Title
Modeling the Opponent Facilitates Adversarial Problem Solving

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 20(0)

Authors
Burns, Bruce D.
Vollmeyer, Regina

Publication Date
1998

Peer reviewed

Modeling the Opponent Facilitates Adversarial Problem Solving

Bruce D. Burns (burns@rz.uni-potsdam.de)
Regina Vollmeyer (vollmeye@rz.uni-potsdam.de)
Institut für Psychologie, Universität Potsdam
14415 Potsdam, Germany

Abstract

Competition can be seen as Adversarial Problem Solving (APS), thus ideas from problem solving research can be applied to it. We tested if better modeling of the opponent led to better performance in APS using a zero-sum game played by pairs, but with no obvious skill component. We replicated earlier results that showed that third-order modeling (i.e., what I think my opponent thinks of me: R3MA), but not second-order model (i.e., what I think about my opponent: R2MA) correlated with performance. We also manipulated who was played (same person as in an earlier game, or a predetermined sequence) and who players were told their opponent was (same or different). Players performed better when they could apply the appropriate model (i.e., what they were told matched the opponent). Therefore, we showed that more accurate modeling of an opponent can lead to better APS. However, the critical aspect of modeling may be third-order modeling accuracy. We also found support for a game theory analysis of the task.

Competition as Adversarial Problem Solving

In May 1940, a British trawler was on its way to a rendezvous with a German boat in the English Channel. Aboard the trawler were two double agents (German agents who were actually British agents) code named "Snow" and "Biscuit". Snow was about to introduce Biscuit as his subagent to be taken to Germany for training, but in the murky world of double-agents, British intelligence thought it unwise for either to know that the other was working for the British. However, both guessed this fact. Biscuit then formed the opinion from Snow's behavior and conversation that he was a genuine German agent who would turn Biscuit in as soon as they met the boat. Meanwhile Snow decided that Biscuit was a genuine German agent who would betray him. So Snow did everything possible to convince Biscuit that he, too, was a genuine German agent, which confirmed Biscuit's fears. Eventually, to avert apparent disaster, Biscuit locked Snow in his cabin and returned the boat to England. By trying to do what appeared to be the right thing each had contributed to the ultimate failure of the mission (related by Watzlawick, Weakland, & Fisch, 1974, pp. xi-xii).

Problem solving research has rarely considered the type of problem that Snow and Biscuit were trying to solve, one in which another person constituted the problem task. Studying static problems simplifies research, however, many tasks which fit the definition of problem solving (i.e., trying to reach a goal) are not static. Unless the definition of problem solving is to be restricted, any general theory of problem solving must incorporate such tasks.

Situated cognition research has emphasized the need to take into account other cognitive agents in the environment (e.g., see Greeno, Moore, & Smith, 1993). However, such research has not considered competition between those agents. Competition between two people may provide a useful situation for the study of interactive tasks, because the goals of the adversaries are relatively clear. Yet in competitive tasks, a critical element is that people react, and react to reaction, just like Snow and Biscuit did.

Competition has been studied within social psychology, but often only in contrast to cooperation (see Axelrod, 1980). Competition has also been studied in economics and game theory (see Rapoport, 1960). Yet much of this research has been normative, and has focused either on explaining or describing aggregate outcomes of competition, or reactions to it. Little of this research has shed light on the cognitive processes people use when trying to *successfully* compete. However, in labeling competition *Adversarial Problem Solving* (APS), it has been suggested that competition can be seen as problem solving (see Gilhooly, 1988; Holding, 1989; Thagard, 1992), and thus may make similar cognitive demands.

Thagard (1992) proposed a set of cognitive processes that could be involved in successful APS. In particular, Thagard argued that across many domains of APS, it is the ability to model the opponent, and further to model how the opponent models you, that is critical for success. The mental model that is formed can then be used to anticipate and counter the opponent's actions. Such models are a result of interpersonal perception, a topic with a long history (see Kenny, 1994). However, Thagard's proposal concerns *accuracy* in interpersonal perception, a topic that has received less attention, partly for methodological reasons (see Kenny, 1994, ch. 7). More precisely though, the implication of Thagard's proposal is that accurate modeling is not the critical determinant of success in any one competitive event; instead, the critical factor is having a *more* accurate model than your opponent has of you.

A form of modeling was studied by Ruscher and Fiske (1990), who showed that individuals in competition focus more than non-competitors on trying to individuate their opponents, particularly on task-relevant attributes (although competing *groups* may tend to stereotype opponents, see Brewer, 1979). However, they did not examine whether trying to individuate the opponent improved performance. In this paper, we attempted to empirically test the proposal that when a player can apply a better model of the opponent, then they will have more APS success (i.e., reach the goal of the competition) than the opponent.

A Methodology for Studying APS

While it seems to make sense that modeling an opponent should play a role in success during APS, there is little empirical support for this intuition. One difficulty with examining APS is that it often involves tasks that are very complex. For example, Holding (1989) pointed out that taking into account an opponent in a task could increase the search space of problem states enormously. Therefore, we needed a simple, manipulatible, competitive task in order to test hypotheses about APS. The *prisoner's dilemma* game would not be appropriate because it involves pressure towards both cooperation and competition. Thus the goals that the players choose can change, making it hard to assess if they have reached them. Therefore, we used a purely adversarial game: a repeated, zero-sum, two-player game in which players simultaneously select the number ONE, TWO, or THREE. A *Chooser* and an *Avoider* player made their selections in secret, then each player was told what the opponent selected. If the two players' selections coincided, then Chooser won the amount of points that corresponds to the number they both selected and Avoider lost the same amount of points. If the players' numbers do not match then, Avoider wins a point and Chooser loses a point. Thus, Chooser could win up to three points, but Avoider could only ever win one point. To offset this, Avoider was expected to win more individual trials of the game. Table 1 presents the pay-off matrix for this game.

Table 1: Pay-off matrix for the game. Outcome for Chooser is the first number in the pair while the second number is the outcome for Avoider.

		Avoider's selection		
		ONE	TWO	THREE
Chooser's selection	ONE	+1 / -1	-1 / +1	-1 / +1
	TWO	-1 / +1	+2 / -2	-1 / +1
	THREE	-1 / +1	-1 / +1	+3 / -3

Assessing Modeling

To assess the participants' models of each other, we reasoned that if players can accurately model their opponent, then they should be able to accurately assess characteristics of their opponent. Therefore, we had participants use seven-point scales to respond to a set of word pairs that indicated their general assessment of their opponent, such as *negative - positive*.

However, there may be different aspects of these models, in particular, there may be an important distinction between your model of what you think your opponent is like, and what you think is your opponent's model of you. Hymans (1989) suggested that the latter may be critical for using deception. We will refer to these two aspects as *second-order* and *third-order* models, following Dennet's (1978, pp. 274-275) proposal of second- and third-order intentions. Therefore, players rated the set of word pairs for three targets: themselves (the *self scale*, representing a first-order model), their opponent (the *opponent scale*, representing a second-order model), and how they thought their opponent would rate them (the *opponent-self scale*, representing a third-order model).

It was not clear a priori which personal characteristics should be critical for successful APS. However, our proposal was that the critical aspect of modeling for successful APS was how accurate you are in what you thought about your opponent, rather than your actual assessment of the opponent. Thus we used the responses from all six scales - the three scales from each player - to derive the accuracy of a player's second- and third-order models of the opponent. To assess a player's second-order accuracy for modeling an opponent we compared the player's rating of the opponent (the opponent scale) with the opponent's rating of him or herself (the self scale), by summing the absolute differences between the ratings of the same item on these two scales. Third-order modeling accuracy for the player was derived by summing the absolute differences between items on the player's rating of 'how you think the opponent would rate you' (i.e., the opponent-self scale) and items on the opponent's actual rating of the player (i.e., the opponent's opponent scale).

If APS merely requires you to be better than your opponent, then it implies that absolute modeling accuracy is less important than *relative* modeling accuracy. Therefore, the unit of analysis in these experiments should be the pair of players, rather than individual players. So we used two measures of relative modeling accuracy calculated for each pair of players: the *relative second-order modeling accuracy* (to be referred to as *R2MA*), which was Avoider's second-order modeling accuracy subtracted from Chooser's second-order modeling accuracy; and, the *relative third-order modeling accuracy* (to be referred to as *R3MA*), which was the Avoider's third-order modeling accuracy subtracted from the Chooser's third-order modeling accuracy.

A Game Theory Analysis

Games like this can be analyzed using game theory (see Rapoport, 1960). Game theory is essentially descriptive as it seeks to analyze a state of affairs that exists, or to predict a future equilibrium state. It does not address what processes bring about this state of affairs, thus, game theory can only predict the equilibrium point the players should tend towards. However, it provided a useful tool for analyzing what should happen in our game. The critical concept was that of a *mixed strategy*. A mixed strategy assumes that on each trial an alternative is chosen stochastically and independent of the previous choices. Von Neumann (1928) showed that for any finite, constant sum, two-person game, there exists a *mixed-strategy equilibrium* that specifies the probability distribution with which each possible choice should be randomly made. The mixed-strategy equilibrium is defined as the optimal probability distribution for both players, such that if both players use the specified distribution for random choices, then neither player can gain by deviating from the distribution. This probability distribution depends on the pay-off matrix of the game. For our game, the distribution was selecting ONE with probability .46 (more precisely, 6 out of every 13 trials), choosing TWO with probability .31 (4/13), and THREE with probability .23 (3/13). With this distribution, Avoider would be expected to win at least +.077 (1/13) points per trial, if Avoider imitates a stochastic process. In this game, the mixed-strategy equilibrium for Chooser is the same set of probabilities as those for Avoider, even though Chooser has an expected outcome of -.077 points per trial.

A Test of this Methodology

Studies by Burns (1993) and Burns and Vollmeyer (in press) applied this methodology to APS. As predicted by game theory, the Avoider players had a clear advantage in these studies. Further, the mean proportions with which Avoider selected ONE, TWO, and THREE were close to the mixed-strategy equilibrium. However, the Chooser players' proportions were almost equal for each selection, although the mixed-strategy equilibrium predicted the same proportions as for Avoiders.

Relative modeling accuracies were calculated as outlined above. There was no evidence that R2MA correlated with performance in the game as measured by the cumulative score after the final trial of the game (in Burns, 1993: $r[48] = .02$, $p = .88$; Burns and Vollmeyer, in press: $r[48] = -.19$, $p = .18$). However, success and R3MA correlated, (in Burns: $r[48] = .29$, $p = .50$; Burns and Vollmeyer: $r[48] = .37$, $p = .006$). Thus, it appeared that modeling was associated with success, but perhaps only the third-order component was critical. For no single item were raw responses associated with success, nor was relative accuracy on any item particularly associated with performance, instead, only the overall accuracy was predictive.

An Experiment: Manipulating Modeling

While Burns (1993) and Burns and Vollmeyer (in press) supported the claim that modeling the opponent was important for success, these were purely correlational studies. To more firmly establish a causal link from modeling to performance, we directly manipulated the usefulness of players' models. This allowed us to generate and test a specific prediction: If modeling the opponent is useful in APS, then when people can use their models more effectively they should do better than when they cannot.

To test this hypothesis, we had participants play the game twice. In Game 1, pairs of participants played 25 trials of the game against each other. In Game 2, they played 30 trials (starting from a score of zero again) but we manipulated who participants played (the *played-opponent*), and who they were told that they were playing (the *told-opponent*). The played-opponent was either *same* (i.e., each other again), or *different* (i.e., both played against a predetermined sequence of numbers). However, the two members of the pair were in opposite told-opponent conditions: One player was told that the opponent was the *same* as in Game 1, and the other player was told that the opponent was *different*. Therefore, one of the players should have had a modeling advantage over the other. As illustrated in Table 2, one player could apply the appropriate model, while the other could not.

Table 2: Illustration of the design of the experiment. Factors of who both players' opponent was, and who Avoider player was told they were playing (Chooser player was told the opposite), were crossed so that Avoider had the opportunity to apply an appropriate or an inappropriate model (Chooser had the opposite opportunity).

	Avoider told *same* opponent	Avoider told *different* opponent
Played *same* opponent	apply *appropriate* model	apply *inappropriate* model
Played *different* opponent (a sequence)	apply *inappropriate* model	apply *appropriate* model

If modeling the opponent is a determinant of success, then the player who could apply the appropriate model should do better than his or her opponent who had inaccurate information, independent of who they actually played and what they were told. Therefore, the told-opponent condition should interact with the played-opponent condition, and the performance by Avoider players in the appropriate model cells of Table 2 should be higher than those in the inappropriate model cells.

Method

Participants. One hundred and ninety-two participants from the University of California, Los Angeles introductory psychology subject pool took part for partial course credit.

Procedure. Participants went to separate rooms on their arrival, in order to eliminate all possibility of communication between players (the game was played via computer), then they were given the instructions for how to play the game. These included the pay-off matrix shown in Table 1, although shown in a way that fit with a participant's assigned role (either Chooser or Avoider, randomly assigned). In Game 1, they played against each other for 25 trials. On each trial, the two players had as much time as they liked to decide between selecting ONE, TWO, or THREE. So that they always had access to their own and their opponent's history of selections, they kept a record of all selections on a scoring sheet. Once Game 1 was completed, they were given the self, opponent, and opponent-self scales to complete. Each scale consisted of the same ten items, and each item consisted of a pair of words which anchored the ends of a seven-point interval. These ten word pairs were: *risk-taking - risk-avoiding; humorous - serious; negative positive; hard - soft; rational - intuitive; foolish - wise; weak - strong; pessimistic - optimistic; severe - lenient; cruel - kind.*

For Game 2, they were told that they had either the same or a different opponent, and had their scores reset to zero. They played the predetermined opponent (same opponent as in Game 1, or different) for 30 trials. The different opponent was a predetermined sequence so that the effects of different distributions could be tested (not reported here). Abric and Kahan (1972) found that playing a sequence did not affect players' strategies in a prisoner's dilemma game, if they thought they were playing a person. Our participants were given no indication that they might not play a person.

After finishing Game 2, participants completed the self, opponent, and opponent-self scales again.

Results

Performance. For the 96 pairs, the first game was won by Avoider 61 times, and the mean score for Game 1 (*score1*) was 1.35 (SD = 8.11) in favor of Avoider. To test if the Game 2 manipulation affected players' performance, we calculated their *relative score* for Game 2: score1 from Game 1 was subtracted from the difference between a pair's absolute scores for Game 2. Such a relative score for Game 2 controlled for individual differences which may have been reflected by Game 1 performance.

Table 3: Mean relative scores for Game 2 (SD in parentheses) in favor of Avoider for each told-opponent (same vs. different) and played-opponent (same person vs. sequence) conditions. For each cell of the table n = 24.

	Avoider told *same* opponent	Avoider told *different* opponent
Played same person	3.25 (10.13)	-2.75 (11.93)
Played sequence	1.67 (8.69)	5.13 (9.80)

The mean relative scores for each group are presented in Table 3. A 2x2 ANOVA on relative scores found no main effects of played-opponent, $F(1, 92) = 2.23$, $p = .13$, or told-opponent, $F(1, 92) = .37$, but as predicted there was a significant interaction between these two factors, $F(1, 92) = 5.16$, $p = .025$. Comparing appropriate verse inappropriate model conditions, the former had a mean relative score of 4.19 (SD = 9.90) while the latter had a mean relative score of -0.54 (SD = 10.56). As predicted, when the appropriate model could be applied, performance was better, $t(94) = 2.26$, $p = .026$.

Modeling Accuracy. We calculated R2MA and R3MA measures as outlined above. In detail, the following calculation was made for a player's absolute second-order modeling accuracy: For each item on a player's opponent scale, the absolute difference between its rating and the same item's rating on the opponent's self scale was calculated, then the differences for the ten items were summed. Therefore, a low sum indicated accurate modeling. In order to calculate R2MA, the absolute second-order accuracy for the Avoider player was subtracted from the absolute accuracy for the Chooser player. Thus, a positive R2MA indicated that Avoider was a more accurate second-order modeler, the higher the better. The R3MA measure was calculated in the same way, except that the absolute differences were calculated between items on a player's opponent-self scale and the opponent's opponent scale. Like R2MA and R3MA, all of our dependent measures were difference scores derived from the Avoiders' and Choosers' individual measures. As a convention, better performance on our measures was always stated in terms of the Avoider being better, so positive scores indicated better performance by the Avoider, and negative scores indicated better performance by the Chooser. Such consistency meant that we always predicted positive correlations.

In this experiment, R2MA and R3MA measures were calculated at two points: after Game 1 and after Game 2. As the modeling accuracy of players who played against a predetermined sequence would not be expected to correlate with performance, we only examined Game 2 correlations for pairs who played each other in Game 2. There were only 46 such pairs because the rating data was incomplete for two pairs.

Again, the R2MA measures were not associated with performance (score1 with R2MA after 20 trials: $r = .059$, $p = .70$; relative score with R2MA after Game 2, $r = -.10$, $p = .50$). However, R3MA after Game 2 was correlated with relative score, $r = .35$, $p = .018$. Further analysis showed again that relative modeling accuracy measure derived for individual word pairs from the scales did not correlate with performance. The correlation between score1 and R3MA for Game 1 was not significant, $r = .19$, $p = .22$. Perhaps it takes time to learn enough about the opponent.

To test whether R3MA was affected by our manipulations, we calculated mean R3MA from after the manipulated Game 2. As we expected, when pairs played against each other in Game 2, R3MA was higher when Avoider was told that the opponent was the same, $M = 3.04$ ($SD = 5.84$, $n = 24$), than when told that the opponent was different, $M = -1.09$ ($SD = 8.56$, $n = 22$), although this difference was not quite significant, $t(44) = 1.93$, $p = .060$ (however, this test had low power). This trend was not inconsistent with modeling differences being responsible for the better performance of players who were given a situation in which the appropriate model could be applied, and it provided some evidence for the validity of the R3MA measure. However, if R3MA measures were some type of artifact, then they should have been affected even when the opponent was not a person. But when pairs played the sequence, there was no difference between mean R3MA for when Avoider was told that they had the same opponent, $M = 1.13$ ($SD = 7.00$, $n = 24$), and for when they were told that the opponent was different, $M = 1.71$ ($SD = 6.13$, $n = 24$), $t(46) = .31$, $p = .76$. Similarly, the correlation between R3MA and relative score for players who competed against the sequence was also not significant, $r(48) = .12$, $p = .43$. This adds some plausibility to our interpretation of R3MA as measuring relative modeling accuracy.

Quality of Distributions of Selections. The mean proportions of trials on which Avoider players gave each selection in Game 1 were again close to the game theory equilibrium. Avoider chose ONE, TWO, and THREE with mean proportions, .42 (SD = .12), .29 (SD = .19), .29 (SD = .11), respectively. However, as before, Chooser made these three selections with proportions, .32 (SD = .13), .35 (SD = .10),

.33 (SD = .13). In Game 2, when players competed against a sequence, the nature of the sequence affected their distributions (the sequence type results are not important for the issues addressed by this paper, so they are not reported here). However, players who continued to compete against each other had proportions like those in Game 1: for Avoider, .43 (SD = .13), .29 (SD = .07), .28 (SD = .11); for Chooser, .32 (SD = .11), .35 (SD = .10), .33 (SD = .12).

Analyzing the distribution of selections allowed us to ask whether the told-opponent manipulation could have affected performance via changes to players' distributions. To test this, we calculated *relative distribution extremity* (*RDEx*) measures for Game 1 and Game 2. To calculate this, we first summed how much a player's proportion for selection of ONE was above .33, and proportions for TWO and THREE were below .33, thus it was a measure of the quality of a player's distribution. RDEx was the Chooser's sum minus that for Avoider. We analyzed RDEx for players that competed against each other in Game 2 with a 2x2x2 ANOVA on RDEx with factors for game (1 or 2), role (Avoider or Chooser) and told-opponent (told same or different opponent). We found a large effect of role, $F(1, 46) = 34.14$, $p < .001$, but no effect of game, $F(1, 46) = 1.90$, $p = .18$, or told-opponent, $F(1, 46) = .16$. No interactions were significant (all Fs < 1.0). Therefore, distributions did not change between games, and RDEx was not affected by who players were told was the opponent.

Could quality of distributions be a third factor that explains the correlation between R3MA and performance? To test this, we did a regression analysis of RDEx and R3MA measures on relative score. We found a significant multiple $R = .49$, $F(2, 43) = 6.79$, $p = .003$. For RDEx, $\beta = .35$ was significant, $t(44) = 2.61$, $p = .013$, and so was the β (.32) for R3MA, $t(44) = 2.39$, $p = .021$. Therefore, RDEx and R3MA both contributed to performance, but separately. Burns and Vollmeyer (in press) also found this.

Discussion

As anecdotal evidence suggests, it appeared that accurately modeling the opponent was associated with success in APS. This was shown in two ways: 1), the relative accuracy of third-order modeling was associated with success, which replicated previous studies; 2), when we manipulated the situation to provide the opportunity to use the appropriate model, players performed better and had greater modeling accuracy, as predicted.

Is Modeling Causal for Success? How strongly can we argue that relative third-order modeling accuracy is causally related to success? We never directly manipulated third-order modeling, which weakens any claim of causality. However, taken together, the results provided good evidence that better modeling led to better performance, because they argue against many otherwise plausible alternative explanations for the relationship between performance and modeling. First, we were able to affect performance and modeling exactly as predicted by manipulating the situation so that modeling could be more or less effectively applied. There would be no reason to expect an advantage of being able to apply the appropriate model in Game 2 of the experiment, if something was not learnt in Game 1 that was specific to the opponent. Further, we found that when they could apply the appropriate model, players tended to be more accurate at modeling their opponent, as would be expected if our manipulations had their effect via modeling. Second, if our results were the product of some artifact of how winners and losers responded, then our experimental manipulations should have affected modeling just as they affected performance, even when the players competed against a fictitious opponent. However, when they played against a predetermined sequence there was no impact of whether they were told that they had the same or a different opponent, and performance did not correlate with R3MA.

We did find that using a better distribution of selections (as measured by RDEx) was also associated with better performance. So it was critical to show that our measure of relative modeling accuracy was not just a substitute for RDEx. Regression analysis showed that the influence of RDEx on performance was separate from that of R3MA. Further, there was no evidence that distribution quality was affected by the manipulations in the experiment, despite these manipulations affecting performance. Therefore, modeling appeared to be the only mechanism through which our manipulations could influence performance.

We have still not completely eliminated the possibility that some unmeasured factor could actually be responsible for the performance-modeling relationship, for example, intelligence or motivation could be proposed as candidates. Epstein and Harackiewicz (1992) did find complex effects of achievement motivation in competition, although they found no effects of their manipulations of motivation on performance. Yet although quality of distribution would appear to be the likely mechanism via which motivation or intelligence could affect performance, distribution quality was not influenced by the manipulations in our experiment. This finding, together with the finding that modeling accuracy and quality of distribution accounted for separate variance in performance, makes it hard to explain the performance-modeling relationship as due to a third factor that correlates with both relative modeling accuracy and performance. The barrier for any speculated third factor is to propose a mechanism through which it could affect performance other than modeling or quality of distribution.

Further Issues. We have not attempted to determine the size of the contribution of modeling accuracy to success in APS. To do so would require developing the best possible measure of modeling. However, our aim was to determine the *existence* of a model effect rather than its size. The importance of modeling may depend on the particular game and its demands, and on how evenly matched the competitors are on other factors. In our game, we tried to match players as evenly as possible by removing all specific skill components from the game, but still found that quality of distribution was a component of success. Yet our game also provided very little information on which to base a model of the opponent. In normal competitive situations, more information is available, and thus the opportunity for those who can exploit it should be greater. Therefore, it is possible we have underestimated the amount of variance in performance accounted for by modeling accuracy.

Although we found that third-order modeling but not second-order modeling accuracy was important, we should be cautious about interpreting this null result as showing that second-order modeling is unimportant. In particular, the R2MA measure relies on accurate self-ratings, but there are debates over the accuracy of self-ratings.

Our findings also raise an intriguing question that we cannot answer at this time: is APS a general skill? Some people may be better competitors because they find competition motivating, but perhaps modeling is a general skill that underlies better competition, independent of task specific factors. Even if higher motivation were responsible for better modeling, then it is still unclear exactly how it would lead to better models. Of course, it could be just that sometimes people happen to hit on the right model in a particular game and then use it to win (unless they do not know that they can use it, as in the inappropriate model conditions), and conversely, they sometimes hit on the wrong model and use it to lose. Thus, having the right model may be responsible for success, but it may not be a general ability. However, it is possible that some people are better at modeling, giving them an advantage in general.

How could modeling accuracy facilitate performance? Perhaps it helps select between safe and risky choices, or it allows players to use deception, or it might help them decide when to deviate from the equilibrium strategy suggested by game theory. This is an issue for further study.

If modeling is a skill, then it might support Hymans' (1989) suggestion that deception might be an evolved skill. Deception is a quintessential human skill, as is implied by its role in the Turing Test for artificial intelligence.

In many ways our results were consistent with the game theory analysis. Avoiders won by about the amount predicted by the mixed-strategy equilibrium, and were close to the frequency distribution of selections predicted by this equilibrium. However, the Chooser players did not match this distribution, and Avoiders did not appear to exploit this. Perhaps Chooser's distribution arose because even if the mixed-strategy equilibrium was the best for the Chooser, it was still a losing strategy. As well as game theory being useful for analyzing the game, the results in turn provided some support for the psychological validity of the game theoretic concepts used.

Conclusions

Given that anecdotal evidence supports the claim that modeling is important in APS, in one sense, our results were not surprising, although the finding that third-order modeling was particularly important goes beyond the anecdotes. However, we have demonstrated a methodology capable of empirically demonstrating this intuition and suitable for beginning to study this phenomenon. Research on competition has often focused on factors that bring about competition rather than cooperation, and how competition is conducted. Rarely have the factors behind successful APS been examined, even though competition is a pervasive aspect of life. Our results suggested that modeling is an important component of successful APS just as interpersonal perception has been seen as an important component of other social situations. Good modeling can even be a way of avoiding competition, as it would have been for Snow and Biscuit. This work presents an opportunity and a challenge to problem solving research and theory: to address interactive tasks.

Acknowledgments

This research was partly supported by a stipend to the first author from the Humboldt Foundation.

References

Abric, J. C., & Kahan, J. P. (1972). The effect of representation and behavior in experimental games. *European Journal of Social Psychology, 2*, 129-144.

Axelrod, R. (1980). *The evolution of cooperation*. New York: Basic Books.

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivation analysis. *Psychological Bulletin, 86*, 307-324.

Burns, B. D. (1993, June). *Modeling the opponent in adversarial problem solving*. Poster session presented at the Fifth Annual Convention of the American Psychological Society, Chicago, Illinois.

Burns, B. D., & Vollmeyer, R. (in press). Modeling the adversary and success in competition. *Journal of Personality and Social Psychology*.

Dennet, D. C. (1978). *Brainstorms: Philosophical essays on mind and psychology*. Montgomery, VT: Bradford Books.

Epstein, J. E., & Harackiewicz, J. M. (1992). Winning is not enough: The effect of competition and achievement orientation on intrinsic interest. *Personality and Social Psychology Bulletin, 18*, 128-138.

Gilhooly, K. J. (1988). *Thinking: Directed, undirected and creative (2nd ed.)*. San Diego, CA: Academic Press.

Greeno, J. G., Moore, J. L., & Smith, D. R. (1993). Transfer of situated learning. In D. K. Detterman & R. J. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, and instruction*, Norwood, N.J.: Ablex.

Holding, D. H. (1989). Adversarial problem solving. In K. J. Gilhooly (Ed.), *Human and machine problem solving*. New York: Plenum Press.

Hymans, R. (1989). The psychology of deception. *Annual Review of Psychology, 40*, 133-154.

Kenny, D. A. (1994). Interpersonal perception: A social relations analysis. New York: Guilford Press.

Rapoport, A. (1960). *Fights, games, and debates*. Ann Arbor, MI: University of Michigan Press.

Ruscher, J. B., & Fiske, S. T. (1990). Interpersonal competition can cause individuating processes. *Journal of Personality and Social Psychology, 58*, 832-843.

Thagard, P. (1992). Adversarial Problem Solving: Modeling an opponent using explanatory coherence. *Cognitive Science, 16*, 123-149.

Von Neumann, J. (1928). Zur Theorie der Gesellschaftspiele [On a theory of games]. *Mathematische Annalen, 100*, 295-320.

Watzlawick, P., Weakland, J. H., & Fisch, R. (1974). *Change*. New York: Norton.