PROCEEDINGS B

# Evidence for strategic cooperation in humans

| | |
|---|---|
| Journal: | *Proceedings B* |
| Manuscript ID | RSPB-2017-0689.R1 |
| Article Type: | Research |
| Date Submitted by the Author: | n/a |
| Complete List of Authors: | Burton-Chellew, Maxwell; University of Oxford, Zoology; Magdalen College,<br>El Mouden, Claire; University of Oxford, Zoology<br>West, Stu; University of Oxford, Dept of Zoology |
| Subject: | Behaviour < BIOLOGY, Evolution < BIOLOGY |
| Keywords: | Altruism, Confusion, Signalling, Rationality, Reputation |
| Proceedings B category: | Behaviour |
| | |

# Evidence for strategic cooperation in humans

Maxwell N. Burton-Chellew,[1,2] Claire El Mouden[1] & Stuart A. West[1,2]

[1]Department of Zoology, University of Oxford, OX1 3PS

[2]Calleva Research Centre for Evolution and Human Sciences, Magdalen College, Oxford, OX1 4AU

## Abstract

Humans may cooperate strategically, cooperating at higher levels than expected from their short-term interests, to try and stimulate others to cooperate. To test this, we experimentally manipulated the extent an individual's behavior is known to others, and hence whether or not strategic cooperation is possible. In contrast to many previous studies, we avoided confounding factors by preventing individuals from learning during the game about either payoffs or about how other individuals behave. We found clear evidence for strategic cooperators – just telling some individuals that their groupmates would be informed about their behavior led to them tripling their initial level of cooperation, from 17 to 50 percent. We also found that many individuals play as if they do not understand the game, and their presence obscures the detection of strategic cooperation. Identifying such players allowed us to detect and study strategic motives for cooperation in novel, more powerful, ways.

## Keywords

Altruism; Confusion; Signalling; Rationality; Reputation

## 1. Introduction

Experiments using economic games have shown that humans routinely cooperate, sacrificing personal earnings for the good of the group (1, 2). A key question is the extent to which human cooperation is driven by a concern for the welfare of others, versus a strategic concern to increase personal success. For example, individuals may be motivated by a concern for fairness and make decisions that attempt to achieve more equitable outcomes (3-5). Alternatively, individuals may play strategically, and initially invest in helping others if they think this will lead to greater help in return (6). For example, if some people increase their level of cooperation in response to the cooperation of others, then individuals can be favoured to strategically cooperate, to induce this response from partners (6-23). While many experiments suggest fairness is important (1-5, Cf. 24), tests for the importance of strategic cooperation have produced mixed results (23, 25-34).

The lack of clear experimental support for strategic cooperation may reflect how it has been tested for, rather than whether it occurs. One issue is that some studies have focused on levels of cooperation over time, which can confound strategic cooperation with learning about payoffs and/or the behavior of others (25, 26, 35). One solution is to deprive individuals of information during a repeated version of the game. Another method is to focus upon initial levels of cooperation, before individuals can be influenced by the behavior of others. Here we use both of these methods.

Another potential problem is that variation between individuals might obscure strategic cooperation. For example, if a fraction of individuals are confused about the game's payoffs then their motivations will be hard to interpret (24). The presence of such 'irrational' players (individuals may also play irrationally for a number of other reasons, including boredom, inattention, and a desire to please or out-smart the experimenter) will obscure strategic cooperation at the aggregate level. One way to ameliorate this problem is to focus upon individual decisions to identify and control for such 'irrational' players (23, 36).

Our experiment allowed us to first identify if individuals were playing irrationally or not, and then test if they cooperated strategically (6). We measured cooperation as the value of voluntary contribution towards a public good that was personally costly but beneficial for the group (37, 38). In this game, the strategy that would give the greatest payoff in a single round is to contribute nothing. We tested for irrationality by first making all individuals play one round of this public goods game with computerized groupmates. Individuals should have no concern for the welfare of the computer, and so, if they understand the game, rational players will contribute nothing to the public good. In contrast, individuals that cooperate with the computer are behaving irrationally within the context of the game. We then examined how players cooperated with humans depending on if they played irrationally or not and if their behaviour was visible or not to their groupmates. If individuals cooperate strategically, then we expect them to cooperate more when told that their behaviour will be visible, and thus will be able influence their groupmates' decisions.

## 2. Materials and Methods

### 2. 1 Participants

The data presented here were collected at the Centre for Experimental Social Science (CESS) Oxford as part of a larger experiment (24, 39). For full methods see the electronic supplementary material (5, 40-44).

### 2.2 Game parameters

Our general setup was to make groups of four anonymous individuals play a public goods game and to vary whether their behaviour was visible or not to their groupmates. Our experiment was conducted in three stages, each involved playing the same public goods game (Table 1, Fig. ESM1). First, everyone played one round with computerized groupmates to test their 'rationality'. Second, everyone played with humans in a repeated game that had no information between rounds. In truth all players were invisible here but only some of them knew this for certain, whereas the rest faced ambiguous invisibility. Third, everyone played again with humans, but some were placed into 'visible' treatments whereby they knew their groupmates would learn about their behavior in some way, thus enabling strategic cooperation. The rest were again placed into an invisible treatment, where they knew for certain that their groupmates could not observe their behavior (see below and Fig.1). Group composition was constant in each stage and all players were told this.

### 2.3 Stage 1: Testing for confused/irrational players

Experiments on cooperation using economic games often assume that the costs a player incurs are done so 'rationally', in order to satisfy his/her preferences for the welfare of others (3). However not all players may play rationally, for example, if a self-interested player is uncertain or confused about how to maximize their earnings then they may also cooperate at some level, even though this may be inconsistent with their preferences, and hence irrational by definition (24). One way to test between these competing possibilities is to make individuals play the same game with computerized groupmates in an 'asocial control' that eliminates any concerns for other players (24, 45-48). If players cooperate both when playing with computers and when playing with humans then one cannot interpret their behavior solely on the basis of which players benefit.

We therefore tested for confused/irrational players by first making all individuals play a single round of the same game with computerized groupmates (with no feedback to prevent any learning). This was after receiving standard instructions and control questions that first made no mention of playing with computers (5, 42). The CESS laboratory forbids deception and participants are reminded of this at the start of experiments, and they were told many times that they were playing computerized groupmates and that only they would be affected by their decision ("*You are the only real person in the group, and only you will receive any money*"). For convenience, we hereafter refer to individuals that did not cooperate (contributed = 0 MU) with computers as 'rational *players*', and individuals that cooperated (contributed 0 MU) with computers as 'irrational *players*'. Our use of 'rational' here is merely a catch all term for players that are not irrational within the context of the game. An alternative nomenclature arguably could be comprehenders and non-comprehenders. However note that if a player does not cooperate with the computer they do not necessarily understand the game and vice versa. We use the term 'players' because we are not making any claims about the general rationality of these 'individuals' outside of the context of the laboratory game, and although interesting, the question of why individuals cooperate with computers is not the focus of this study (24, 45-

47, 49). Instead we show that controlling for such players facilitates the detection of strategic cooperation.

**2.4 Stage 2: Testing cooperation over time**

We then had all our participants play the same public goods game in constant groups in a 'No information' environment for six rounds. Here all players were invisible, but only some knew this for certain.

Specifically, we told all the players *"You WILL NOT receive any information about the decisions of the other players, nor about your earnings in these rounds"*. In addition, players in 3 of our 12 sessions (72 individuals) were told explicitly that their groupmates would also not receive any information, *"**nor will anyone else at any time** except for the experimenter after the experiment"*. Therefore their behaviour was 'certainly invisible', in so far as they trust experimenter instructions. In contrast, for the other 216 individuals in the other 9 sessions the invisibility of their behavior was ambiguous because they were not told what information other players' received. Comparing behavior between these treatments enables us to test if ambiguity about ones visibility to others is enough to trigger strategic cooperation. The lack of information between rounds means we can also compare how potential strategic cooperation changes over time while controlling for learning or interactions between players.

**2.5 Stage 3: Testing for strategic cooperation**

In the next stage we directly tested if providing information that makes behavior visible induces strategic cooperation. After stage 2, we restarted the game with new randomly formed groups and manipulated the extent to which strategic cooperation was possible across three conditions that varied in how visible they made behaviour to groupmates. We had two visible treatments, one where behaviour was 'fully' visible, and one where it was only 'partly' visible, and we had a new invisible treatment, where players could only observe the behavior of outside groups (not their own). Contrary to stage two, in these three treatments all players received information during the game that could affect their behavior. We therefore only compared their levels of *initial cooperation*, which occurred after our manipulation (telling players what information their groupmates will receive) but before any learning or in-game dynamics could occur.

Specifically, to make behaviour fully visible, we told players that they and their groupmates would be told after each round the individual contributions of each player. To make behaviour partly visible, we told players that they and their groupmates would be told their personal earnings from each round. This made an individual's behavior 'partly visible', because the act of cooperation always informs groupmates that there is at least one cooperator in the group, and information on earnings reliably informs individuals of the average level of cooperation in their group. We told players in both visible treatments: "*You and everybody else in your group will receive the SAME INFORMATION.*" Players in the fully visible treatment were then told: "*The information each person will receive is what the decisions were of each player in the group and what their own earnings are from each round*" and players in the partially visible were told: "*The information each person will receive is what their own earnings are from each round.*" We randomly assigned 48 players, regardless if they were irrational or not, to each of these two visible treatments. However, because rational players were so rare (see results), our analyses mostly combined the fully and partially visible treatments into one 'visible' treatment (N=96).

To make behaviour invisible within groups, and thus to prevent strategic cooperation from working, we again did not provide any information on either personal earnings or the behavior of same-group members. However, in contrast to stage two, individuals knew they and their groupmates would receive information, but that it would only be about the behavior of other groups that they had no connection to (for the purposes of another study, we provided these individuals with one of four types of information about other group(s) (39) (ESM Methods and Results). Crucially, the flow of information between groups was strictly unidirectional, thus an individual's behavior was neither visible to his/her groupmates, nor to the groups he/she was observing. This meant an individual's cooperation could still be motivated by a concern for the welfare of others, but not sensibly by strategic concerns. Here we are interested in how individuals initially cooperated when knowing that their behaviour will be visible or not to their groupmates. Specifically, we told all players in our invisible treatment the following:

"You and everybody else in your group will receive the SAME INFORMATION
The information you and the others in your group will receive WILL NOT COME FROM YOUR GROUP, but instead will come from (an)other group(s).
*Your decisions will only affect the earnings of people in your group."* [original emphases shown]

In all our treatments, all members of the same group were assigned to the same treatment, and the treatment design was common knowledge to all members of the group. Our focus here is on how individuals played in the opening round of these treatments, before they have had a chance to learn about the game or other players, depending on whether their groupmates could potentially learn something about how they had behaved. We focus on the first round because, at this stage, the only difference between individuals is the information that they have been told they and their groupmates *will receive in the future*. At this point, individuals cannot have been influenced by the way in which other individuals play. This allows us to test for strategic motives while controlling for learning about the game's payoffs and the nature of other players.

## 3. Results

### 3.1 Irrational players

We first tested if players cooperated with computerized groupmates (stage one). We found that 185 (64%) of our players cooperated with the computer and thus were 'irrational' in the context of our experiment. This meant that, at most, only 103 (36%) of our players, who did not cooperate with the computer, could be classified as rational. Overall the average cooperation towards computers was 36% (7.1 MU ±0.42 S.E.M., Median = 5 MU, Mode = 0 MU, Table 1, Fig. ESM2). The rarity of rational players made it difficult to compare their behaviour across the partly and fully visible treatments (N = 23 & 20), and so we combined these data into one 'visible' treatment for all analyses in the main text (ESM Results, Fig. ESM3).

**3.2 Strategic cooperation**

Our main test is how levels of cooperation (number of MU contributed) varied, in the opening round of the repeated game with information (stage three), depending on if this information made an individual's behavior visible or not to their groupmates, and if we had classified them as rational or not. Overall, we found that 'rational' and 'irrational' players responded differently to whether their behavior was visible or not, with only rational players cooperating significantly more when visible (generalized linear model with a binary-logistic link set to 20 trials, hereafter referred to as GzLM: level of cooperation ~ rationality*visibility: $F_{1,284} = 12.4$, $P < 0.001$, Fig. 2ab).

Specifically, the rational players cooperated three times as much in the treatments where behavior was visible. The mean level of cooperation was 17% (3.3 ±0.79 MU, N = 60) in the invisible treatment, and 50% (9.9 ±1.35 MU, N = 43) in the visible treatments, a significant difference (GzLM: $F_{1,101} = 17.9$, $P < 0.001$, $\beta = 1.6 \pm 0.39$). This difference is even more striking if we examine the initial modal cooperation, which was 0% (0 MU) when invisible and 100% (20 MU) when visible (Fig. 2a). The general timeline of cooperation for rational players from stage two to stage three can be seen in figure 3.

In contrast, the irrational players did not significantly vary their behavior depending upon whether they were visible or not. Their mean level of cooperation when invisible was 47% (9.3 ±0.58 MU, N = 132), and when visible was 51% (10.2 ±0.88 MU, N = 53), a non-significant difference (GzLM: $F_{1,183} = 0.6$, $P = 0.444$, $\beta = 0.2 \pm 0.21$). Their modal cooperation was 50% (10 MU) in both cases (Fig. 2b). These results were robust to various forms of analysis (ESM Results, Fig. ESM4).

Although the average behaviour of rational and irrational players was the same in the visible treatment, cooperating at 50 and 51% respectively, the variances of their decisions were significantly different (respective variances = 78.3 and 41.0, Levene's test of homogeneity of variances = 14.4, df = 1,94, $P < 0.001$). The contributions of rational players exhibited a bimodal distribution of 0 and 100% (0 and 20 MU) whereas the irrational players tended to contribute 50% (10 MU) (Fig. 2a,b), suggesting that the behavior and cognition among rational and irrational players differed in the visible treatment, despite similar levels of average cooperation.

**3.3 Strategic cooperation under ambiguity**

We found further support for strategic cooperation in the prior no-information game (stage 2). Again we found a significant interaction between how rational and irrational players cooperated depending on how visible their behavior (potentially) was (GzLM, initial cooperation in the first round of the no-information game ~ rationality*visibility: $F_{1,284} = 5.7$, $P = 0.017$, Table 1, Fig. ESM5).

Merely being potentially visible was enough to make rational players, but not irrational players, initially cooperate more. When rational players could be certain their behavior was invisible they initially only cooperated at 14% (2.8 ±1.27 MU), but when their visibility was ambiguous they initially cooperated at 32% (6.3 ±0.88 MU)(GzLM: $F_{1,101} = 3.9$, $P = 0.052$, $\beta = 1.0 \pm 0.57$, Fig. ESM5a). In contrast, the initial cooperation of irrational players did not significantly vary depending on visibility, cooperating at 53% (10.5 ± 0.80 MU) when their invisibility was certain, and 49% (9.8 ± 0.53 MU) when their visibility was ambiguous (GzLM: $F_{1,183} = 0.5$, $P = 0.468$, $\beta = 0.1 \pm 0.20$, Fig. ESM5b). For an overview of levels of initial cooperation by all treatments see Table 1.

**3.4 Strategic decline**

If individuals are cooperating in order to stimulate others to cooperate in the future, then they might cooperate less as the final round approaches. In support of this idea, we found that rational players faced with ambiguous visibility in stage two significantly decreased their cooperation over the six rounds, from an initial 32% (6.3 ±0.88 MU) to 13% (2.6 ±0.66 MU) (GzLMM: $F_{1,412}$ = 91.4, P < 0.001, β = -0.48 ± 0.050, Fig. 3). In contrast, if they had certain invisibility, they cooperated at a low, constant, level of around 13% (2.6 ±0.49 MU mean of six rounds) (GzLMM: $F_{1,105}$ = 1.2, P = 0.276, β = -0.09 ± 0.084). This led to a significant interaction for rational players between the degree of invisibility (ambiguous versus certain) and cooperation over time (GzLMM: $F_{1,511}$ = 13.9, P < 0.001, Fig. ESM6). In contrast, this same interaction was not significant for irrational players, as their cooperation over time did not depend on if their invisibility was certain or ambiguous (GzLMM: $F_{1,885}$ = 1.0, P < 0.322, Fig. ESM6). Therefore many rational players facing ambiguous visibility initially cooperated at a higher level, before generally decreasing their cooperation as the end of the game approached (Table ESM1). As no information was available between rounds this decrease could not possibly have been due to individuals learning or responding to the behaviour of others.

## 4. Discussion

We found experimental support for strategic cooperation in humans. We categorized players depending upon whether they cooperated (irrational) or not (rational) with a computer in a one shot public goods game. Telling rational players, who did not cooperate with a computer, that their groupmates could potentially learn about their behaviour, led to them tripling their level of initial cooperation, from 17 to 50% (Fig. 2a). This result is even more striking when examining the modal level of cooperation, which increased from 0 to 100%! In contrast, irrational players, who cooperated with the computer, did not vary their level of initial cooperation depending upon whether others could potentially learn about their contributions (Fig. 2b).

Our results provide a potential explanation for why previous studies often failed to find conclusive support for strategic cooperation (25, 30, although see refs. 33, 34). We found clear evidence for strategic cooperation, but only in 'rational' players. Overall, we found that at most 36% of individuals played rationally, whereas the other 64% of individuals played irrationally, or just didn't fully understand how to best play the game. Consequently, the irrational players who do not appear to cooperate strategically were relatively common (64%) although this number presumably could be reduced depending on what instructions are used (47, 49). This variation across individuals produces 'noise' that unless controlled for will reduce the chances of detecting strategic cooperation and potentially other interesting behaviours.

Are strategic cooperators self-interested individuals that take time to calculate what is best for them, or do they just intuitively cooperate in situations that potentially favour cooperation? Our study cannot differentiate between these possibilities. There is growing evidence that many humans show intuitive cooperation, cooperating without considering the costs and benefits of the situation (50-52). Strategic cooperation could also be intuitive if natural selection has favoured heuristics that activate cooperation accordingly, such as when behavior is visible and reputational mechanisms are likely (53-59). However such heuristics or priming effects would be expected to also apply to irrational players, and we found they did not, although this could perhaps be because they were generally less attentive to the instructions and less focused on the game and their environment.

A number of competing explanations for our data can be rejected. One potential hypothesis is that our test for rationality merely separates cooperators from non-cooperators. However this is clearly not the case, because when we told rational players that their behaviour would be visible to their groupmates they generally adjusted their behaviour from non-cooperative to cooperative. In contrast, irrational players did not behave differently when their behaviour was visible or not. Another potential hypothesis therefore is that irrational players cooperate for purely pro-social reasons. However the fact that they also cooperated with computers, and thus paid costs when no other players would benefit, means that their motivation to pay costs cannot be interpreted purely on the basis of who benefits. Instead, the social benefits of their behaviour in this experiment appear to be byproducts of their irrationality, which may be driven by confusion surrounding the game or their intuition (24, 51). Consistent with our interpretation, the modal behaviour of irrational players across all treatments was 50%, suggesting the (potentially rational) use of a bet-hedging strategy by players who were uncertain how best to play the game.

Finally, because our no information version of the game always preceded our games with information, another potential hypothesis is that the contrast in games communicated to players an experimenter demand to respond differently (60). We do not reject the idea of experimenter demand, but point out that again it would have to not apply to irrational players whose behavior was unchanged. Furthermore, the overall level of cooperation in the visible treatments is consistent with published behavior in public-goods games. Instead the novel result that illuminates strategic cooperation comes from the lack of cooperation by rational players if placed into an invisible treatment (17%), suggesting that their cooperation in visible treatments (50%) is largely strategic.

## 5.0 Conclusion

Overall, the high number of irrational players along with the low level of cooperation by rational players when invisible suggests that measures of altruistic motives may have been biased in many previous studies. Aside from this, our results have two potentially important real-world behavioral implications. First, when cooperative relationships fail, e.g. among housemates, colleagues, teams or nations, this may be because individuals perceive a reduction in future opportunities for cooperation. Attempts to maintain cooperation may therefore have more success by emphasizing the future benefits of mutually beneficial relationships. Second, perceived ambiguity and comprehension significantly affected how individuals behaved in response to opportunities to influence social behavior. Understanding this will be essential if one wants to influence behavior successfully, while avoiding harmful unintended consequences. This is becoming increasingly important as public policy makers seek to apply behavioral insights that attempt to influence or 'nudge' people's behaviour in socially useful ways (61).

## Competing interests

We have no competing interests

**Authors' contributions**

MNBC conceived, designed, programmed, conducted, analyzed and wrote the study; CEM conceived, designed, conducted, analyzed and wrote the study; SAW conceived, designed, analyzed and wrote the study.

**Data accessibility**

Data available from the Dryad Digital Repository:
http://dx.doi.org/10.5061/dryad.s612m

**Acknowledgements**

CESS for running experiments. Miguel dos Santos, Daniel Friedman & Larry Samuelson for discussion.

**Funding**

MNBC & CEM: John Fell Fund Oxford; MNBC & SAW: Calleva Research Centre for Evolution and Human Sciences, Magdalen College.

Figure 1. Our experimental design, hypothesis, and general result. In our invisible treatment, players were told that their groupmates would not be able to observe their behavior nor their own earnings (which can be used to calculate the average cooperation of groupmates). In our visible treatments, players were told that their groupmates would see their own earnings, and half the players were also told that their groupmates would see their individual decisions. Rational players were those that did not cooperate with computers in a preceding version of the same public goods game. Irrational players did cooperate with the computer.

Figure 2ab. Rational players are strategic cooperators. The figures show the distribution of the level of contribution in a public goods game (cooperation), distinguishing between individuals who (a) had not cooperated with computers (rational players) or (b) had previously cooperated with computers (irrational players). The light grey bars are for individuals whose behavior would not be revealed to their groupmates (invisible), and the dark grey bars are for individuals whose behavior would be revealed to their groupmates (visible). Part (a) shows that rational players cooperate at higher levels when their behavior is visible.

Figure 3. The timeline of strategic cooperation. This figure examines the behavior of the rational players in two repeated public goods games. For simplicity, only the rational players that experienced the ambiguously visible treatment in the no-information game are shown (N=82). All players first played six rounds of the game with no information, not knowing if their behavior was visible or not, before being randomly assigned to play another six rounds, in new groups, where their behavior would either be observable in some form (visible, dark grey diamonds) or not (invisible, light grey circles) to their groupmates. Dashed lines show 95% confidence intervals.

## References

1. Fehr E, Fischbacher U. The nature of human altruism. Nature. 2003;425(6960):785-91.

2. Henrich J, Boyd R, Bowles S, Camerer C, Fehr E, Gintis H, et al. "Economic man" in cross-cultural perspective: behavioral experiments in 15 small-scale societies. Behav Brain Sci. 2005;28(6):795-815; discussion -55.

3. Fehr E, Schmidt KM. A theory of fairness, competition, and cooperation. Q J Econ. 1999;114(3):817-68.

4. Fischbacher U, Gachter S, Fehr E. Are people conditionally cooperative? Evidence from a public goods experiment. Econ Lett. 2001;71(3):397-404.

5. Fischbacher U, Gachter S. Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments. Am Econ Rev. 2010;100(1):541-56.

6. Kreps DM, Milgrom P, Roberts J, Wilson R. Rational Cooperation in the Finitely Repeated Prisoners-Dilemma. J Econ Theory. 1982;27(2):245-52.

7. Trivers RL. Evolution of Reciprocal Altruism. Quarterly Review of Biology. 1971;46(1):35-&.

8. Alexander RD. The Biology of Moral Systems. New York: Aldine de Gruyter; 1987.

9. Roberts G. Competitive altruism: from reciprocity to the handicap principal. Proc R Soc Lond B. 1998;265:427-31.

10. Roberts G, Sherratt TN. Development of cooperative relationships through increasing investment. Nature. 1998;394:175-9.

11. Sylwester K, Roberts G. Cooperators benefit through reputation-based partner choice in economic games. Biology Letters. 2010;6(5):659-62.

12. Nowak MA, Sigmund K. Evolution of indirect reciprocity by image scoring. Nature. 1998;393:573-7.

13. Milinski M, Semmann D, Bakker TCM, Krambeck HJ. Cooperation through indirect reciprocity: image scoring or standing strategy? P Roy Soc Lond B Bio. 2001;268(1484):2495-501.

14. Milinski M, Semmann D, Krambeck HJ. Donors to charity gain in both indirect reciprocity and political reputation. P Roy Soc Lond B Bio. 2002;269(1494):881-3.

15. Milinski M, Semmann D, Krambeck HJ. Reputation helps solve the 'tragedy of the commons'. Nature. 2002;415(6870):424-6.

16. Yoeli E, Hoffman M, Rand DG, Nowak MA. Powering up with indirect reciprocity in a large-scale field experiment. P Natl Acad Sci USA. 2013;110:10424-9.

17. Leimar O, Hammerstein P. Evolution of cooperation through indirect reciprocity. P Roy Soc Lond B Bio. 2001;268(1468):745-53.

18. Barclay P. Trustworthiness and competitive altruism can also solve the "tragedy of the commons". Evol Hum Behav. 2004;25(4):209-20.

19. Barclay P, Willer R. Partner choice creates competitive altruism in humans. Proceedings of the Royal Society B-Biological Sciences. 2007;274(1610):749-53.

20. Barclay P. Biological markets and the effects of partner choice on cooperation and friendship. Curr Opin Psychol. 2016;7:33-8.

21. McCullough ME, Pedersen EJ. The Evolution of Generosity: How Natural Selection Builds Devices for Benefit Delivery. Social Research. 2013;80(2):387-410.

22. Wubs M, Bshary R, Lehmann L. Coevolution between positive reciprocity, punishment, and partner switching in repeated interactions. Proceedings of the Royal Society B-Biological Sciences. 2016;283(1832).

23. Cox CA, Jones MT, Pflum KE, Healy PJ. Revealed reputations in the finitely repeated prisoners' dilemma. Econ Theor. 2015;58(3):441-84.

24. Burton-Chellew MN, El Mouden C, West SA. Conditional cooperation and confusion in public-goods experiments. P Natl Acad Sci USA. 2016;113(5):1291-6.

25. Andreoni J. Why Free Ride - Strategies and Learning in Public-Goods Experiments. J Public Econ. 1988;37(3):291-304.

26. Croson RTA. Partners and strangers revisited. Econ Lett. 1996;53(1):25-32.

27. Engle-Warnick J, Slonim RL. The evolution of strategies in a repeated trust game. J Econ Behav Organ. 2004;55(4):553-73.

28. Dal Bo P. Cooperation under the shadow of the future: Experimental evidence from infinitely repeated games. Am Econ Rev. 2005;95(5):1591-604.

29. Muller L, Sefton M, Steinberg R, Vesterlund L. Strategic behavior and learning in repeated voluntary contribution experiments. J Econ Behav Organ. 2008;67(3-4):782-93.

30. Andreoni J, Croson R. Partners versus strangers: random rematching in public goods experiments. In: Plott CR, Smith VL, editors. Handbook of experimental economics results. 1. Amsterdam: North-Holland; 2008. p. 776-83.

31. Engelmann D, Fischbacher U. Indirect reciprocity and strategic reputation building in an experimental helping game. Games and Economic Behavior. 2009;67(2):399-407.

32. Ambrus A, Pathak PA. Cooperation over finite horizons: A theory and experiments. J Public Econ. 2011;95(7-8):500-12.

33. Reuben E, Suetens S. Revisiting strategic versus non-strategic cooperation. Exp Econ. 2012;15(1):24-43.

34. Dreber A, Fudenberg D, Rand DG. Who cooperates in repeated games: The role of altruism, inequity aversion, and demographics. J Econ Behav Organ. 2014;98:41-55.

35. Andreoni J, Croson R. Partners versus strangers: random rematching in public goods experiments. In: Plott CR, Smitt VL, editors. Handbook of Experimental Economics Results. 1. Amsterdam: North-Holland; 2008. p. 776.

36. Cooper R, DeJong DV, Forsythe R, Ross TW. Cooperation without reputation: Experimental evidence from prisoner's dilemma games. Games and Economic Behavior. 1996;12(2):187-218.

37. Ledyard J. Public goods: a survey of experimental research. In: Kagel JH, Roth AE, editors. Handbook of Experimental Economics. Princeton: Princeton University Press; 1995. p. 253-79.

38. Chaudhuri A. Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. Exp Econ. 2011;14(1):47-83.

39. Burton-Chellew MN, El Mouden C, West SA. Social learning and the demise of costly cooperation in humans. P Roy Soc B-Biol Sci. 2017;In Press.

40. Greiner B. Subject pool recruitment procedures: organizing experiments with ORSEE. J Econ Sci Assoc. 2015;1(1):114-25.

41. Fischbacher U. z-Tree: Zurich toolbox for ready-made economic experiments. Exp Econ. 2007;10(2):171-8.

42. Fehr E, Gachter S. Altruistic punishment in humans. Nature. 2002;415(6868):137-40.

43. Gosling SD, Rentfrow PJ, Swann WB. A very brief measure of the Big-Five personality domains. J Res Pers. 2003;37(6):504-28.

44. Simmons JP, Nelson LD, Simonsohn U. False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. Psychol Sci. 2011;22(11):1359-66.

45. Houser D, Kurzban R. Revisiting kindness and confusion in public goods experiments. Am Econ Rev. 2002;92(4):1062-9.

46. Shapiro DA. The role of utility interdependence in public good experiments. Int J Game Theory. 2009;38(1):81-106.

47. Ferraro PJ, Vossler CA. The Source and Significance of Confusion in Public Goods Experiments. Be J Econ Anal Poli. 2010;10(1).

48. Burton-Chellew MN, West SA. Prosocial preferences do not explain human cooperation in public-goods games. P Natl Acad Sci USA. 2013;110(1):216-21.

49. Bigoni M, Dragone D. Effective and efficient experimental instructions. Econ Lett. 2012;117(2):460-3.

50. Rand DG, Greene JD, Nowak MA. Spontaneous giving and calculated greed. Nature. 2012;489(7416):427-30.

51. Rand DG, Peysakhovich A, Kraft-Todd GT, Newman GE, Wurzbacher O, Nowak MA, et al. Social heuristics shape intuitive cooperation. Nat Commun. 2014;5.

52. Rand DG, Kraft-Todd GT. Reflection does not undermine self-interested prosociality. Front Behav Neurosci. 2014;8.

53. Gigerenzer G. Rationality for Mortals: how people cope with uncertainty. New York: OUP; 2008.

54. Burnham TC, Hare B. Engineering human cooperation: Does involuntary neural activation increase public goods contributions? Hum Nature-Int Bios. 2007;18(2):88-108.

55. Bateson M, Nettle D, Roberts G. Cues of being watched enhance cooperation in a real-world setting. Biology Letters. 2006;2(3):412-4.

56. Sparks A, Barclay P. Eye images increase generosity, but not for long: the limited effect of a false cue. Evol Hum Behav. 2013;34(5):317-22.

57. Nettle D, Harper Z, Kidson A, Stone R, Penton-Voak IS, Bateson M. The watching eyes effect in the Dictator Game: it's not how much you give, it's being seen to give something. Evol Hum Behav. 2013;34(1):35-40.

58. Delton AW, Krasnow MM, Cosmides L, Tooby J. Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. P Natl Acad Sci USA. 2011;108(32):13335-40.

59. Krasnow MM, Delton AW, Tooby J, Cosmides L. Meeting now suggests we will meet again: Implications for debates on the evolution of cooperation. Sci Rep-Uk. 2013;3.

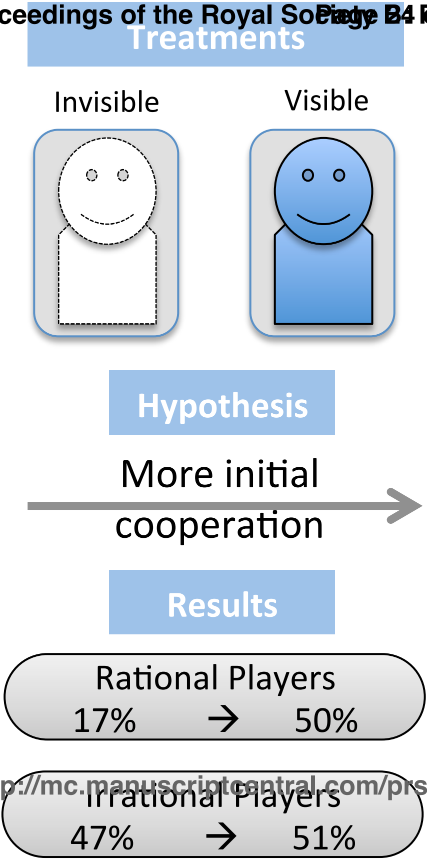
60. Zizzo DJ. Experimenter demand effects in economic experiments. Exp Econ. 2010;13(1):75-98.

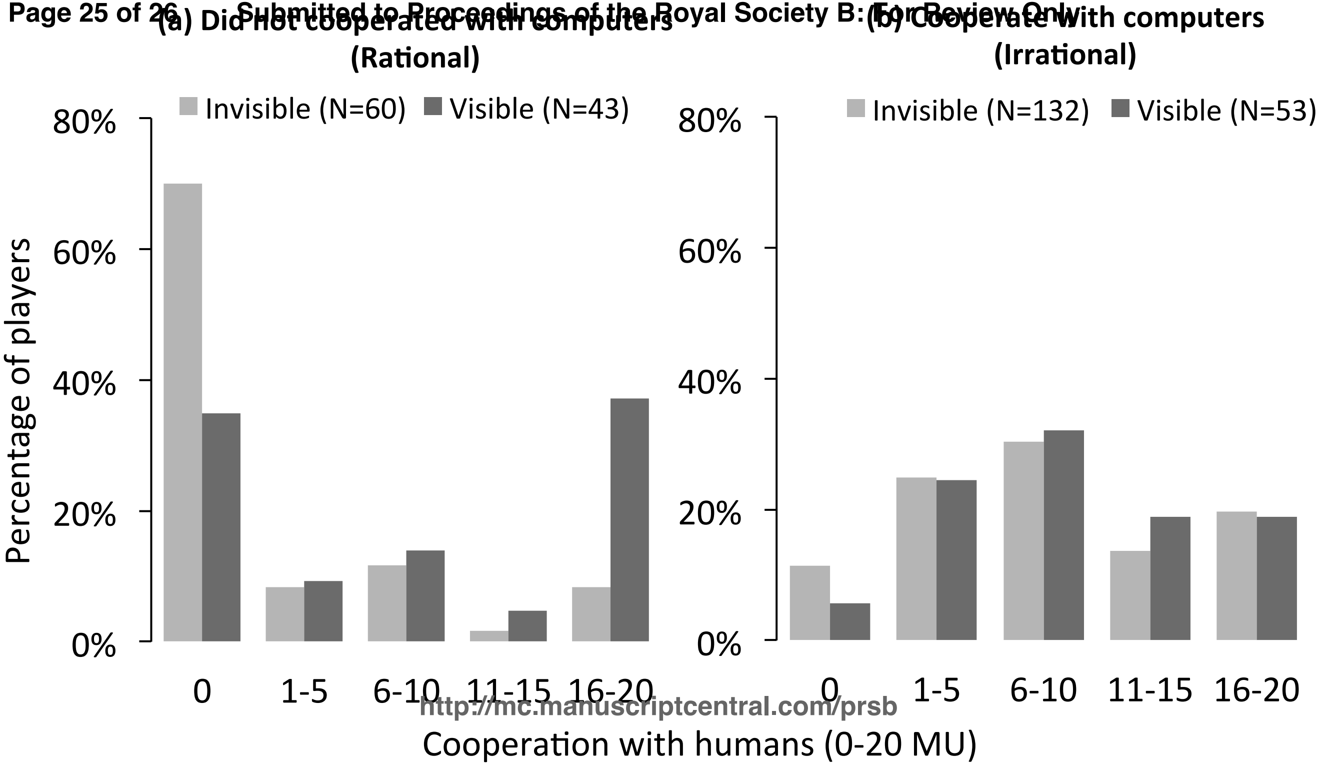
61. Thaler RH. Behavioral Economics: Past, Present, and Future. Am Econ Rev. 2016;106(7):1577-600.

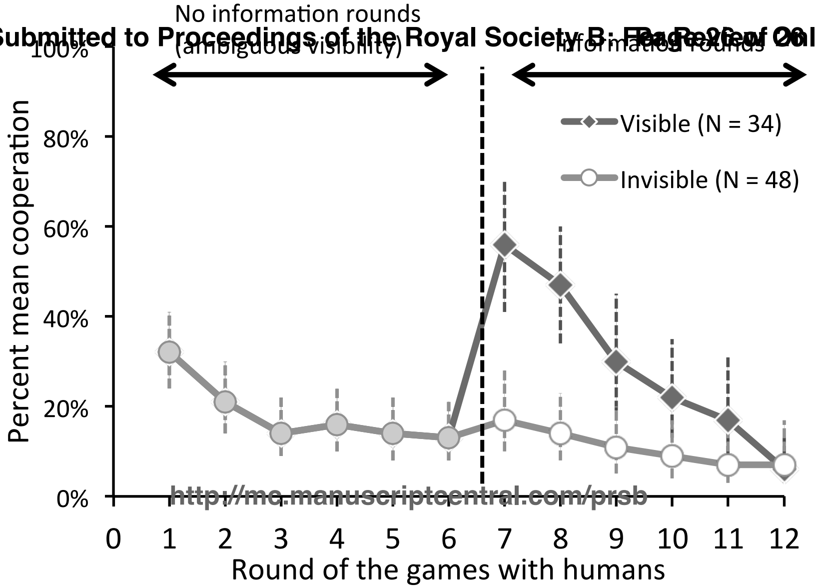
Table 1. A summary of the mean initial cooperation of rational (did not cooperate with computers) and irrational (cooperated with computers) players for the different treatments

| Stage and treatment | Told groupmates will see: Their own payoffs | Told groupmates will see: Your behavior | Told groupmates will see: Behavior of other groups | Mean initial cooperation (0-20 MU): Rational players (N=103) | Mean initial cooperation (0-20 MU): Irrational players (N=185) |
|---|---|---|---|---|---|
| Computer play[1] (Stage 1) | Na | Na | Na | 0 by definition | 11.1 ±0.44 |
| Certain invisibility[2] (Stage 2) | No | No | No | 2.8 ±1.27, N= 21 | 10.5 ±0.80, N = 51 |
| Ambiguous visibility[3] | Ambiguous | Ambiguous | Ambiguous | 6.3 ±0.88, N = 82 | 9.8 ±0.53, N = 134 |
| Certain invisibility[4] (Stage 3) | No | No | Yes | 3.3 ±0.79, N = 60 | 9.3 ±0.58, N = 132 |
| Partly visible[5] | Yes | No | No | 9.2 ±1.89, N = 23 | 9.2 ±1.47, N =25 |
| Fully visible[6] | Yes | Yes | No | 10.8 ±1.96, N = 20 | 11.1 ±1.03, N = 28 |

[1] 1 round of play. Players told they were playing computerized groupmates
[2] 6 rounds. Players received no information and told no one else would either.
[3] 6 rounds. Players received no information but not told no one else would either.
[4] 6 rounds. Players told groupmates could only observe other groups
[5] 6 rounds. Players told that everyone would learn their personal payoffs each round.
[6] 6 rounds. Players told that everyone would learn payoffs and everyone's decisions. Partly and Fully visible treatments were combined for analyses in main text.


Treatments
Invisible
Visible
Hypothesis
More initial cooperation
Results
Rational Players
17% → 50%
Irrational Players
47% → 51%


(a) Did not cooperated with computers (Rational)
Invisible (N=60)
Visible (N=43)
(b) Cooperate with computers (Irrational)
Invisible (N=132)
Visible (N=53)
80%
60%
40%
20%
0%
Percentage of players
0
1-5
6-10
11-15
16-20
Cooperation with humans (0-20 MU)


No information rounds
(ambiguous visibility)
Information rounds
Visible (N = 34)
Invisible (N = 48)
Percent mean cooperation
100%
80%
60%
40%
20%
0%
0
1
2
3
4
5
6
7
8
9
10
11
12
Round of the games with humans

**Evidence for strategic cooperation in humans**

Maxwell N. Burton-Chellew,[1,2] Claire El Mouden[1] & Stuart A. West[1,2]

[1]Department of Zoology, University of Oxford, OX1 3PS
[2]Calleva Research Centre for Evolution and Human Sciences, Magdalen College, Oxford, OX1 4AU

List of Electronic Supplementary Material Files

File 1
ESM Materials & Methods & Results

File 2
ESM Table 1

File 3
Fig. ESM 1. Overview of the experiment. The experiment was conducted in three stages. Stage one identified if players were rational or not. Diagnosis was based on if players cooperated with computers. Stage two was a repeated public-goods game with constant groups but no information between rounds. Only some players knew that their behavior was invisible to their groupmates. Stage three was another repeated public-goods game with new constant groups and information between rounds that either revealed a player's behavior to their groupmates fully, or partly (via payoffs), or not. Results show levels of initial cooperation.

File 4
Fig. ESM 2. Cooperation with computers. The figure shows the distribution of the level of contributions in the public goods game (cooperation) towards computerized groupmates. Players were told that their groupmates were computers that would be choosing contributions randomly. Contributions above 0 MU made a loss and therefore cannot be rationalized, in the context of this game, with either selfish or prosocial preferences.

File 5
Figure. ESM 3. Strategic cooperation when partly versus fully visible. The figures show the distribution of the level of contribution in a public goods game (cooperation), distinguishing between individuals who (a) had not cooperated with computers (rational players) or (b) had previously cooperated with computers (irrational players). The light grey bars are for individuals who knew their groupmates would be told their earnings and thus could infer the average cooperation of their groupmates (partly invisible). The dark grey bars are for individuals who knew their individual behavior would be revealed to their groupmates (fully visible).

File 6
Figure ESM 4. Analyzing players that pass the control questions gives the same results. The figures show the behavior of only those players that passed the control questions. Part (a) shows that among players that passed control

questions, the rational players cooperate at higher levels when their behavior is visible.

File 7

Fig. ESM 5. Rational players facing uncertainty are strategic cooperators. The figures show the distribution of the level of contribution in a public goods game (cooperation), distinguishing between individuals who (a) had not cooperated with computers (rational players) or (b) had previously cooperated with computers (irrational players). The light grey bars are for individuals who knew for certain their behavior would not be revealed to their groupmates (invisible), and the dark grey bars are for individuals whose visibility was ambiguous because they were not told what information their groupmates would receive (potentially visible). Part (a) shows that rational players cooperate at higher levels when their behavior is potentially visible.

File 8

Fig. ESM 6. Strategic cooperation over time. This figure examines the behavior of players in a repeated public goods games with no information between rounds. The figure distinguishes between individuals who (a) had not cooperated with computers (rational players) or (b) had previously cooperated with computers (irrational players). The light grey bars are for individuals who knew their behavior would not be revealed to their groupmates (invisible), and the dark grey bars are for individuals whose visibility was ambiguous because they were not told what information their groupmates would receive (potentially visible). Dashed lines are 95% confidence intervals.

**Strategic cooperation in humans: Electronic Supplementary Material**

Maxwell N. Burton-Chellew,[1,2] Claire El Mouden[1] & Stuart A. West[1,2]

[1]Department of Zoology, University of Oxford, OX1 3PS
[2]Calleva Research Centre for Evolution and Human Sciences, Magdalen College, Oxford, OX1 4AU

## ESM Materials & Methods

### Participants

The data presented here were collected at the Centre for Experimental Social Science (CESS) Oxford as part of a larger experiment (24, 38). CESS recruited participants using ORSEE (39) and obtained participant consent forms. Experiments were conducted using z-Tree (40). We tested 288 participants (151 females, 137 males, mean self-reported age = 27 years, range = 18-74, S.D. = 11.97) across 12 sessions of 24 individuals.

For the purposes of best practice we report the following details (43). The overall sample size was as large as possible given the costs of the experiment and the value of a small-project grant from the John Fell Fund Oxford (maximum £7,500). No participants or sub-groups have been omitted from our analyses. All participants completed a questionnaire after the experiment recording their gender, age, and a short 10-item personality test (42). The primary purpose of this questionnaire was to occupy the participants while we arranged their cash payments. We have not analyzed any of these demographic variables, as they formed no part in our a priori hypotheses, although we may use them later for exploratory analyses.

### Game parameters

In each round, we gave each individual 20 Monetary Units (MU) and made them contribute 0-20 MU to a group fund. We multiplied the group fund by an efficiency factor of 1.6 before sharing it among the four group members equally, regardless of differences in contribution. Therefore the marginal per capita return on each contributed MU was 0.4. The strategy that would maximize income in a single round (payoff dominant strategy) was therefore to contribute 0 MU. MUs from each round were banked and non-transferrable between rounds. The unique equilibrium for all rounds is also to contribute 0 MU. This can be deduced for a population of rational self-interested individuals that each assumes everyone to be rational, via backward induction from rational play in the final round (6). However, it can still be rational for a selfish individual to cooperate before the final round if individuals are uncertain about the rationality or selfishness of their groupmates and can be influenced in some way (6).

We gave our participants a copy of standard instructions and control questions routinely used in an attempt to ensure comprehension (5, 41). However, in contrast to other studies, we did not force our participants to correctly answer all questions before progressing to the experiment. We analyze all participants and then repeat our main analyses on the subset of participants that only answered all questions correctly. The results are qualitatively the same (ESM Results). Our experiment was conducted in three stages: first playing computerized groupmates with no information, second playing human groupmates with no information, and third playing human groupmates with information between rounds that either allowed or prevented strategic cooperation.

**No information game**

Another advantage of the no information game is that by comparing the behavior of individuals in this no information setting, depending on how they are later on randomly assigned to either visible or invisible treatments, we can test if our treatment differences are due to different effects or just different participant pools. We avoided spillover effects because individuals cannot learn about either game or about other players in the no information treatment, and because all individuals play this no information treatment, it cannot differentially affect the later treatments.

**Information games**

In our invisible treatment, for the purposes of another study, we told 48 players the following:

*Let us call this other group, Group A, and call your group, Group B.*

*All the players in both Group A and B will only receive information about the decisions in Group A.*

*Nobody in either group will learn anything about Group B.*

*Everybody in your group will only ever see information about the other group.*

And we told 144 of these 192 participants:

*The people in your group will receive information about 3 other groups. Let us call these Groups A, B and C.*

*Your group will receive information describing the decisions within these groups.*

*Your group will not receive any information about players in your own group.*

*Specifically, your group will be told*

[and]

*what the average decisions were in each group.*

[or] *what the decisions were of the player that made the most money in each of groups A, B, and C.*

[or] *what the decisions were of the three players that made the most money out of all the players in Groups A, B, and C combined.*

How individuals responded to these different sources of social information over time is the focus of another study comparing how people respond to information on common or successful behaviours (38). We combined the data from these four sub-treatments into one 'invisible' treatment (N=4x48=192, randomly assigned).

**Statistical methods**

We compared average levels of cooperation using the generalized linear model (GzLM) function in IBM SPSS version 22. We ran our models with a binary-logistic link set to 20 trials (because contributions were bounded between 0 and 20 MU). We set the parameter estimation method to hybrid; the scale parameter method to Pearson chi square, and the covariance matrix used a model-based estimator. The chi square statistics used the likelihood ratio and the confidence interval type was profile likelihood. All other settings used the default. We compared levels of cooperation over time using the generalized linear mixed model (GzLMM) function to control for pseudo replication arising from repeated measures on the same individuals over time. We fitted random intercepts for each participant. We estimated degrees of freedom with the more conservative option to reduce false positives (Satterthwaite method), and all other options were set to the default. We compared the frequencies of cooperators using Fisher's exact test (FET) and all significance values are two tailed calculations.

**ESM Results**

**Irrational players**

For the distribution of cooperation towards computers see Fig. ESM 2.

**Cooperation when partly visible versus fully visible**

For the distribution of cooperation towards humans, when either partly or fully visible in the games with information, see Fig. ESM 3. We aimed to compare the cooperation of rational players across treatments. However the rarity of rational players meant that we only had limited samples to compare rational behaviour in the fully visible (N = 20) and partly visible (N = 23) treatments. Therefore for most analyses we combine the data from the partly visible and fully visible treatments, because in both treatments players are told their behavior will be visible, in some form, and have the potential to influence their groupmates' behavior.

For completeness, here we compare behavior in the partly and fully visible treatments anyway. Rational players significantly varied their cooperation across all three treatments (GzLM: $F_{2,100} = 9.0$, $P < 0.001$) whereas irrational players did not (GzLM: $F_{2,182} = 0.8$, $P = 0.430$). However, the difference in cooperation levels of both rational and irrational players in the partly and fully visible treatments was not significant. Rational players cooperated at 46% (9.2 ± 1.88 MU, N = 23) when partly visible and 54% (10.8 ± 1.96 MU, N = 20) when fully visible, a non-significant difference (GzLM: $F_{1,41} = 0.3$, $P = 0.568$). Irrational players cooperated at 46% (9.2 ± 1.47 MU, N = 25) when partly visible and 55% (11.1 ± 1.02 MU, N = 28) when fully visible, a non-significant difference (GzLM: $F_{1,51} = 1.2$, $P = 0.284$).

**Strategic cooperation & robustness checks**

The increased cooperation among rational players when visible was due to both more players cooperating (28/43, versus 18/60, FET: P = 0.0006) and significantly higher levels of cooperation from those that cooperated (mean cooperation of rational players that contributed > 0 when invisible = 55% [11.0 ±1.5 MU]; when visible = 76% [15.2 ±1.17 MU]; GzLM: $F_{1,44} = 4.6$, $P = 0.039$, $\beta = 1.0 \pm 0.45$, Fig. 2a). Among irrational players, the frequency of players that cooperated with humans (contributed > 0 MU) did not significantly differ between the visible and invisible treatments either (117/132 versus 50/53, Fisher's Exact Test: P = 0.2852).

For robustness we repeated the main analyses in the text on only those players that passed all the control questions (N = 94 of 288, 33%). We found the same qualitative results, with rational players nearly quadrupling their cooperating when visible (54% versus 14%), shifting their modal cooperation from 0% to 100%. In contrast, irrational players did not significantly adjust their behavior (48% versus 41% respectively). These differences lead to a significant interaction between how individuals behaved with the computer (rational versus irrational), and if their behavior was visible or not to their groupmates (GzLM rationality*visibility: $F_{1,90} = 5.4$, $P = 0.022$, ESM results Fig. ESM4).

We also confirmed that our rational players assigned to visible treatments were not merely of a more cooperative nature. We were able to do this, by examining their behavior in the prior game played with no information between rounds, before they were assigned to either the invisible or visible treatments in the later game with information. If the two groups of players had different levels of 'baseline' cooperation, then we would have detected it here, but we did not.

Across all six rounds the levels of cooperation were similar between those rational individuals that would later on be assigned to visible or invisible treatments (GzLM comparing cooperation in round 1 of the no-information game depending on what treatment player was assigned to later on in the game with information: $F_{1,101} = 1.2$, $P = 0.266$; R2: $F_{1,101} = 1.3$, $P = 0.251$; R3: $F_{1,101} = 0.0$, $P = 0.926$; R4: $F_{1,101} = 0.5$, $P = 0.482$; R5: $F_{1,101} = 0.1$, $P = 0.734$; R6: $F_{1,101} = 0.9$, $P = 0.336$).

We tested how rational and irrational players changed their cooperation when moving from the no-information game to the game with information, depending on if they had been assigned to the visible or invisible treatments. This way we compare their response to learning that their behavior will be visible, whilst controlling for how they as individuals behaved previously when they were invisible in the no-information game. This basically replicates our main analysis but controls for an individual's baseline cooperation. Overall, the increase in cooperation from the end of the no-information game to the start of the games with information depending on visibility was significantly greater for rational players (GLM, testing interaction between individual change in cooperation level ~ rationality*visibility: $F_{1,284} = 7.7$, $P = 0.006$). Specifically, rational players assigned to a visible treatment increased their cooperation from 15% (3.0 ±1.00 MU) to 50% (9.9 ±1.35 MU) whereas those assigned to the invisible treatment only increased from 10% (1.9 ±0.67 MU) to 17% (3.3 ±0.79 MU). This meant there was a significant interaction for rational players between visibility and round (end of first game or start of second game) (GzLMM: $F_{1,112} = 9.6$, $P = 0.003$). In contrast, this same interaction was not significant for irrational players, who increased from 43% (8.6 ±1.00 MU) to 51% (10.2 ±0.88 MU) when restarting into the visible treatment, and similarly from 43% (8.6 ±0.61 MU) to 47% (9.3 ±0.58 MU) if assigned to the invisible treatment (GzLMM: $F_{1,144} = 0.8$, $P = 0.369$).

Finally, we tested if *how much* more than 0 MU a player cooperated with a computer inversely predicted how strategic they were (thank you to a reviewer for this suggestion). For each player we calculated how much they increased their cooperation from the end of the no information game in stage 2 to the start of the information game in stage 3, depending on if assigned to a visible treatment or not. We then regressed this 'strategic response' against how much individuals cooperated with computers to test if there was a negative correlation, which may suggest that players that cooperate more with computers understand the game less or are less strategic than those that cooperate a little with computers. However there was no significant correlation for irrational players assigned to a visible treatment (linear model: $F_{1,51} = 1.5$, $P = 0.231$, $\beta = -0.13 \pm 0.108$).

**Strategic decline**

A perfect strategic cooperator would initially cooperate and then decrease their cooperation without ever increasing it. We therefore analyzed how individuals changed their cooperation over time (Table ESM1). We took an individual's cooperation for each of the six rounds in each of the no-information game and the information game, depending on the visibility of their behavior and if they were irrational or not (had previously cooperated with computers or not). We considered four mutually exclusive patterns of individual behavior: (1) a player's cooperation decreased at least once and never increased; (2) conversely, a player's cooperation increased at least once and never decreased; (3) a player's cooperation never changed; and (4) a player's cooperation both increased at least once and decreased at least once. The first pattern of behavior is most in keeping with ideas of strategic cooperation, as it requires a player to initially cooperate and then to decrease their cooperation without ever increasing it.

We found that the first behavioural pattern, whereby an individual initially cooperates and then decreases it without ever increasing it, was most common in rational players in the visible treatments. Overall, around 25% of rational players showed this behavior when visible (24% when potentially visible in the no-information game, and 26% when visible in the games with information). In contrast only 10% of rational players behaved this way when they were invisible. Although these differences are not significant on their own, possibly due to small sample sizes, they are when combined. Specifically, 24% of rational players (20/82) conformed to this pattern of initially cooperating and then reducing but never increasing their cooperation when potentially visible in the no information game, versus just 10% (2/21) when invisible (FET significance value = 0.2311). Likewise, in the games with information, 26% of rational players (11/43) behaved this way when visible and only 10% (6/60) when invisible (FET significance value = 0.0576). Pooling these together gives 25% (31/125) versus 10% (8/81, FET = 0.0064).

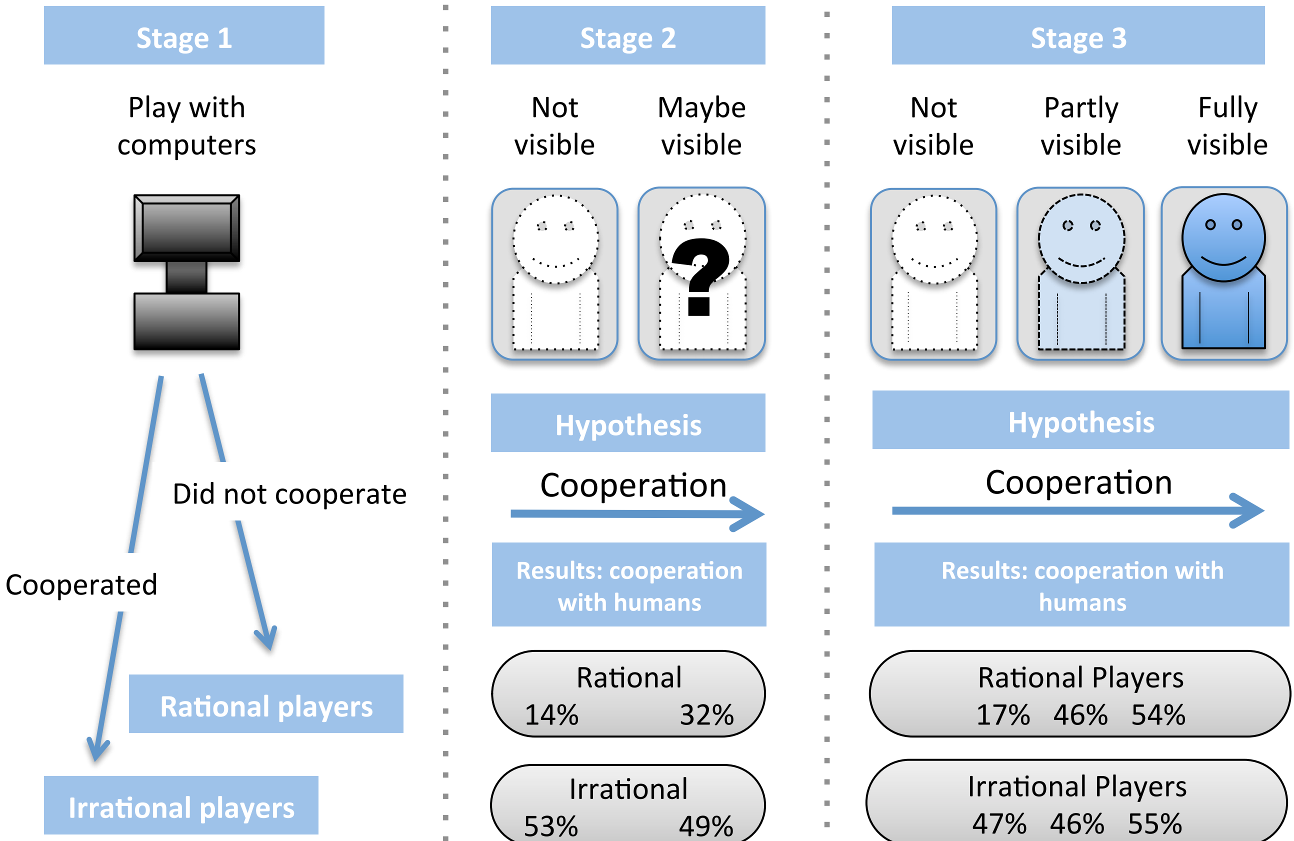
Consistent with the above, among rational players, we found they were significantly less constant when visible in the information game (16 v 58%, FET significance value = 0.0001) and less constant, although not significantly so, in the no-information game (50 v 62%, FET = 0.4633). Rational players, when visible or potentially visible, show more changes in behavior, suggesting strategic motivations. In contrast, we found that irrational players were very volatile, with 69-74% of them both increasing and decreasing their cooperation at least once, even in the no-information versions of the game.

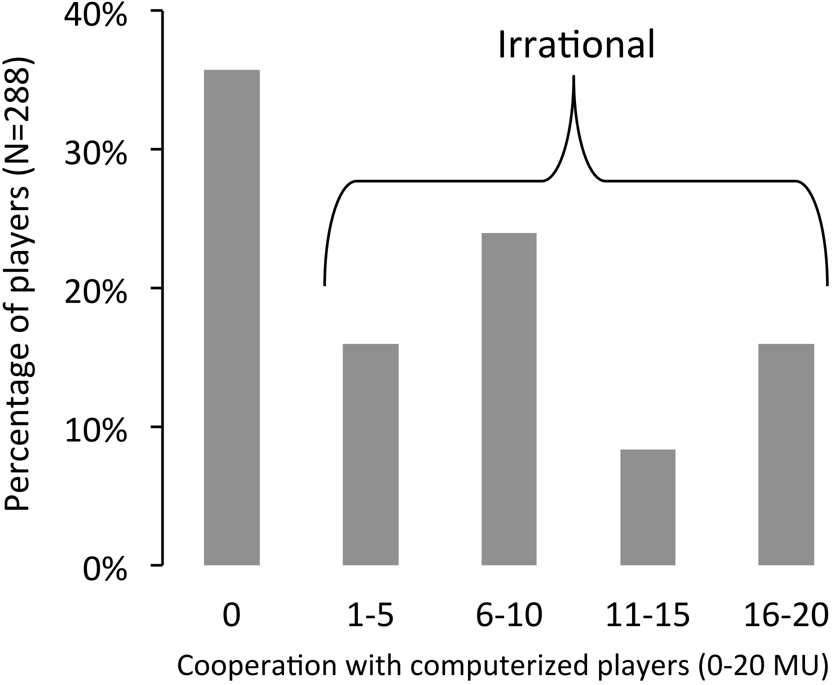
Finally we tested how many players could be classified as potential strategic cooperators, cooperating in the first round of the game (contribute > 0 MU) but then completely free riding in the final round (contribute 0 MU) (Table ESM1). Rational players were more likely to show this pattern of behavior in the visible treatments. In the no-information game, 32% (26/82) of rational players exhibited this behavior when partly visible, but only 10% (2/21) when invisible (FET: P = 0.0539). In the games with information, 49% (21/43) of rational players exhibited this behavior when visible, but only 18% (11/60) when invisible (FET: P = 0.0013).

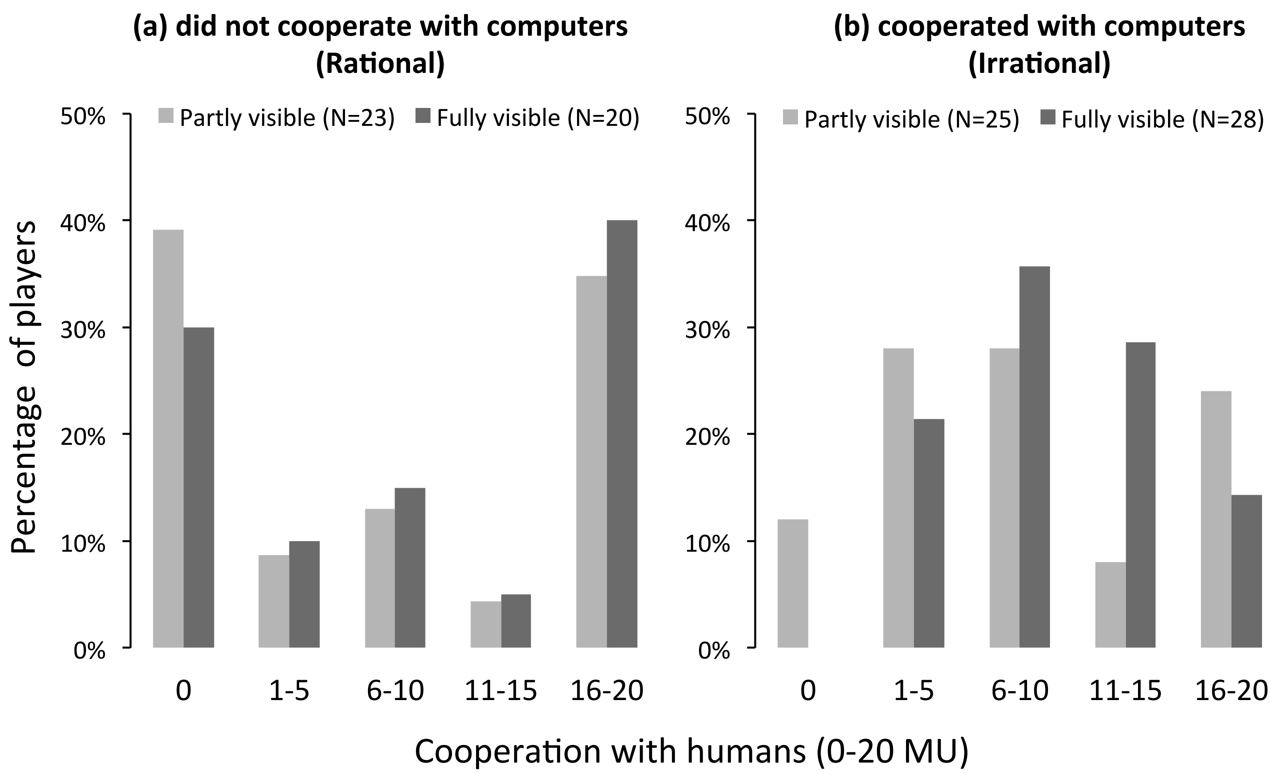
**Table ESM1:** An individual level analysis of patterns of cooperation over six rounds of the public goods game depending on rationality and treatment visibility.

| Rationality | Information | Visibility | N | Only decrease[1] | Never change[2] | Only increase[3] | Volatile[4] | PSC[5] |
|---|---|---|---|---|---|---|---|---|
| Rational | No-information | Invisible | 21 | 10% | 62% | 5% | 24% | 10%[6] |
| | | Ambiguous | 82 | 24% | 50% | 1% | 24% | 32% |
| | Information | Invisible | 60 | 10% | 58% | 2% | 30% | 18% |
| | | Visible | 43 | 26% | 16% | 2% | 56% | 49% |
| Irrational | No-information | Invisible | 51 | 4% | 16% | 6% | 75% | 12% |
| | | Ambiguous | 134 | 10% | 17% | 4% | 69% | 16% |
| | Information | Invisible | 132 | 22% | 8% | 1% | 70% | 35% |
| | | Visible | 53 | 17% | 8% | 6% | 70% | 36% |

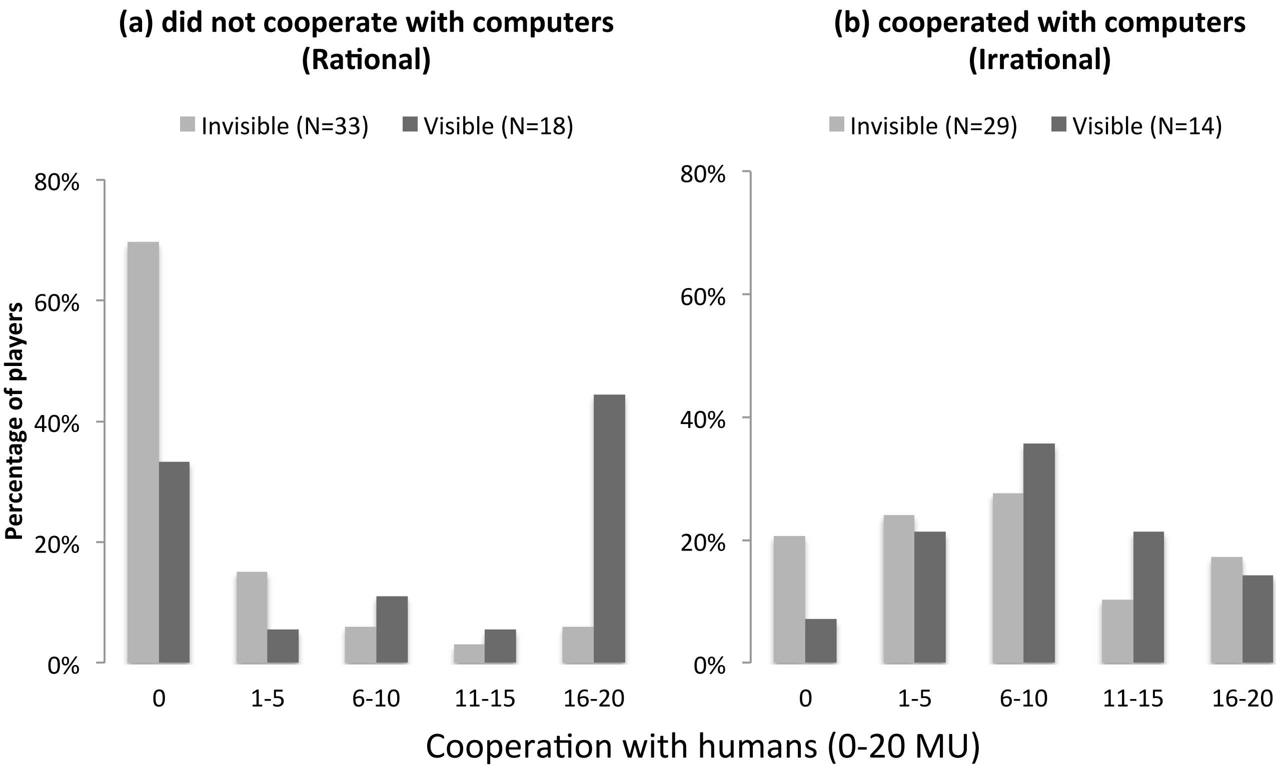
[1] A player's cooperation decreases at least once and never increases.
[2] A player's cooperation is constant.
[3] A player's cooperation increases at least once and never decreases.
[4] A player's cooperation both increases at least once and decreases at least once.
[5] Potential strategic cooperation: a player’s cooperation was greater than 0 in round 1 and equalled 0 in final round.
[6] Values within a row may sum to more than 100% due to rounding approximations.

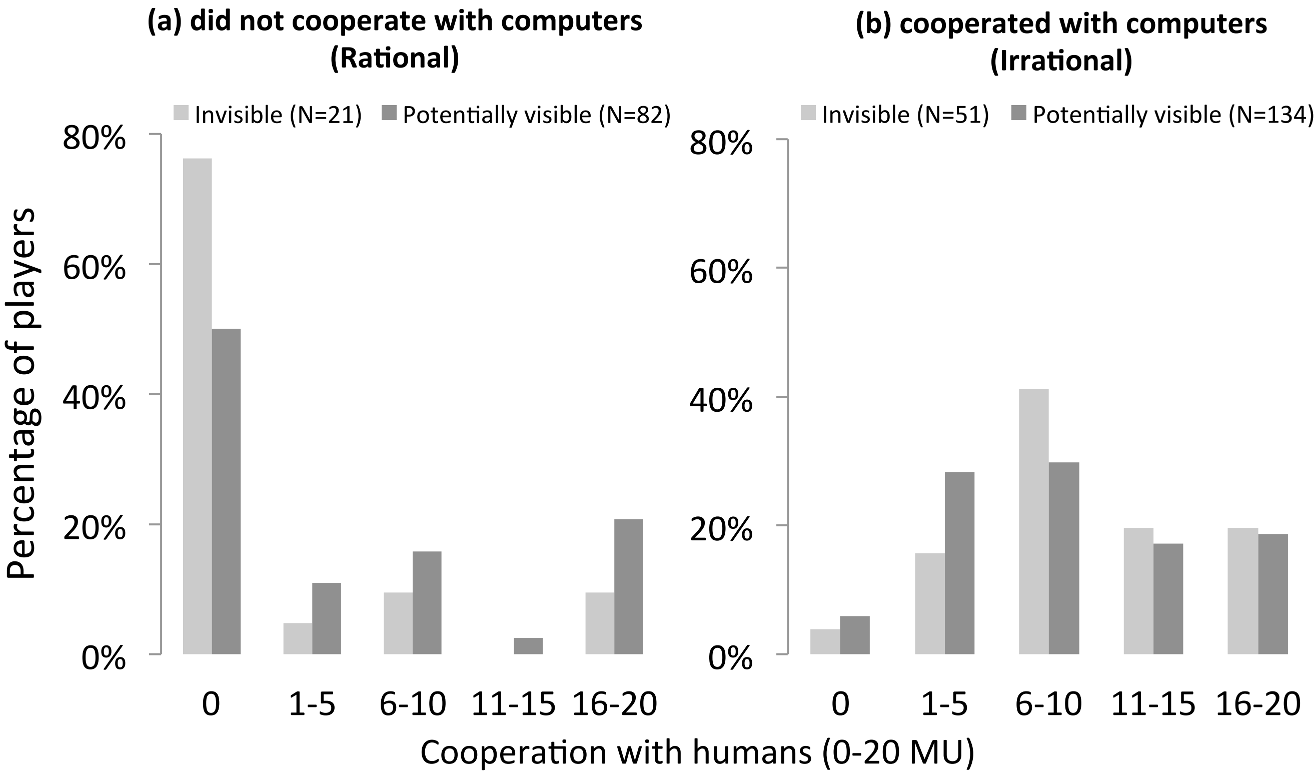

Stage 1
Play with computers
Did not cooperate
Cooperated
Rational players
Irrational players
Stage 2
Not visible
Maybe visible
Hypothesis
Cooperation
Results: cooperation with humans
Rational
14% 32%
Irrational
53% 49%
Stage 3
Not visible
Partly visible
Fully visible
Hypothesis
Cooperation
Results: cooperation with humans
Rational Players
17% 46% 54%
Irrational Players
47% 46% 55%


Irrational
40%
30%
20%
10%
0%
Percentage of players (N=288)
0
1-5
6-10
11-15
16-20
Cooperation with computerized players (0-20 MU)

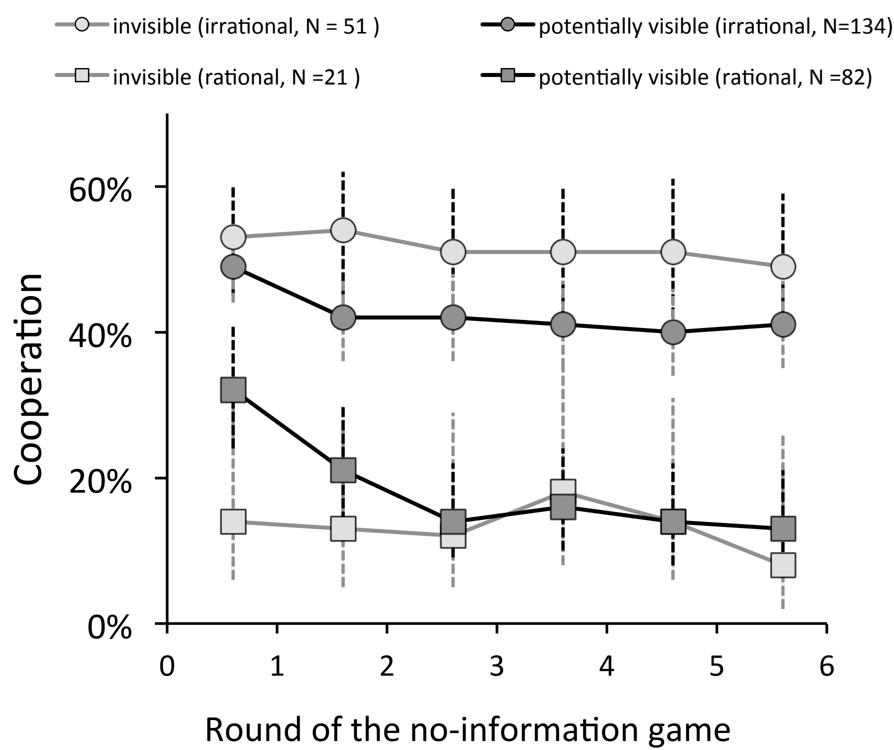


(a) did not cooperate with computers
(Rational)
(b) cooperated with computers
(Irrational)
Partly visible (N=23)
Fully visible (N=20)
Partly visible (N=25)
Fully visible (N=28)
Percentage of players
0%
10%
20%
30%
40%
50%
0
1-5
6-10
11-15
16-20
Cooperation with humans (0-20 MU)


(a) did not cooperate with computers
(Rational)
(b) cooperated with computers
(Irrational)
Invisible (N=33)
Visible (N=18)
Invisible (N=29)
Visible (N=14)
Percentage of players
80%
60%
40%
20%
0%
0
1-5
6-10
11-15
16-20
Cooperation with humans (0-20 MU)



(a) did not cooperate with computers (Rational)
Invisible (N=21)
Potentially visible (N=82)
(b) cooperated with computers (Irrational)
Invisible (N=51)
Potentially visible (N=134)
Percentage of players
80%
60%
40%
20%
0%
0
1-5
6-10
11-15
16-20
Cooperation with humans (0-20 MU)



invisible (irrational, N = 51 )
potentially visible (irrational, N=134)
invisible (rational, N =21 )
potentially visible (rational, N =82)
60%
40%
20%
0%
Cooperation
0
1
2
3
4
5
6
Round of the no-information game