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 examined using a competition of cognitively motivated agents, which inherit properties of human memory and decision-making, and simplistic, yet effective agents. We argue that metacognition constitutes the unique attribute that allows sophisticated agents to adapt to unforeseen conditions, cooperators and competitors.*

1. Introduction

Unlike other species, humans are not optimized to any specific natural environment or task, but they are very good at many things. At least in the long run, generalists agents like humans seem to be superior to specialist ones. Agents that are optimized to a particular ecological niche might succeed in current conditions, but once their environment changes they are likely to be suboptimal and soon extinct. 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 the notion that the flexibility and adaptivity that metacognition affords us is our main evolutionary advantage.

The same arguments can be applied to artificial as well as biological agents. In particular, the focus on optimality that dominates many fields of the cognitive sciences can be seen as counterproductive, and indeed as the very source of their controversial pattern of reaching short-term objectives while making little or no progress toward their overall goal. Artificial Intelligence has met a number of high-profile challenges (a world champion chess player, or a vehicle that can drive itself semi-autonomously) 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 none of these models can exhibit robust behavior in unforeseen situations. Finally, Machine Learning has produced algorithms that can use large amounts of data to adapt their performance, but only within the boundaries of their specific 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. In this game, three agents try to locate a fictional lemonade stand one of 12 possible locations (arranged in a circle and referred to as 0 through 11). The reward for each agent is the sum of the distances from the other two. A complete game consists of 100 consecutive trials. At the beginning of each trial, the three agents independently and synchronously decide the locations of their respective stands. The positions and rewards of all the agents are then calculated and revealed.

Many similar simple games feature either zero-sum competition (e.g., paper rock scissors; Billings, 2000) or the possibility of choosing between either cooperation or and competition (e.g., the prisoner's dilemma; Rapoport, Guyer & Gordon, 1976). The A unique feature of interest of this game is that it features permits a simultaneous combination of both cooperation (between two agents) and competition (against the third). As we will see, the emerging dynamics are quite interesting and prevent any notion of optimality. In order to succeed, the agents must adapt to the others' strategies, 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 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

These agents are "self-centered," in the sense that they ignore the actions of the other players. They correspond to basic approaches to the problem that can be used in isolation.

The ***Random*** agent chooses a random location independent of previous situations. The random agent is maximally unpredictable. This strategy can be successful in many games (e.g. zero-sum games such as in paper-rock-scissors (West & Lebiere, 2001) or adversarial games such as in the Prisoner's Dilemma (Lebiere, Wallach, & West, 2000). In the Lemonade Game, however, randomness precludes cooperation and effectively ensures poor results. Indeed, the random agent often received the poorest score in our tournaments.

The ***Sticky*** agent selects its initial position at random, and them maintains it throughout the game. This agent is designed to be maximally predictable. In the lemonade game, predictability is a powerful invitation to cooperation; as a result the sticky agent outperforms the others, even when its opponents are much more sophisticated agents. The ***Roll*** agent is also easily predictable. At each trial i, it chooses a position $p_i=p_{i-1}+c$ *(modulo 12))*, with c being an arbitrary constant. Similarly, the ***SquareRoot*** agent chooses $p_{i-1}^{0.5}+c$.

2.1 Evaluation

When self-centered agents play against each other, they do comparably well. No self-centered agent is clearly superior to the others. In particular, neither being maximally

predictable (sticky) nor maximally unpredictable (random) is inherently advantageous when playing against similarly self-centered agents, as shown in Table 1.

Table 1: Simple Agent Tournament Results

RANDOM	8.002
STICKY	8.002
ROLL	7.996

3. Metacognitive approaches

The term *Metacognition* refers to benefiting from awareness of each players performance and limitations, including one's own.

3.1 Basic Metacognitive Agents

Extending the basic agents with rudimentary metacognitive abilities created an initial set of 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 your current location: the further your opponents are from yourself the higher your score. Under the minimization assumption, maintaining one's current location is catastrophic: the closer one's opponents are to yourself the lower one's 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 tries to copy it. Thus, CopyCat picks an opponent and always chooses its previous choice plus an increment *c*. The increment is needed to avoid the special case the opponent plays sticky, and thus both agents end up in the same location. The best constant increment is *c*=6, which ensures that a loss is avoided in case the opponent plays sticky, and it is neutral in other cases. ***CopyBest*** is a variation that also monitors whether copying an opponent is working; when it is not, it switches to copying the other opponent.

Cooperator takes a more active and constructive approach, and assumes that cooperation is the key to success. In order to establish a cooperative relationship, Cooperator initially issues a request for cooperation by making itself maximally predictable (i.e., playing "sticky") and waits for an opponent to pick up the offer and cooperate (thus, 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. Thus, 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 opponent "helps the poor", that is, cooperates with the lower-scoring player.

Statistician maintains a record of its opponents' moves uses it to predict their subsequent moves. It then selects a location that is maximally distant from its opponents' predicted moves. Its predictions are 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, Statistican plays aggressively rather than cooperatively.

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

Cognitive models usually implement strategies to solve specific problems. The term *metacognition* stems from the realization that human problem-solvers have multiple strategies at their disposal, choosing and adapting them while carrying out the task: they are aware of their limitations. In the context of the Lemonade game, metacognition is especially relevant as strategies depend on the constellation of the players in the game. Some opponents may be willing to cooperate, or (at minimum) they are predictable and exploitable. For example, Statistician reliably outperforms Random because it can predict and cooperate with the third player, but it is defeated in games where this player is Roll.

We decompose the actions of metacognitive agents in each Lemonade trial into two steps. In 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. A prediction can be 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, in light of the expected payoff at each location, which can be calculated given the locations of all three stands. This step may be as simple as maximizing utility (joint probability and payoffs), but it may also include a strategy to induce future cooperation with a player or to hurt a specific player that may be performing too well.

Metacognitive agents can compare different strategies for both prediction and action. Each strategy's evaluation is updated immediately after each trial. We distinguish two possible monitoring mechanisms. Prediction strategies can be evaluated in parallel: all strategies may be used to predict each opponent's move, and they can all be evaluated after each trial. Action strategies, however, can only be evaluated one at a time if their long-term effects are to be considered. 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 use the decision history of that agent within the current game.

The prediction strategies use n-gram representation, where the opponent's moves there are recorded as series of n consecutive locations. This representation has been successfully used in sequence learning models (e.g., Lebiere & West, 1999) We provided a range of different algorithms by encoding relative and absolute movements of the agents separately. The Meta model, included different strategies are obtained by encoding series of n = 1, 2, or 3 choices, and encoding locations in absolute terms as well as relative movements from the previous agent location.

Action Strategies

An action strategy uses the predictions (a probability distribution for each opponent) in order to determine the agent's move. We considered the following *elementary action strategies*.

Utility optimization: This strategy chooses the location with the highest immediate expected payoff. Assuming the point of view of player *a*, and its opponents as *b* and *c*, then the utility of a being at location l_a would be

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

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. *p*' are the probability estimates for one agent choosing a specific location.

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 Agent

The *Meta* agent implements a hybrid combination of the elementary strategies. The 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 reward to updated their overall payoff. 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 both 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.

IBL 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*), p_l (probability predicted by *s* that *l* would be chosen in the next step). We then calculate a blend of the episodes, in which episodes are weighed by their relevance (did the strategy yield a high probability of the actual location?), their recency (a temporal decay is applied) and frequency.

We calculate a base-level activation value (as in ACT-R) for each episode, taking temporal decay into account. The activation is applied to the predicted probability for the chosen location in that episode:

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

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 pareto distribution. 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, n)* 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, n=0.004*.

4. Evaluation

We evaluated the strategies in a tournament that ran games with 100 rounds each, running every combination of three different agents. (We aggregated data from several repetitions of each combination.) 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 analyze three outcomes of agent pairings: 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. A + sign indicates that the Scored Agent (x-axis), on average, reaches higher payoffs than the 1st opponent (y-axis). Circle size indicates the payoff that the Scored Agent achieves on average when the 1st opponent is present in a game (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 third agents. A column of large dark circles marks a strong, reliable agent.

Consider *CopyCat* as our target (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: cooperating with *Sticky* makes *CopyCat* equally predictable. This may be exploited by a third agent, which may choose to destroy *CopyCat's* ambitions. In a game against *Random*, the winnings are more reliable.

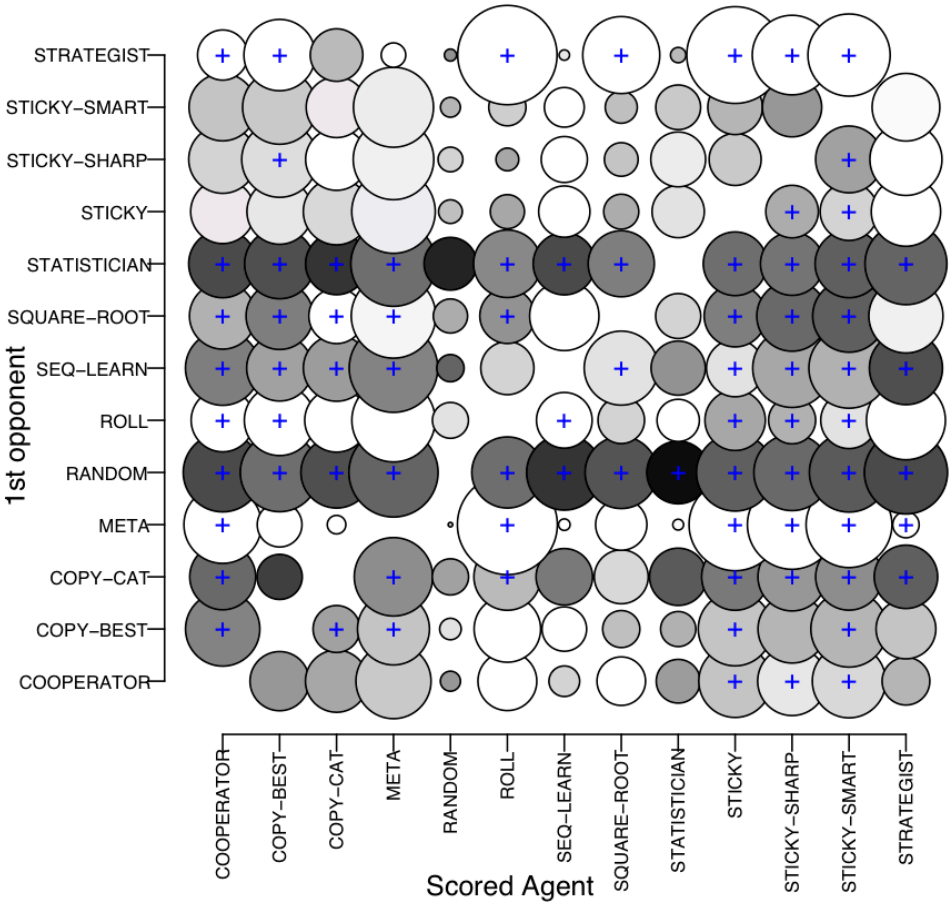

Figure 1: Performance of the strategies (x axis) when playing against other strategies (y axis). Sizes of circles indicate points achieved, while color of circles indicates variability of success depending on third player (dark: less variable). Plus signs indicate a numeric win of the scored agent over the 1st opponent.

Meta as well as some cooperating agents (Stick&friends, Cooperator) achieve high and reliable results. The development of *Meta* showed that its cooperative action strategy was crucial to its success; it differs from *Cooperator* only in its monitoring of the success of other players, cooperating with the more successful ones if predictable.

Meta as well as some cooperating agents (Stick&friends, 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, cooperating with the more successful ones if predictable.

Monitoring also plays a role in several of the strategies, including *CopyCat* and *StickySmart*. *StickySmart* outperformed the non-metacognitive *Sticky*.

Table 2 gives the aggregated tournament results (250 rep.). *Meta* consistently outperforms all other agents. The Meta strategy was further evaluated by removing all but two basic prediction mechanisms (uni- and bigram models) and all action strategies except *Cooperation*. In a further tournament (200 rep.) did the resulting simplified agent perform worse than the full *Meta* strategy (8.205 vs. 8.432). This shows that the hybridization of strategies is beneficial.

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 one agent's chances to succeed. A series of

Table 2: Tournament results (average score)

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

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; performance is crucial for making profitable choices. 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 meta 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). Still, 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. *International 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 *Proceedings of the 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