In: Proceedings of the 19th Behavior Representation in Modeling & Simulation (BRIMS), Charleston, SC, 2010

Resistance is Futile: Winning Lemonade Market Share through Metacognitive Reasoning in a Three-Agent Cooperative Game

David Reitter, Ion Juvina, Andrea Stocco and Christian Lebiere
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA
reitter@cmu.edu, ijuvina@cmu.edu, stocco@cmu.edu, cl@cmu.edu

Keywords:
Metacognition, Cognitive Modeling, Games, Cooperation

ABSTRACT: *The Lemonade Game is a three-player game in which players have to pick locations on a circular board, which are as far away as possible from those chosen independently by other players. Players may observe other player's moves and infer their strategies. The game was studied using a competition of cognitively motivated agents, which inherit properties of adaptivity and stochasticity from human memory and decision-making, and simplistic, yet effective agents implementing fixed strategies. We argue that metacognition is the unique attribute that allows sophisticated agents to adapt to unforeseen conditions, cooperators and competitors.*

1. Introduction

Unlike other species, humans are not optimized for a specific natural environment or task, but are instead good at many things. Agents optimized to a particular ecological niche might succeed at first, but once their environment changes they are likely to be suboptimal and become extinct: Generalists beat specialists. While there is no doubt that we owe our superior adaptability to cognitive rather than physical attributes, the precise source of that superiority has been the subject of some debate, and proposals have been made to precisely formulate and measure that capability (e.g., Anderson & Lebiere, 2003). Here we provide support for flexibility and adaptivity afforded by *metacognition* as our main evolutionary advantage.

Our argument applies to artificial as well as biological agents. In particular, the focus on optimality and specialization that dominates much of the cognitive sciences can be seen as counterproductive, and indeed as the very source of a controversial pattern of reaching short-term objectives while making little or no progress toward the overall goal. Artificial Intelligence has met high-profile challenges (a world champion chess player, or a semi-autonomous vehicle) but it seems no closer to the original dream of a generally intelligent artifact. Cognitive Psychology has seen the development of high-fidelity models that reproduce human behavior in highly controlled tasks, but rarely do these models exhibit robust behavior in unforeseen situations. Finally, Machine Learning algorithms can use large amounts of data to adapt their performance, but only within the boundaries of their representations. The common thread of these approaches is narrow optimality within limited circumstances, and often disastrous behavior outside these confines.

1.1 The Lemonade Game

The question that arises is how to study the flexibility and adaptivity that might be the true magic of human cognition. One possibility is to adopt open-ended challenge

tasks where agents are exposed to unforeseen situations. That was the approach chosen for the Dynamic Stocks and Flow Model Comparison Challenge (Lebiere, Gonzalez, & Warwick, 2009). Another possibility is to select an environment that highlights the complexity of the interactions of the agents that inhabit it. One such deceptively simple but subtly complex task is the Lemonade Game used in a recent challenge by Martin Zinkevich of Yahoo Research[1]. In this game, three agents have to simultaneously place their fictional lemonade stands at one of 12 possible locations, arranged in a circle and referred to as 0 through 11. Just like a real lemonade stand sells more beverages when competitors are far, so an agent maximizes its payoff by selecting a location that is maximally distant from the others. In particular, each agent's payoff is calculated as the sum of the distances from the other two. A full game consists of 100 consecutive such trials, where the three agents independently and synchronously decide the locations of their respective stands. After communicating their locations, the positions and payoffs of every agent is calculated and revealed.

Many similar simple games either enforce zero-sum competition (e.g., Paper Rock Scissors; Billings, 2000) or present an alternative between cooperation and competition (e.g., the Prisoner's Dilemma; Rapoport, Guyer & Gordon, 1976). A unique feature of the Lemonade Game is that it permits a combination of both cooperation (between two agents) and competition (against the third). As we will see, the emerging dynamics are quite interesting and make any easy optimal strategy impossible. In order to succeed, the agents must adapt to the others' strategies, find ways to communicate their intent to cooperate and detect a similar willingness in others, and more generally encounter and adapt to patterns of behavior that cannot be derived from the environment but instead arise from the agents themselves and their interaction. We will start by outlining simple agents to play the game and their limitations. Then, we will describe a more complex approach that depends upon a combination of action strategies, sequence-detection abilities, and, most importantly, meta-cognitive supervision that continually oversees the behavior of the agent.

2. Basic Decision-making Agents

The simplest possible agents are "self-centered," in the sense that they entirely ignore the actions of the other players. Four basic agents were evaluated.

The ***Random*** agent picks a random location at every trial. Thus, its behavor is maximally unpredictable. This strategy can be a successful baseline in many games (e.g., Paper-Rock-Scissors, West & Lebiere, 2001) or adversarial games like the Prisoner's Dilemma (Lebiere, Wallach, & West, 2000). In the Lemonade Game, however, randomness has the side effect of precluding cooperation. Indeed, the random agent often received the poorest score in our tournaments when playing against non-basic agents.

The ***Sticky*** agent was designed to be maximally predictable. It randomly selects an initial position only the first trial, and maintains it for the rest the game. In the Lemonade Game, predictability invites cooperation; as a result the sticky agent often outperforms much more sophisticated competitors. The remaining two agents changed positions at every trial, but following simple and predictable rules. At each trial i, the ***Roll*** agent chooses a position $p_i = (p_{i-1}+c) \% 12$, where c is an arbitrary constant. Similarly, the ***SquareRoot*** agent chooses $pi=p_{i-1}^{1/2}+c$.

2.1 Evaluation of Basic Agents

When self-centered agents play against each other, their performances are comparable,

[1] http://tech.groups.yahoo.com/group/lemonadegame/

with no agent being superior to the others. This implies neither being predictable (sticky) nor unpredictable (random) is inherently advantageous when playing against similarly self-centered agents. In a tournament (n=10,000 repetitions, see also Section 4), the Random, Sticky, Roll and Square Root strategies each scored 8.000 points.

3. Metacognitive approaches

In addition to the basic agents, we designed and tested a set of simple metacognitive players. The term *Metacognition* refers to benefiting from awareness of each player's performance and limitations, including one's own. The intial set of metacognitive agents was created by extending the basic agents with rudimentary metacognitive abilities. Five such agents were produced.

3.1 Basic Metacognitive Agents

StickySmart, an extension of Sticky, assumes that its opponents try to either maximize or minimize the distance from itself. Under the maximization assumption, it pays off to maintain its current location: the further your opponents are from itself the higher its score. Under the minimization assumption, maintaining its current location is catastrophic: the closer its opponents are to itself the lower is its score. In this case, StickySmart moves to the opposite location (over the diagonal), which restores the situation under the maximization assumption.

CopyCat assumes that at least one of its opponents has an effective strategy, and it simply copies it. Thus, CopyCat initially selects an opponent and subsequently just chooses the opponent' previous location, incremented by a constant c. A non-zero value of c is needed to avoid the special case the opponent plays sticky, and thus both agents end up in the same location.

CopyBest is a variation that also monitors whether copying an opponent is paying off.. When CopyBest is outperformed by both the other agents, it switches to copying the second opponent.

Cooperator assumes that cooperation is the key to success, actively trying to establish a partnership with another agent. In order to do so, Cooperator initially makes itself maximally predictable by playing like Sticky, It waits for an opponent to cooperate with its behavior and become a partner. Two partners are said to cooperate if they maximize the clock-distance between themselves, that is, they select locations that lay on the opposite sides of a diameter. Cooperator plays "sticky" as long as it does not repeatedly lose points. Otherwise, it switches partners.

StickySharp is an extension of StickySmart. When the two opponents of StickySmart cooperate, any sticky agent will lose. StickySharp tries to find a way out by issuing an alternative cooperation offer toward its opponents by playing Roll. StickySharp succeeds if one of its opponents breaks the existing cooperation agreement and enters a new cooperation agreement with StickySharp

Statistician minimizes cooperation but maximizes the efforts to predict the other agents. It maintains a history of its opponents' location choices and predicts the next location based on a weighted average of each opponents' previous locations, where most recent choices are weighted more then less recent ones. Because it maximizes only its own payoff, Statistician plays aggressively rather than cooperatively.

Finally, ***Strategist*** extends Cooperator: it preserves cooperation and adds altruism. First, Strategist assesses its opponents' predictability. If none of the two opponents is predictable, Strategist plays "sticky", assuming that at least one opponent will accept the offer to cooperate, which in turn makes the behavior of this opponent predictable. If only one opponent is predictable, Strategist cooperates with it, while continuing to assess the predictability of the other opponent. If both opponents are predictable, Strategist cooperates with either

the weaker or the stronger of its two opponents depending on its own performance. If Strategist's performance has been consistently good, the weaker opponent is chosen; otherwise, the stronger opponent is chosen to cooperate with. This discretionary selection ensures that both principles of cooperation and altruism are enforced. Note that Strategist cannot always be altruistic without affecting its commitment to cooperation. Due to the zero-sum nature of the game, helping the weaker opponent would weaken the stronger opponent, which would eventually force Strategist to switch partners. These repeated switches make Strategist's behavior look less predictable to its potential partners, thus making it less attractive as a partner and therefore less capable of cooperating.

3.2 A General Model of Metacognition

These agents as well as many cognitive models in the literature implement fixed strategies to solve specific problems. The term *metacognition* stems from the realization that human problem-solvers have the capability to choose among a repertoire of different strategies, monitoring and refining, monitoring and refining them while carrying out a task. In the context of the Lemonade game, metacognition is especially relevant as the value of any specific strategy depends on the configuration of the players in the game. For example, Statistician outperforms Random if it can predict and cooperate with the third player (which is often the case, see Figure 1), but as it cannot predict the Roll strategy, it is defeated in games against Random and Roll (it can't predict Roll).

In designing a truly metacognitive agent, we divided its action cycle in two steps. During the first step, predictions are generated for the other players in the game. These predictions depend on previously observed behavior of those players within the same game. Predictions are represented as a probability distribution over locations, indicating the estimated probability of a given opponent placing their lemonade stand at the given location in the next trial. The second step consists of making a decision about where to place one's own lemonade stand in the next iteration, iteration, based on the opponents predicted moves and the ensuing payoffs. This step may be very simple and simple consists of maximizing the expected payoff, but it might include more complex strategies to induce future cooperation with a player or to hurt a specific player that may be performing too well.

The key feature of the metacognitive agent is that it possesses different strategies for both the prediction and action steps, and it keeps monitoring them throughout the game. In particular, a measure of each strategy's utility is updated immediately after each trial. There are two different monitoring mechanisms. Prediction strategies can be evaluated in parallel: all strategies can be used to predict each opponent's location, and the true location provides can be used to evaluate all the predictions at the same time. On the other hand, action strategies that maximize long-term payoffs (e.g., focusing on making oneself predictable), can only be evaluated one at a time. As a consequence, it is easier to converge on prediction strategies than on action strategies.

Prediction Strategies

Prediction strategies produce a probability distribution $P(a)$ over the 12 locations for a given opponent. They maintain and access a record of the decision history of a particular agent within the current game.

All the prediction strategies rely on a n-gram representation, where the opponent's moves are recorded as chunks of n consecutive locations. This representation has been successfully used in sequence learning models (e.g., Lebiere & West, 1999). A range of different different strategies were generated by varying the size of n from 1 to 3, and encoding locations in absolute terms or in term of relative movements from the previous location.

Action Strategies

An action strategy uses the probability distributions of each opponent in order to determine the agent's move. We considered the following *elementary action strategies*.

Utility Optimization: This strategy simply consists in choosing the location that yields the highest payoff for that particular trial. Assuming the point of view of player *a*, and indicating its opponents as *b* and *c*, then the utility of a being at location l_a would be

$$u(a,l_a) = \sum_{l_b=0}^{11}\sum_{l_c=0}^{11} p'(b,l_b)p'(c,l_c)payoff(a,b,c)$$

where $payoff(l_a, l_b, l_c)$ is the reward that *a* receives if players *a,b,c* are in positions l_a, l_b, l_c, respectively, while $p'(x, l)$ is the estimated probability of an agent *x* choosing a specific location *l*.

The Sequence Learning agent in the tournament uses utility optimization as its action strategy.

Offer to Cooperate: This class of strategies is designed to be as predictable as possible. It includes two instances of the *Sticky* action strategy that choose different, but constant, locations. Note that these strategies offer to cooperate, but do not cooperate themselves; the action meta-layer will switch strategy if one of them proves unreliable.

Cooperation: This action strategy identifies the opponent that is best performing while being predictable. Predictability is measured as a single location being predicted with probability > 0.85. If the better-performing opponent is not predictable enough, the worse performing opponent is chosen if any prediction is available. The strategy then cooperates by choosing the location opposite the predicted of that opponent. If no reliable prediction can be made (during the initial steps), the cooperator plays consistently the same location in order to offer cooperation to another agent. Cooperation is the most successful one of the action strategies.

Imitation: As a further action strategy, we included the *Copy Cat* as described above.

The Metacognitive Layer

The *Meta* agent implements a hybrid combination of the described elementary strategies: A metacognitive layer combines all predictions and chooses an action strategy. This agent has a principled approach to choosing strategies, it is cognitively motivated, and was not optimized by hand to succeed in the task.

The agent's metacognitive layer evaluates both types of strategies using immediate feedback; in the case of prediction strategies, we evaluate the reliability of the estimates for the chosen location. In the case of action strategies, we use their immediate payoff to update their overall utility. To make the agent adaptive to changes in a strategy's payoff over time, we adopted a cognitively motivated approach known as *instance-based learning* (IBL, Gonzalez & Lebiere, 2003). This approach balances frequency and recency of the observed strategy performance. This approach is derived from the learning mechanisms in the ACT-R cognitive architecture. It has been applied with success to both sequence learning paradigms (Lebiere & Wallach, 2001) and games like paper rock scissors (Lebiere & West, 1999) and baseball (Lebiere, Gray, Salvucci & West, 2003). The key intuition behind this approach is that more frequent and more recent memories provide more reliable information, since the environment is less likely to have changed since the memory was formed. In the Lemonade Game, this means that opponents are more likely to follow the same strategies within short periods of time.

Our implementation of IBL is based on the learning mechanisms in the ACT-R cognitive architecture (Anderson et al, 2004) and involves memorizing an *episode* every time a strategy *s* is evaluated for a specific agent *a*. The episodes encode *t* (time step at which it occurred), *l* (actual location chosen by *a*), and p_l (predicted probability that *l* would be

chosen in the next step). A blend of the episodes is calculated, in which episodes are weighted by their relevance (did the strategy yield a high probability of the actual location?), their recency (a temporal decay is applied) and their frequency.

As in ACT-R, each episode has a base-level activation value that decays with time. For each agent *a* and strategy *s*, a confidence value *c(a,s)* is calculated as follows:

$$c(a,s) = \sum_{<t1,p_1>}^{episodes_{a,s}} \mathrm{p}_1 e^{\frac{b_c + \ln((t_0 - t)^{-d})}{T}} + \varepsilon$$

where b_c is an ACT-R base-level constant (held at 4.0), t_0 is the current time, *T* the Boltzmann temperature. *d* is a decay coefficient (*0.5* in ACT-R models). ε is a term for noise, sampled from a logistic distribution and scaled by a coefficient *k*. We arrive at a confidence value *c(a,s)* for given strategy *s* and opponent agent *a*.

To create a final, blended probability distribution *P'(a)* for an opponent agent *a*, the distributions from each prediction strategy *P(a,s)* are weighted by their confidence.

$$P'(a) = \frac{\sum_{s}^{strategies} c(a,s) * P(a,s)}{\sum_{s}^{strategies} c(a,s)}$$

The same method was used to evaluate the action strategies, except that rather than p_l we use the payoff as quality criterion for the strategy that is stored in each episode.

Parameters (*T, d, k)* as well as the subset of action strategies were fit to optimize the Meta agent's performance against the basic and advanced agents discussed above. The final parameter values were *T=0.2, d=0.7, k=0.004*.

4. Evaluation

All the described agents were evaluated in a simulated tournament that ran 100 rounds per game, running *n* repeated games for each combination of three different agents because of stochastic factors in many agents. We set *n* high enough such that scores were reliable (we do not give hypothesis tests for this reason). Agents were not allowed to retain information across games, and any adaptation (learning) occurred only within the 100 rounds of each game.

The outcome of each game strongly depends on the configuration of players. For instance, a combination of two agents may or may not end up cooperating, winning over the third player. We measured agent performance in three ways: the relative strength of the agents, their absolute performance, and the reliability of their performance with respect to changing third players. Figure 1 visualizes these measures. It plots the results of each of 13 strategies (Scored Agent, x-axis), playing a specific first opponent strategy (1st Opponent, y-axis). We aggregate over all 2nd Opponents, but show the variability.[3] Each circle's size is proportional to the payoff that the Scored Agent achieved when playing against a particular 1st opponent; large circles indicate higher payoffs. The shade of the circle visualizes the reliability of the Scored Agent's performance: dark circles indicate low variance across the different 2nd Opponents. A column of large dark circles marks a strong, reliable agent. A + sign indicates that the Scored Agent, on average, reaches higher payoffs than the 1st Opponent.

For example, let us consider *CopyCat* as our Scored Agent. It defeats both *Statistician and Random*. *CopyCat* also tends to reach high scores when *Sticky* is present, exploiting Sticky's predictability. However, it is also very susceptible to intervention by the third agent, as cooperating with *Sticky* makes *CopyCat* equally predictable. This may be exploited by a third agent, which may choose to hurt *CopyCat's* ambitions in a Kamikaze strategy, leaving *Sticky* as the winner. In a

[3] Each strategy combination played will yield data contributing to six circles in the figure.

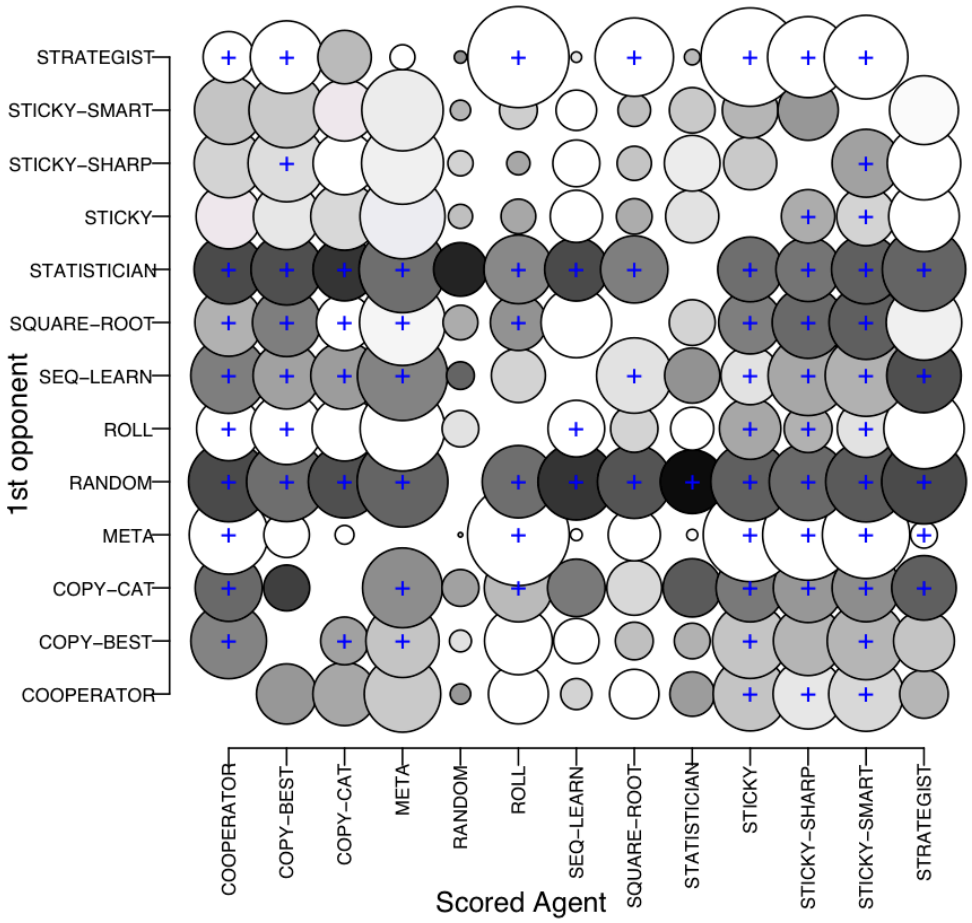

Figure 1: Performance of the strategies (x axis) when playing against other strategies (y axis). Sizes of circles indicate points achieved, while color of circles indicates variability of success across third players (dark: less variable).

game against *Random, CopyCat's* winnings are more reliable.

Meta as well as some cooperating agents (*e.g.*, *Cooperator*) achieve high and reliable results. The development of *Meta* showed that its cooperative action strategy was crucial to its success; that strategy differs from *Cooperator* only in its monitoring of the success of other players.

The data show how metacognition enables the success of different strategies, including *CopyCat* and *StickySmart*. It is interesting to notice , for instance, how the metacognitive *StickySmart* outperformed the non-metacognitive *Sticky*.

Table 1 reports the aggregated tournament results (all agent combinations, n=250 rep.) *Meta* consistently outperforms all other agents. The robustness of *Meta* was further tested by removing all but two basic prediction mechanisms (uni- and bigram models) and all action strategies except *Cooperation*. The resulting agent performed worse than the full *Meta* strategy (8.205 vs. 8.432). This shows that not only metacognitive abilities, but also a larger repertoire of strategies is beneficial in these type of tasks.

5. Conclusion

From the viewpoint of cognitive modeling, this paper examined agent collaboration in a three-player game known as the Lemonade Game. The Lemonade Game differs from other paradigms (e.g., Paper, Rock, Scissors) in that both being predictable and collaborating with an opponent improves an agent's chances to succeed. The series of simulations has shown that most successful strategies include offers to collaborate by making oneself predictable (*Sticky*) or more direct forms of collaboration (*CopyBest, Cooperate, Collaborate*). We found that monitoring of one's own and the opponents'

Table 1: Tournament results (average score, n=250)

Meta	8.432
Sticky Smart	8.311
Sticky	8.238
Sticky Sharp	8.222
Cooperator	8.214
Strategist	8.172
CopyBest	8.152
Roll Clock	8.039
CopyCat	7.948
SquareRoot	7.824
Sequence Learning	7.673
Statistician	7.602
Random	7.172

performance is crucial for making profitable choices: a predictable *Sticky* agent may be sandwiched between two exploiting opponents, and only monitoring allows it to escape them.

Yet, comparing the meta-cognitive Meta agent to some high-performing alternative agent, one would expect it to do slightly worse in some cases. Because of the inefficiency of its metacognitive analysis, it will be worse than the fixed strategy in the cases when that one is appropriate (which could be many, if it is very good). Again, any fixed strategy is likely to be poor for at least some combinations of opponents, and that is where *Meta* profits. The overhead of *Meta* over the fixed strategy can be kept small, while the price of a fixed strategy in a poor match can be very high. That tends to favor *Meta* overall, even if those cases are few. This can be seen as a special case of a general argument against narrow optimization in the development of cognitive agents, since that optimization is only meaningful within limited circumstances and its cost in loss of robustness outside of those circumstances is often left unspecified.

The key to robustness in unforeseen situations, such as being matched with an agent that one has never encountered, is the ability for an agent to evaluate the effectiveness of *all* its strategies, modify them as needed and select them accordingly.

References

Anderson, J. R., D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Quin. An integrated theory of mind. *Psychological Review*, 111:1036–1060, 2004.

Anderson, J. R. & Lebiere, C. L. (2003). The Newell test for a theory of cognition. *Behavioral & Brain Sciences 26*, 587-637.

Billings, D. (2000). The first international RoShamBo programming competition. *Int'l Computer Games Association Journal 23(1)*, 42-50.

Gonzalez, C. and C. Lebiere. Instance-based cognitive models of decision making. In D. Zizzo and A. Courakis, editors, *Transfer of knowledge in economic decision making*. Palgrave McMillan, New York, 2005.

Lebiere, C., Gonzalez, C., & Warwick, W. (2009). A Comparative Approach to Understanding General Intelligence: Predicting Cognitive Performance in an Open-ended Dynamic Task. In *Proc. Second Artificial General Intelligence Conference (AGI-09)*. Amsterdam-Paris: Atlantis Press.

Lebiere, C., Gray, R., Salvucci, D. & West R. (2003). Choice and Learning under Uncertainty: A Case Study in Baseball Batting. In *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*. pg 704-709.

Lebiere, C., & Wallach, D. (2001). Sequence learning in the ACT-R cognitive architecture: Empirical analysis of a hybrid model. In Sun, R. & Giles, L. (Eds.) *Sequence Learning: Paradigms, Algorithms, and Applications*. Springer LNCS/LNAI, Germany.

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 2000*, pp. 185-193. NL: Universal Press.

Lebiere, C., & West, R. L. (1999). A dynamic ACT-R model of simple games. In *Proceedings of the Twenty-first Conference of the Cognitive Science Society*, pp. 296-301. Mahwah, NJ: Erlbaum.

Rapoport, A., Guyer, M. J., & Gordon, D. G. (1976). *The 2X2 game.* Ann Arbor, MI: The University of Michigan Press.

Reitter, D. Metacognition and multiple strategies in a cognitive model of online control. *Journal of General Artificial Intelligence*, under review.

West, R. L., & Lebiere, C. (2001). Simple games as dynamic, coupled systems: Randomness and other emergent properties. Journal of Cognitive Systems Research, 1(4), 221-239

The authors acknowledge funding for this work from the Air Force Office of Scientific Research (FA95500810356) and the Defense Threat Reduction Agency (DTRA) (HDTRA1-09-1-0053).