HUMANS *VS.* BOTS: INVESTIGATING MODELS OF BEHAVIOR IN THE ITERATED PRISONER'S DILEMMA

Samarth Swarup
Mark G. Orr
Gizem Korkmaz

Biocomplexity Institute
University of Virginia
Charlottesville, VA, USA
{swarup,mo6xj,gkorkmaz}@virginia.edu

Kiran Lakkaraju

Sandia National Laboratories
Albuquerque, NM, USA
klakkar@sandia.gov

ABSTRACT

We present an approach to studying human behavior in the Iterated Prisoner's Dilemma (IPD) game, where human participants play against software bots in an online environment. This setting allows precise control of the strategies faced by the human subjects. Our goal is to test models of human behavior and the means by which cooperation can be promoted. Results from our experiments with subjects from Amazon Mechanical Turk show that the leading models of human behavior in the IPD are incomplete, as they can explain our data only partially. We propose a new model of behavior, Majority Wins, which provides a better fit to the data, which we test and demonstrate through a simulation.

Keywords: game theory, human behavior, iterated prisoner's dilemma, moody conditional cooperation, aspiration learning, majority wins.

1 INTRODUCTION

Situations of cooperation and conflict are ubiquitous in social interactions. Social dilemma games formalize many of these situations using game theoretical representations (Rand and Nowak 2013). These games are generally used to analyze situations where an individual can attain an immediate payoff by acting selfishly but, if everyone acts selfishly, the group suffers a loss. A motivating example is interactions in social media. Constructive, empathetic interactions with people who oppose one's views require effort, whereas antagonistic "point-scoring" interactions can be easy. However, antagonistic interactions lead to increasing polarization over time (Mason 2015) and society as a whole suffers. This kind of interaction can be codified as a prisoner's dilemma.

The Prisoner's Dilemma (PD) is the most-studied example of a social dilemma game (see Section 2). It offers a succinct model of cooperation and conflict in the context of a two-player interaction. Each player can either *cooperate* or *defect*. If both players (moving simultaneously) cooperate, they each get a reward R. If they both defect, they each get a (lower) reward P. If one cooperates and the other defects, the cooperator gets a reward S and the defector gets a reward T. In the prisoner's dilemma, we have the relation $T > R > P > S$. The paradox of the prisoner's dilemma is that while both players would be better of cooperating, the only Nash equilibrium is that both defect.

In the iterated version of the PD (the IPD), two players play the PD repeatedly against each other. Despite its simplicity, the IPD has a very rich strategy space which has been extensively studied both theoretically (Press and Dyson 2012, Hilbe et al. 2017) and through computational simulation (Mathieu and Delahaye 2017). There have also been several studies of humans playing the IPD, that aim to understand how people make decisions in situations of cooperation and conflict. These studies have shown that people exhibit "moody conditional cooperation" (MCC) while playing the IPD (Grujić et al. 2014). This means that the probability that a person will cooperate is a function of two factors: their "mood", which is simply the action they have taken in the last step, and the proportion of *cooperate* actions taken by their opponents in the recent past. This phenomenological regularity is explained in more detail in Section 2. In an attempt to explain the MCC phenomenon, it has been suggested that people are doing "aspiration learning", which is a form of reinforcement learning (Ezaki et al. 2016) where an agent reinforces actions which give a reward that is above a threshold (its aspiration level) and penalizes actions for which the reward falls below the aspiration level. We evaluate this model in Section 6.1.

In the current work, we present a new approach to assessing human behavior in social dilemma games by having human subjects play the IPD against multiple opponents in an online environment. Unknown to the subjects, their opponents are all bots with precisely specified strategies. This allows us to test whether humans play in the same way against different strategies. We tested human behavior against bots that play randomly, bots that play "tit-for-tat" (TFT), and bots that play "reverse tit-for-tat" (RTFT). These are explained further in Section 3. We find that our data only match the MCC pattern in the random case, whereas human behavior shows different patterns in the TFT and RTFT cases. We then create a simulation to test if the aspiration learning model can match our observations. We find that it cannot for the TFT and RTFT cases. Finally, we propose a new model, called Majority Wins, which can match all of our observed data.

In the following, we begin by providing some background of previous work in modeling human behavior in the IPD (Section 2). Then we describe our human subject experiments (Section 3). Results and prior models are described in 4. That is followed by the simulation that we use to test the aspiration learning model (Sections 5 and 6.1). Then we present the Majority Wins model and evaluate it using the same simulation (Section 6.2). Our data set from the human subject experiments and our simulation code (in Java) are available through GitHub at https://github.com/samarthswarup/IPD_network_sim.

2 PREVIOUS WORK

There is much theoretical and computational work focused on understanding the strategy space and finding winning strategies in the IPD; see (Press and Dyson 2012, Hilbe et al. 2017, Mathieu and Delahaye 2017) and references therein. This is not the focus of our work. We are interested in investigating how humans make decisions in the IPD, with the larger goal of understanding human decision-making in networked contexts. Therefore, we focus our remarks in this section on human subject experiments with the IPD.

Earlier work has shown the relevance of working memory constraints (Milinski and Wedekind 1998), which will be relevant for us in Section 6.2. More recently, Grujić et al. (2014) did a meta-analysis of multiple experiments with human subjects to show the same pattern, known as moody conditional cooperation, emerges in each case, as explained in Section 1. This was followed by attempts to explain this phenomenon, and the aspiration learning model was proposed as potential solution (Ezaki et al. 2016). However, all the experimental data to which this model has been applied have been experiments with human subjects only, which do not allow any kind of control of the strategies faced by the subjects. To do a more precise and rigorous investigation of human behavior in the IPD, we did experiments with single subjects playing against a population of bots, as we describe next.

3 HUMAN SUBJECT EXPERIMENTS

An online experiment was conducted using the Controlled, Large, Online Social Experimentation (CLOSE) platform (Lakkaraju et al. 2015). CLOSE was built to help researchers build and deploy online experiments, including experiments that can study social influence among subjects. Subjects were recruited and rewarded through Amazon Mechanical Turk (AMT) an online marketplace for tasks. Workers can login and perform tasks posted by requestors. Requestors provide payment to the workers through the AMT interface.

AMT has become a popular site to deploy tasks, called "Human Intelligence Tasks (HITS)", that are easy for humans but difficult for machine. Increasingly, AMT and other sites like it are being used as a way to recruit subjects for experiments that were previously conducted in the lab. AMT has several benefits, it is inexpensive, provides a diverse subject pool (Lakkaraju 2015, Paolacci and Chandler 2014), and is fast. Replication of results between the lab and AMT (Crump et al. 2013, Berinsky et al. 2012) suggest that AMT can be an useful proxy for laboratory experiments.

3.1 Experiment Design

This experiment was designed to identify the strategies employed by humans when playing the Iterated Prisoners Dilemma. Subjects were told they would be playing against bots or other players (though all subjects were playing against bots). Each subject was assigned a group with eight bots. A subject played 60 rounds (the first 10 were trial rounds where the score did not count). In each round a subject was paired with a bot and then had to choose between two actions, "A" and "B" within 30 seconds. If no action was chosen within 30 seconds, the subjects received 0 points for that round.

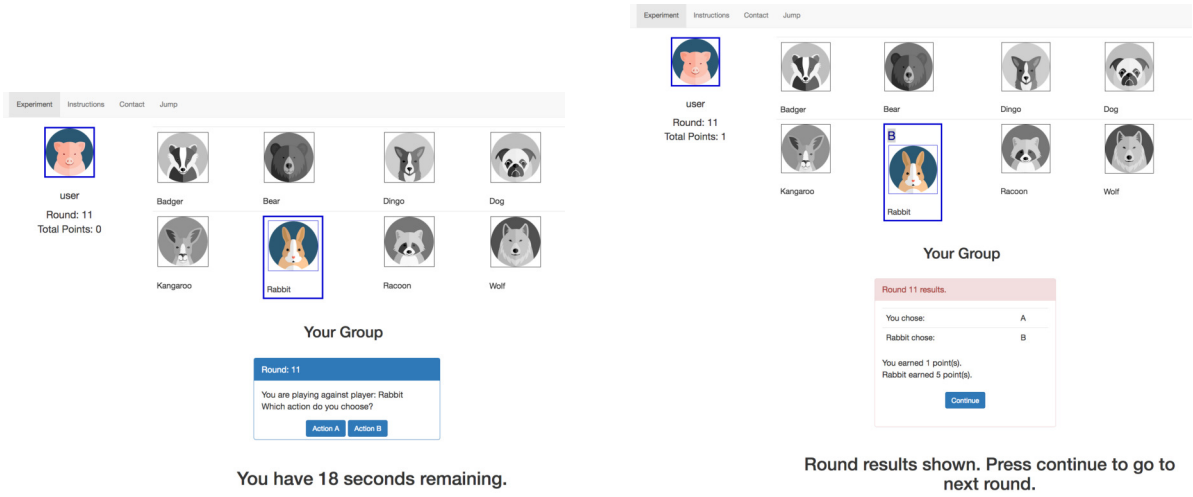
Actions were anonymized; cooperate and defect were replaced with "A" and "B" to remove any potential emotional reactions to the original names. Subjects were merely told they were playing a game and shown the payoff matrix in Figure 1. In each round, the subject played against one randomly-chosen bot. The points earned by the subject accumulated over all 50 non-trial rounds. After 60 rounds the experiment stopped and the subject was notified of how much they earned. Subjects earned $1.50 (USD) for completing the experiment and an extra $.01 for every point they could earn. The maximum they could earn was an extra $2.50.

Figure 1: Payoff matrix shown to subjects.

All bots played a single strategy which depended upon the condition the subject was (uniformly randomly) assigned to:

Control Condition: Bots would randomly pick between A and B.

Tit-for-Tat Condition: Bots will play tit-for-tat. The bot will play the same action as the one the subject played against the bot the last time they played. If they have not played before, the bot will choose randomly.



(a) Decision-making screen for subject.

(b) Response screen for subject.

Figure 2: The decision-making screen presented to the subject.

Reverse Tit-for-Tat Condition: Exactly the same as tit-for-tat, except that bots will play the opposite action from the subject from the last time they played (Nachbar 1992). Figure 2a shows the decision making screen presented to the subject. The avatars of the eight bots are shown at the top. For every subject a random set of unique bots are chosen. The current opponent is highlighted and the other bots are in grey.

The subject can choose an avatar to represent themselves, in this case it is a pig. Total current points are always displayed, along with the payoff matrix. Figure 2b shows the response phase of the round. The decision of the opponent is highlighted above them. The subject's choice and reward is summarized in the box below. The subject can view their response to a round as long as they wish.

4 RESULTS

The marginal mean proportion cooperate by experimental condition were as follows (standard error in parentheses): Control, 0.30 (0.09); TFT, 0.65 (0.08); RTFT, 0.20 (0.08). This broadly matches expectations. TFT is known to promote cooperation and, conversely, RTFT is expected to promote defection. The 30% probability of cooperation against the random strategy also matches the overall level of cooperation seen in other human subject experiments (Grujić et al. 2014).

4.1 Analysis of Moody Conditional Cooperation

Grujić et al. (2014) develop the "Moody Conditional Cooperation" strategy and show that it fits the data from three different human subject IPD experiments. The essence of the MCC strategy is that a player's actions (i.e., either cooperate or defect) are based on their opponent's recent decisions as well as the subject's previous decision. The MCC was identified in games of humans playing other humans; so no control of the strategies of subjects was possible. Our goal is to study how humans play IPD in response to particular strategies.

Our experimental context varies in one other way. Unlike the experiments studied by Grujić et al. (2014), in our case, subjects only play a single other neighbor per round. We therefore use a window approach to calculate the neighbor decisions. Let $o_0, \ldots, o_n$ be the decisions made by the opponent bots in round 0 to

n. Let $x_0, \ldots, x_n$ be the decisions made by a single subject in rounds 0 to n. Let w be the window size – the number of decisions before the decision at time t that we consider. The fraction of cooperating neighbors at time t with window size w is calculated as:

$$c(t,w) = \frac{\Gamma(o_{t-1}, o_{t-2}, \ldots, o_{t-w})}{w},$$

where $\Gamma(.)$ is the number of cooperate decisions made by the opponents. The "moody probability of cooperating" is the probability of a subject choosing *cooperate* at time step t given a fraction of cooperating neighbors $c(t,w)$ and previous decision x_{t-1}:

$$P_A(t) := P(x_t = \text{“}cooperate\text{”} | c(t,w), x_{t-1}).$$

We calculate P_A using the experimental data we gathered. We will focus our study on $w = 5$. Results are similar with other values of w, though sample size gets smaller. Figure 3 shows the results.

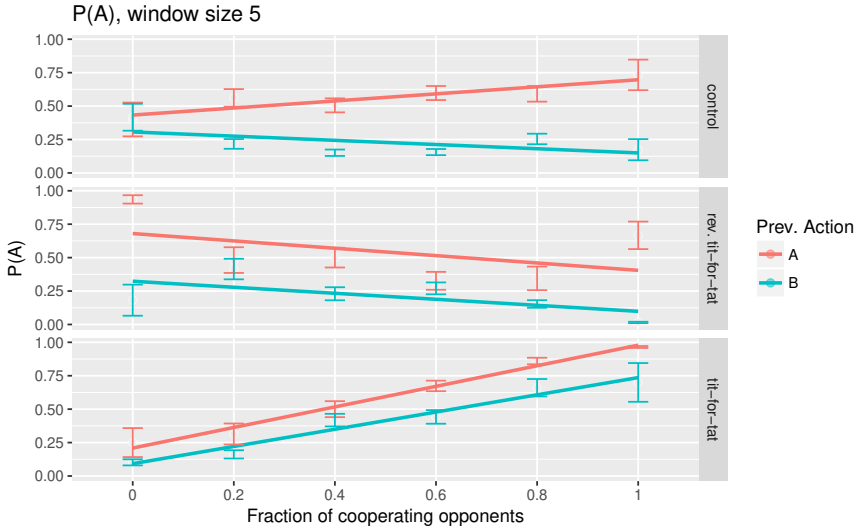


Figure 3: Probability of cooperating, window size = 5, grouped by condition. Error bars are standard deviations of a binomial distribution $\sqrt{p(1-p)/n}$.

The three subgraphs are for different conditions (all bots playing the control strategy, all bots playing the reverse tit-for-tat strategy, and all bots playing the tit-for-tat strategy). The two lines correspond to the previous decision made by the subject. The x-axis is the fraction of opponents that cooperated in the past 5 rounds. The y-axis is the probability of a subject playing "A".

The results for the control condition exhibit moody conditional cooperation as described by Grujić et al. (2014). As the fraction of cooperating opponents increases, a subject was more likely to cooperate if they had cooperated in the timestep before. Conversely, a subject was more likely to defect if they had defected in the time step before. This aligns with the MCC model.

However, MCC does not hold in the reverse tit-for-tat and tit-for-tat condition. Interestingly in the reverse tit-for-tat behavior is conditioned on past behavior of the opponent, however the past action does not cause a divergence in the probability to cooperate. In fact, the P(A) decreases nearly the same as the fraction of cooperating opponents increases. The pattern is the same for the tit-for-tat condition, except the probability

to cooperate increases. In both of these conditions, MCC does not hold. Thus the MCC model is not a complete description of human behavior in the IPD.

4.2 The Aspiration Learning Model

It has been hypothesized that moody conditional cooperation emerges from a deeper underlying model known as aspiration learning (Ezaki et al. 2016). This model updates the current probability of cooperation as follows:

$$p_t = \begin{cases} p_{t-1} + (1-p_{t-1})s_{t-1} & \text{if } x_{t-1} = A, s_{t-1} \geq 0, \\ p_{t-1} + p_{t-1}s_{t-1} & \text{if } x_{t-1} = A, s_{t-1} < 0, \\ p_{t-1} - p_{t-1}s_{t-1} & \text{if } x_{t-1} = B, s_{t-1} \geq 0, \\ p_{t-1} - (1-p_{t-1})s_{t-1} & \text{if } x_{t-1} = B, s_{t-1} < 0, \end{cases}$$

where p_t is shorthand for $P(x_t = \text{“}cooperate\text{”})$, and s_{t-1} is known as the *stimulus* at the previous time step. The stimulus is calculated as,

$$s_{t-1} = \tanh\left[\beta(r_{t-1} - \theta)\right], \tag{1}$$

where r_{t-1} is the reward at time $t-1$ and θ is the aspiration level (or threshold) for acceptable reward. β is a parameter that controls the slope of the tanh function and thereby controls how sharply s_{t-1} changes as the reward crosses the threshold. It is also assumed that a player can “mis-implement”, or flip, its action with probability ε. This means that the effective probability of cooperation of an aspiration learning player is $\tilde{p}_t = p_t(1-\varepsilon) + (1-p_t)\varepsilon$.

The aspiration learning model is based on the Bush-Mosteller model of reinforcement learning (Macy and Flache 2002), and is essentially a stochastic version of the Win Stay Lose Shift (WSLS) strategy (Nowak and Sigmund 1993). At each time step, it promotes the probability of the action based on its success. For example, if the aspiration level is chosen such that $T > R > \theta > P > S$, then s_{t-1} will be positive for the cases BA and AA, and negative for the cases BB and AB, where the first action in each pair is the subject's, and the second the opponent's (in a two-player game). Thus, as long as the opponent keeps cooperating, the aspiration learning (AL) agent will continue doing whatever action it has been doing.

The flipping probability ε allows exploration in the reinforcement learning sense. This can help the AL player improve its payoff. For example, if the players are doing AA, the AL player will continue increasing its probability of generating A, but with probability ε is will flip to B, allowing it to discover a higher payoff. It will then reduce the probability of generating A in successive rounds. This allows the AL player to do better than strategies such as TFT in certain cases, e.g., against the all-cooperate strategy. It also allows the AL player to escape mutual-defect scenarios against TFT players.

In order to examine whether aspiration learning can explain our experimental data, we created a simulation described below.

5 SIMULATION

We created a simulated version of our human subject experiments, so that we could assign various strategies to the focal agent (corresponding to the human subject) and evaluate its behavior against other strategies.

The simulation can accept a network structure for the agent interactions and implements multiple strategies of individual behavior, including TFT, RTFT, and aspiration learning. At each timestep, the simulation randomly chooses an agent from the population and then randomly chooses one of its neighbors on the

network. The two chosen agents then play one round of the PD and can update their payoff and strategy based on the outcome. This process is repeated for a specified number of rounds and the agent IDs, actions, and other parameters are output to file. It is, thus, a simple and general simulation model, and can easily be altered for other strategies and other games.

We use this platform to test the aspiration learning model against our human subject experiments as follows. We created a star network with the focal agent in the center, with eight neighbors, to mimic our experimental condition. We ran three different simulations with the eight neighbors playing the random strategy, TFT, and RTFT respectively. The focal agent followed the aspiration learning model in each case. Results are presented below.

6 SIMULATION RESULTS

6.1 The Aspiration Learning Model

In these simulations, we set the aspiration level to 1.5, and we have $T = 5, R = 4, P = 2, S = 1$ to match the values used in our human subject experiments. The agent maintains the average of the reward obtained over the last five rounds as the value of r_{t-1} in Equation 1. After every 60 rounds, we clear the focal agent's history (i.e., it starts computing the award reward from scratch) to simulate the appearance of a new subject. The overall simulation was run for 100,000 time steps so that we could get good probability estimates. We set $\varepsilon = 0.2$ and $\beta = 0.4$, following Ezaki et al. (2016).

Figure 4 shows the outcome of the simulation for the case where the opponents play the random strategy with $P(A) = 0.5$. We see that a clear MCC pattern emerges and looks very similar to our experimental results for the control condition.

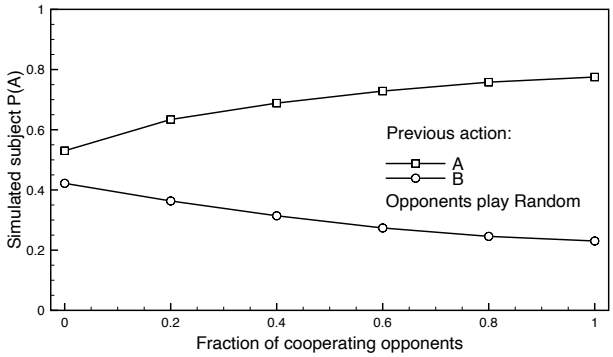
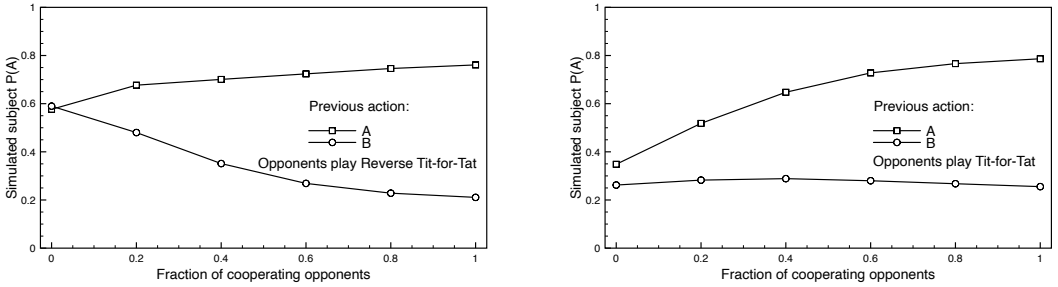


Figure 4: Simulated Moody Conditional Cooperation effect with opponents playing the random strategy and the focal agent using aspiration-based reinforcement learning.

We then repeated the simulation for the RTFT and TFT cases, with exactly the same parameters for the focal agent. These results are shown in Figure 5. We see that in these cases, aspiration learning once again produces MCC-like patterns, which are at odds with our experimental observations in Figure 3. In particular, the probability of cooperation when the focal agent has just cooperated in the RTFT case (this is the upper line in Figure 5a, marked with squares) increases slightly with increasing fraction of cooperating neighbors, whereas our experimental results show a decrease. Similarly, the probability of cooperation when the focal agent has just defected in the TFT case (this is the lower line in Figure 5b, marked with circles) stays about constant with increasing fraction of cooperating neighbors, whereas our experimental results show an increase.



(a) Reverse-Tit-for-Tat opponents.

(b) Tit-for-Tat opponents.

Figure 5: Simulation results with opponents playing tit-for-tat (a) and reverse tit-for-tat (b), with the focal agent using aspiration-based reinforcement learning as before. These results are very different from the empirical data.

This difference cannot be reconciled using the aspiration learning model. To see why, recall that aspiration learning reinforces successful actions. Consider the TFT case where $c(t,w) = 1$ and the focal agent has just defected. This corresponds to the right-most point on the lower line in Figure 5b. Two cases are possible: either the focal agent has been cooperating and just switched to defection by chance in the last step, or has been defecting for more than one time step. In the former case, the agent sees an increase in reward by switching to defection and will reinforce that action. In the latter case, the agent is already defecting and will further reinforce that action. This is possible against the TFT neighbors because the neighbors keep their own history of interactions with the focal agent, so we can see a sequence of BA interactions because the focal agent is playing different opponents in each round.

In either case, therefore, the agent is likely to produce the defect action again, which is what we see in Figure 5b. The probability of cooperation is just above 0.2 (which is due the flipping probability ε). In the human subject experiment results, we see that this probability is above 0.7 (lowest subgraph in Figure 3). The aspiration learning model cannot reproduce this experimentally-observed result. A similar argument applies to the RTFT case for the downward trend in the experimental data for the probability of cooperation after the subject has just cooperated. Aspiration learning does not show this trend.

Thus the aspiration learning model cannot completely account for human behavior in the IPD.

6.2 The Majority Wins Model

We introduce a new model to account for the experimental data. We call this the Majority Wins (MW) model. It is an instance of a memory-n model (Hilbe et al. 2017), where an agent keeps track of the last n games it has played. In the MW model, an *AA* or *BA* game (i.e., where the opponent cooperates and the agent either cooperates or defects) is counted as a "win". Let c_A be the count of *AA* games in the last n, and let c_B be the count of *BA* games in the last n, then the action chosen by the agent at time t is,

$$x_t = \begin{cases} A & \text{if } c_A \geq c_B, \\ B & \text{if } c_B > c_A. \end{cases}$$

We include a default probability of cooperation, $p_{def} = 0.3$, for the case $c_A = c_B = 0$. The simulation was run in the same way as before. The memory of the agent is cleared after every 60 rounds, to simulate the introduction of a new subject into the experiment. The length of the memory was set to 5. Similar results are obtained with other choices. Figure 8 shows that the MW model reproduces the MCC effect. Note that the error bars are very tight because of the large number of simulated rounds. We also show the trendlines in blue

to emphasize that the probability of cooperation generally goes up with increasing number of cooperators when the agent has just cooperated, and goes down in the other case.

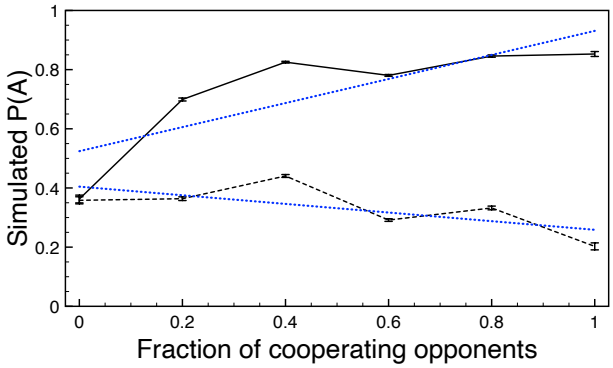
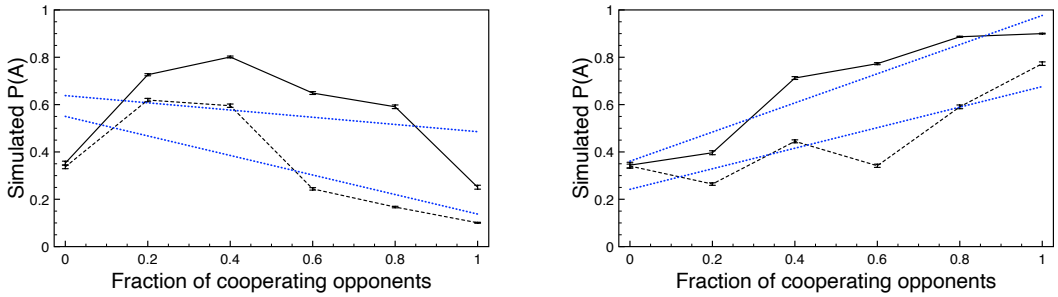


Figure 6: Simulated Moody Conditional Cooperation effect with opponents playing the random strategy and the focal agent using the majority wins model. Error bars are standard deviations of a binomial distribution $\sqrt{p(1-p)/n}$.

Next we ran the simulation for the RTFT and TFT cases. These results are shown in Figure 7. We see that these broadly match the results from the human subject experiments (Figure 3). The trendlines are now pointing in the right direction. In particular, the probability of cooperation when the agent has just cooperated is decreasing for the RTFT condition, and the probability of cooperation when the agent has just defected in increasing for the TFT condition. A fine detail, worth noting, is that the probability of cooperation in each case (including random) exhibits a zig-zag pattern that is also present in the experimental data (compare with the points in Figure 3), though it is more pronounced in the simulation.

The only places where the MW model fails to match the data is for the probability of cooperation when the fraction of cooperating neighbors is zero, for the RTFT case. The experimental data show a clear divergence, with a high probability of cooperating when the agent has just cooperated and a low probability when the agent has just defected. In the MW model, we see a low probability of cooperating even if the agent has just cooperated (because it uses the default probability in this case). The experimental data also show a higher than expected probability of cooperation at the other end for this case, when the fraction of cooperating neighbors is 1, but the sample size is very small in this case.



(a) Reverse-Tit-for-Tat.

(b) Tit-for-Tat.

Figure 7: The Majority Wins model: Simulated probability of cooperating, window size = 5, grouped by condition. Error bars are standard deviations of a binomial distribution $\sqrt{p(1-p)/n}$.

Finally, we tested the model for the condition that all agents play the majority wins strategy. This corresponds to the case of human subject experiments, such as those modeled by Grujić et al. (2014), in which there are no bots. In this case also, the MCC pattern is reproduced, as shown in Figure 8, which is further evidence for the correctness of the MW model.

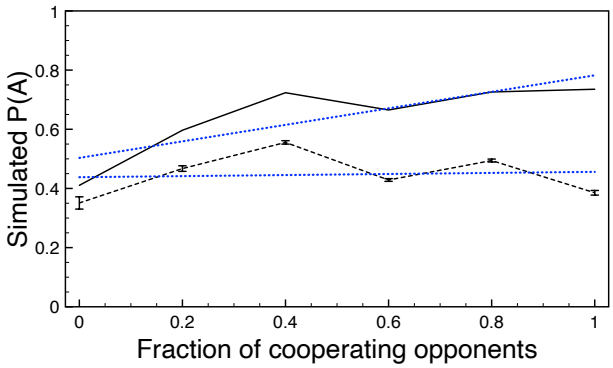


Figure 8: Simulated Moody Conditional Cooperation effect with all agents playing the majority wins strategy. Error bars are standard deviations of a binomial distribution $\sqrt{p(1-p)/n}$.

7 CONCLUSION

To summarize, we have shown that the two main existing models of human behavior in the IPD are incomplete. We see that the moody conditional cooperation phenomenon is only shown by humans when playing against the random strategy and that the aspiration learning model also only models human data successfully for our experiments against random bots. Both the models fail to reproduce the observed changes in behavior when humans are playing against TFT and RTFT bots.

It is important to note that, subjectively, it is not easy to figure out the strategy that the bots are playing because there are eight bots who play against the human subject in random order and each use their own histories of interaction to inform their action selection. If the human subject were playing against a single TFT bot, say, then it might be reasonable to believe that the subject can determine the bot's strategy and respond accordingly. If that were the case in our experiments, we would expect the humans to use the always-cooperate strategy against TFT bots and the always-defect strategy against the RTFT bots. In fact, several subjects are eventually able to figure out this strategy in the RTFT case (but not in the TFT case, interestingly). However, the initial exploration of strategy space they do results in the observed pattern of probabilities we see in the experiments. There is further work to be done to understand the mechanism by which strategy convergence happens in one case but not the other. We hypothesize that this may be due to a sensitive dependence on initial conditions, i.e., whether the human subject starts out by cooperating or defecting, and the random choices initially made by the opponents.

We have introduced a new model of human behavior in the IPD, which we term the Majority Wins model. We showed that this model provides a better qualitative match for all the observed data. However, further testing should be done against other common strategies. We believe that the MCC pattern emerges whenever human subjects are faced with reasonably complex strategies, i.e., ones which are hard for the subject to predict (random falls in this category, as does MW, but not TFT or RTFT). This hypothesis can be tested if we can create a single-parameter family of strategies where the parameter controls the complexity of the strategy. This is an important direction for future research to understand human behavior in this domain.

Finally, we offer some comments on the broader implications of this work. Our subject population did not have prior exposure to the IPD, as is generally the case in experiments of this type. Thus, the strategies they

employ are presumably general behavioral strategies that might give us some insight into broad mechanisms of decision-making in situations of cooperation and conflict. It is interesting, therefore, that the Majority Wins strategy is a very poor strategy for the IPD. This is because "success" in any single game of the PD depends entirely on the opponent's choice; if the opponent cooperates, the subject wins, and if the opponent defects, the subject loses. In the iterated setting, it thus makes sense to try to ascertain what is likely to make the opponent cooperate in the *next* round, since both players have to move simultaneously. MW, on the other hand, counts the actions that were "successful" in the same round, even though the action of the subject had no influence on the action of the opponent in the same round.

What can we conclude from this? To us, the most straightforward conclusion from this is that humans have not evolved behavioral strategies to deal with IPD-type situations. Another closely related possibility is that the situations humans (and our ancestors) did generally encounter evolutionarily did not include IPD-type situations. However, there is one important caveat: it could be the case that humans are adapted to deal with IPD-type situations and that adaptation is reputation (which enables TFT-type strategies). However, in situations where it is hard to keep track of reputation, such as when playing against multiple opponents in random order (or at the same time as in the prior experiments), MW might be a fallback strategy.

Coming back to our motivating example about interactions in social media, this suggests that one reason why we might find it hard to maintain constructive interaction over time is that the platforms create IPD-type situations to which we are ill-adapted. Social media interactions are between large numbers of people, and it is hard to keep track of "reputation," though various platforms have tried to implement some version of it (e.g., karma on Reddit). This is precisely the kind of situation in which our primary mechanism for ensuring constructive interaction fails and our fallback strategy (MW) is not particularly useful. This suggests that the main challenge to be solved in order to improve interactions on social media is to come up with a better way of tracking reputation, which cannot easily be gamed.

ACKNOWLEDGMENTS

This material is based upon work supported in part by the National Science Foundation (NSF) under Grant No. SMA-1520359, by the Defense Advanced Research Projects Agency (DARPA), via the Air Force Research Laboratory (AFRL) Contract No. FA8650-19-C-7923, and by Air Force Office of Scientific Research under award number FA9550-17-1-0378. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views, official policies or endorsements, either expressed or implied, of NSF, DARPA, AFRL, AFOSR, USAF, or the U.S. Government.

REFERENCES

Berinsky, A. J., G. A. Huber, and G. S. Lenz. 2012, July. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk". *Political Analysis* vol. 20 (3), pp. 351–368.

Crump, M. J. C., J. V. McDonnell, and T. M. Gureckis. 2013, March. "Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research". *PLoS ONE* vol. 8 (3), pp. e57410.

Ezaki, T., Y. Horita, M. Takezawa, and N. Masuda. 2016, 07. "Reinforcement Learning Explains Conditional Cooperation and Its Moody Cousin". *PLOS Computational Biology* vol. 12 (7), pp. 1–13.

Grujić, J., C. Gracia-Lázaro, M. Milinski, D. Semmann, A. Traulsen, J. A. Cuesta, Y. Moreno, and A. Sánchez. 2014. "A Comparative Analysis of Spatial Prisoner's Dilemma Experiments: Conditional Cooperation and Payoff Irrelevance". *Sci Rep* vol. 4, pp. 4615.

Hilbe, C., L. A. Martinez-Vaquero, K. Chatterjee, and M. A. Nowak. 2017. "Memory-n strategies of direct reciprocity". *Proceedings of the National Academy of Sciences* vol. 114 (18), pp. 4715–4720.

Lakkaraju, K. 2015, March. "A Study of Daily Sample Composition on Amazon Mechanical Turk". In *Social Computing, Behavioral-Cultural Modeling, and Prediction*, edited by N. Agarwal, K. Xu, and N. Osgood, pp. 333–338. Springer International Publishing.

Lakkaraju, K., B. Medina, A. N. Rogers, D. M. Trumbo, A. Speed, and J. T. McClain. 2015, March. "The Controlled, Large Online Social Experimentation Platform (CLOSE)". In *Social Computing, Behavioral-Cultural Modeling, and Prediction*, edited by N. Agarwal, K. Xu, and N. Osgood, pp. 339–344. Springer International Publishing.

Macy, M. W., and A. Flache. 2002. "Learning dynamics in social dilemmas". *Proceedings of the National Academy of Sciences* vol. 99 (suppl 3), pp. 7229–7236.

Mason, L. 2015. ""I Disrespectfully Agree": The Differential Effects of Partisan Sorting on Social and Issue Polarization". *American Journal of Political Science* vol. 59 (1), pp. 128–145.

Mathieu, P., and J.-P. Delahaye. 2017. "New Winning Strategies for the Iterated Prisoner's Dilemma". *Journal of Artificial Societies and Social Simulation* vol. 20 (4), pp. 12.

Milinski, M., and C. Wedekind. 1998. "Working memory constrains human cooperation in the Prisoner's Dilemma". *Proceedings of the National Academy of Sciences* vol. 95 (23), pp. 13755–13758.

Nachbar, J. H. 1992, December. "Evolution in the finitely repeated prisoner's dilemma". *Journal of Economic Behavior & Organization* vol. 19 (3), pp. 307–326.

Nowak, M., and K. Sigmund. 1993. "A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game". *Nature* vol. 364, pp. 56–58.

Paolacci, G., and J. Chandler. 2014, June. "Inside the Turk Understanding Mechanical Turk as a Participant Pool". *Current Directions in Psychological Science* vol. 23 (3), pp. 184–188.

Press, W. H., and F. J. Dyson. 2012. "Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent". *Proceedings of the National Academy of Sciences* vol. 109 (26), pp. 10409–10413.

Rand, D. G., and M. A. Nowak. 2013. "Human cooperation". *Trends in Cognitive Sciences* vol. 17 (8), pp. 413 – 425.

AUTHOR BIOGRAPHIES

SAMARTH SWARUP is a Research Associate Professor in the Biocomplexity Institute at the University of Virginia. He holds a PhD in Computer Science from University of Illinois at Urbana-Champaign. His email address is swarup@virginia.edu.

MARK G. ORR is a Research Associate Professor in the Biocomplexity Institute at the University of Virginia. He holds a PhD in Cognitive Psychology from the University of Illinois at Chicago. His email address is mo6xj@virginia.edu.

GIZEM KORKMAZ is a Research Associate Professor in the Biocomplexity Institute at the University of Virginia. She holds a PhD in Economics from the European University Institute (Italy). Her email address is gkorkmaz@virginia.edu.

KIRAN LAKKARAJU is a Senior Member of Technical Staff at Sandia National Laboratories, California in the Systems Research & Analysis III group. He holds a M.S. and PhD in Computer Science from the University of Illinois at Urbana-Champaign. His email address is klakkar@sandia.gov.