–4 are prosocial (\( N = 27 \), or 35 if including others). Despite the opportunity to appear prosocial at no cost, and in contrast to the low frequency of noncooperators in our incentivized strategy method (\( n = 15, 21\% \)), the most common response, chosen by 50% of players, was a self-interested motivation, “making myself the maximum money possible” (\( n = 36 \) of 72). Table S6 reveals how these responses depended on behavioral type.

**SI Discussion**

**Why Might Players Think That Contributing to the Public Good Is a Profitable Decision?** It has been claimed that the income-maximizing strategy is “simple to figure out” (13). However, this claim cannot be asserted; it has to be scientifically ascertained. We argue that there are several reasons why players may reach mistaken conclusions about the game. Although we do not wish to criticize the difficult job of making instructions for economic games, we think it possibly significant that the instructions we copied from ref. 2 used the word invest when describing contributions. The word invest has clear financial connotations to do with risk and gain (as confirmed in ref. 12). The word earn has no such connotations and may depend on other factors, such as what other people choose to invest. Also, the fact that the experimenter then offers players the opportunity to explicitly condition on the behavior of others would only further reinforce this belief (27, 28). Furthermore, the instructions specify that each MU contributed returns 0.4 MU, but this
does not clearly specify that they incur a cost of ~0.6 MU, and judging from the answers to the control questions, it seems that some participants thought that each MU contributed returned 1.4 MU (i.e., the 0.4 MU return was pure profit). Evidence of this potential confusion comes from a study showing that cooperation is significantly reduced in the public-goods game when players are explicitly informed that they “lose money on contributing” (63). Finally, one reason that people may contribute is that they may think contributing something is a necessary requirement to benefit from the public good (e.g., paying to become a member of a club/group in the world). These mixed reasons may explain why many of our players varied their contributions during the six rounds of play with humans. Such within-individual variation suggests some players used a bet-hedging strategy in response to a risky or uncertain decision environment.

However, why would the standard control questions (SI Appendix) not prevent these erroneous beliefs? An examination of the specific questions will show that the first two scenarios do little other than reinforce these false beliefs. Specifically, they show that if no one contributes, then no one makes any additional money (scenario 1), but then conversely if everyone contributes fully, then everyone makes a lot of money (scenario 2). The only hint that contributions are costly comes from scenario 3, whereby players can see that for the case where the rest of the group invests 30 MU, to contribute 0 MU is better than 10 or 15 MU. However, scenario 4 then shows them that they benefit the more others contribute, specifically if they invest 8 MU, they do better if the others contribute 22 MU instead of 12 or 7 MU. Tellingly, players struggle the most with scenario 3, with only 32% of player answering correctly, compared with 65%, 54%, and 51% for scenarios 1, 2, and 4, respectively. This significant difference means that if experimenter assistance is provided, it is most likely to be for scenario 3, which is the only scenario that demonstrates there is a social dilemma.

Why Might Players Show a Correlation Between Contributions and Either Expectations of Others or the Actions of Others? Again, without wishing to criticize the instructions, they were perhaps not ideal, as they specifically state to the participants that they can “condition their contribution on that of the other group members” (2). In addition, the input screen asks them to enter their “conditional contribution to the project” (2). This suggestive language implies that this is something they may wish to consider doing and therefore may act as a demand characteristic for the game (27, 28).

There are also other reasons why contributions may correlate with either expectations or the decisions of others, without reflecting concerns for fairness or inequity aversion (59). Conditional cooperators are considered to have social preferences because their decisions correlate with their knowledge and beliefs about the contributions of their groupmates. However, this correlation does not automatically imply social preferences. If we relax the assumption that players perfectly understand the game, with full confidence, then there are three alternative reasons, none of which require any social preferences, as to why this correlation may exist: (i) if a player thinks that “contributing X MU is best,” then it is reasonable for them to conclude that many others will think likewise, making their expectations match their own behavior, rather than vice versa; (ii) if a player is uncertain but knows that others are contributing X MU, then it is reasonable for them to conclude that others know better what to do and to copy their behavior in the hope of a better payoff, making their behavior match their knowledge of other players’ behavior; and (iii) if a player thinks that the income-maximizing response to what others contribute changes depending on what they contribute, then they will in some sense condition their contributions on these other contributions. This will lead to their behavior matching both their knowledge and expectations of others, even when others are merely computers playing randomly, as evidenced by our results.

A revealing insight from the strategy method is the presence of humped and negative cooperators. These players reduce their contributions the more their groupmates contribute. This is not consistent with a sense of fairness or social preferences. It is consistent with the appropriate strategy for a threshold, or nonlinear, public-goods game (62). In such games, a self-interested player can benefit by compensating for the lack of contributions by others players. Clearly, in such games, the income-maximizing strategy does depend on knowing the behavior of others.

It is probably reasonable to consider that some of these behaviors may be unique artifacts of the strategy method. For instance, negative and humped cooperators may not occur in unconditional versions of the game, although they might. Due to their rarity, it would require a huge sample size to capture enough of them to detect their behavior in the unconditional game. It would also require repeated games that control for learning, through the omission of feedback between rounds, to sample enough decisions from each individual. However, conditional cooperators are much more common and show consistency between play in the unconditional games and the strategy method. This means that even if the strategy method inflates the frequency of apparent conditional cooperators, concerns about the misinterpretation of their motives are not just restricted to experiments using the strategy method.
Fig. S1. Behavioral type predicts play with humans in each round. Shown are the mean contributions (0–20 MU) to the public good, grouped by behavioral type. We followed previous studies in dividing our players into separate types on the basis of their pattern of contributions when playing the strategy method with computers (Fig. 1). Different types contributed different amounts in unconditional games, but for each type, there was no significant difference between the mean contributions to groups with computers for groupmates (dark gray) and groups with humans (light gray). Dashed lines represent 95% binomial CIs.

Fig. S2. People respond to computers and humans alike. Shown are the individual contributions (0–20 MU, n = 72) to the group project of players playing either with computers (one time, black circles) or humans (six times, gray circles). The x axis is the individual's reported expectation of what they thought their groupmates would contribute. Players' contributions correlated with their expectations regardless of playing with computers or humans. The color-coded slopes are predicted values from the parameter estimates of a GLMM fitting contributions as dependent on expectations and are not significantly different.
Fig. S3. No prosocial shift in response to playing with humans. For each player we used a record of what they had contributed to the group project (0–20 MU) in the first round of an unconditional game and of what they had expected their groupmates’ mean contribution to be. This allowed us to compare their unconditional contribution as a response to their expectations, with what they had contributed in the prior strategy method (Fig. 1) for the same mean contribution of groupmates. If the contributions were the same, then the player scored a discrepancy value of 0. If they contributed more in the unconditional game than in the strategy method game, then they scored a positive discrepancy. For example, if they contributed 10 MU in the unconditional game and 5 MU in the strategy method, then they scored a discrepancy of +5 and vice versa. Overall, the mean discrepancy was identical, at +0.9028 MU (black line), when playing with computers (black circles) or humans (light gray circles). This was not significantly different to 0 (black/gray dashed lines show 95% CI).
Table S1. Separate distribution of behavioral types reported from 14 trials presented in eight studies (including this one) measuring conditional cooperation using the strategy method of ref. 1

<table>
<thead>
<tr>
<th>Type</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>1–12</th>
<th>13–14</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study code</td>
<td>F01</td>
<td>FG10</td>
<td>K08</td>
<td>K08</td>
<td>HT09</td>
<td>HT09</td>
<td>HT09</td>
<td>HT09</td>
<td>M09</td>
<td>M09</td>
<td>M13</td>
<td>FV10</td>
<td>BC16</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Groupmates</td>
<td>Human</td>
<td>Human</td>
<td>Human</td>
<td>Human</td>
<td>Human</td>
<td>Human</td>
<td>Human</td>
<td>Human</td>
<td>Human</td>
<td>Human</td>
<td>Computer</td>
<td>Computer</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Country</td>
<td>Switzerland</td>
<td>Switzerland</td>
<td>Austria</td>
<td>Japan</td>
<td>United States</td>
<td>Russia</td>
<td>Russia</td>
<td>Russia</td>
<td>Russia</td>
<td>Colombia</td>
<td>Colombia</td>
<td>Colombia</td>
<td>Vietnam</td>
<td>United States</td>
<td>United Kingdom</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Conditional</td>
<td>50%</td>
<td>55%</td>
<td>44%</td>
<td>42%</td>
<td>81%</td>
<td>48%</td>
<td>57%</td>
<td>60%</td>
<td>60%</td>
<td>63%</td>
<td>54%</td>
<td>50%</td>
<td>53%</td>
<td>50%</td>
<td>55%</td>
<td>51%</td>
<td>0.470*</td>
</tr>
<tr>
<td>Humped</td>
<td>14%</td>
<td>12%</td>
<td>11%</td>
<td>11%</td>
<td>0%</td>
<td>5%</td>
<td>5%</td>
<td>10%</td>
<td>12%</td>
<td>8%</td>
<td>8%</td>
<td>8%</td>
<td>13%</td>
<td>10%</td>
<td>9%</td>
<td>11%</td>
<td>0.600*</td>
</tr>
<tr>
<td>Unclassified</td>
<td>7%</td>
<td>10%</td>
<td>22%</td>
<td>11%</td>
<td>11%</td>
<td>43%</td>
<td>27%</td>
<td>20%</td>
<td>27%</td>
<td>25%</td>
<td>13%</td>
<td>38%</td>
<td>13%</td>
<td>19%</td>
<td>20%</td>
<td>17%</td>
<td>0.602*</td>
</tr>
<tr>
<td>Free rider</td>
<td>30%</td>
<td>23%</td>
<td>22%</td>
<td>36%</td>
<td>8%</td>
<td>5%</td>
<td>11%</td>
<td>10%</td>
<td>2%</td>
<td>4%</td>
<td>25%</td>
<td>4%</td>
<td>23%</td>
<td>21%</td>
<td>16%</td>
<td>21%</td>
<td>0.168*</td>
</tr>
<tr>
<td>Study sample</td>
<td>44</td>
<td>140</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>44</td>
<td>44</td>
<td>20</td>
<td>52</td>
<td>48</td>
<td>48</td>
<td>48</td>
<td>40</td>
<td>72</td>
<td>596</td>
<td>112</td>
<td>0.458†</td>
</tr>
</tbody>
</table>

Frequencies are presented as percentages; there may be some errors in totals due to rounding. For a summary statistics and comparisons of the distributions by studies using computers instead of humans, see the final three columns. F01, Fischbacher et al., 2001 (1); FG10, Fischbacher and Gachter 2010 (2); K08, Kocher et al., 2008 (3); HT09, Herrmann and Thoni 2009 (4); M09, Martinsson et al., 2009 (5); M13, Martinsson et al., 2013 (6); FV10, Ferraro and Vossler 2010 (50); BC14, this study. Trials 3–5 were one study investigating cross-cultural distributions. Trials 6–9 were one study, conducted in four Russian locations: two rural (6, 7) and two urban (8, 9). Trials 10 and 11 were one study, conducted in Colombia, with individuals of low-middle socioeconomic status (10) and high socioeconomic status (11). Trials 12 and 13 are separate studies, both conducted using computerized groupmates instead of humans. NA, not applicable.

*Fisher’s exact test of the corresponding absolute counts summed across all studies with either human or computer groupmates.
†χ² test of overall distributions. The absolute counts for each type were summed across all studies either with humans or with computers.
Table S2. Correlations and paired t tests between individual contributions to the computer and contributions to humans across six rounds (rounds 1–6; n = 72; Fig. S1)

<table>
<thead>
<tr>
<th>Round</th>
<th>Mean contribution</th>
<th>Mean increase</th>
<th>Correlation*</th>
<th>Paired t(df)</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8.3</td>
<td>0.6</td>
<td>0.73</td>
<td>1.0(71)</td>
<td>0.339</td>
</tr>
<tr>
<td>2</td>
<td>8.3</td>
<td>0.6</td>
<td>0.69</td>
<td>1.0(71)</td>
<td>0.307</td>
</tr>
<tr>
<td>3</td>
<td>8.0</td>
<td>0.3</td>
<td>0.64</td>
<td>0.3(71)</td>
<td>0.695</td>
</tr>
<tr>
<td>4</td>
<td>8.3</td>
<td>0.6</td>
<td>0.58</td>
<td>0.8(71)</td>
<td>0.427</td>
</tr>
<tr>
<td>5</td>
<td>8.1</td>
<td>0.4</td>
<td>0.63</td>
<td>0.5(71)</td>
<td>0.598</td>
</tr>
<tr>
<td>6</td>
<td>7.4</td>
<td>−0.3</td>
<td>0.69</td>
<td>−0.4(71)</td>
<td>0.710</td>
</tr>
<tr>
<td>1–6</td>
<td>8.1</td>
<td>0.4</td>
<td>0.78</td>
<td>0.7(71)</td>
<td>0.471</td>
</tr>
</tbody>
</table>

Mean increase, mean contribution with humans minus the mean contribution to computers (7.7 MU).

*Significance was below 0.001 in all cases.

Table S3. Strategies and understanding about the game were correlated

<table>
<thead>
<tr>
<th>Type (vs. computers)</th>
<th>Answer: no*</th>
<th>Answer: yes</th>
<th>Answer: sometimes</th>
<th>Answer: unsure</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noncooperators</td>
<td>10</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>Unconditional cooperators</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Conditional cooperators</td>
<td>6</td>
<td>18</td>
<td>7</td>
<td>5</td>
<td>36</td>
</tr>
<tr>
<td>Negative cooperators</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Humped cooperators</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Unclassified</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Totals</td>
<td>21</td>
<td>33</td>
<td>11</td>
<td>7</td>
<td>72</td>
</tr>
</tbody>
</table>

Response to question asking if players understood that the income maximizing strategy does not depend on what others contribute. Question asked: “In the game, if a player wants to maximize his or her earnings in any one particular round, does the amount they should contribute depend on what the other people in their group contribute?”

*This is the correct answer.

Table S4. Results of six logistic GLMs on the probability of understanding the game depending on behavioral type

<table>
<thead>
<tr>
<th>Classification scheme</th>
<th>Number of types</th>
<th>(\chi^2(\text{df}))</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expansive (Fig. 1)</td>
<td>6</td>
<td>15.3(5)</td>
<td>0.009</td>
</tr>
<tr>
<td>F01 + negative</td>
<td>5</td>
<td>14.8(4)</td>
<td>0.005</td>
</tr>
<tr>
<td>F01</td>
<td>4</td>
<td>12.9(3)</td>
<td>0.005</td>
</tr>
<tr>
<td>Three-way</td>
<td>3</td>
<td>12.5(2)</td>
<td>0.002</td>
</tr>
<tr>
<td>Binary (cooperator/non)</td>
<td>2</td>
<td>11.9(1)</td>
<td>0.001</td>
</tr>
<tr>
<td>Binary (constant/inconstant)</td>
<td>2</td>
<td>12.1(1)</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Each model uses a different scheme to classify types according to their contribution schedule in the strategy method. Expansive (Fig. 1) uses the classification scheme presented in Fig. 1, which is five categories plus unclassified. F01 + negative uses the four-way classification scheme published by ref. 1 plus the addition of negative cooperators that show a negatively conditional strategy. F01 uses the four-way classification scheme published by ref. 1. Three-way is the same as F01 but combines conditional and humped cooperators. Binary (cooperator/non) is a binary scheme comparing all players that contributed something to those that contributed 0 MU. Binary (constant/inconstant) is a binary scheme comparing those that contributed a constant amount (including 0 MU) or not.
Table S5. Individual contributions to human groupmates of the 13 players that did not cooperate with computers

<table>
<thead>
<tr>
<th>Player</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>R5</th>
<th>R6</th>
<th>$t_{05}$</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>6</td>
<td>16</td>
<td>0</td>
<td>10</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>2.2</td>
<td>0.084</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>8</td>
<td>3</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>1.9</td>
<td>0.113</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>12</td>
<td>5</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.0</td>
<td>0.363</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Contribution was 0–20 MU. Player number is arbitrary.

*Tests separately for each player if their six contributions to humans are different from 0.

Table S6. Measuring motivations

<table>
<thead>
<tr>
<th>Response*</th>
<th>NonC</th>
<th>ConC</th>
<th>HumC</th>
<th>Unc.</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myself</td>
<td>11</td>
<td>15</td>
<td>5</td>
<td>5</td>
<td>36</td>
</tr>
<tr>
<td>Other people</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Everyone</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>The group</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>More than others</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>None of the above</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>15</td>
<td>36</td>
<td>7</td>
<td>14</td>
<td>72</td>
</tr>
</tbody>
</table>

Responses chosen from a multiple-choice question asking “What was your most important motivation in the games with real people? Please select the answer that best describes your motivations.” NonC, noncooperator; ConC, conditional cooperator; HumC, humped cooperator; Unc., unclassified.

*Full answers available were as follows: “Making myself the maximum money possible”; “Making other people the maximum money possible”; “Making everyone the maximum money possible”; “Making the group the maximum money possible”; “Making myself more money than other people”; and “Other.”

Other Supporting Information Files

SI Appendix (PDF)
APPENDICES

Appendix 1: The experimental design and order.

Appendix 2: Copy of our on-screen instructions adapted from reference (2).

Appendix 3: Copy of our printed hand-out of instructions.

Appendix 4: Standard control questions, copied from reference (2).

Correct answers shown in (brackets).

Appendix 5: Screen shots.

Appendix 6: The individual contribution schedules for all 72 participants when completing the strategy-method version of the public goods game with computerized groupmates.
Appendix 1: The experimental design and order (see next page).

All 72 participants interacted through computer terminals. In stage 1 they received online instructions at their own pace (Appendix 2). They also received a print-out of the instructions which they kept for the duration of the experiment (Appendix 3). The instructions did not yet mention that they would play with computers at some stage. Stage 2: participants answered 10 control questions asking them to calculate player payoffs in various hypothetical scenarios (Appendix 4). Stage 3: players were told that they were now going to play with computerized groupmates and made to complete a contribution schedule for all possible average contributions of their three computerized groupmates (Appendix 6). Stage 4: participants now played a one-shot game (one unconditional contribution plus recorded expectation of groupmates’ mean contribution) still with computerized groupmates. Stage 5: players now repeated the one-shot game but human groupmates for six rounds, again recording their expectations of groupmates’ mean contributions. In all cases our players did not learn about their payoffs or the decisions of others so no learning or strategic play was possible. Stage 6: players answer a question testing if they understand that the income-maximizing strategy is independent of the behaviour of others in a one-shot game.
Appendix 1. The experimental order.

1) Instructions
2) Control questions
3) Strategy method with computers
4) One shot game with computerized groupmates (contribution and expectations of groupmates)
5) 6 x One shot games with human groupmates (contributions and expectations of groupmates)
6) Question testing understanding of game: does the best strategy depend on behaviour of others?

Strategy Method: Computer opponents

Your conditional contribution to the project (contribution schedule).

If computer players contribute, on average, 0. Your contribution is?
If computer players contribute, on average, 1. Your contribution is?
If computer players contribute, on average, 2. Your contribution is?
If computer players contribute, on average, 3. Your contribution is?
Appendix 2: Copy of our on-screen instructions adapted from reference (2).

You are now taking part in an economic experiment, which has been financed for purely academic research purposes.

If you read the following instructions carefully, you can, depending on your decisions, earn a reasonable amount of money.

It is therefore very important that you read these instructions with care.

The instructions, which we have distributed to you, are solely for your private information.

You are not allowed to communicate during the experiment.

If you have any questions, please ask us. Violation of this rule will lead to your exclusion from the experiment and all payments.

If you have questions, please raise your hand. A member of the experimenter team will come to you and answer them in private.

During the experiment we shall not speak of £ Pounds but rather of MU (Monetary Units).

During the experiment your entire earnings will be calculated in MU.

At the end of the experiment the total amount of MU you have earned will be converted to £ Pounds at the following rate:

1 MU = 3 pence, so 20 MU = 60 pence and 33 MU = 99 pence.

At the end of the experiment your entire earnings from the experiment will be immediately paid to you in cash.

We describe the experimental process below.

The decision situation

You will learn how the experiment will be conducted later. We first introduce you to the basic decision situation. You will find control questions at the end of the instructions that help you to understand the decision situation.

You will be in a group consisting of 4 people. Each group member has to decide on the allocation of 20 MU. You can put these 20 MU into your private account or you can invest them fully or partially into a project.

Each point you do not invest into the project, will automatically remain in your private account.

Your income from the private account:

You will earn one MU for each MU that you put into your private account.

For example, if you put 20 MU into your private account (and therefore do not invest into the project) your income will amount to exactly 20 MU out of your private account.
If you put 6 MU into your private account, your income from this account will be 6 MU. **No one except you earns something from your private account.**

[new screen]

**Your income from the project:**

Each group member will profit equally from the amount you invest into the project.

On the other hand, you will also get a payoff from the other group members' investments.

The income for each group member will be determined as follows:

\[
\text{Income from the project} = \text{sum of all contributions} \times 0.4
\]

If, for example, the sum of all contributions to the project is 60 MU, then you and the other members of your group each earn 60 \times 0.4 = 24 MU out of the project.

If four members of the group contribute a total of 10 MU to the project, you and the other members of your group each earn 10 \times 0.4 = 4 MU.

**Total income:**

Your total income is the sum of your income from your private account and that from the project:

\[
\text{Income from your private account} = 20 - \text{contributions to the project} \\
+ \text{Income from the project} = 0.4 \times \text{sum of all contributions to the project} \\
= \text{Your Total Income}
\]
Appendix 3: Copy of our printed hand-out of instructions.

This information is the same as the on-screen instructions and can be used as a reminder.

We describe the experimental process below.

The decision situation
You will be in a group consisting of 4 people. Each group member has to decide on the allocation of 20 monetary units (MU). You can put these 20 MU into your private account or you can invest them fully or partially into a group project. Each point you do not invest into the group project, will automatically remain in your private account.

Your income from the private account:
You will earn one MU for each MU that you put into your private account.
For example, if you put 20 MU into your private account (and therefore do not invest into the project) your income will amount to exactly 20 MU out of your private account.
If you put 6 MU into your private account, your income from this account will be 6 MU.
No one except you earns anything from your private account.

Your income from the project:
Each group member will profit equally from the amount you invest into the project.
On the other hand, you will also get a payoff from the other group members' investments.
The income for each group member will be determined as follows:

Income from the project = sum of all contributions x 0.4

If, for example, the sum of all contributions to the project is 60 MU, then you and the other members of your group each earn 60 x 0.4 = 24 MU out of the project.
If four members of the group contribute a combined total of 10 MU to the project, you and the other members of your group each earn 10 x 0.4 = 4 MU.

Your total income:
Your total income is the sum of your income from your private account plus that from the group project:
Your Total Income = Your income from your private account (which = 20 – your contribution to the project) + Your income from the group project (which = 0.4 x the sum of all contributions to the group project)
Appendix 4: Standard control questions, copied from reference (2).
Correct answers shown in (brackets).

Please answer the following control questions. They will help you to gain an understanding of the calculation of your income, which varies with your decision about how to distribute your 20 MU.

There is a calculator available.
To access the calculator, click on the image in the bottom right of your screen.
Please enter your answers into the spaces provided and press Continue when done.

Question 1: Each group member has an endowment of 20 MU. Nobody (including yourself) contributes any MU to the project.
What is your total income (in MU)?
ANSWER BOX (20 MU)

What is the total income of each of the other group members?
ANSWER BOX (20 MU)

[New screen]

Question 2: Each group member has an endowment of 20 MU. You invest 20 MU in the project. Each of the other three members of the group also contributes 20 MU to the project. What is your total income (in MU)?
ANSWER BOX (32 MU)

What is the total income (in MU) of each of the other group members?
Question 3: Each group member has an endowment of 20 MU. The other three group members contribute a total of 30 MU to the project.

What is your total income (in MU), if you – in addition to the 30 MU – invest 0 MU into the project?

**Answer Box (32 MU)**

What is your total income (in MU), if you – in addition to the 30 MU – invest 10 MU into the project?

**Answer Box (26 MU)**

What is your total income (in MU), if you – in addition to the 30 MU – invest 15 MU into the project?

**Answer Box (23 MU)**

[New screen]

Question 4: Each group member has an endowment of 20 MU. Assume that you invest 8 MU to the project.

What is your total income (in MU), if the other group members together – in addition to your 8 MU – contribute another 7 MU in total to the project?

**Answer Box (18 MU)**

What is your total income (in MU), if the other group members together – in addition to your 8 MU – contribute another 12 MU in total to the project?

**Answer Box (20 MU)**

What is your total income (in MU), if the other group members together – in addition to your 8 MU – contribute another 22 MU in total to the project?
210

ANSWER BOX (24 MU)

211
Appendix 5: Screen shots.

Stage 1: Instructions (three screens). Also presented on paper and kept during experiment.

You are now taking part in an economic experiment, which has been framed purely for academic research purposes. If you read the following instructions carefully, you can, depending on your decisions, earn a reasonable amount of money. It is therefore very important that you read these instructions with care.

The instructions, which we have distributed to you, are solely for your private information. You are not allowed to communicate during the experiment.

If you have any questions, please ask us. Unprompted disclosure will lead to your exclusion from the experiment and all payments.

If you have questions, please take your hand in member of the experimenter team will come to you and answer them in private.

During the experiment we shall not speak of Pounds but rather of MU (monetary units). During the experiment your entire earnings will be calculated in MU. At the end of the experiment the total amount of MU you have earned will be converted to £ Pounds at the following rate: 1 MU = 3 pence, 20 MU = 50 pence and 30 MU = 90 pence.

At the end of the experiment your entire earnings from the experiment will be immediately paid to you in cash.

We describe the experimental process below.

The decision situation:

You will be shown how the experiment will be conducted later. We first introduce you to the basic decision situation. You will first control questions at the end of the instructions that help you to understand the decision situation.

You will be in a group consisting of 3 people. Each group member has to decide on the allocation of 20 MU. You can put these 20 MU into your private account or you can invest them fully or partially into a project. Each point you do not invest into the project, will automatically remain in your private account.

Your income from the private account:

You will earn one MU for each MU that you put into your private account. For example, if you put 20 MU into your private account (and therefore do not invest into the project) your income will amount to exactly 20 MU out of your private account.

If you put 6 MU into your private account, your income from the account will be 6 MU.

No one except you earns something from your private account.
Your income from the project:

Each group member will profit equally from the amount you invested into the project.

On the other hand, you will also get a payoff from the other group members' investments.

The income for each group member will be determined as follows:

Income from the project = sum of all contributions x 0.4

If, for example, the sum of all contributions to the project is 60 MU, then you and the other members of your group each earn 60 x 0.4 = 24 MU out of the project.

If four members of the group contribute a total of 15 MU to the project, you and the other members of your group each earn 15 x 0.4 = 6 MU.

Total income:

* Your total income is the sum of your income from your private account and that from the project.

Income from your private account = 24 - contribution to the project

+ Income from the project = 0.4 x sum of all contributions to the project

= Your Total Income
Stage 2: Control Questions (four screens).

Question 1: Each group member has an endowment of 20 MU. Nobody (including yourself) contributes any MU to the project. What is your total income (in MU)?

Question 2: Each group member has an endowment of 20 MU. You invest 20 MU in the project. Each of the other three members of the group also contributes 20 MU to the project. What is your total income (in MU)?

What is the total income of each of the other group members?
Question 3. Each group member has an endowment of 20 MU. The other three group members contribute a total of 30 MU to the project. What is your total income (in MU) if you, in addition to the 20 MU, invest 5 MU into the project?

What is your total income (in MU) if you, in addition to the 20 MU, invest 10 MU into the project?

What is your total income (in MU) if you, in addition to the 20 MU, invest 15 MU into the project?

Question 4. Each group member has an endowment of 20 MU. Assume that you invest 5 MU into the project. What is your total income (in MU) if the other group members together, in addition to your 5 MU, contribute another 10 MU in total to the project?

What is your total income (in MU) if the other group members together, in addition to your 5 MU, contribute another 12 MU in total to the project?

What is your total income (in MU) if the other group members together, in addition to your 5 MU, contribute another 15 MU in total to the project?
Stage 3: Playing the computer in the strategy method (four screens).
Your conditional contribution to the project (Contribution schedule).

<table>
<thead>
<tr>
<th>Computer players contribute, on average,</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer players contribute, on average, 2</td>
<td></td>
</tr>
<tr>
<td>Computer players contribute, on average, 3</td>
<td></td>
</tr>
<tr>
<td>Computer players contribute, on average, 4</td>
<td></td>
</tr>
<tr>
<td>Computer players contribute, on average, 5</td>
<td></td>
</tr>
<tr>
<td>Computer players contribute, on average, 6</td>
<td></td>
</tr>
<tr>
<td>Computer players contribute, on average, 7</td>
<td></td>
</tr>
<tr>
<td>Computer players contribute, on average, 8</td>
<td></td>
</tr>
<tr>
<td>Computer players contribute, on average, 9</td>
<td></td>
</tr>
<tr>
<td>Computer players contribute, on average, 10</td>
<td></td>
</tr>
<tr>
<td>Computer players contribute, on average, 11</td>
<td></td>
</tr>
<tr>
<td>Computer players contribute, on average, 12</td>
<td></td>
</tr>
<tr>
<td>Computer players contribute, on average, 13</td>
<td></td>
</tr>
<tr>
<td>Computer players contribute, on average, 14</td>
<td></td>
</tr>
<tr>
<td>Computer players contribute, on average, 15</td>
<td></td>
</tr>
<tr>
<td>Computer players contribute, on average, 16</td>
<td></td>
</tr>
<tr>
<td>Computer players contribute, on average, 17</td>
<td></td>
</tr>
<tr>
<td>Computer players contribute, on average, 18</td>
<td></td>
</tr>
<tr>
<td>Computer players contribute, on average, 19</td>
<td></td>
</tr>
<tr>
<td>Computer players contribute, on average, 20</td>
<td></td>
</tr>
</tbody>
</table>

Enter the amount which you want to contribute to the project. If the other players (the computer players) make the average contribution which it stands to the left of the entry field. This decision is binding and final. When you have completed your entries press the “Confirm” button.

Computer 1 randomly picked 5 MU.
Computer 2 randomly picked 10 MU.
Computer 3 randomly picked 10 MU.

Therefore the average computer contribution was 12 MU.
Therefore your conditional contribution was 2 MU.

Your profits are being calculated.
There is no further feedback for this round but any money you have earned has gone to your final income.
Stage 4: Playing the computer in the unconditional round (three screens).
Computer 1 randomly picked 10 MU.
Computer 2 randomly picked 8 MU.
Computer 3 randomly picked 11 MU.

Therefore the average computer contribution was 13 MU.

Your contribution was 12 MU.

Your profits are being calculated.
There is no further feedback for this round but any money you have earned has gone to your final income.
Stage 5: Playing humans without feedback for six rounds (three screens, screens 2 and 3 repeated six times).
Stage 6: Our control question for comprehension (one screen).

In the game, if a player wants to maximize his or her earnings in any one particular round, does the amount they should contribute depend on what the other people in their group contribute?
1. Yes
2. Sometimes
3. No
4. Unsure

Please enter the corresponding number (from 1 to 4) for your answer in the box.
Appendix 6: The individual contribution schedules for all 72 participants when completing the strategy-method version of the public goods game with computerized groupmates.

Each panel shows on the y-axis one participant’s responses in the strategy method for each and every possible mean computer contribution on the x-axis. The background colour indicates how each participant was classified, with purple = conditional-cooperator (N=36, pages 1-3); green = negative-cooperator (N=3); blue = humped-cooperator (N=4); yellow = unclassified (including constant or unconditional-cooperators, N=14); and pink = non-cooperator (N=15).