Modeling Social Information in Conflict Situations through Instance-Based Learning Theory

Varun Dutt, Jolie M. Martin, Cleotilde Gonzalez
varundutt@cmu.edu, jolie@cmu.edu, coty@cmu.edu
Dynamic Decision Making Laboratory, Carnegie Mellon University
4609 Winthrop Street, Pittsburgh, PA 15213

Abstract: Behavior in conflict situations can be influenced by the social information that individuals have about their opponents. This paper tests whether an existent Instance-based Learning (IBL) model, built using the Instance-based Learning Theory (IBLT) to explain behavior in a single-person binary-choice task (BCT), can predict behavior in a two-player iterated prisoner's dilemma (IPD) game. The same IBL model is generalized to two conditions in the IPD: Social, where individuals have information about their opponents and their choices; and Non-social, where individuals and opponents lack this information. We expect the single-person IBL model to predict behavior in the Non-social condition better than in the Social condition. However, due to the structural differences between BCT and IPD, we also expect only moderately good model predictions in the Non-social condition. Our results confirm these expectations. These findings highlight the need for additional cognitive mechanisms to account for social information in conflict situations.

Keywords: conflict; cooperation; social information; 2x2 games; iterated prisoner's dilemma; instance-based learning theory; cognitive modeling; generalization.

Introduction

One objective of a participant in 2x2 games (between two players, each of whom has two available choice options), is to maximize personal economic benefit by cooperating or competing with an opponent. A popular game called the prisoner's dilemma (PD) (Axelrod, 1980; Rapoport & Chammah, 1965) has been widely used to investigate such conflict situations. In the PD, each of two participants chooses simultaneously whether to cooperate (C) or defect (D). If both cooperate, they obtain an equal outcome that is larger than if both had chosen to defect. (In Figure 1, each player's outcomes are higher for C-C than for D-D.)

		Opponent's Action	
		D	C
Your Action	D	You get -1 point, Opponent gets -1 point	You get 10 points, Opponent gets -10 points
	C	You get -10 points, Opponent gets 10 points	You get 1 point, Opponent gets 1 point

Figure 1. The matrix of outcomes in the prisoner's dilemma game. "Your Action" and "Opponent's Action" refer to the actions of the two players. "D" and "C" are labels used for "defection" and "cooperation," respectively.

However, if one participant defects while the other cooperates, the defector obtains an even larger outcome while the cooperator suffers a loss (shown by the C-D and D-C outcomes in Figure 1). In a one-trial PD, the standard finding is a larger proportion of D choices than C choices when aggregated over several participants (Rapoport & Chammah, 1965). However, in the iterated PD (IPD), where people are asked to repeatedly make C or D decisions, the proportion of D choices are shown to decrease over time (Rapoport & Chammah, 1965). Thus, the PD represents a tradeoff between short-term individual gain of defection and long-term individual gain of sustained mutual cooperation (Baker & Rachlin, 2002).

Despite the general focus on maximizing personal benefits in the PD, researchers have argued that the economic perspective alone is oftentimes insufficient to capture the social aspects of such games, including the amount of information that is shared between participants (Dawes, Van De Kragt, & Orbell, 1988; Gonzalez & Martin, 2011; Schuster & Perelberg, 2004). For example, if a participant does not know that he is actually playing with another human opponent in the IPD, is not provided with the matrix in Figure 1, and is asked to maximize his benefits by repeatedly choosing between the C and D buttons from experience (Non-social condition), then he might strive to strictly maximize his own observed outcomes. However, when the participant knows that he is playing a human opponent and has descriptive information about how his own and opponent's choices will affect one another's outcomes (e.g., Figure 1) (Social condition), he might be inclined to take the other's perspective. In the Social condition, he chooses to cooperate or defect not only to maximize his own outcomes, but also to uphold his preferences regarding fairness and trust (Baker & Rachlin, 2002; Gonzalez & Martin, 2011).

Furthermore, recent literature in decisions from experience has shown that human behavior is primarily driven by experience when people are presented with *both* the descriptive and experiential information (like that in the Social condition in the form of knowledge of human opponents and outcomes in Figure 1 along with repeated choices to defect and cooperate) (Jessup, Bishara, & Busemeyer, 2008; Lejarraga & Gonzalez, 2011). This human behavior could be evaluated in the form of alternations from a choice in the current trial (to defect or cooperate) to the choice in the next trial. One expects that early on in the task humans show exploration and thus alternate much more between C and D choices; but with repeated experience their alternations decay over trials (due to exploitation of learned choices) (March, 1991). If the IPD represents decisions from experience, then we expect that human alternations in both Social and Non-social conditions to show the exploration-exploitation tradeoff where in Social condition humans rely on experience rather than the descriptive information just like they would do in the Non-social condition.

Many laboratory studies have evaluated human behavior in different social information conditions in the IPD (Baker & Rachlin, 2002; Gallagher, Jack, Roepstorff, & Frith, 2002; Martin, Gonzalez, Juvina, & Lebiere, in-prep; McCabe, Houser, Ryan, Smith, & Trouard, 2001). Moreover, there have been a number of mathematical and cognitive attempts to model human behavior in the presence of social information in the IPD (Bordini, Bazzan, Vicari, & Campbell, 2000; Cho & Schunn, 2002; Erev & Roth, 1998; Erev & Roth, 2001; Kim & Taber, 2004; Lebiere, Wallach, & West, 2000; Ritter & Wallach, 1998; West, Lebiere, & Bothell, 2006). Among the mathematical attempts the two common approaches that have been used include agent-based modeling and reinforcement learning (Bordini, Bazzan, Vicari, & Campbell, 2000; Erev & Roth, 1998; Erev & Roth, 2001). Among the cognitive attempts there has been both a single memory-based account in ACT-R architecture (Lebiere, Wallach, & West, 2000) and several procedural accounts in the ACT-R and SOAR architectures (Cho & Schunn, 2002; Ritter & Wallach, 1998; Kim & Taber, 2004). The mathematical attempts have lacked cognitive explanations of the human behavior like memory and recall. Moreover, the cognitive-procedural attempts have mainly relied on fixed strategies that often compete to reproduce the effects of social information (Gonzalez & Martin, 2011). The single cognitive memory-based attempt assumes a single shared memory for two humans and makes no distinction between recalling an outcome from memory for the first time and experiencing it repeatedly (Lebiere, Wallach, & West, 2000). Thus, this modeling approach might be unrealistic in explaining the effects of social information in both the Social and Non-social conditions.

In this paper, we investigate how an existent memory-based model based upon the Instance-based Learning theory (IBLT) to capture individual behavior, is able to account for behavior in the Social and Non-social conditions of the IPD. Most recently, memory-based models of experiential-learning, derived from IBLT have shown robust generalization to novel conditions in a single-person binary-choice task (BCT), and have also performed well at predicting behavior in a complex multi-person BCT (e.g., a market entry game) (Gonzalez, Dutt, & Lejarraga, 2011). In these BCTs, participants are told to maximize their outcomes, lack information about how their outcomes are generated, and can only gather this information through experience. Thus, the BCTs are the closest to the Non-social condition in the IPD where a participant does not know that he is actually playing with another human opponent and, like in a BCT, is instructed to maximize his own outcomes from experience, without knowing the outcome matrix ahead of time.

IBLT (Gonzalez et al., 2003) proposes that people make decisions by storing and retrieving instances from memory, where an instance serves as the basic unit of experience. The use of instances in memory in IBLT depends on a gradual transition from implicit exploration to exploitation processes that account for the exploration-exploitation tradeoff as more and more similar instances accumulate in memory. The theory reflects a generic decision making process that includes recognition, judgment, choice, execution, and feedback steps that affect decisions with accumulated instances in memory, and according to the interaction of a decision maker with a decision task.

In this paper, we test whether the same model based upon IBLT (hereafter, IBL model) used in the BCTs, can explain human behavior in two IPD conditions, Non-social and Social. Because the existent IBL model was built for single-person tasks where participants lacked social information, we hypothesize that the model when generalized to the IPD conditions will be able to explain human behavior in the Non-social condition better than in the Social condition. The generalization process involves using the model developed for the BCT, with identical parameters (from Lejarraga et al., 2010) and generalizing it in the two IPD conditions. This generalization process is a standard procedure to test the robustness of cognitive models (Busemeyer & Wang, 2002). The behavioral data we use in the generalization of the IBL model to the IPD is reported in a separate manuscript (Martin et al., in-prep). First, we briefly describe the Social and Non-social conditions of the IPD. Then, we describe how we generalized an existent IBL model and compared its performance to observed behavior in these two conditions. Finally, we discuss results of comparison and describe potential future directions in this research.

Iterated Prisoner's Dilemma with and without Social Information

In Martin et al. (in prep), the experimental procedures and human data results are presented in detail. Here, we summarize their methods and some of their findings. One-hundred and twenty participants were randomly paired with one another and assigned to one of two between-subjects conditions, Social and Non-social, to play the IPD. The two conditions fall near opposite ends of the Hierarchy of Social Information (HSI) framework (Gonzalez & Martin, 2011) with least social information available to a participant in the Non-social condition and the most social information available to a participant in the Social condition. Participants in both conditions in the IPD played a total of 200 repeated trials (which were unnumbered with no known endpoint to participants) and made repeated "C" and "D" decisions (See Figure 1). In the Non-social condition, participants did not know they played another person, and thus, only knew the decisions they took and their own outcomes (they were essentially maximizing their own outcomes in this condition). In the Social condition, participants were informed that they played another person, were given the outcome matrix similar to Figure 1 from the outset of the game, and they saw the decisions and outcomes of the other player throughout the interaction. Participants received a base pay of $10 and could earn additional pay based upon points earned in the IPD. In both conditions, participants who were randomly paired to play the IPD were anonymous and did not see or talk to each other. Two standard dependent measures were used to compare the IBL model results to human data: 1. Average proportion of defections (D-rate) over trials (as a measure of overall human behavior); and, 2. Average proportion of alternations (from cooperate (C) to defect (D) and vice-versa) (A-rate) over trials (an overall measure of human learning or exploration-exploitation). These proportions were computed over 30 pairs of human participants and 30 pairs of model participants in each condition over the 200 trials (i.e., averaged over all participants). The behavioral results will be summarized below together with the results from the IBL model.

The IBL model

We used an existent model based upon IBLT that was built to explain human behavior in single-person BCT (Lejarraga, Dutt & Gonzalez, 2010). An instance, i.e., smallest unit of experience, in the IBL model consists of three parts: a situation in a task (a set of attributes that define the decision situation), a decision in a task, and an outcome resulting from making that decision in that situation. Different parts of an instance are built through a general decision process: creating a situation from attributes in the task, a decision and expectation of an outcome when making a judgment, and updating the decision's outcome in the feedback stage when the actual outcome is known. In the IBL model, instances accumulated in memory over time are retrieved from memory and are used repeatedly according to their availability in memory. This availability is measured by a statistical mechanism called *Activation*, originally implemented in the ACT-R architecture (Anderson and Lebiere, 1998). In this paper, we extend the IBL model of a single-person BCT to the two-player IPD by simply allowing the same two single-person models with their own memories as opponents to interact with each other in the IPD (having an independent memories for each model does away with the assumption of a single shared memory by Lebiere, Wallach, & West, 2000). Next, we summarize the single-player IBL model and explain the extensions to the IPD.

In the IBL model, each instance consists of a label that identifies an option in the IPD (i.e., to cooperate or defect) and the outcome obtained (e.g., 10 points). Thus, the structure of an instance is simply, (option, outcome) (e.g., defect, 10). There are four instance-types one for each of the four possibilities in Figure 1. In each trial *t* of the IPD, the option with the highest blended value is selected (Equation 1 below). The blended value of an option depends on outcomes observed in the option and the probability of retrieval of instances from memory corresponding to outcomes (Equation 2 below). Furthermore, the probability of retrieval of instances from memory is a function of their activation in memory, governed by the recency and frequency of retrieval of instances from memory (Equation 3 below).

The IBL model for Iterated Prisoner's Dilemma

In the IBL model the selected option in a trial is one with the highest *blended value, V* (Lebiere, 1999) resulting from all instances belonging to options. The blended value of option *j* is defined as:

$$V_j = \sum_{i=1}^{n} p_i x_i \qquad [1]$$

where x_i is the value of the observed outcome in the outcome slot of an instance *i* corresponding to the option *j* and p_i is the probability of that instance's retrieval from memory (for the IPD, the value of *j* is either to defect or to cooperate and x_i could be -10, -1, +1, +10 depending upon the respective decision choices in Figure 1). The blended value of an option is the sum of all observed outcomes x_i for the option in the corresponding instances in memory, weighted by their probability of retrieval. In any trial *t*, the probability of retrieval of instance *i* from memory is a function of that instance's activation relative to the activation of all other instances corresponding to that option, given by

$$P_{i,t} = \frac{e^{A_{i,t}/\tau}}{\sum_j e^{A_{j,t}/\tau}} \qquad [2]$$

where τ is random noise defined as $\tau = \sigma \times \sqrt{2}$, and σ is a free noise parameter. Noise in equation 3 captures the imprecision of retrieving instances from memory.

The activation of each instance in memory depends upon the *Activation* mechanism (Anderson & Lebiere, 1998). A simplified version of the activation mechanism that relied on recency and frequency of use of instances in memory was sufficient to capture human choice behavior in several BCTs (Lejarraga, Dutt, & Gonzalez, 2010) and has been used in the IBL model reported in this paper. For each trial *t*, *Activation* $A_{i,t}$ of instance i is:

$$A_{i,t} = ln\left(\sum_{t_i \in \{1,\ldots,t-1\}} (t - t_i)^{-d}\right) + \sigma \,.\, ln\left(\frac{1-\gamma_{i,t}}{\gamma_{i,t}}\right) \qquad [3]$$

where, d is a free decay parameter and t_i is the time period of a previous trial where the instance *i* was created or its activation was reinforced due to an observed outcome in the task corresponding to the instance's outcome in memory. Thus, the model only reinforces instances when a corresponding outcome is observed in the task and not when instances are retrieved from memory (an assumption in the model by Lebiere, Wallach, & West, 2000). The summation includes a number of terms that coincides with the number of times that an outcome has been observed in previous trials and that the corresponding instance i's activation has been reinforced in memory. Therefore, the activation of an instance corresponding to an observed outcome increases with the frequency of observation of the outcome (i.e., by increasing the number of terms in the summation) and with the recency of those observations (i.e., by small differences in $t_i \in \{1, \ldots, t-1\}$ of outcomes that correspond to that instance in memory). The decay parameter d affects the activation of the instance directly, as it captures the rate of forgetting. The higher the value of the *d* parameter, the faster is the decay of memory.

The $\gamma_{i,t}$ term is a random draw from a uniform distribution bounded between 0 and 1, and the $\sigma \,.\, ln\left(\frac{1-\gamma_{i,t}}{\gamma_{i,t}}\right)$ term represents Gaussian noise important for capturing the variability of human behavior. The higher the σ value, the more variability there will be in the retrieval of information from memory. Lejarraga, Dutt, and Gonzalez (2010), found the optimized value of d=5.0 and σ = 1.5 in the IBL model of the BCT by minimizing the dependent measure (maximization-rate) between the model and human data. The high value of d and σ parameters assumes a high rate of decay of memory instances and considerable variability in model's performance over trials. As we use the same model in this paper, we keep both d and σ parameters at values determined by Lejarraga, Dutt, and Gonzalez (2010).

First Trial

Given that in the first trial there are no past instances from which to calculate blended values of the two options, the model makes a selection between two pre-populated instances in memory. Each pre-populated instance corresponds to one of the two options, cooperating or defecting, with a value of +30 pre-assigned to the instance's outcome slot. These pre-populated instances in memory may represent the expectations that participants bring to the laboratory (Lejarraga, Dutt, & Gonzalez, 2010). The choice of a +30 value is the same as that assumed by Lejarraga, Dutt, and Gonzalez (2010). As the +30 value is higher than any of the possible outcomes in the task (Figure 1), it will trigger an initial exploration of the two options. Since both pre-populated instances have the same outcome, in practice the model makes a random selection of the two options in the first trial. Because the +30 values are never observed as outcomes in the IPD according to its matrix (Figure 1), thus the activation of these pre-populated instances decays quickly enough that they cease to affect decisions in the model after the first few trials in the IPD.

Implementation and execution of the IBL model in the IPD

The same single-person model (described above) was duplicated to form the two players in the IPD (called M1 and M2) and these acted as a pair of participants interacting repeatedly for 200 trials in the IPD, just as human participants did in two conditions, Social and Non-social (Martin et al., in prep). Both M1 and M2 used identical mechanisms and the same parameter values. The outcomes for each model in a given trial were determined as a consequence of both their decisions, as for human players (See Figure 1). The same IBL model with M1 and M2 players was generalized to the Social and Non-Social conditions separately to determine how the model that represents individual behavior in the BCT would perform in each of the conditions of the IPD. The performance of the model was determined by computing the mean squared distance (MSD) over 200 trials between the D-rate and A-rate predictions from the model and that from the human data in each condition. Because we expect the model to generalize better in the Non-social condition compared to the Social condition, the MSDs in the Non-social condition should be smaller than those in the Social condition. Also, according to IBLT, regardless of the learning situation, the gradual transition from exploration to exploitation occurs according to the consistency and similarity of the problem and outcomes in the IPD (Gonzalez et al., 2003). Since the same payoff values are maintained throughout the learning process and in both conditions, we expect similar transitions from exploration to exploitation as measured by the A-rate in human data. Thus, the A-rates should gradually decrease over time, and the MSDs for the A-rate should be similar in both conditions.

Results

Table 1 summarizes the MSDs obtained by generalizing the same IBL model from Lejarraga et al. (2010) to the two conditions of the IPD, Social and Non-social. The MSD for the D-rate was considerably higher in the Social condition compared to that in the Non-social condition; however, the MSDs for the A-rate were about the same in the two conditions (fractionally better in the Social condition compared to the Non-social condition). These results seem to meet our expectation that the IBL model originally created for a single-person BCT would perform better in the Non-social condition compared to the Social condition and that humans would gradually transition from exploration to exploitation regardless of the condition and due to the consistency of the problem and outcomes as predicted by IBLT.

Table 1. The MSDs in the different conditions.

Condition	MSD (D-rate)	MSD (A-rate)
Non-Social	0.0201	0.0071
Social	0.1415	0.0049

Figure 2 presents the D-rate and A-rate in the model and human data over 200 trials of IPD in the Non-social condition (panel A) and the Social condition (panel B), respectively. The model's predictions for the D-rate seem to overestimate the D-rate over time in both conditions. Moreover, the overestimation of the model's predictions is exacerbated in the Social condition due to a drop in the human D-rate over trials. Furthermore, although there is a marked change in behavior of the D-rate in human data between the two conditions, the exploration-exploitation (reflected by the A-rate) is very similar in both conditions as the IBL model predicts. Also, the model's A-rate is high initially and low in the later trials and this behavior reflects the predicted gradual transition from exploration to exploitation. The initial instances with +30 values of utility drive the initial exploration and gradually moves to the actual values of the payoffs obtained from the game.

Discussion

In this paper, we expected that an IBL model, developed for a single-person BCT without any social information and based upon decisions from past experience, would be able to make more accurate predictions in conflict situations where social information is absent compared to where it is present. Furthermore, we expected that regardless of the social condition, human exploration-exploitation in both conditions would be similar. Based upon results in this paper, our expectations were met. The IBL model that accounts of individual behavior in binary choice, performed reasonably well in the Non-social condition compared to the Social condition according to the D-rate in human data. This happens because the IBL model is experience-based, uses retrieval from memory, and it has been shown to do well in single-person BCT similar to the IPD's Non-social condition (Lejarraga, Dutt, & Gonzalez, 2010). Moreover, because the model seems to essentially rely on gained experience in the IPD, the model moves gradually from exploration to exploitation in the IPD (Gonzalez et al., 2003). However, it is valuable to note that humans in the Non-social condition might be primarily exploring the outcome distribution (like the model); whereas, humans in the Social condition might be primarily exploring the choice strategy of their opponent (unlike the model). These differences between model and humans explain reasons for fits in the two conditions.

Although the model performs reasonably well it also seems to overestimate the human D-rate. We believe that the overestimation of the D-rate in the non-social condition can only be due to the difference in the dynamics from the single-person BCT (Lejarraga, Dutt, & Gonzalez, 2010) to the IPD. In the IPD, considering an ambivalent 50-50 chance of defections and cooperation, the expected value of the defection option (10*0.5-1*0.5 = 4.5) for a player is much higher than that for the cooperation option (= -4.5). Because blended values of the two options approach the expected value over many trials, the model, that seems to be driven to maximize the blended value in each trial, yields a high D-rate over trials. In addition, in the BCT, the probability of occurrence of outcomes remains fixed for all trials in both options, whereas in the IPD, the probability changes dynamically as a function of the other player's actions. The fact that the overestimation of the D-rate in the Social condition is much larger and that the gap increases over time supports these explanations.

In evaluating an existing IBL model in different social information conditions, we have highlighted the challenge that social information brings to computational model of individual choice behavior. Our next step is to calibrate the IBL model in the Non-social and Social conditions and in conditions that are in between these two extremes with same and different parameters for both model participants to evaluate the highest potential of the model to explain human behavior. Furthermore, we would like to gain insight into exactly what the existing IBL model

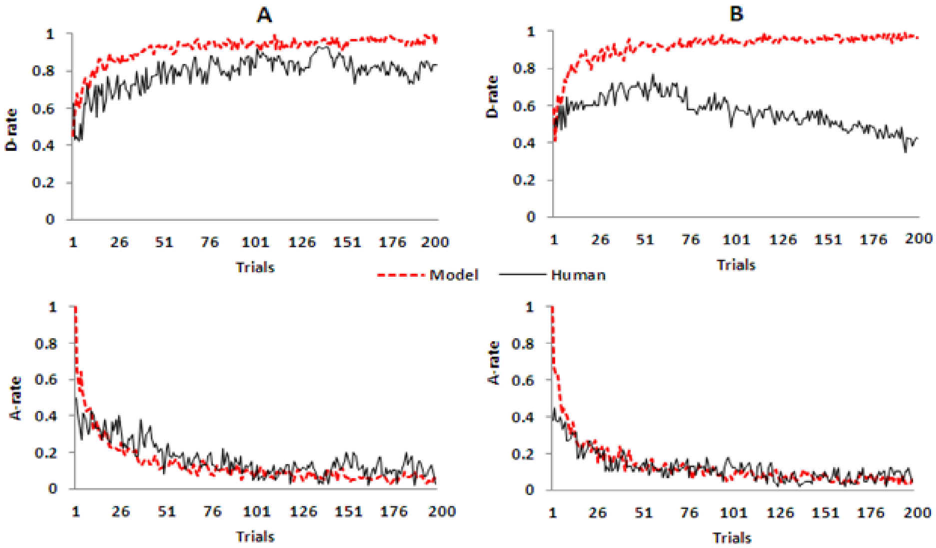

Figure 2. The D-rate (top row) and the A-rate (bottom row) in the Non-social condition (panel A) and Social condition (panel B).

lacks. Gaining this insight that accounts for the effects of social information in conflict situations will be an ongoing focus of this research.

Acknowledgements

This research was supported by a Defense Threat Reduction Agency grant to Cleotilde Gonzalez and Christian Lebiere (HDTRA1-09-1-0053).

References

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Hillsdale, NJ:Lawrence Erlbaum Associates.

Axelrod, R. (1980). Effective choice in the Prisoner's Dilemma. *Journal of Conflict Resolution*, *24*(1), 3-25.

Baker, F., & Rachlin, H. (2002). Teaching and learning in a probabilistic prisoner's dilemma. *Behavioural Processes*, *57*, 211–226.

Bordini, R. H., Bazzan, A. L. C., Vicari, R. M., & Campbell, J. A. (2000). Moral Sentiments in the Iterated Prisoner's Dilemma and in Multi-Agent Systems. *Brazilian Electronic Journal of Economics*, *3*(1).

Cho, K., & Schunn, C. (2002). Strategy Shift in Prisoner's Dilemma through Utility Learning. In *Proceedings of the 2002 ACT-R Summer Workshop*. Carnegie Mellon UNiversity, Pittsburgh, PA.

Dawes, R. M., Van De Kragt, A. J. C., & Orbell, J. M. (1988). Not me or thee but we: The importance of group identity in eliciting cooperation in dilemma situations: Experimental manipulations. *Acta Psychologica*, *68*(1-3), 83-97.

Erev, I., & Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *The American Economic Review*, *88*(4), 848-881.

Erev, I., & Roth, A. E. (2001). Simple reinforcement learning models and reciprocation in the Prisoner's Dilemma game. In G. Gigerenzer & R. Selten (Eds.), *Bounded rationality: The adaptive toolbox* (pp. 215-231). Cambridge, MA: MIT Press.

Gallagher, H., Jack, A., Roepstorff, A., & Frith, C. D. (2002). Imaging the intentional stance in a competitive game. *NeuroImage*, *16*, 814– 821.

Gonzalez, C., Dutt, V., & Lejarraga, T. (2011). A loser can be a winner: Comparison of two instance-based learning models in the market entry competition. Invited submission to *Games*, open access journal. Unpublished manuscript under review.

Gonzalez, C., & Dutt, V. (2010). Instance-based Learning: Integrating Decisions from Experience in Sampling and Repeated Choice Paradigms. Manuscript submitted for publication.

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

Gonzalez, C. & Martin, J. M. (2011). Scaling up Instance-Based Learning Theory to account for social interactions. Manuscript submitted for publication.

Jessup, R. K., Bishara, A. J., & Busemeyer, J. R. (2008). Feedback produces divergence from prospect theory in descriptive choice. *Psychological Science*, *19*, 1015–1021.

Kim, S.-Y. & Taber, C. (2004). A cognitive/affective model of strategic behavior-2-person repeated prisoner's dilemma game. In *Proceedings of the sixth International Conference on Cognitive Modeling* (pp. 360-361). Pittsburgh, PA: Carnegie Mellon University/University of Pittsburgh.

Lebiere, C. (1999). Blending. In *Proceedings of the Sixth ACT-R Workshop*. George Mason University, Fairfax, VA.

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*, 185-193. NL: Universal Press.

Lejarraga, T., Dutt, V., & Gonzalez, C. (2010). Instance-based learning: A general model of decisions from experience in repeated binary choice. Manuscript in press in *Journal of Behavioral Decision Making*.

Lejarraga, T., & Gonzalez, C. (2011). Effects of feedback and complexity on repeated decisions from description. Manuscript submitted for publication.

March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, *2*(1), 71–87.

Martin, J. M., Gonzalez, C., Juvina I., Lebiere, C. (in-prep). Effects of social information in conflict situations. Unpublished manuscript in preparation.

McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences*, *98*, 11832-11835.

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

Rapoport, A., Erev, I., Abraham, E. V., & Olson, D. E. (1997). Randomization and adaptive learning in a simplified poker game. *Organizational Behavior and Human Decision Processes, 69*(1), 31-49.

Ritter, F. E., & Wallach, D. P. (1998). Models of two-person games in ACT-R and Soar. *In Proceedings of the Second European Conference on Cognitive Modelling.* 202-203. Nottingham: Nottingham University Press.

Schuster, R., & Perelberg, A. (2004). Why cooperate? An economic perspective is not enough. *Behavioural Processes*, *66*, 261-277.

West, R. L., Lebiere, C. & Bothell, D. J. (2006). Cognitive architectures, game playing, and human evolution. In R. Sun (ed.), *Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation*. Cambridge University Press; New York, NY, 103-123.