Conditional cooperation and confusion in public-goods experiments

Maxwell N. Burton-Chellew[a,b,c], Claire El Mouden[a,c], and Stuart A. West[a,b,1]

[a]Department of Zoology, University of Oxford, Oxford OX1 3PS, United Kingdom; [b]Calleva Research Centre for Evolution and Human Sciences, Magdalen College, Oxford OX1 4AU, United Kingdom; and [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.

altruism | strategy method | inequity aversion | reciprocity | social preferences

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 alterative 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

Significance

The finding that people vary in how they play economic games has led to the conclusion that people vary in their preference for fairness. Consequently, people have been divided into fair cooperators that make sacrifices for the good of the group and selfish free-riders that exploit the cooperation of others. This conclusion has been used to challenge evolutionary theory and economic theory and to guide social policy. We show that variation in behavior in the public-goods game is better explained by variation in understanding and that misunderstanding leads to cooperation.

Author contributions: M.N.B.-C., C.E.M., and S.A.W. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. Email: stuart.west@zoo.ox.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1509740113/-/DCSupplemental.

hence test whether misunderstanding is correlated with greater cooperation. Although behavior with computers has been examined previously, it is not known how such behavior correlates with understanding and play with humans (33, 34).

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/fully 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 (χ^2 test comparing our distribution to an amalgam of the distributions reported in refs. 1 and 2: $\chi^2_{(3)} = 5.2$, $P = 0.156$; 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).

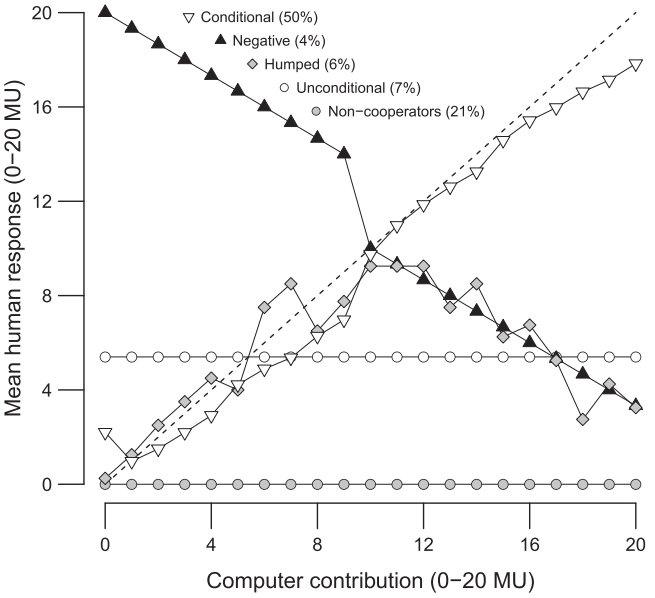

Fig. 1. Cooperating with computers. The average public-good contribution when playing with computers, grouped by behavioral type, for each possible mean contribution of their three computerized groupmates in the strategy method (n = 72). Dashed line equals perfect matching of contributions. We followed previous studies by dividing individuals into types on the basis of their contribution pattern (1–6) (*SI Results*). The distribution of types is not significantly different to previous experiments with humans (Table 1 and Table S1).

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 not 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

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

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

*Fisher's exact test per type, χ^2 test for totals.
†The thee negative cooperators are classified as humped and the five unconditional cooperators as unclassifiable (*SI Results*).

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 ~ 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 ~ type: $F_{3,68} = 6.9$, $P < 0.001$, R^2_{adj} from a linear model = 0.12; Fig. 2*A* 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. 2*A* 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: $F_{1,70} = 17.0$, $P < 0.001$, $\beta = 0.173 \pm 0.046$; 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.

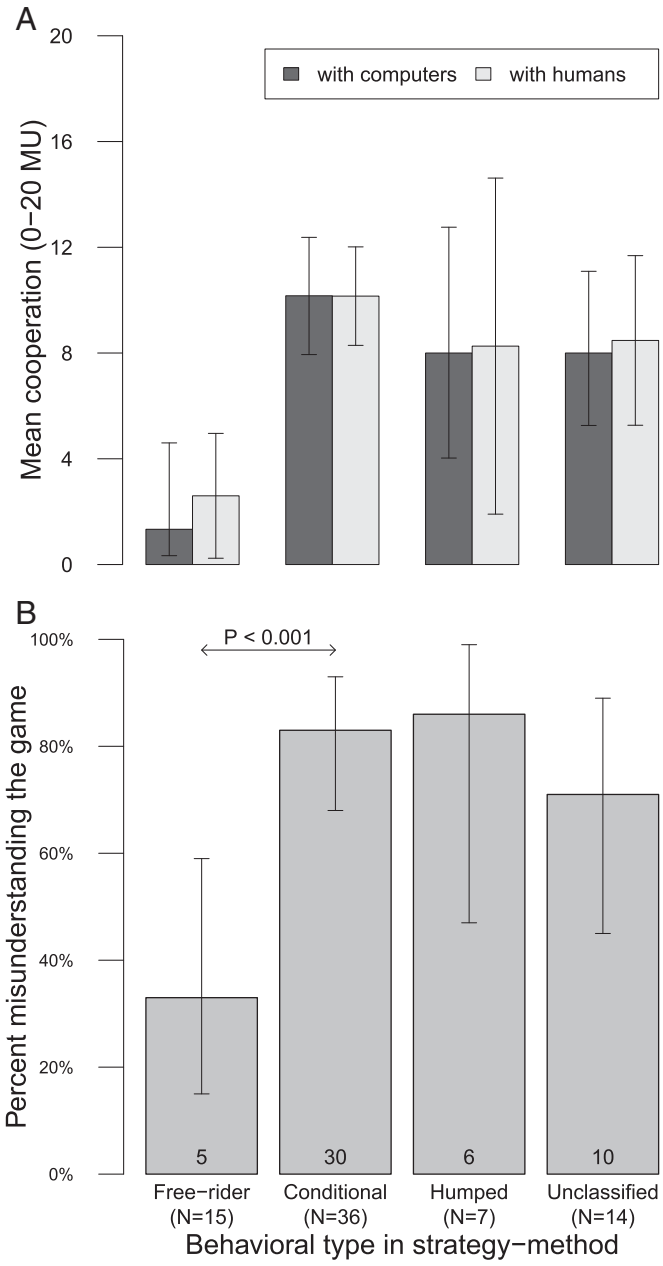

Fig. 2. Play with computers predicts play with humans and conditional cooperators misunderstand the game. (*A*) The mean contribution (±95% CIs) to the public-good grouped by behavioral type (Fig. 1). For all types, the mean levels of cooperation were not significantly different when playing with computers (dark gray) vs. when playing with humans (light gray). (*B*) The percentage (±95% modified Wald method CIs) 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 contribute. 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).

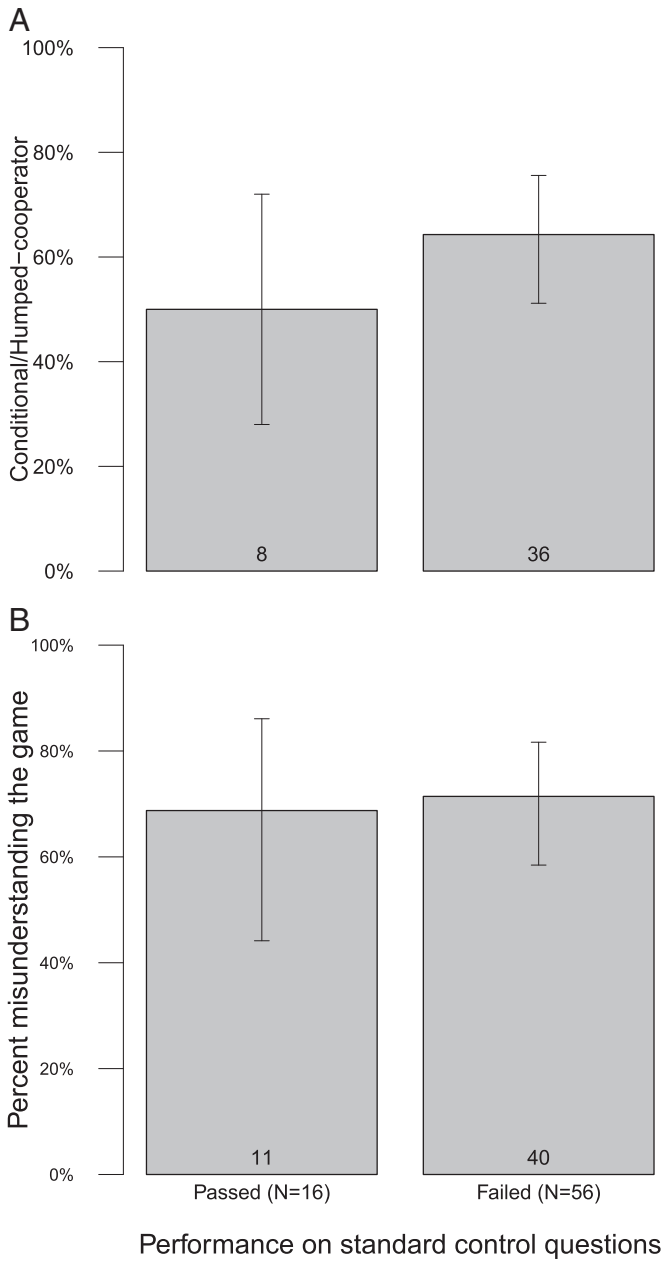

Fig. 3. Standard control questions fail to control for understanding. We divided individuals into those that correctly answered all 10 standard control questions (n = 16) and those that did not (n = 56). The control questions involved calculating the payoffs in four hypothetical scenarios. Those that passed were (*A*) just as likely to play as a conditional or humped cooperators when playing with computers and (*B*) no more likely to report that the income-maximizing decision did not depend on the contributions of others.

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. 2*B*, 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. 2*B*). 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. 2*B*]. Refs. 1 and 2 suggested that their 6–10% of unclassified players may have been confused players, yet Fig. 2*B* 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 physiologically between humans and computers when playing simpler games (19, 56–58); (*ii*) not explain why behavior significantly correlated with understanding (Fig. 2*B* 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 everyday expectations 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 humped cooperators (8 of 16, 50%) than those that failed the standard control questions (36 of 56, 64%, FET: $P = 0.744$; Fig. 3*A*). 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. 3*B*).

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 (1–6).

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 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 humped 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 computerized 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 instructions and control questions (*SI Appendix*), which were copied 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.

1. Fischbacher U, Gachter S, Fehr E (2001) Are people conditionally cooperative? Evidence from a public goods experiment. *Econ Lett* 71(3):397–404.
2. Fischbacher U, Gachter S (2010) Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *Am Econ Rev* 100(1):541–556.
3. Kocher MG, Cherry T, Kroll S, Netzer RJ, Sutter M (2008) Conditional cooperation on three continents. *Econ Lett* 101(3):175–178.
4. Herrmann B, Thoni C (2009) Measuring conditional cooperation: A replication study in Russia. *Exp Econ* 12(1):87–92.
5. Martinsson P, Villegas-Palacio C, Woolbrant C (2009) Conditional cooperation and social group: Experimental results from Colombia. *Environment for Development*, Discussion Paper Series, EfD DP 09-16. Available at www.rff.org/files/sharepoint/WorkImages/Download/EfD-DP-09-16.pdf. Accessed January 6, 2016.
6. Martinsson P, Nam PK, Villegas-Palacio C (2013) Conditional cooperation and disclosure in developing countries. *J Econ Psychol* 34:148–155.
7. Burlando RM, Guala F (2005) Heterogeneous agents in public goods experiments. *Exp Econ* 8(1):35–54.
8. Camerer CF, Fehr E (2006) When does "economic man" dominate social behavior? *Science* 311(5757):47–52.
9. Ones U, Putterman L (2007) The ecology of collective action: A public goods and sanctions experiment with controlled group formation. *J Econ Behav Organ* 62(4):495–521.
10. Rustagi D, Engel S, Kosfeld M (2010) Conditional cooperation and costly monitoring explain success in forest commons management. *Science* 330(6006):961–965.
11. Kosfeld M, von Siemens FA (2011) Competition, cooperation, and corporate culture. *Rand J Econ* 42(1):23–43.
12. Volk S, Thoni C, Ruigrok W (2012) Temporal stability and psychological foundations of cooperation preferences. *J Econ Behav Organ* 81(2):664–676.
13. Camerer CF (2013) Experimental, cultural, and neural evidence of deliberate prosociality. *Trends Cogn Sci* 17(3):106–108.
14. Cheung SL (2014) New insights into conditional cooperation and punishment from a strategy method experiment. *Exp Econ* 17(1):129–153.
15. Nielsen UH, Tyran JR, Wengstrom E (2014) Second thoughts on free riding. *Econ Lett* 122(2):136–139.
16. Hartig B, Irlenbusch B, Kolle F (2015) Conditioning on what? Heterogeneous contributions and conditional cooperation. *J Behav Exp Econ* 55:48–64.
17. Mertins V, Schote AB, Hoffeld W, Griessmair M, Meyer J (2011) Genetic susceptibility for individual cooperation preferences: The role of monoamine oxidase A gene (MAOA) in the voluntary provision of public goods. *PLoS One* 6(6):e20959.
18. Mertins V, Schote AB, Meyer J (2013) Variants of the monoamine oxidase A gene (MAOA) predict free-riding behavior in women in a strategic public goods experiment. *J Neuroscience Psychology Econ* 6(2):97–114.
19. Suzuki S, Niki K, Fujisaki S, Akiyama E (2011) Neural basis of conditional cooperation. *Soc Cogn Affect Neurosci* 6(3):338–347.
20. Dawes CT, et al. (2012) Neural basis of egalitarian behavior. *Proc Natl Acad Sci USA* 109(17):6479–6483.
21. Gintis H (2000) Beyond Homo economicus: Evidence from experimental economics. *Ecol Econ* 35(3):311–322.
22. Bowles S, Gintis H (2002) Homo reciprocans. *Nature* 415(6868):125–128.
23. Bowles S, Hwang SH (2008) Social preferences and public economics: Mechanism design when social preferences depend on incentives. *J Public Econ* 92(8-9):1811–1820.
24. Gsottbauer E, van den Bergh JCJM (2011) Environmental policy theory given bounded rationality and other-regarding preferences. *Environ Resour Econ* 49(2):263–304.
25. Gachter S (2007) Conditional cooperation: behavioural regularities from the lab and the field and their policy implications. *Psychology and Economics: A Promising New Cross-Disciplinary Field*, eds Frey BS, Stuzter A (MIT Press, Cambridge, MA), pp 19–50.
26. Martinsson P, Villegas-Palacio C, Wollbrant C (2015) Cooperation and social classes: Evidence from Columbia. *Soc Choice Welfare* 45(4):829–848.
27. Bardsley N (2008) Dictator game giving: Altruism or artefact? *Exp Econ* 11(2):122–133.
28. Zizzo DJ (2010) Experimenter demand effects in economic experiments. *Exp Econ* 13(1):75–98.
29. Henrich J, et al. (2005) "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behav Brain Sci* 28(6):795–815.
30. Rand DG, et al. (2014) Social heuristics shape intuitive cooperation. *Nat Commun* 5:3677.
31. Heintz C (2013) What can't be inferred from cross-cultural experimental games. *Curr Anthropol* 54(2):165–167.
32. Andreoni J (1995) Cooperation in public-goods experiments: Kindness or confusion. *Am Econ Rev* 85(4):891–904.
33. Houser D, Kurzban R (2002) Revisiting kindness and confusion in public goods experiments. *Am Econ Rev* 92(4):1062–1069.
34. Ferraro PJ, Vossler CA (2010) The source and significance of confusion in public goods experiments. *The B.E. Journal of Economic Analysis & Policy*, 10.2202/1935-1682.2006.
35. Fehr E, Gächter S (2002) Altruistic punishment in humans. *Nature* 415(6868):137–140.
36. Fischbacher U, Follmi-Heusi F (2013) Lies in disguise: An experimental study on cheating. *J Eur Econ Assoc* 11(3):525–547.
37. Falk A, Heckman JJ (2009) Lab experiments are a major source of knowledge in the social sciences. *Science* 326(5952):535–538.
38. Trivers R (2004) Mutual benefits at all levels of life. *Science* 304(5673):964–965.
39. Smith VL (2005) Sociality and self interest. *Behav Brain Sci* 28(6):833.
40. Heintz C (2005) The ecological rationality of strategic cognition. *Behav Brain Sci* 28(6):825.
41. Burnham TC, Johnson DP (2005) The biological and evolutionary logic of human cooperation. *Anal Kritik* 27(1):113–135.
42. Hagen EH, Hammerstein P (2006) Game theory and human evolution: A critique of some recent interpretations of experimental games. *Theor Popul Biol* 69(3):339–348.
43. Burnham TC, Hare B (2007) Engineering human cooperation: Does involuntary neural activation increase public goods contributions? *Hum Nat* 18(2):88–108.
44. Delton AW, Krasnow MM, Cosmides L, Tooby J (2011) Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *Proc Natl Acad Sci USA* 108(32):13335–13340.
45. Pedersen EJ, Kurzban R, McCullough ME (2013) Do humans really punish altruistically? A closer look. *Proc R Soc B* 280:20122723.
46. Raihani NJ, Bshary R (2015) Why humans might help strangers. *Front Behav Neurosci* 9:39.
47. Burton-Chellew MN, West SA (2013) Prosocial preferences do not explain human cooperation in public-goods games. *Proc Natl Acad Sci USA* 110(1):216–221.
48. Fischbacher U, Gachter S, Quercia S (2012) The behavioral validity of the strategy method in public good experiments. *J Econ Psychol* 33(4):897–913.
49. Herrmann B, Thöni C, Gächter S (2008) Antisocial punishment across societies. *Science* 319(5868):1362–1367.
50. Fischbacher U, Schudy S, Teyssier S (2014) Heterogeneous reactions to heterogeneity in returns from public goods. *Soc Choice Welfare* 43(1):195–217.
51. Gintis H, Smith EA, Bowles S (2001) Costly signaling and cooperation. *J Theor Biol* 213(1):103–119.
52. Trivers RL (1971) Evolution of reciprocal altruism. *Q Rev Biol* 46(1):35.
53. Burton-Chellew MN, Nax HH, West SA (2015) Payoff-based learning explains the decline in cooperation in public goods games. *Proc Biol Sci* 282(1801):20142678.
54. Smith A (2013) Estimating the causal effect of beliefs on contributions in repeated public good games. *Exp Econ* 16(3):414–425.
55. Gachter S, Renner E (2010) The effects of (incentivized) belief elicitation in public goods experiments. *Exp Econ* 13(3):364–377.
56. van 't Wout M, Kahn RS, Sanfey AG, Aleman A (2006) Affective state and decision-making in the Ultimatum Game. *Exp Brain Res* 169(4):564–568.
57. Rilling JK, Sanfey AG, Aronson JA, Nystrom LE, Cohen JD (2004) The neural correlates of theory of mind within interpersonal interactions. *Neuroimage* 22(4):1694–1703.
58. Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD (2003) The neural basis of economic decision-making in the Ultimatum Game. *Science* 300(5626):1755–1758.
59. Fehr E, Schmidt KM (1999) A theory of fairness, competition, and cooperation. *Q J Econ* 114(3):817–868.
60. Smith EA (2005) Making it real: Interpreting economic experiments. *Behav Brain Sci* 28(6):832.
61. Gerkey D (2013) Cooperation in context public goods games and post-Soviet collectives in Kamchatka, Russia. *Curr Anthropol* 54(2):144–176.
62. Croson R, Marks M (2000) Step returns in threshold public goods: A meta- and experimental analysis. *Exp Econ* 2(3):239–259.
63. Tinghög G, et al. (2013) Intuition and cooperation reconsidered. *Nature* 498(7452):E1–E2, discussion E2–E3.
64. Gurven M, Winking J (2008) Collective action in action: Prosocial behavior in and out of the laboratory. *Am Anthropol* 110(2):179–190.
65. Fischbacher U (2007) z-Tree: Zurich toolbox for ready-made economic experiments. *Exp Econ* 10(2):171–178.
66. Greiner B (2015) Subject pool recruitment procedures: Organizing experiments with ORSEE. *J Econ Sci Assoc* 1(1):114–125.