In D. Reitter & F. E. Ritter (Eds.), Proceedings of the 14th International Conference on Cognitive Modeling (ICCM 2016). University Park, PA: Penn State.

Learning the dynamics of Prisoner's Dilemma: Lessons from modeling and simulation

Michael Yu (msyu@cmu.edu)
Department of Social and Decision Sciences, 5000 Forbes Ave
Pittsburgh, PA 15213 USA

Cleotilde Gonzalez (coty@cmu.edu)
Department of Social and Decision Sciences, 5000 Forbes Ave
Pittsburgh, PA 15213 USA

Abstract

Learning to deal with social dilemmas can be difficult as outcomes depend not only on a person's decisions but also on other people's decisions and on how past decisions have changed that environment. We investigate how people might learn about social dilemmas by studying how simulated players using a cognitive model known as instance-based learning (IBL) interact with each other and with a set of fixed strategies in the Prisoner's Dilemma (PD). The current simulation study presents systematic variations in the payoff structure and the other player's strategy. Results indicate that the IBL model can reproduce predicted patterns of cooperation based on the payoff structure and that the model is sensitive to the strategies with which it is matched. The simulations offer explanations of how cognitive processes handle social dilemmas and how the environment of social dilemmas can influence this process.

Keywords: Instance-Based Learning; Cognitive Modeling; Prisoner's Dilemma; Cooperation.

Introduction

Instance-based learning (IBL) is a cognitively-inspired, descriptive model of how we make decisions in dynamic environments, i.e., environments that change over time and in which earlier decisions can inform and influence future actions (Gonzalez, Lerch, & Lebiere, 2003). Dynamic decision making often occurs in repeated social dilemmas, when people are asked to decide between actions that benefit themselves at a cost to the group or benefits the group at a cost to themselves. Past responses can influence how members of the group respond in the future, creating a dynamically complex learning environment. The Prisoner's Dilemma (PD) is a commonly studied social dilemma that instantiates this type of situation in a two-person game.

Classical game theory assumes that players understand explicit information about outcomes, reason about the other player's strategy, and solves for the best solution. Behavioral game theory assumes a similar understanding and adds preferences about the other player's actions and outcomes (e.g., Fehr & Schmidt, 1999; Rabin, 1993). Evolutionary game theory and related models of simulation assume no understanding of the game nor consideration of the other player, but rather that players follow pre-determined strategies (Axelrod & Hamilton, 1981; Danielson, 1992; Messick & Liebrand, 1995). The game is resolved by finding strategies that get better outcomes when different combinations of strategies interact in different ways. Approaches that consider learning including reinforcement learning (C. F. Camerer & Ho, 1998; C. Camerer & Ho, 1999) and applying cognitive models to the PD (Gonzalez, Ben-Asher, Martin, & Dutt, 2015; Gonzalez & Ben-Asher, 2014; Lebiere, Wallach, & West, 2000; Stevens, Taatgen, & Cnossen, 2016) ask how learning responds to changes in the decision-making environment.

This paper integrates approaches from these various traditions to develop a different perspective on learning in the PD. Our primary approach stems from cognitive modeling, using an IBL model to study the learning process. This contrasts with the common methods in classical and behavioral game theory in which the PD can be solved before the players interact. Similar to evolutionary game theory, IBL models interact with and respond to the environment (including payoffs and the strategies of other players), prompting questions of how the environment can change this interaction. Contrasted with evolutionary game theory, the focus of IBL is on learning by individual agents – rather than populations – using paradigms that are more similar to those of repeated interaction in classical and behavioral game theory. Under IBL, changes in behavior are a consequence of dynamic learning rather than population dynamics. A more fundamental assumption with the IBL models that contrasts with much of the classical and behavioral economic research is a focus on descriptive models rather than optimization. While IBL models try to make better decisions, the goal of this paper is *not* to find the "best" way to solve the PD but to understand how the environment influences the nature of decision-making in the PD as well as the stability of those decisions.

The present focus on environmental variation contrasts with much of the previous cognitive modeling research, as well. Similar to evolutionary game theory, the current work uses simulations to support studying a broad range of environments. We adopt general IBL models with robust parameters derived from previous research, but do not emphasize creating a specific model to fit a specific set of human data, as is commonly the case with past cognitive modeling research. The emphasis of the current work is *not* to develop the most comprehensive or best-fitting model of human decision-making in the PD, but to develop a clearer understanding of how the environment interacts with learning dynamics. Thus, our comparisons to human data are on general trends rather than absolute fit, which may require, as in the case of some of the more sophisticated cognitive models, integration of additional features beyond what a basic IBL learning model provides.

In brief, this paper asks how systematic changes in payoffs and partner strategies influence the learning process of an IBL model. We examine how an IBL model interacts with other simulated players in a repeated PD under various payoff conditions.

Repeated Prisoner's Dilemma

In the Prisoner's Dilemma players pick between two options: cooperate or defect. Payoffs are operationalized as follows. If both players cooperate, each receives a payoff of R (Reward). If one cooperates and the other defects, the cooperator receives S (Sucker) and the defector receives T (Temptation). If both defect, each receives P (Punishment). The PD is defined by the following relationship: $T > R > P > S$, and the best response for each player is to defect. However, if both players cooperate, they would each receive more than they would have had both defected. Repeating the PD with the same partner makes the game more complex, as current decisions can influence future outcomes. Over time, players not only learn what outcomes are likely for any given decision, but also influence the behavior of the other player, changing the likelihood of the different outcomes.

Payoff sensitivity. A common finding in empirical studies of the PD is that the player's likelihood of cooperation is related to the values of T, R, P, and S. One of the strongest representations of this was proposed by Rapoport (1967) as the K-index, with a higher K-index predicting higher cooperation. The K-index was defined as: $K = \frac{R-P}{T-S}$. However, rational economic models always predict defection and no correlation between cooperation and the K-index. According to these models, players can reason that defection is best in the last round, since reputation has no effect after the last round; that defection is the best in the second to last round, since reputation has no effect in the last round; and so on in a process known as backward induction.

In contrast, learning models, such as IBL, expect players to explore their options and learn through trial-and-error. Past research predicts that two IBL models acting independently of each other would result in a decrease of cooperation from early to later rounds (Gonzalez et al., 2015), and IBL models have also been found to be sensitive to the payoffs of the PD, consistent with Rapaport's K-index (Gonzalez & Ben-Asher, 2014). In this research we explore these predictions systematically in a wider range of payoff values of the PD and when the IBL model is paired against various known strategies.

Strategic sensitivity. Outcomes depend not only on payoffs, but on the likelihood of each outcome from each decision. These likelihoods depend on the partner's behavior in the game. In contrast to the reasoned approach taken by rational economic theory, where each player attempts to predict what other will do, evolutionary game theorists assume that players follow predetermined strategies (Axelrod & Hamilton, 1981). Strategies that are successful against other strategies are replicated, whereas unsuccessful strategies are removed from the population, leaving the most 'robust' strategies. However, this approach focuses on population characteristics (groups of agents) with 'learning' occurring over generations, rather than on learning of individual members of the population.

Two better-known strategies in evolutionary game theory are tit-for-tat (TFT) and win-stay-lose-shift (WSLS). TFT cooperates in the first round and, in all following rounds, copies the other player's action from the previous round. Thus, it cooperates if the other player cooperated and defects if the other player defected. WSLS also cooperates in the first round. In all subsequent rounds, WSLS will stick with the decision it made in the last round if it received either R or T; or will change its decision if it received either S or P.

In contrast, IBL models focus on how individuals learn – but how IBL models interact with other strategies has not been explored. As IBL models learn over time, they may adjust to other strategies within one game of the repeated PD.

Instance-based Learning

In IBL models decisions are stored in memory as a unique combination of actions and outcomes. Each pair of action and outcome is referred to as an *instance*. When facing a choice, the model estimates a *blended value* for each action being considered. The action with the highest blended value is selected. The blended value, V, for an action, x, at a point in time, t, is: $V_{xt} = \sum_o p_{xot} u_{xo}$ [Eq. 1], where p_{xot} is the retrieval probability of an outcome, o, associated with the action, x; and u_{xo} is the utility associated with the action, x, and outcome of o. In the PD, the actions can be represented as either "cooperate" or "defect," and the outcomes as T, R, P, and S.

The *retrieval probability* of an instance is influenced by the activation of that instance relative to the sum of instances which include the same action. The retrieval probability, p_{xot}, for action, x, and outcome, o, at a specific time, t, is: $p_{xot} = e^{\frac{A_{xot}}{\sigma\sqrt{2}}} / \sum_x e^{\frac{A_{xot}}{\sigma\sqrt{2}}}$ [Eq. 2], where σ is a noise parameter, A_{xot} is the activation of the instance with action, x, outcome, o, and at time, t.

Activation is higher for instances that were more frequent or more recently observed. The *activation*, A_{xot} for an option, x, and outcome, o, at time t_i is: $A_{xot} = \ln \sum (t - T_{xo})^{-d} + \sigma \ln\left(\frac{1-\gamma_{xot}}{\gamma_{xot}}\right)$ [Eq. 3], where d is a decay parameter, σ is the same noise parameter as in Eq. 2, T_{xo} is the set of all times in which the instances with action, x, and outcome, o were observed, and γ_{xot} is a draw from a uniform distribution bounded by 0 and 1 for the current action, x, at time t. For example, if "cooperate" is more often and more frequently met with the sucker outcome (S) than the reward outcome (R), then the cooperate/sucker instance would be higher in activation and retrieval probability, and the sucker payoff would more strongly influence the blended value of the "cooperate" option than the reward payoff.

The IBL model often takes standard parameter values from ACT-R, the cognitive architecture from which the Activation mechanism comes from (Anderson & Lebiere, 1998): σ is set to 0.25 and d is set to 0.5. As people are not expected to approach decisions with an empty memory, a common approach is to create *prepopulated instances* that represent initial beliefs about the decisions they expect to experience (Lejarraga et al., 2012). The utility associated with these prepopulated instances is typically set to some value higher than the highest possible observable value from the actual decisions to allow for initial exploration and are entered into memory with a time of 0.

The IBL model presented above is a model of an individual, aware of only (Gonzalez et al., 2015) interacts with different PD environments while controlling for broader social considerations including other-regarding preferences and beliefs.

Simulation Overview

In our simulations, we have an IBL model play the repeated PD with either another IBL model or another strategy over the course of 100 rounds. We simulate 400 pairs of players and focus on three measures: individual cooperation rates; alternation rates (switches from cooperate to defect or defect to cooperate); and how models behave as pairs (mutual cooperation, mutual defection, and mixed cases).

The IBL model follows the definition noted above, with σ set to 0.25 and d set to 0.5. Prepopulated instances for cooperate and defect are included with utilities set higher than the temptation payoff to promote exploration. The utility is set arbitrarily at 1.5 times T, i.e., at 15. When matched with other IBL models or other strategies, the IBL model receives information only about its actions and outcomes and no information about the other model's actions or outcomes.

To examine a range of payoffs, we adapt a method used by Moisan, et al. (2015) and inspired by Rapaport and Chammah (1965) and Axelrod (1967), where the payoffs of the PD are normalized with a fixed value for T and S. In the present simulation, T is fixed at 10, S is fixed at 0, and R and P vary between 0 and 10 in intervals of 1 such that $R \geq P$. Our simulations include boundary cases that are not strictly version of the PD: $T = R = 10$, $P = S = 0$, and $R = P$. These boundary cases give a sense as to how the models may behave close to the limits as $R \to T$, $P \to S$, and $P \to R$. This method also provides different payoff structures that have the same K-index, which occurs for any payoff structure in which R and P have the same difference.

To examine strategic sensitivity, we consider two simple, unconditional and two sophisticated, conditional strategies. For simple strategies, we match the IBL model with models that unconditionally cooperate (All-C) or unconditionally defect (All-D). For sophisticated strategies, we match the IBL model with another IBL model, a TFT and WSLS strategy.

Simulation Results

Figure 1 presents the results of simulations in which two IBL models are paired with varying levels of R and P. The panels indicate the cooperation and alternation rates of one of the IBL models from each pair, with the top of each panel representing 100% cooperation/alternation and the bottom representing 0% cooperation/alternation; and the left of each panel representing the 1st round and the right representing the 100th round. Panels that are lightly shaded represent the boundary conditions and are not 'true' PD games.

Across all games, the average cooperation rate starts at 50% (prepopulated instances cause the models to randomly decide to cooperate and defect in the first round), but drift towards increased cooperation or increased defection over time. Behavior does not necessarily change consistently towards cooperation or defection, e.g., the environment in which $R = 9$ and $P = 1$, shows a slight increase in cooperation before settling at a lower rate of cooperation. Final round behavior (at the far right of each panel, discussed more below) includes a variety of cooperation rates across panels. While this behavior may stabilize at a certain cooperation rate, this does not imply that the players have settled on cooperation or defection. While alternation rates also tend to decrease, for many payoff environments, the alternation rates *do not* trend towards 0. End-of-game behaviors include environments where players may switch back and forth between cooperation and defection.

Payoff sensitivity

Cooperation Rate A logistic regression of the final round cooperation relative to the K-index of each simulation indicates that cooperation by the IBL models increases as the K-index increases, B = 4.225, 95% CI [4.097, 4.355]. The coefficient on an ordinary least squares (OLS) regression provides a more intuitive estimate of the effect of K-index on cooperation, B = 0.681, 95% CI [0.664, 0.698], implying a 68.1 percentage point increase from a K-index of 0 to 1 (or a roughly 6.8% increase for each increase of 0.1). These findings are consistent with predictions that cooperation should increase with a higher K-index (Rapoport, 1967). This is visually confirmed in Figure 1, as each diagonal (bottom left to top right) indicates environments with the same K-index. For example, R = 2, P = 1 and R = 9, P = 8 have the same K-index (0.1) and relatively lower cooperation than both R = 5, P = 1 and R = 9, P = 5 (K-index = 0.4).

The K-index appears to account for most but not all of the influence of payoffs on cooperation. A logistic regression of the final round cooperation on the K-index and the R payoff[1] shows that the effect of the K-index is an order of magnitude larger than that of the R payoff, $B_{K\text{-}index} = 4.830$, $B_R = -0.109$; or with OLS, $B_{K\text{-}index} = 0.732$, $B_R = -0.010$.

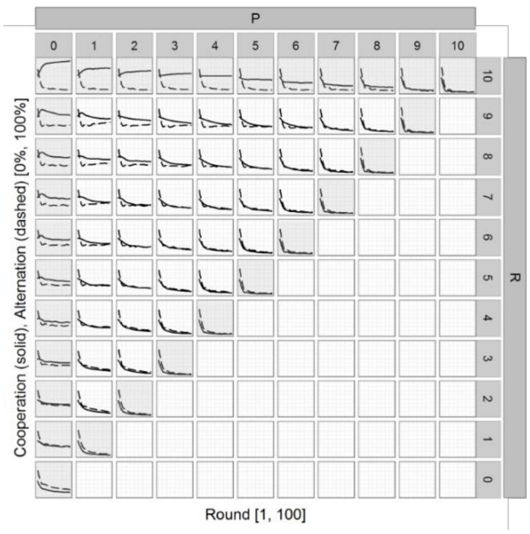

Figure 1: Development of cooperation and alternation for paired IBL models with varying levels of R and P (T = 10, S = 0) across rounds; boundary conditions shaded in gray

Alternation While cooperation rates indicate the general degree of cooperation between players, alternation provides an indication of how stable that cooperation is. Alternation rate may be compared to the concept of evolutionarily stable strategies (ESS) from the evolutionary game theory literature – which suggest how resistant strategies are to changes in the environment (and in particular, invasion by other strategies). In contrast to the idea of ESS, however, higher alternation, while seemingly less stable, may suggest a more robust learning *mechanism*, allowing agents to adapt more quickly from environmental changes.

While distinct from cooperation, alternation is constrained by the cooperation rate. As cooperation rates trend towards 0 or 1, alternation rates trend towards 0. For example, if everyone cooperates in both Round 99 and Round 100, it is impossible for anyone to have switched from defect to cooperate (since all cooperated in Round 99) or switched from cooperate to defect (since all cooperated in Round 100). The highest possible alternation exists where cooperation rates are 50 percent. Nonetheless, cooperation and alternation rates are not perfectly correlated. Environments which produce similar cooperation rates (e.g., R = 10, P = 3, cooperation = 0.293; R = 9, P = 3, cooperation = 0.305) can produce different alternation rates (e.g., R = 10, P = 3, alternation = 0.058; R = 9, P = 3, alternation = 0.210).

Given the non-linear constraints placed by cooperation on the alternation rate, it is unsurprising that alternation is less well predicted by the K-index, although an effect still exists. A logistic regression of alternation between the second-to-last and last rounds on the K-index suggests a positive effect of K-index on alternation, $B = 1.400$, 95% CI [1.277, 1.525]; or by OLS, $B = 0.179$, 95% CI [0.164, 0.195]. Including both the K-index and R payoff as parameters in the regressions we find coefficients for the logistic regression of $B_{K\text{-}index} = 2.728$, $B_R = -0.223$; and by OLS, $B_{K\text{-}index} = 0.285$, $B_R = -0.021$.

Visual inspection of Figure 1 suggests that the pattern of the K-index predicting alternation seems stronger for our 'proper' PDs (non-shaded panels) than for the boundary conditions (shaded panels). As we move towards the upper left panel and higher K-indices, we see that alternation increases. The trend breaks as we approach the boundary condition of $R = 10$ (top row), which shows lower alternation relative to $R = 9$. At this point, the R and T payoffs are identical and there is no temptation motive to draw players from mutual cooperation towards defection. This suggests that the dynamics between temptation and reward may be particularly critical in driving alternation.

Table 1: Pattern of individual and paired behaviors at corner points within simulated Prisoner's Dilemma

K-index (R, P)	0.8 (9, 1)	0.1 (9, 8)	0.1 (2, 1)
C	44.50	4.25	7.50
Alt	29.50	3.75	13.25
CC	25.50	1.50	2.00
DD	38.25	93.75	87.75
CD/DC	36.25	4.75	10.25

Paired behaviors. Looking at the strategies of *both* players in a pair provides further insight into the social dynamics of the PD. For example, a 50% cooperation rate could be achieved if half of the pairs are engaged in mutual cooperation (CC) and half in mutual defection (DD); or if all pairs include one cooperator and one defector (CD/DC).

Table 1 provides a deeper analysis of the final round behavior for the cases of (R, P) ∈ {(9, 1), (9, 8), (2, 1)}, i.e., the simulations involving the highest/lowest K-indices that are not the boundary conditions. The top half provide results for a single player in each pair, with "C" as the probability of cooperation and "Alt" as the probability of alternation; and the bottom half provides results for the pair of players. The table highlights the differences in the two rightmost payoffs, which have identical K-indices, but with an implicitly painful sucker payoff but little temptation, (9, 8) or a relatively painless sucker payoff but high temptation (2, 1). While there is low cooperation in both cases, the case with high temptation and a low sucker payoff shows more alternation that seems to pull people away from mutual defection in favor of increased mixed pairs ("CD/DC").

[1] Only R or P can be included in addition to the K-index due to multicollinearity.

Strategic sensitivity

Figures 2 and Figure 3 present simulation results of IBL models paired with non-IBL strategies. As with the IBL models paired with other IBL models, the models received no information about the other player's actions or outcomes, but the IBL models outcomes were influenced by the other player's choices. The first set of non-IBL strategies look at unconditional strategies, i.e., strategies that are not influenced by the IBL model's decision; whereas the latter set of non-IBL strategies look at conditional strategies, i.e., strategies which are influenced by the IBL model's decisions either directly (TFT) or indirectly (WSLS). They include panels for different simulations with varying P and R. However, they focus on a subset of P and R (with values of 0, 1, 8, 9 and 10) to allow for a more detailed view of the panels themselves.

All-C/All-D. Figure 2 shows that IBL models paired with strategies that always cooperate or always defect learn to defect quickly. Defection yields the best payoff with no risk of retaliation from these strategies. In contrast, TFT and WSLS would only defect when playing with All-D, but would cooperate with a partner who played All-C.

Exceptions occur only at certain boundary conditions. When paired with All-C, this occurs at $T = R = 10$; that is, when there is no temptation to defect. Cooperating or defecting in such an environment yields the same utility ($T = R$), making indifference reasonable. When paired with All-D, this occurs at $P = S = 0$. Again, cooperating or defecting yields the same utility ($P = S$), and indifference is reasonable. This indifference is reflected in the alternation rate which approaches 50% towards round 100 and is consistent with cooperating and defecting at random.

These findings are consistent with the behavior of individualist human players -- who also defect more against unconditional strategies (Kuhlman & Marshello, 1975), and contrast with strategies, such as TFT or WSLS which would not naturally learn to be opportunistic in these cases.

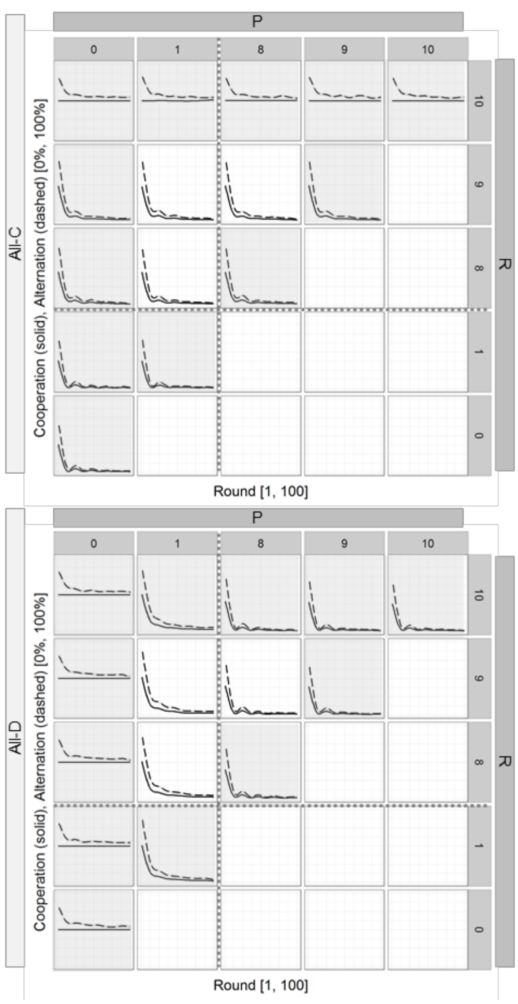

Figure 2: Development of cooperation and alternation for IBL models paired with All-C (top) and All-D (bottom) with select levels of R and P (T = 10, S = 0) across rounds; boundary conditions shaded in gray

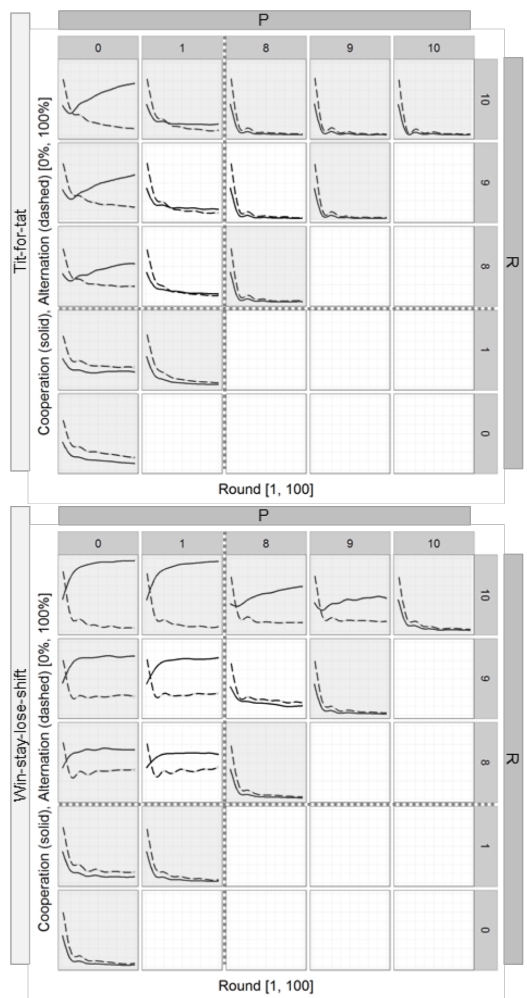

Figure 3: Development of cooperation and alternation for IBL models paired with Tit-for-tat (top) and Win-stay-lose-shift (bottom) with select levels of R and P (T = 10, S = 0) across rounds; boundary conditions shaded in gray

TFT/WSLS Figure 3 highlights the different response of the IBL model when paired with two popular strategies from evolutionary game theory, TFT and WSLS.

Surprisingly, TFT tends towards defection. High defection may be associated with research suggesting that TFT does not do well when their partner behaves inconsistently (Imhof, Fudenberg, & Nowak, 2007), as might be expected when IBL models explore their options. The exception at $P = S = 0$ suggests that the sucker payoff makes cooperation more prohibitive when learning to play with TFT. Not being penalized for moving away from mutual defection, provides an opportunity for the players to arrive at mutual cooperation.

Results for WSLS are similar to those observed in paired IBL models, which may be explained by WSLS having been developed as a simple *learning* model. When paired with a WSLS strategy, cooperation is more greatly affected by reducing the difference in the temptation and reward payoffs (with highest cooperation appearing at the boundary condition of $T = R = 10$) compared to reducing the difference in the punishment and sucker payoffs.

A comparison of Figures 1 and 3 suggest that alternation is higher when an IBL model is paired with WSLS than with other strategies. This is clearer in Figure 4, which shows the relationship between alternation and cooperation in the last round for the different simulations of IBL models partnered with TFT, WSLS, and a second IBL strategy. The relationship shows an upside-down U-shaped curve with all partners, consistent with earlier observations that high and low cooperation rates decrease maximum possible alternation. However, models partnered with a WSLS strategy shows higher alternation at almost all levels of cooperation relative to the TFT and IBL strategies.

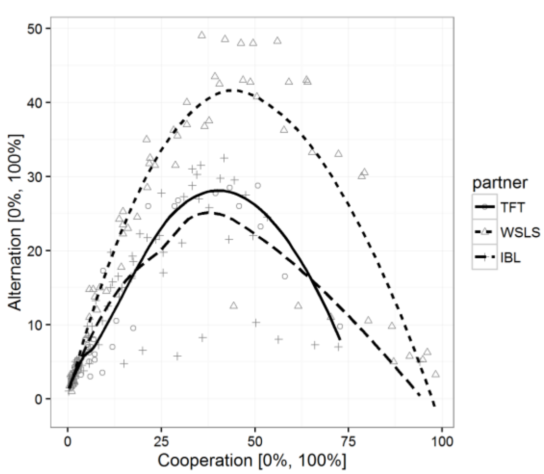

Figure 4: Relationship between cooperation and alternation of the IBL model in final round, when partnered with TFT (circles/solid), WSLS (triangles/short dash), and second IBL (plus/long dash)

Conclusion

Applying simulated cognitive models to social dilemmas helps us to understand how features of a social dilemma specifically impact the learning processes in the absence of other factors, such as other regarding preferences and expectations. The focus on learning differs from classical, behavioral, and evolutionary game theory which do not treat individual learning as a mechanism. The focus on simulation allows us to concentrate on learning more cleanly. For example, previous research finds that cooperation weakly decreases over time when people are paired with a strategy that always cooperates (Lave, 1965; Oskamp, 1971). As this decrease is slight, understanding the nature of this change, or if it is simply noise, can be challenging. By focusing on learning, the results in this paper provides clearer evidence that cooperation may decrease as a result of learning.

Simulations also provide some clearer insight into not only cooperation but the stability of cooperation as highlighted by alternation rates. Our findings suggest that the impact of the strategic environment can influence cooperation and alternation differently. In the case of WSLS, more alternation might draw players out of the ‘basin of attraction’ represented by mutual defection. Future work might investigate whether high alternation can help players adapt to a changing payoff or strategic environment, given that higher alternation suggests consistent exploration.

The present application of basic cognitive models, such as IBL, to the PD is not intended as a substitute for research using human data or for more complex models that try to fit this data. Indeed, the current research can serve as valuable baseline for such models to better highlight the contribution of specific mechanisms, such as information (Gonzalez et al., 2015), surprise and meta-cognition (C. Camerer & Ho, 1999; Gonzalez & Ben-Asher, 2014; Stevens et al., 2016), and initial beliefs (Lebiere et al., 2000). Altogether, the current research suggests additional areas of investigation and potential boundary conditions under which those models might be tested.

The present work can also be seen as an application of the methods from evolutionary game theory into cognitive modelling, in which we study how varying environments can impact learning rather than population dynamics. Similar to some of that research in that work, we can use simulations across multiple levels of variables to develop a map of sorts in terms of understanding potential areas of investigative interest and guiding our expectations of what results may be likely. The use of basic cognitive models to social dilemmas can help us better understand how we learn and how our approach to games develops over time.

Acknowledgments

This research was supported by the Network Science Collaborative Technology Alliance sponsored by the U.S. Army Research Laboratory under Cooperative Agreement No. W911NF-09-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory or the U.S. Government.

References

Axelrod, R. (1967). Conflict of Interest: An Axiomatic Approach. *The Journal of Conflict Resolution*, *11*(1), 87–99.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*, 1390–1396.

Camerer, C. F., & Ho, T. (1998). Experience-Weighted Attraction Learning in Coordination Games: Probability Rules, Heterogeneity, and Time-Variation. *Journal of Mathematical Psychology*, *42*(2/3), 305–26.

Camerer, C., & Ho, T.-H. (1999). Experience-weighted attraction learning in normal form games. *Econometrica*.

Danielson, P. (1992). *Artificial morality: Virtuous robots for virtual games*. Routledge.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868.

Gonzalez, C., & Ben-Asher, N. (2014). Learning to cooperate in the Prisoner's Dilemma: Robustness of predictions of an instance-based learning model. In *35th Annual Neeting of the Cognitive Science Society (CogSci2014)* (pp. 2287–2292).

Gonzalez, C., Ben-Asher, N., Martin, J. M., & Dutt, V. (2015). A cognitive model of dynamic cooperation with varied interdependency information. *Cognitive Science*, *39*(3), 457–495.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, *27*(4), 591–635.

Imhof, L. A., Fudenberg, D., & Nowak, M. A. (2007). Tit-for-tat or win-stay, lose-shift? *Journal of Theoretical Biology*, *247*(3), 574–580.

Kuhlman, D. M., & Marshello, A. F. (1975). Individual differences in game motivation as moderators of preprogrammed strategy effects in prisoner's dilemma. *Journal of Personality and Social Psychologyocial Psychology*, *32*(5), 922–931.

Lave, L. B. (1965). Factors affecting Co-operation in the Prisoner's Dilemma. *Behavioral Science*, *10*(1), 26–38.

Lebiere, C., Wallach, D., & West, R. L. (2000). A memory based account of the Prisoner's Dilemma and other 2x2 games. In *Proceedings of International Conference on Cognitive Modeling* (pp. 185–193).

Messick, D. M., & Liebrand, W. B. G. (1995). Individual heuristics and the dynamics of cooperation in large groups. *Psychological Review*, *102*(1), 131–145.

Oskamp, S. (1971). Effects of programmed strategies on cooperation in the Prisoner's Dilemma and other mixed-motive games. *Journal of Conflict Resolution*, *15*(2), 225–259.

Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. *The American Economic Review*, *83*(5), 1281–1302.

Rapoport, A. (1967). A note on the "index of cooperation" for Prisoner's Dilemma. *Journal of Conflict Resolution*, *11*(1), 100–103.

Rapoport, A., & Chammah, A. M. (1965). *Prisoner's dilemma: A study in conflict and cooperation*. University of Michigan Press.

Stevens, C. A., Taatgen, N. A., & Cnossen, F. (2016). Instance-Based Models of Metacognition in the Prisoner's Dilemma. *Topics in Cognitive Science*, *8*(1), 322–334.