PNAS

Supporting Information for

Confusion cannot explain cooperative behavior in public goods games

Guangrong Wang[a,b], Jianbiao Li[b,c], Wenhua Wang[b,c], Xiaofei Niu[b], Yue Wang[b]

[a] *Neural Decision Science Laboratory, School of Economics & Management, Weifang University, Weifang 261061, China*

[b] *School of Economics/Institute for Study of Brain-like Economics, Shandong University, Jinan 250100, China*

[c] *China Academy of Corporate Governance/Reinhard Selten Laboratory, Nankai University, Tianjin 300071, China*

Corresponding authors:
Jianbiao Li, Ph.D, School of Economics, Institute for Study of Brain-like Economics, Shandong University, 27 Shanda Nanlu, Jinan, China. Email: biaojl@sdu.edu.cn, biaojl@126.com.
Wenhua Wang, Ph.D, School of Economics, Institute for Study of Brain-like Economics, Shandong University, 27 Shanda Nanlu, Jinan, China. Email: whua_wang@126.com.

This PDF file includes:

Supporting Methods, Results and Discussion
Figures S1 to S7
Tables S1 to S12
SI References

Supporting Methods

Methods of Experiment 1

A total of 240 participants (mean age: 21.19 years; 123 females and 117 males) who had never participated in a public goods experiment were recruited to take part in one session each. There were 12 participants in each session. All 20 sessions were conducted at the Institute for Study of Brain-like Economics, Shandong University, China. The experiment was programmed and conducted using z-Tree. The monetary reward had a mean of 46 Chinese Yuan (around $ 6.8) and ranged from 32 to 58 Yuan.

The underlying decision situation in our experiments was a standard linear public goods game, with four members in each group. Each group member received an initial endowment of 20 monetary units (MU) and had to decide either to keep 20 MU or to contribute a fraction (0 – 20 MU) to a group project. The payoff function is given by

$$\pi_i = 20 - g_i + 0.5\sum_{j=1}^{4} g_j \quad (1)$$

where g_i is the amount contributed to the group project by participant i and g_j is the amount contributed to the group project by the four group members. The amount contributed to the project was doubled and was shared equally among the four group members. While the contribution of each MU to the group project yielded a private marginal return of 0.5 MU and a social marginal benefit of 2 MU, the amount of each MU kept for oneself was worth 1 MU to the participant. Here, we use 0.5 as the marginal per capita return (MPCR) to simplify the computational difficulty for participants, a parameter that has been widely used in previous literature.

After the participants were informed about the rules of the experimental task through an explanation of the written instructions, they had to answer 10 standard control questions. The experimenter would then check their answers and provide a verbal explanation if someone repeatedly failed to answer correctly. Before starting the experiment, we asked the following question: "In a one-shot game, given that the amount contributed to the project by the other three group members in your group is 30 MU, if you want to maximize your own benefit, how much should you contribute to the project (of course, your actual contribution may be different)?" If a participant answers "zero," this means they understood the game; if otherwise, this means they did not understand it.

Participants had to make two set of decisions, a "conditional contribution" schedule and an "unconditional" decision. For conditional contributions, participants had to decide how much they would contribute to the group project, given the average contribution of the other group members. Specifically, the contribution schedule of the five possible average contributions of the other three group members (0, 5, 10, 15, and 20) was shown, and participants had to make their corresponding contribution for each of the five values. For unconditional contributions, participants simultaneously and privately contributed money to a group project (**Fig. S1**). Experiment 1 was one-shot, and participants were aware of this. Thus, the participants' preferences were elicited without mixing preferences with strategic considerations.

Within this basic setup, we conducted two treatments and applied a between-subjects design. In the human treatment, all participants were randomly divided into groups of four; in the computer treatment, each participant was matched with three computer players in a group, with only the human participant receiving real monetary rewards. To prevent learning, we did not provide participants with any information about their earnings in the game.

There were some differences between Experiment 1 and Burton-Chellew, El Mouden, and West (hereinafter “BEW”)’s experiment in the instructions, parameter settings and procedures, as follows. First, the instructions for the computer treatment in Experiment 1 were explicitly explained with a computer framing, whereas those for BEW’s experiment were explained with a human framing at first, participants were told that they would be playing with computers until just before the actual decision was made. Second, to simplify the computational difficulty for participants, the MPCR is 0.5 instead of 0.4 of BEW’s Experiment. Third, to avoid excessive burden or fatigue for the participants, they were asked to report their contribution conditional on others’ contributing on average 0, 5, 10, 15, or 20 MU, respectively, but not for every possible average contribution from others within the range 0 to 20 (i.e., for each contribution ∈ [0, 1, 2, ..., 20]). Fourth, we tested participants’ understanding of the nature of the game immediately after ensuring that they correctly answered the 10 standard control questions, rather than at the end of the experiment as BEW did.

Methods of Additional Experiment 1

For the additional experiment 1, which we explored the beliefs about appropriate or inappropriate behavior when playing with computers, 72 participants were recruited (41 females, 31 males, mean age: 22.32 years, SD = 2.55) in 6 sessions. All sessions were conducted at the Institute for Study of Brain-like Economics, Shandong University, China. The monetary reward had a mean of 43 Chinese Yuan (around $ 6.0) and ranged from 33 to 60 Yuan.

The basic decision situation was a standard linear public goods game, with four members in each group and MPCR = 0.5, as in Experiment 1. Similarly, after being informed of the rules of the experiment through an explanation of the written instructions, participants answered 10 standard control questions. The experimenter would then check their answers and provide a verbal explanation if someone repeatedly failed to answer correctly. All participants were aware that each of them would be matched with three computer players in a group, with only the human participant receiving real monetary rewards.

Participants first completed a “conditional contribution” schedule, in which they decided how much they would contribute to the group project, given the average contribution of the other computer members. Specifically, participants were shown the contribution schedule of the five possible average contributions of the other three computer members (0, 5, 10, 15, and 20) and had to make their corresponding contribution for each of the five values. Next, participants made an “unconditional” decision. Then, participants were asked to evaluate the social appropriateness of actions on a six-point scale ranging from 1: “Very socially inappropriate” to 6: “Very socially appropriate”. In this part, they first evaluated how socially appropriate they think it is to contribute the amount ∈ [0, 5, 10, 15, 20] MU conditional on other computer players’ contributing on average 0, 5, 10, 15, or 20 MU, respectively. They then evaluated how socially appropriate they think it is to contribute the amount ∈ [0, 5, 10, 15, 20] without knowing how much others contribute.

The evaluation of actions was incentivized. Participants were told that, at the end of the experiment, one of the possible scenarios would be selected at random, and that their response in this situation would be compared to those of all other participants. If a participant’s appropriateness rating was the same as the modal response, then that participant would earn 10 MU, otherwise they would earn nothing.

Methods of Additional Experiment 2

We collected data from a new sample of 72 participants (39 females, 33 males, mean age: 21.04 years, SD = 2.11) in 6 sessions. All sessions were conducted at the Institute for Study of Brain-like Economics, Shandong University, China. The monetary reward had a mean of 38 Chinese Yuan (around $ 5.3) and ranged from 19 to 46 Yuan.

The basic decision situation and the experimental procedure are identical to those in Additional Experiment 1. Upon reading the instructions and answering standard control questions, participants initially made conditional contribution decisions followed by an unconditional decision. Then each participant was asked three types of questions: first, what they personally thought that one should do in the public goods game, which they answered by indicating how much one ought to contribute conditional on other computer members' contributing on average 0, 5, 10, 15, or 20 MU, respectively. This question measures participants' personal beliefs about what is most appropriate in the game. Next, participants were presented questions about either their normative or empirical expectations. In the former case, participants were asked to guess the most common response to the personal belief question by other participants in previous experiments. For empirical expectations, participants were asked to guess the most frequent choice that subjects actually made in previous experiments conducted with a separate group of subjects. Normative and empirical expectations were incentivized: for both, a correct response yielded a 10 MU bonus, respectively.

Methods of Experiment 2

We collected data for 72 participants (38 females, 34 males, mean age: 21.10 years, SD = 1.79) over 6 sessions. All sessions were conducted in the laboratory of Institute for Study of Brain-like Economics at Shandong University in China. The monetary reward had a mean of 52 Chinese Yuan (around $7.2) and ranged from 43 to 64 Yuan.

Upon arrival, participants were seated at their computer terminal and given a random ID to ensure their anonymity. They received a set of printed general instructions. Unlike BEW who asked participants to read the experimental instructions themselves and answer the control questions without being told the correct answers, we read the experimental instructions aloud to the participants and explained the experiment. Participants were actively encouraged to ask questions by raising their hands, and their questions were answered privately. After participants completed 10 standard control questions, they received information for each question about the correct answer and how the correct answer was calculated.

Our experimental procedures were then rigorously aligned with those of BEW, after making sure that all participants answered the questions correctly (**Fig. S2**). Participants first played in a special case with computer players using the strategy method, then played a typical public goods game, again with computers. After that, participants played the typical public goods game, but with humans, for six rounds. The details of the experimental procedure are consistent with those of BEW, including that participants learned from the online instructions that they would play with the computer players only after reading the initial instructions and completing the control questions, that participants were told about the strategy method only after being told to play with the computer, and so on. After participants completed the entire experimental procedure replicating BEW, i.e., after they

answered the questions testing their understanding of the game in Stage 6, we added another stage. This ensures that the participants' previous decisions were not affected. In Stage 7, participants were asked to answer the following questions: "In a one-shot game, given that the amount contributed to the project by the other three group members in your group is 30/10/60 MU, if you want to maximize your own benefit, how much should you contribute to the project (of course, your actual contribution may be different)?"
The experimental instructions and detailed screenshots of experimental procedures can be found in https://osf.io/v7y8b/.

Supporting Results

Results of Experiment 1

Distribution of Behavior Types

Participants are separated into four categories based on participants' pattern of contributions in the strategy method, including conditional cooperators, humped cooperators, free riders and others (**Fig. S3**).

Conditional Cooperation Level

Overall, the contributions in the human treatment were significantly greater than those in the computer treatment (**Table S1**).

Underlying Motives for Conditional Cooperation

Participants first answered the question, "Did your decision depend on the contributions of other members of your group?". The results showed that 56% and 43% of the participants in the human and computer treatments, respectively, answered "yes." When they answered "yes," they further chose the most important motive from the following statements in computer treatment. "I think most people would do the same thing if they face the same situation as me." "If I contribute too little, I will feel guilty or ashamed." "If I contribute too little, people might think I'm too selfish." "Other." The first three options correspond to social norms, self-image concern and social image concern, respectively. In human treatment, the added option is "I consider the interests of three group members", which corresponds to altruism. If participants answered "no", three options were provided and one had to be selected in both the computer and human treatments. The first option is "I only care about my own interests." The second option is "I think most people would do the same thing if they face the same situation as me." The third option is "Other." These options correspond to self-interest, social norms, and other motivations. The six confused participants were not included in the analysis of motives.
In the human treatment, 91% (48) of the 53 free riders answered "no," that is, the contributions of other group members did not influence their contributions. Their motives were mainly self-interest and social norms, each accounting for 50% and 42%. Only 9% (5) of the 53 free riders answered "yes" and their motive was social norms. 96% (47) of the 49 conditional cooperators reported that other members' contributions influenced their decisions. Their most common motive was social norms (47%), followed by altruism and self-image concerns (19% and 21%, respectively). Only 4% (2) of the 49 conditional cooperators answered "No," but they were motivated by social norms rather than self-interest. Almost all humped cooperators indicated that their decisions were affected by

others' contributions and that the main underlying motive was social norms (75%, **Tables S2 and S3**).
In general, there were 47% (25 of 53) of free riders chose social norm as their main motive, 45% (24 of 53) chose self-interest and 8% chose other motives. For conditional cooperators, 49% (24 of 49) reported social norm as their most important motive, 20% (10 of 49) concerned their self-image, 18% (9 of 49) considered others' interests, 2% (1 of 49) concerned their social image and the remaining 10% (5 of 49) due to other motives. Among humped cooperators, 15% (2 out of 13) said their main motive was to take others' interests into account, 69% (9 of 13) out of social norm, 8% (1 of 13) out of self-interest and 8% (1 of 13) reported other motives. For two that showed other behavior patterns, one chose altruism and another chose social norm. Differences in altruism, self-image concerns, and self-interest are significant when comparing free riders and cooperators (FET: altruism: $p < 0.01$; social norm: $p = 0.58$; self-image: $p < 0.01$; social image: $p = 1.00$; self-interest: $p < 0.01$).
In the computer treatment, 84% (67) of the 80 free riders reported that their decisions were not influenced by others' contributions, and social norms and self-interest were the most important motives, accounting for approximately 70% and 25%, respectively. 16% (13) of the 80 free riders answered "yes," with social norms being the main motive, accounting for 77%. All of conditional and humped cooperators answered "yes," and the social norm motive accounted for approximately 67% and 60%, respectively (**Tables S4 and S5**).
In sum, in the computer treatment, there were 34% (27 of 80) of free riders reported social norm as their most important motive, 59% (47 of 80) contributed zero out of their self-interest. The remaining 7% (6 of 80) chose other motives. Among conditional cooperators, 67% (18 of 27) chose social norm as their main motive, 22% (6 of 27) and 4% were most concerned with self-image and social image, respectively. For humped cooperators, social norm was the most frequently selected motive, accounting for 60% (6 of 10), self-image and social image each account for 10%. Differences in social norms, self-image concerns, social image concerns, and self-interest are significant when comparing free riders and cooperators (FET: social norm: $p < 0.01$; self-image: $p < 0.01$; social image: $p = 0.10$; self-interest: $p < 0.01$).

Unconditional Cooperation of Different Behavior Types
After completing the conditional contribution schedule, participants were asked to make a one-shot unconditional contribution decision in both the computer and human treatments. When participants played the typical public goods game with computers, they contributed on average 20% of their endowment (4.0 MU, median = 0 MU). However, the cooperation level was significantly higher when playing with humans (mean = 29% of endowment, 5.8 MU, median = 5 MU, t test $t = -2.49$, $p < 0.01$).
Among free riders, there were around 71% (57 of 80, 39 of 55 in the computer and human treatment, respectively) of participants showed consistency, i.e., kept contribute zero in unconditional game. Their main motives were maximizing self-interest (accounting for 68% in the computer treatment, 41% in the human treatment) and adhering to social norm (26% in the computer treatment, 49% in the human treatment). For those who contribute non-zero in unconditional game, their main motives were social norm (43% in the computer treatment, 50% in the human treatment) and altruism (25% in the human treatment). For those cooperators defined by strategy methods (including conditional cooperators, humped

cooperators and others), there were 83% in the computer treatment (33 of 40) and 85% in the human treatment (55 of 65) remained cooperative in the unconditional game. The main motives for cooperative behavior toward both computer and human were conformity to social norms (52% and 55%, respectively), self-interest (18% and 11%, respectively), social image (9% and 2%, respectively), altruism (7% in the human treatment) and other undisclosed factors. For those who contribute zero in unconditional game, 57% in the computer treatment and 80% in the human treatment reported social norms as their most motive.

Results of Additional Experiment 1

There were 52 free riders, 16 conditional cooperators and 4 humped cooperators in additional experiment 1. This distribution of behavior types was similar with that of the computer treatment in Experiment 1 (FET: $p = 0.71$).

Following the approach of Krupka and Weber (2013), mean appropriateness ratings are calculated by transforming participants' responses into evenly-spaced numerical scores using the following scale: very socially inappropriate = −1; inappropriate = −0.6; somewhat socially inappropriate = −0.2; somewhat socially appropriate = 0.2; socially appropriate = 0.6; very socially appropriate = 1.

When we divided the participants into free riders and cooperators based on their contribution in the conditional contribution schedule, it was suggested that the perceived social norm of these two types of individuals was distinct. To be specific, free riders always rated contributing 0 as the most socially appropriate, regardless of the amount contributed by the three computers (**Fig. S4**). In contrast, cooperators rated a contribution of 10 as the most socially appropriate when the contribution of the computer players is unknown. If the computer players' average contributions were 0, 5, 10, 15, 20, respectively, they correspondingly rated 0, 5, 10, 15, 20 as most socially appropriate (**Fig. S5**). These results thus provide evidence for that while both free-riders and cooperators adhere to social norms, their perceived social norms are distinct. The social norm that free-riders adhere to involves no contribution, while the social norm that cooperators adhere to involves non-zero cooperation.

Results of Additional Experiment 2

There were 50 free riders, 16 conditional cooperators and 6 humped cooperators in additional experiment 2. This distribution of behavior types was similar with that of the computer treatment in Experiment 1 (FET: $p = 0.94$) and the additional experiment 1 (FET: $p = 0.88$).

When we divided the participants into free riders and cooperators based on their contribution in the conditional contribution schedule, it was suggested that the three beliefs of these two types of individuals was distinct. For free riders, when the computer players' contributions were 0, 5, 10, 15, and 20 MU, respectively, their personal normative beliefs were 0, 0.76, 1.38, 2.4 and 3.2 MU; their empirical expectations were 0.02, 0.14, 0.72, 0.52 and 1.04 MU; their normative expectations were 0.02, 0.78, 1.92, 2.54 and 3.34 MU. For cooperators, when the computer players' contributions were 0, 5, 10, 15, and 20 MU, their personal normative beliefs were 0, 2.36, 5.23, 8.41 and 10.82 MU; their empirical expectations were 0, 1.68, 6.14, 8.86 and 12.05 MU; their normative expectations were 0.05, 2.23, 5.05, 8.45 and 10.95 MU (**Table S6**). These three beliefs exhibited significant differences between free-riders and cooperators (*t*-test: all *p*s < 0.01 for an average of 5, 10, 15 and 20 MU contribution by computers; *p*s > 0.1 for 0 MU contribution by computers).

Results of Experiment 2

Distribution of Behavior Types

11 out of our 72 participants (15%) were conditional cooperators. 10 participants (14%) were humped cooperators. The remaining 51 participants (71%) were free riders (**Table S7, Fig. S6**).

Play with Computers versus Humans

When participants played the typical public goods game again with computers, they contributed on average 11% of their endowment (2.2 MU, median = 0 MU). However, their cooperation levels were significantly higher when playing with humans (mean = 26% of endowment, 5.2 MU, median = 4.3 MU, paired t test $t(71) = 5.1$, $p < 0.01$). Although our replicated results showed that the behavior types predicted the level of cooperation in the subsequent unconditional games, the predictive power is much greater in the computer game than in the human game. This differs from the results of BEW who found similar predictive power in both cases. If we ran a generalized linear model (GLM) of contributions in the six human rounds separately, the three-way classification scheme significantly predicted contributions for five of the six human rounds (R1: $F_{2,69} = 3.6$, $p = 0.03$, R^2_{adj} from a linear model = 0.08; R2: $F_{2,69} = 2.2$, $p = 0.12$, R^2_{adj} from a linear model = 0.05; R3: $F_{2,69} = 4.0$, $p = 0.02$, R^2_{adj} from a linear model = 0.08; R4: $F_{2,69} = 7.6$, $p < 0.01$, R^2_{adj} from a linear model = 0.17; R5: $F_{2,69} = 4.1$, $p = 0.02$, R^2_{adj} from a linear model = 0.09; R6: $F_{2,69} = 7.2$, $p < 0.01$, R^2_{adj} from a linear model = 0.16). Furthermore, at the individual level, contributions made in the one-shot game with computers were significantly lower than the mean unconditional contributions with humans in each round (**Table S8**).

We also examined how individuals conditioned their contributions on their beliefs about the behavior of their groupmates. When playing with computers, the relationship between participants' contributions and the amount that they expected their groupmates to contribute is not significant [GLM, contribution ~ expectation: $F_{14,57} = 0.62$, $p = 0.83$]. When playing with humans, our result [generalized linear mixed model (GLMM) on six rounds of data, contribution ~ expectation: $F_{1,430} = 364.34$, $p < 0.01$] is consistent with BEW's, i.e., players' contributions were positively correlated with the amount that they expected their human groupmates to contribute. For each of the six rounds with humans, this correlation was true [GLM, R1: $F_{13,58} = 3.78$, $p < 0.01$, $R^2_{adj} = 0.40$; R2: $F_{15,56} = 6.08$, $p < 0.01$, $R^2_{adj} = 0.46$; R3: $F_{12,59} = 4.97$, $p < 0.01$, $R^2_{adj} = 0.46$; R4: $F_{16,55} = 5.46$, $p < 0.01$, $R^2_{adj} = 0.49$; R5: $F_{15,56} = 4.87$, $p < 0.01$, $R^2_{adj} = 0.47$; R6: $F_{13,58} = 5.02$, $p < 0.01$, $R^2_{adj} = 0.46$]. Pooling all data, the analysis results show that the correlation between participants' contributions and the amount that they expected their groupmates to contribute significantly depended on the nature of groupmates [GLMM: interaction between nature of groupmates and expectations, $F_{1,500} = 47.71$, $p < 0.01$]. For free riders and cooperators separately, the above conclusions are valid [GLMM: interaction between expectations and groupmates on contributions, free riders: $F_{1,353} = 59.05$, $p < 0.01$; cooperators: $F_{1,143} = 10.44$, $p < 0.01$].

In line with BEW, we 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. On average, individuals contributed 2.4 (mean of six rounds) less than they expected of their human groupmates and 6.8 MU less than they expected of the computer. These decisions were significantly different in magnitude (paired t tests, all *p*s < 0.01, see **Table S9**). Participants, regardless of their behavior types, did show significant differences between their expectations of the computer and their own contributions. In contrast, only free riders contributed significantly less than their expectations of their human groupmates.
We then examined whether participants' contributions in the unconditional rounds matched their behavior in the strategy method. In the unconditional computer game, 57% (41 of 72 participants) were perfectly consistent, contributing the same amount as in the strategy method, 8 participants contributed less (mean = −3.4 MU), and the remaining 23 participants contributed more (mean = +4.0 MU). In contrast, in the first round of play with humans, 26% (19 of 72 participants) showed consistency, while 2 participants contributed less (mean = −2.0 MU) and 51 participants contributed more in the direct response (mean = +6.5 MU). 32%, 35%, 51%, 43%, and 46% in the second, third, fourth, fifth, and sixth rounds of 72 participants, respectively, showed consistency. The difference in the proportion of participants who are consistent is significant between when their groupmates are humans and when their groupmates are computers in most rounds (FET, R1: $p < 0.01$; R2: $p < 0.01$; R3: $p = 0.01$; R4: $p = 0.62$; R5: $p = 0.13$; R6: $p = 0.24$). These results differ from those of BEW and suggest that there are significantly more contributing behaviors in response to humans than to computers. The deviations are significant both in absolute terms (mean absolute discrepancy; vs. computers: 1.7 MU and vs. humans: 4.1 MU, paired t test: $t_{(71)} = 4.34$, $p < 0.01$) and in net terms (mean net discrepancy; vs. computers: 0.9 MU, and vs. humans: 3.9 MU, paired t test: $t_{(71)} = 4.93$, $p < 0.01$). In summary, about 50% of the participants change their behavior when switching from the strategy method to the direct response game with computer or human. Their behavioral shift does indeed depend on whether their groupmates are computers or humans, and they show a prosocial shift, being more favorable toward their human groupmates (mean net discrepancy of 3 MU, significantly different from 0 MU; **Fig. S7**).

Stability in Contributions when Playing Six Rounds with Humans
BEW claimed that many players were uncertain about how to play the game, thus showing a lack of stability in contributions when playing six rounds with humans. Based on our experimental data, 19 participants (26%) remained perfectly constant in six rounds, among them 10 participants always contribute 0 MU, while 4 participants always contribute 20 MU. 52 participants (72%) changed their contributions within the range of 5 MU from the previous round. On average, the 72 players changed their contributions by 2.3 MU from the prior round (mean absolute changes for rounds 2, 3, 4, 5, and 6, respectively: 2.5, 2.9, 2.7, 1.7, and 1.7 MU). We believe that this fluctuation range is reasonable. Even though participants know their payoff-maximizing strategy, they tend to vary their contributions within a small range to show that they made a deliberated decision.

Comprehenders Can Be Cooperators
By examining the individuals that: (i) contributed 0 MU in both the strategy method and in

the one-shot game with the computer (N = 13 of 72, 18%); (ii) answered all of the standard control questions correctly (N = 16, 22%); and (iii) that passed their beliefs test (N = 21, 29%), BEW ruled out the possibility that a substantial number of people who do understand the game still choose to cooperate.

In the Experiment 2, first, the 38 individuals who contributed 0 MU in both the strategy method and the one-shot game with the computer contributed significantly more in rounds with humans. The mean contributions by round were as follows: 5.0; 4.1; 4.0; 3.5; 3.6; and 2.9 MU for the six rounds, respectively. Each of these were significantly different from 0 MU (paired t tests, R1: $t_{(37)} = 5.8, p < 0.01$; R2: $t_{(37)} = 3.7, p < 0.01$; R3: $t_{(37)} = 3.9, p < 0.01$; R4: $t_{(37)} = 3.7, p < 0.01$; R5: $t_{(37)} = 3.4, p < 0.01$; R6: $t_{(37)} = 3.1, p < 0.01$), as well as across all six rounds combined (paired t test: $t_{(37)} = 4.4, p < 0.01$). Second, the total 72 participants who answered all of the standard control questions correctly contributed more when playing with humans than with computers (see **Table S9**). Third, the 69 participants that passed all of our control questions correctly showed prosocial bias toward humans, giving more to humans in each round than they did to computers (2.1 MU). The mean contributions by round with humans were as follows: 5.7; 5.2; 5.6; 4.9; 4.9; and 4.5 MU for the six rounds, respectively. Each of these were significantly different from 2.1 MU (paired t tests, R1: $t_{(68)} = 5.9, p < 0.01$; R2: $t_{(68)} = 4.1, p < 0.01$; R3: $t_{(68)} = 4.9, p < 0.01$; R4: $t_{(68)} = 4.2, p < 0.01$; R5: $t_{(68)} = 3.9, p < 0.01$; R6: $t_{(68)} = 3.6, p < 0.01$), as well as across all six rounds combined (paired t test: $t_{(68)} = 5.1, p < 0.01$).

In summary, contrary to BEW's claim, comprehenders who correctly understand the incentive structure of the game may be cooperators, free riders, or other types, and they show prosocial tendencies toward human players.

Economic Games and Social Preferences

Based on their findings, BEW propose that "a common assumption in behavioral economics experiments, that choices reveal motivations, will not necessarily hold" (5, p. 1291). However, by providing increased training through the standard control questions, we find solid evidence for the social preference explanation of cooperation in the public goods game.

In fact, ensuring that participants fully understand the experiment is the basic premise of experimental research. Leaving room for confusion is not a problem, but allowing confusion to be completely uncontrolled in an experiment and using it as the main basis for interpreting the results is suspected of deliberately exaggerating its role.

Measuring Motivations

Following BEW, we asked, "What was your most important motivation in the games with real people? Please select the answer that best describes your motivations?" after testing whether their income-maximizing strategy is independent of the behavior of others. 53 (74%) of 72 participants in our Experiment 2 selected "Making myself the maximum money possible" or "Making myself more money than other people" (**Table S10**). The response distribution differs from that of BEW (FET: $p = 0.01$). However, it is important to note that this question asks for the "most important motivation" and only one option can be selected, so those participants who exhibit cooperative behavior due to social preferences may still firstly choose to maximize their own interests.

Results Comparison of Experiment 1, 2, and BEW's Experiment

Table S11 provides a comprehensive comparison of the results from Experiment 1, Experiment 2, and BEW's Experiment according to BEW's paper. It can be seen that both our Experiment 1 and 2 do not reproduce their results.

Supporting Discussion

Literature Reviews on Confusion and Cooperation in Public Goods Games

The question of why people cooperate in social dilemmas has been a key topic of interest not only in economics but in all social sciences for decades. Two main competing explanations for over-contribution have been proposed: the first is social preference (e.g., altruism, warm-glow, or other forms of kindness, 1); the second is noise due to decision error or confusion about the instructions or the incentives in the experiment (2). Scholars have attempted to disentangle contributions due to confusion from those motivated by social preferences. Here, we review this line of literature and report their experimental results (see **Table S12**).

The first branch of the literature identifies confusion amidst multiple motives by modifying the design or the parameters of the standard public goods game. Andreoni's experimental design ingeniously distinguishes between cooperation resulting from social motivations and that resulting from confusion (2). There were three conditions. The first was a standard public goods game to provide a baseline for comparison. In the second condition, participants also played a standard public goods game, but they were paid based on their experimental earnings rank in their group, creating a zero-sum structure. The amount of cooperation in this condition provided a measure of participants who were confused. The third condition was used to measure the difference in cooperation due to information on rank. Contributions due to social factors can be calculated by subtracting the contributions in the second condition from those in the third condition. Based on their results, Andreoni concluded that cooperation is half kindness and half confusion. Confusion decreases over time and kindness increases over time, so that total cooperation remains fairly stable over 10 rounds. Ferraro and Vossler (3) also conducted an experiment in which they varied returns and group size to detect contributions due to confusion. According to their hypothesis, individuals who are confused about the incentives of the game will believe the variation in the returns and group size is a signal about the payoff-maximizing allocation of tokens. They found that about half (53.6%) of the altruism detected in the all-human treatment is noise generated by confused subjects.

The second branch of the literature, represented by Houser and Kurzban (4), eliminates other-regarding motives for contributions by replacing other group members with pre-programmed computers. Participants were aware that they were playing against computers and could not benefit other participants, so off-equilibrium play should not be associated with social motives toward other players. By subtracting contributions in the computer condition from those in the regular public goods experiment, the contributions due to social motives can be calculated. Their finding that confusion accounts for a substantial portion of contributions is in line with the findings of Andreoni (2). Following the basic design of the virtual player method, there are many scholars who have further explored the effect of confusion in public goods experiments, such as Ferraro and Vossler (3), BEW (5) and Yamakawa et al. (6).

The third branch of the literature investigated confusion in public goods games by varying the information available to participants. Bayer, Renner, and Sausgruber (7) conducted an experiment in which they deliberately withheld information about the incentive structure of the game from the subjects. The design of different levels of information across the conditions allows them to control for the experimenter demand effect. They suggested that the common claim that decision errors due to confused participants bias estimates of cooperation upward is not necessarily correct.

It can be seen that the existing literature often disentangles the effects of confusion and social preferences on contributions by using data from repeated public goods games. However, in games where payoffs are dependent on the collective actions of all players, repetition introduces additional complexities. Participants not only learn about the mechanics of the game, but also use the feedback from each round to predict how others might behave in future rounds. This poses the challenge of disentangling strategic responses to information (how players adjust their strategies based on what they learn about others' behaviors) from learning behavior associated with overcoming initial confusion about the rules and structure of the game. Moreover, these two factors—strategic response and learning behavior—may not be entirely distinct and could interact. For instance, a player's strategic response might be influenced by their assessment of how confused other group members are. If they believe others are still confused, they might adopt a different strategy than if they perceive others to have a clear understanding of the game. This potential interaction makes it even more difficult to parse out the effects of strategic response and learning behavior. In addition, participants in repeated public goods experiments may experience frustration when they see others free riding (8). Such emotional responses and possible changes in attitudes can also confound the analysis of learning processes.

In fact, confusion might be especially relevant in one-shot interactions or at the start of repeated interactions. The observed trend of declining contribution levels over time, if attributed to decreased confusion, also aligns with this premise. Similarly, the use of the strategy method can also help avoid the issue of strategic response and changes in attitudes towards others' earnings related to good or bad experiences. From this perspective, we do acknowledge that the basic experimental design of BEW (5) is desirable. Using the strategy method elicitation, BEW reported that participants showed the same conditional contribution pattern regardless of whether their groupmates were humans or computers. Furthermore, they reported that the behavior types from the strategy method significantly predicted the level of cooperation in the subsequent one-shot unconditional games, both with computers and with humans (six rounds without any feedback), and that there was no significant difference in the mean unconditional contributions between games with computers or with humans. Based on these results, BEW claim that variation in levels of cooperation can be explained by variation in understanding of the experiment, and further conclude that "a common assumption in behavioral economics experiments, that choices reveal (prosocial) motivations, will not necessarily hold." (5, p. 1291) However, their uncontrolled implementation of the experiment actually exaggerates the role of confusion.

To critically examine the two competing explanations for over-contribution in the public goods game, we conducted an experiment applying the virtual player method in a between-subjects design. We also replicated the experiment performed by BEW. Overall, our results do not provide the same evidence as BEW. The significantly different cooperative behavior with computers versus humans supports the social preference explanation.

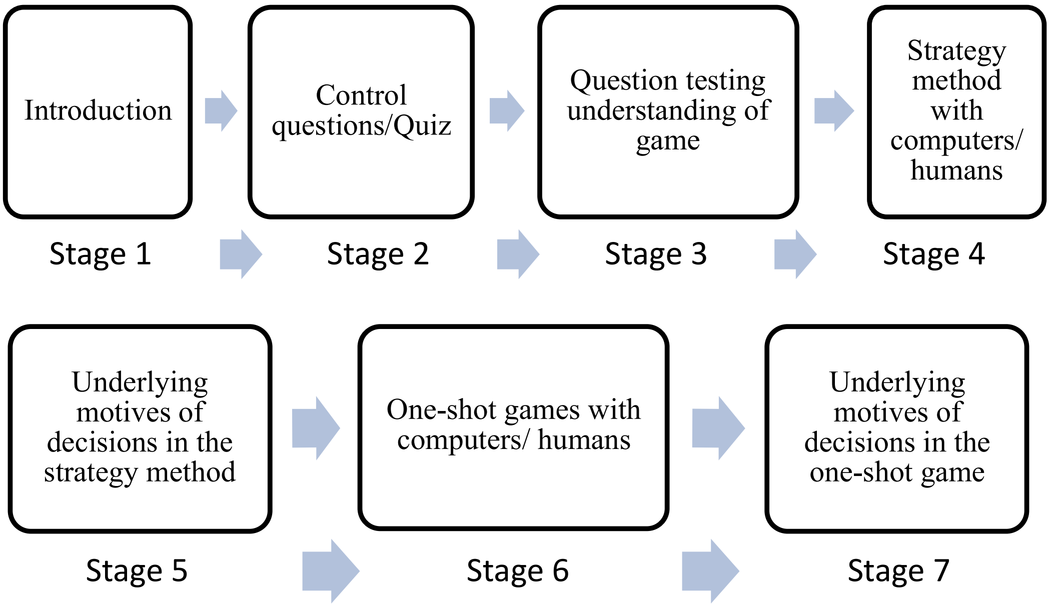

Fig. S1. The experimental order of Experiment 1.

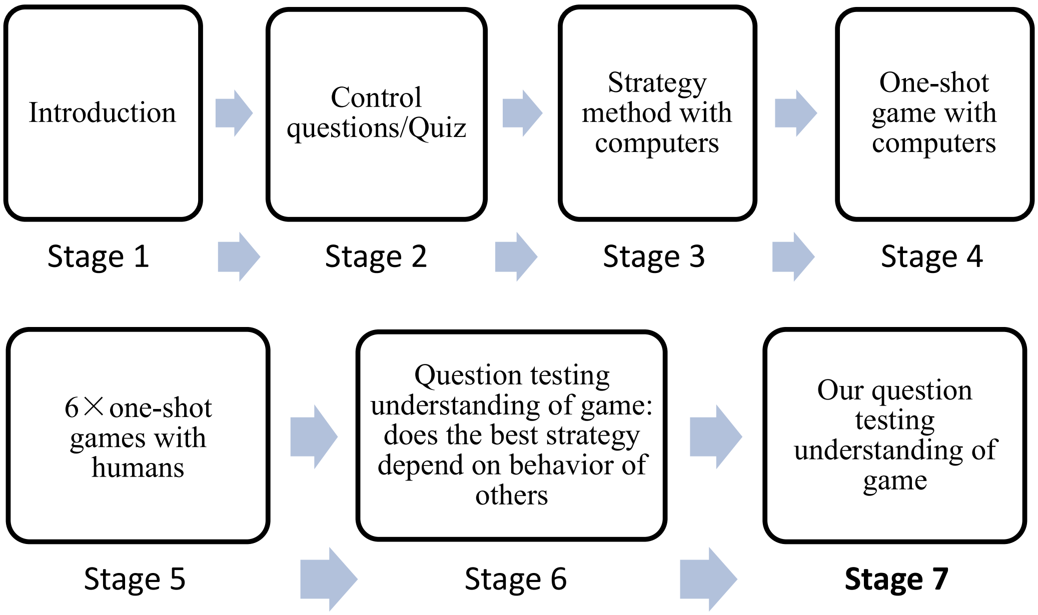

Fig. S2. The experimental order of Experiment 2.

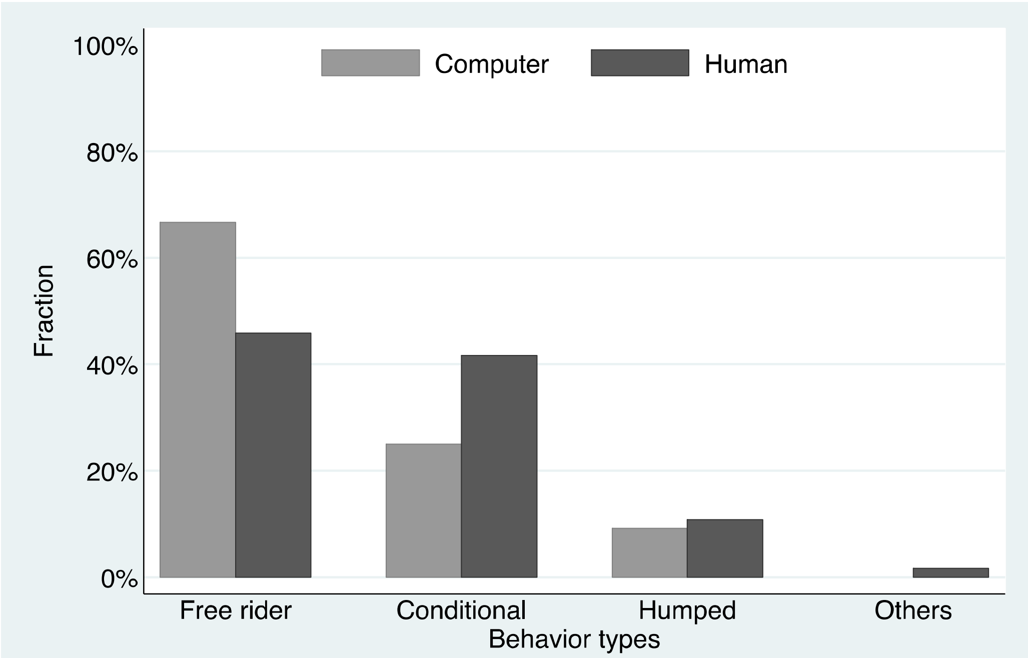

Fig. S3. Distribution of behavior types in the computer and human treatments of Experiment 1.

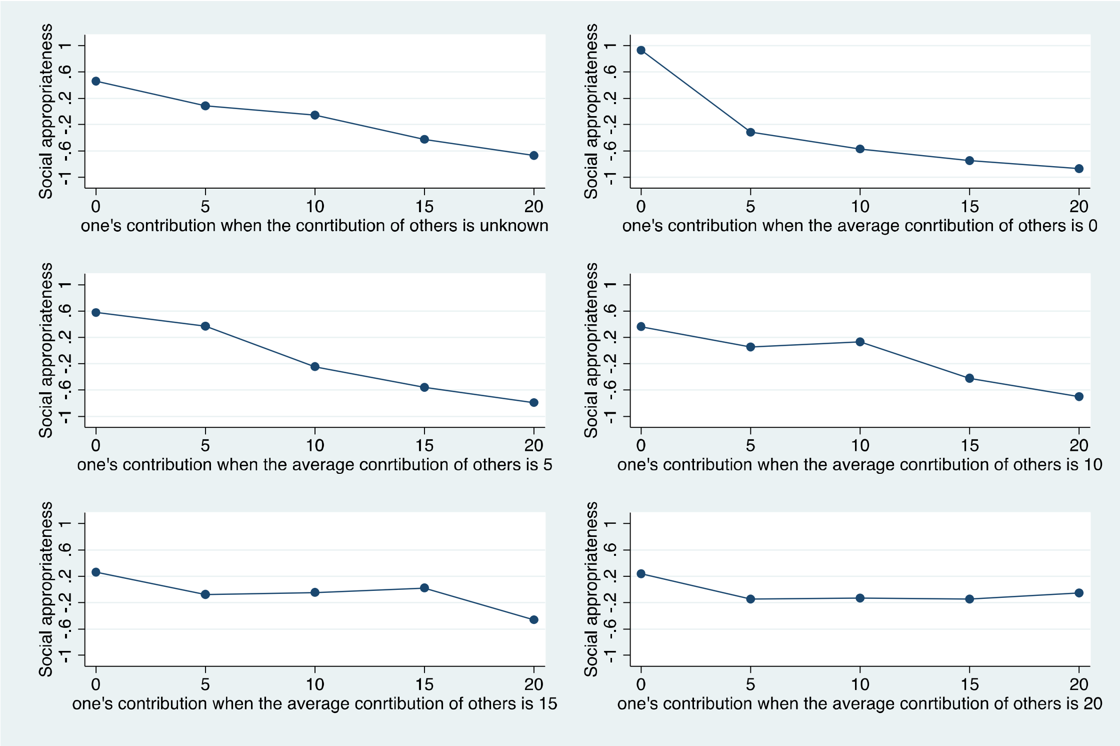

Fig. S4. Social appropriateness of one's contribution in unconditional and conditional games (free riders, N = 52).

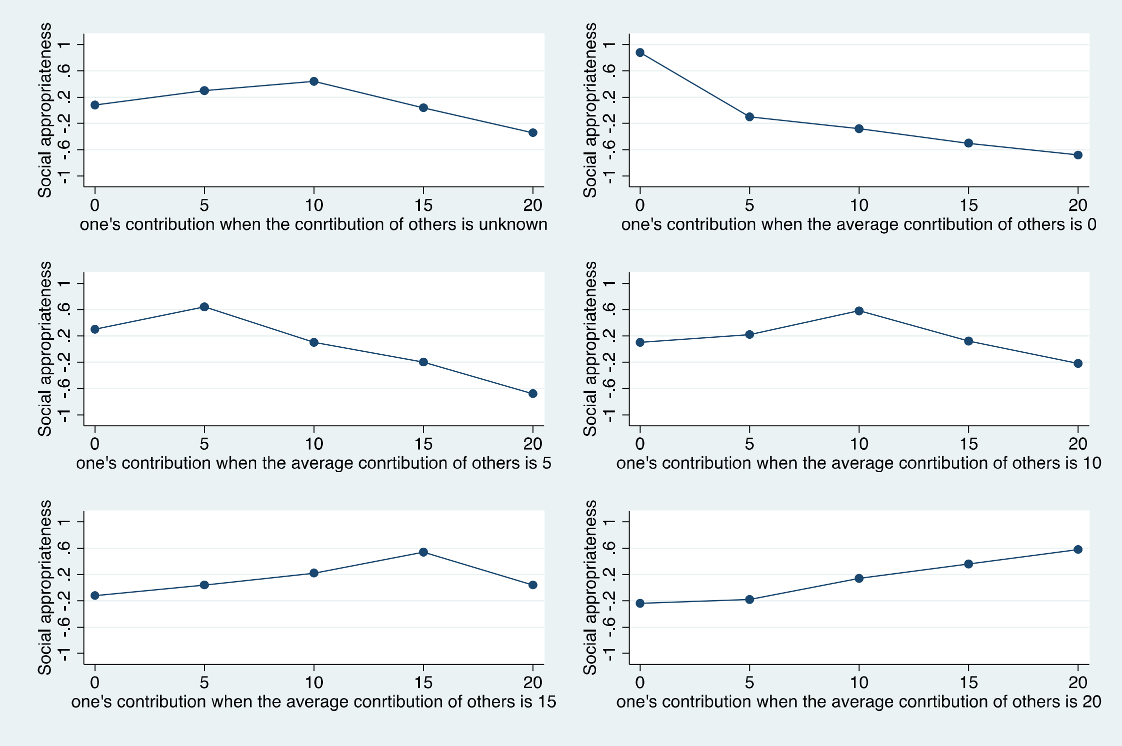

Fig. S5. Social appropriateness of one's contribution in unconditional and conditional games (cooperators, N = 20).

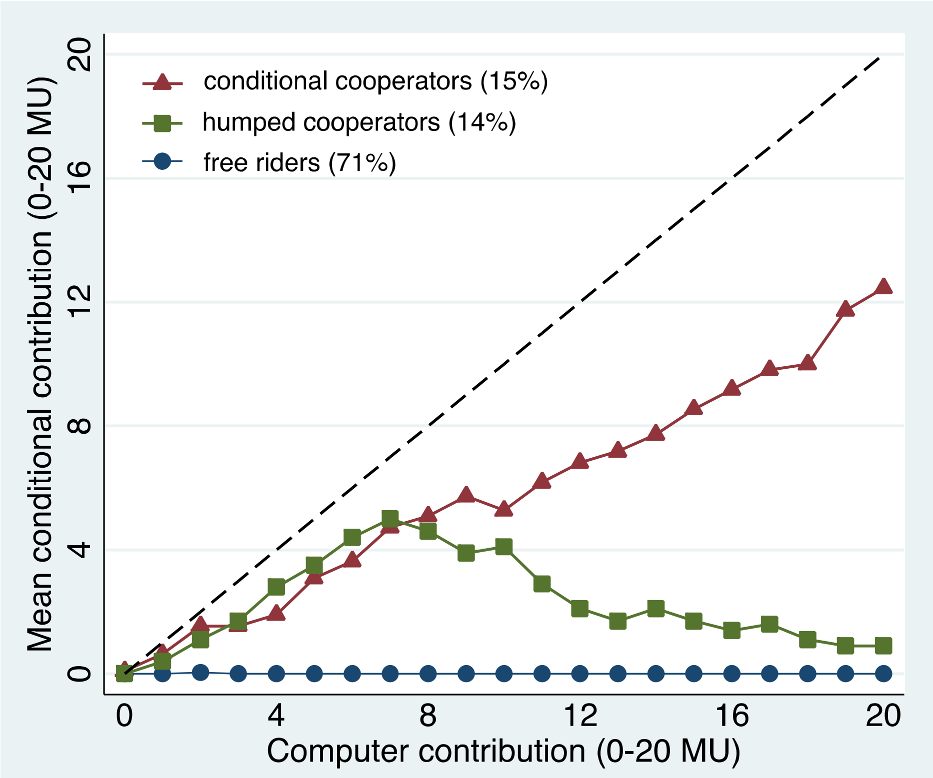

Fig. S6. Cooperating with computers. The average public good contribution when playing with computers, grouped by behavior type, for each possible mean contribution of their three computerized groupmates in the strategy method (N = 72). Dashed line equals perfect matching of contributions.

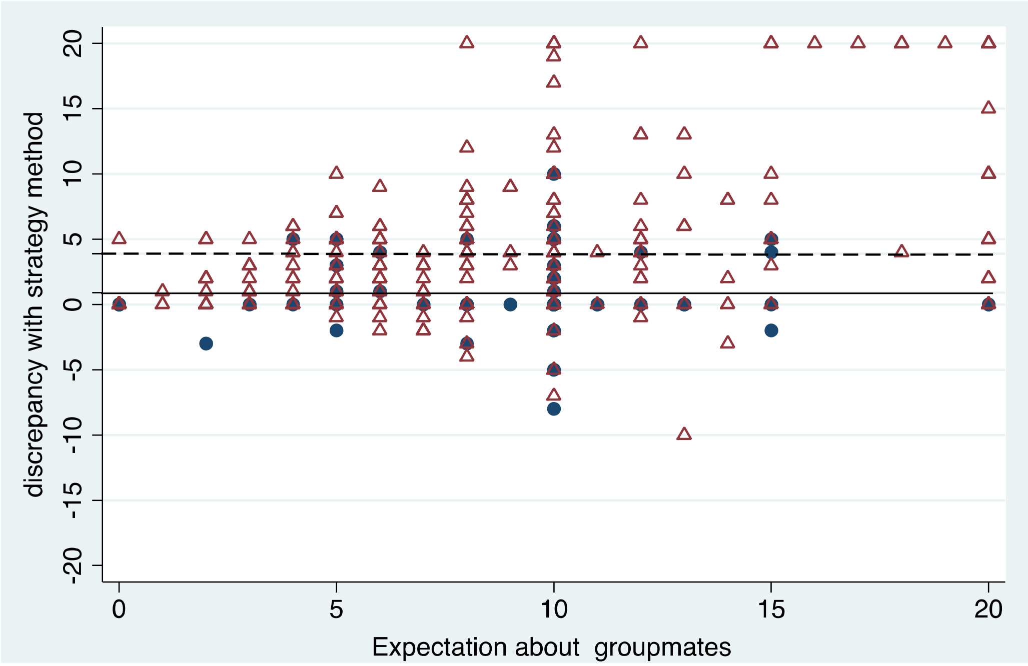

Fig. S7. Prosocial shift in response to playing with humans. For each player we used a record of what they had contributed to the group project (0–20 MU) in the first round of an unconditional game and of what they had expected their groupmates' mean contribution to be. This allowed us to compare their unconditional contribution as a response to their expectations, with what they had contributed in the prior strategy method 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 different, when playing with computers (black circles) at 0.9 MU (black line) or humans (hollow navy circles) at 3.9 MU (Navy dotted line).

Table S1 Average conditional contributions in computer and human treatments of Experiment 1

Conditions	Mean (SE)		t test	
	Computer	Human	*t*	*p*
0	0.0(0.00)	0.3(0.19)	−1.54	0.13
5	0.9(0.18)	2.2(0.36)	−3.16	<0.01
10	1.6(0.31)	3.7(0.48)	−3.59	<0.01
15	2.5(0.46)	5.2(0.61)	−3.55	<0.01
20	3.6(0.64)	6.6(0.79)	−2.90	<0.01

Table S2 Distribution of motives for answering "yes" * in human treatment

motive / behavior	altruism	social norm	self-image	social image	other	total
Free rider	0	5	0	0	0	5
Conditional Cooperator	9	22	10	1	5	47
Humped Cooperator	2	9	0	0	1	12
Other	1	1	0	0	0	2
Total	12	37	10	1	6	66

*Response to the question of whether participants' contributions depended on others' contributions. Question asked: "In the decision you just made, did your contribution depend on the contributions of other members of your group?" The three confused participants were not included in the analysis of motives.

Table S3 Distribution of motives for answering “no” * in human treatment

behavior \ motive	self-interest	social norm	other	total
Free rider	24	20	4	48
Conditional Cooperator	0	2	0	2
Humped Cooperator	1	0	0	1
Other	0	0	0	0
Total	25	22	4	51

*Response to the question of whether participants’ contributions depended on others’ contributions. Question asked: “In the decision you just made, did your contribution depend on the contributions of other members of your group?” The three confused participants were not included in the analysis of motives.

Table S4 Distribution of motives for answering "yes"* in computer treatment

behavior \ motive	social norm	self-image	social image	other	total
Free rider	10	0	0	3	13
Conditional Cooperator	18	6	1	2	27
Humped Cooperator	6	1	1	2	10
Total	34	7	2	7	50

*Response to the question of whether participants' contributions depended on others' contributions. Question asked: "In the decision you just made, did your contribution depend on the contributions of other members of your group?" The three confused participants were not included in the analysis of motives.

Table S5 Distribution of motives for answering "no"* in computer treatment

motive / behavior	self-interest	social norm	other	total
Free rider	47	17	3	67
Conditional Cooperator	0	0	0	0
Humped Cooperator	0	0	0	0
Total	47	17	3	67

*Response to the question of whether participants' contributions depended on others' contributions. Question asked: "In the decision you just made, did your contribution depend on the contributions of other members of your group?" The three confused participants were not included in the analysis of motives.

Table S6 Personal normative belief, empirical expectation and normative expectation of one's contribution in unconditional and conditional games

Condition	Type	Personal normative belief	t	empirical expectation	t	normative expectation	t
0	Free riders Cooperators	0 0	0	0.02 0	0.66	0.02 0.05	0.60
5	Free riders Cooperators	0.76 2.36	3.17^{***}	0.14 1.68	4.22^{***}	0.78 2.23	2.85^{***}
10	Free riders Cooperators	1.38 5.23	4.37^{***}	0.72 6.14	7.06^{***}	1.92 5.05	3.20^{***}
15	Free riders Cooperators	2.4 8.41	4.76^{***}	0.52 8.86	9.41^{***}	2.54 8.45	4.25^{***}
20	Free riders Cooperators	3.2 10.82	4.84^{***}	1.04 12.05	7.99^{***}	3.34 10.95	4.65^{***}

Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table S7 Distribution of behavior types differs between games with computer players

Type	Experiment 2	BEW
Free riders	71% (N = 51)	21% (N = 15)
Conditional cooperators	15% (N = 11)	50% (N = 36)
Humped cooperators	14% (N = 10)	10% (N = 7)
Others	0 (N = 0)	19% (N = 14)
Total	100% (N = 72)	100% (N = 72)

Table S8 Correlations and paired t tests between individual contributions to the computer and contributions to humans across six rounds

Participants	Round	Mean contribution	Mean increase	Correlation	Paired t
All (N=72)	1	5.8	3.6	0.25^{**}	5.85^{***}
	2	5.2	3.0	0.18	4.02^{***}
	3	5.7	3.5	0.33^{***}	5.04^{***}
	4	5.0	2.8	0.34^{***}	4.17^{***}
	5	4.9	2.7	0.29^{**}	3.81^{***}
	6	4.5	2.3	0.37^{***}	3.50^{***}
	1-6	5.2	3.0	0.34^{***}	5.07^{***}
Free riders (N=51)	1	4.8	3.8	0.06	5.09^{***}
	2	4.3	3.3	0.03	3.51^{***}
	3	4.7	3.7	0.16	4.07^{***}
	4	3.5	2.5	0.05	3.10^{***}
	5	3.6	2.6	0.06	2.99^{***}
	6	3.0	2.0	0.07	2.52^{**}
	1-6	4.0	3.0	0.09	4.05^{***}
Conditional cooperators (N=11)	1	8.9	2.9	0.17	2.10^{*}
	2	8.4	2.4	0.01	1.44
	3	10.3	4.3	0.25	2.75^{**}
	4	10.5	4.5	0.07	2.45^{**}
	5	8.5	2.5	0.17	1.74
	6	9.1	3.1	0.20	2.12^{*}
	1-6	9.3	3.3	0.16	2.32^{**}
Humped cooperators (N=10)	1	6.9	2.8	0.08	1.83^{*}
	2	5.8	1.7	0.14	1.32
	3	5.8	1.7	0.55^{*}	1.93^{**}
	4	6.5	2.4	0.37	1.54
	5	7.1	3.0	0.26	1.64
	6	7.2	3.1	0.34	1.47
	1-6	6.6	2.5	0.37	2.01^{*}

Mean increase = mean contribution with humans − the mean contribution to computers (2.2 MU for all; 1.0 MU for free riders; 6.0 MU for conditional cooperators; 4.1 MU for humped cooperators).
Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table S9 Paired t tests between individual contributions and expectation across the round with computers and the six rounds with humans

Participants	Round	Mean contribution	Mean expectation	Paired t
All (N=72)	Unconditional game with computer	2.2	9.0	12.1^{***}
	1	5.8	8.1	4.7^{***}
	2	5.2	7.8	4.98^{***}
	3	5.7	7.3	3.02^{***}
	4	5.0	7.4	4.52^{***}
	5	4.9	7.6	5.01^{***}
	6	4.5	7.5	5.46^{***}
	1-6	5.2	7.6	6.42^{***}
Free riders (N=51)	Unconditional game with computer	1.0	9.0	11.9^{***}
	1	4.8	8.0	5.22^{***}
	2	4.3	7.5	4.51^{***}
	3	4.7	6.7	2.91^{***}
	4	3.5	6.7	4.88^{***}
	5	3.6	6.9	4.57^{***}
	6	3.0	6.7	5.23^{**}
	1-6	4.0	7.1	6.36^{***}
Conditional cooperators (N=11)	Unconditional game with computer	6	8.5	2.3^{**}
	1	8.9	10	1.36
	2	8.4	10	2.42^{**}
	3	10.3	9.9	0.38
	4	10.5	10	0.40
	5	8.5	9.4	1.24
	6	9.1	10.3	3.13^{**}
	1-6	9.3	9.9	1.68
Humped cooperators (N=10)	Unconditional game with computer	4.1	9.7	5.85^{***}
	1	6.9	6.3	0.51
	2	5.8	6.7	1.09
	3	5.8	7.6	1.73
	4	6.5	8.2	1.57
	5	7.1	9.2	1.93^{*}
	6	7.2	8.2	1.00
	1-6	6.6	7.7	1.50

Table S10 Measuring motivations

Response	Free riders	Conditional cooperators	Humped cooperators	Total
Myself	35	5	5	45
Other people	0	0	0	0
Everyone	3	4	3	10
The group	5	1	0	6
More than others	6	1	1	8
None of the above	2	0	1	3
Total	51	11	10	72

Table S11 The difference between BEW's results and our results

No.	Investigated questions	BEW's Results	Results of Experiment 1	Results of Experiment 2
1	Is the distribution of behavior types similar in the computer and human conditions?	Yes	No	No
2	Is the level of contribution playing with computers the same as playing with humans in unconditional one-shot games?	Yes	No	No
3	Do conditional cooperators misunderstand the game?	Yes	No	No
4	Do standard control questions fail to control for understanding?	Yes	No	No
5	Are comprehenders not cooperators?	Yes	No	No
6	Distribution Motivations: Self-interest (%) vs considering others (%)	51 vs 38	–	74 vs 22
7	Can social preferences explain contributions in public goods games?	No	Yes	Yes

Table S12 Literature reviews on the effects of confusion on over-contribution in public goods game

Publication	Definition of Confusion	Experimental Design	Repected/One-shot	Measuring Confusion	Main Findings
Andreoni (1995)	Subjects have somehow not grasped the true incentives.	The experiment has three conditions. The first (Regular condition) is the standard public-goods experiment. The second (Rank condition) is the same as the first, but subjects are paid based on their rank, which makes a zero-sum game out of the standard positive sum public-goods game. The third (RegRank condition) provides feedback about rank, but subjects are paid according to the experimental earnings, just as under the Regular condition.	20 rounds, each round has feedback, including their investment decision, the group's investment in the public goods, their experimental earnings, and their monetary earnings.	Contributions in the Rank condition are considered to be due to confusion because the condition's zero-sum payoff structure left no incentive for cooperation.	On average about 75 percent of the subjects are cooperative, and about half of these are confused about incentives, while about half understand free-riding but choose to cooperate out of some form of kindness.
Houser and Kurzban (2002)	Subjects are "confused" in the sense that they make errors or do not understand the game's incentives.	The experiment has two conditions. The first condition (human condition) was a standard linear public goods game. The second condition (computer condition) was identical to the first except that each group consisted of one human player and three computer "players."	10 rounds, each round has feedback, including the number of tokens contributed by the subject, the total number of tokens contributed to the Group Exchange by the other three players, and the subject's payoff for each round.	All "cooperation" in the computer condition was attributed to confusion.	Confusion accounts for about half, 54 percent, of all tokens contributed to the public good in our standard public goods game.

Publication	Definition of Confusion	Experimental Design	Repected/One-shot	Measuring Confusion	Main Findings
Ferraro and Vossler (2010)	"Confusion" is used to characterize behavior that stems from subjects' inability to discern the relationships between the choices made and the game's incentives.	There are all-human treatment and virtual-player treatment in Experiment 1 and 3. Experiment 1 is a static experiment that allows one to explore how contributions correspond with changes in the MPCR and group size. Experiments 2 and 3 are static and dynamic experiments, respectively. Experiment 2 applies 'strategy method' design of Fischbacher et al., where subjects were matched with virtual players. Experiment 3 is a standard repeated-round VCM game. Experiment 4 uses associative framing and a complete payoff matrix. The experiment begins with 25 rounds under the all-human treatment, followed by 5 founds under the virtual-player treatment.	Experiment 1 is one-shot. Experiment 2 uses strategy method. Experiment 3 has 25 rounds with payoff feedback, including their investment in the group exchange, the aggregate investment of the other group members, their payoff from the group exchange, and their payoff from their private exchange. Experiment 4 has 30 rounds with payoff feedback.	Contributions in the virtual-player treatment stem from confusion.	Experiment 1: 50% of contributions stem from confusion; Half of measured altruism stems from confusion. Experiment 2: Confused players behave like conditional cooperators, 53% of our sample is classified as conditional cooperators. Experiment 3: 50% of contributions stem from confusion. Confused players behave like conditional cooperators. There is little evidence of learning. Experiment 4: Alternative framing and a complete payoff matrix substantially reduce confusion and change the distribution of contributions.
Bayer, Renner and Sausgruber (2013)	Confusion refers to subjects' inability to understand the incentives of the game or their incapability to deduce the dominant strategy.	The first condition is the Standard Condition which replicates a standard linear public goods game. The second condition is the Learning Condition where information about the incentive structure of the game was deliberately withheld from the subjects. In the Minimum Information Condition subjects learn nothing about the game and its incentive structure. In the Limited Information Condition, subjects still do not learn the incentive structure but the instructions contain information about group membership and use some of the terminology that is also used in the instructions of the standard VCM like "project" or "contribution."	To allow for a within-subject comparison, the experiments in both Learning Conditions consisted of two phases: in the first phase subjects made decisions in the 20-times repeated game under limited or minimum information. Then they participated in the Standard Condition and made 20 decisions. Subjects are informed only about their own payoff at the end of each period.	The behavior of subjects from Learning Condition is used as an approximation of the behavior of a confused subject in a standard linear public goods game.	The existence of confusion in the public goods game does not necessarily lead to an upward biased estimate of cooperation levels. The decay in the public goods game that is typically attributed to conditional cooperation is not an artefact of learning of confused subjects.

Publication	Definition of Confusion	Experimental Design	Repected/One-shot	Measuring Confusion	Main Findings
Burton-Chellew et al. (2016)	Individuals might misunderstand the game.	Participants first complete a contribution schedule for all possible average contributions of their computerized groupmates, and play a one-shot unconditional game with computerized groupmates. Participants then repeated the one-shot game, but with human groupmates, for 6 rounds without payoff feedback.	Participants play 6 rounds with human groupmates without feedback.	In addition to comparing contributions with the computer players and contributions with the human players, participants answered the question of whether or not the income-maximizing decision depends on what others do to test their understanding of the game's essential social dilemma.	Individuals divide into the same behavioral types when playing with computers. Behavior across games with computers and humans is correlated and can be explained by variation in understanding of how to maximize income. Misunderstanding correlates with higher levels of cooperation; and standard control questions do not guarantee understanding.
Yamakawa et al. (2016)	Confusion includes the misinterpretation of instructions, unfamiliarity with game rules, and so on. It also includes the situation in which a subject understands the incentive structure of the game, but chooses cooperative behavior mistakenly, jokingly, or for some other reasons.	There are three conditions. The first condition, called the ‘‘H condition,’’ is the standard linear public goods experiment with each group containing two subjects. In the second condition, called the ‘‘C condition,’’ each group consists of one human player and one computer player. The third condition, named the ‘‘HC condition,’’ is similar to the C condition, but the experimental earnings of the computer are paid to a real subject.	20 rounds. The feedback included their own choice, their group member’s choice, and their payoff for that round.	The motives that induce cooperative behavior in the C condition are attributed as confusion.	About 80 % of cooperation is attributable to multi-round motives, while confusion and one-shot motives account for only 2 and 18%, respectively.

Publication	Definition of Confusion	Experimental Design	Repected/One-shot	Measuring Confusion	Main Findings
Goeschl and Lohse (2018)	Confusion refers to a selfish subject's cognitive failure to understand that free-riding is his pay-off maximizing strategy.	There are four different combinations of treatments: human condition-baseline setting with unconstrained decision time; human condition-under time pressure; computer condition-baseline setting with unconstrained decision time; computer condition-under time pressure.	The main experiment is one-shot. To assess the role of confusion over time, in each treatment condition subjects also interacted in nine rounds of a repeated public goods game with feedback.	The behavioral measure of confusion replicates all central elements of the Houser and Kurzban (2002) design.	Subjects in the linear PGG with human partners to time pressure are significantly more likely to contribute zero and make weakly lower contributions on average. The confounding effect of confusion operates through a heterogeneous treatment effect. time pressure selectively affects unconfused subjects by reducing their contributions.

SI References

1. Ledyard J (1995) Public goods: A survey of experimental research. Handbook of Experimental Economics, eds Kagel J, Roth A (Princeton Univ Press, Princeton), pp253–279.
2. Andreoni J (1995) Cooperation in public goods experiments: Kindness or confusion. Am Econ Rev 85(4):891–904.
3. Ferraro PJ, Vossler CA (2010) The source and significance of confusion in public goods experiments. BE J Econ Anal Poli 10(1): 53.
4. Houser D, Kurzban R (2002) Revisiting kindness and confusion in public goods experiments. Am Econ Rev 92(4):1062–1069.
5. Burton-Chellew MN, EI Mouden C, West SA (2016) Conditional cooperation and confusion in public goods experiments. Proc Natl Acad Sci USA, 113(5): 1291-1296.
6. Yamakawa T, Okano Y, Saijo T (2016) Detecting motives for cooperation in public goods experiments. Exp Econ 19:500–512.
7. Bayer RC, Renner E, Sausgruber R (2013) Confusion and learning in the voluntary contributions game. Exp Econ 16: 478-496.
8. Lerner JS, Li Y, Valdesolo P, Kassam KS (2015) Emotion and decision making. Annu rev psychol 66: 799-823.
9. Goeschl T, Lohse J (2018) Cooperation in public good games. Calculated or confused? Eur Econ Rev 107: 185-203.