In Proceedings of the Sixth International Conference on Cognitive Modeling, 70-75.
Mahwah, NJ: Lawrence Earlbaum.

A Model of Human Behavior in Coalition Formation Games[1]

Alex K. Chavez (achavez@sas.upenn.edu)
University of Pennsylvania; 4110 Baltimore Avenue
Philadelphia, PA 19104 USA

Steven O. Kimbrough (kimbrough@wharton.upenn.edu)
University of Pennsylvania; 3730 Walnut Street; Suite 500, Room 565
Philadelphia, PA 19104 USA

Abstract

Coalition formation is a type of mixed-motive game in which n players strategically negotiate to secure positions in advantageous contracts. In this study, we find that systems of agents which learn by a simple linear updating rule successfully can model the outcomes of human players across five coalition formation games studied experimentally by Kahan and Rapoport (1974). "Greedy" agents, which are deterministic and maximizing in their selection of whom to offer to, achieve outcomes on par with humans within a few hundred trials. In comparison, "Matching" agents which use *probability matching*[2] for selecting whom to offer to achieve overall outcomes qualitatively similar to those of humans, but not as closely as the Greedy agents do.

Introduction

There recently has been interest in modeling human learning in strategic games, both in economics (Roth and Erev, 1995; Mookherjee and Sopher, 1997; Erev and Roth, 1998; Camerer and Ho, 1999) and in cognitive science (e.g., Ritter and Wallach, 1998; Lebiere et al., 2003). Previous efforts have focused on modeling learning in repeated, extensive and normal form games. Here, we extend the literature by modeling the outcomes reported in an experimental study of a characteristic function form game (Kahan and Rapoport, 1974).

Our goals in this paper are to see if agents endowed with a minimal learning rule can model the play by humans in the coalition formation game of Kahan and Rapoport (1974), and to compare the two selection rules–Greedy, and Matching–to be described in detail later. Because the human data available is limited, our approach is not to produce entire trajectories of play, but to see if the overall behavior of our agents corresponds to that of humans.

Previous work of Dworman et al. (1995a,b,c, 1996) used genetic programming to explore coalition games similar to those in this paper, where agents were explicit strategies in the space of offers and players. However, such agents cannot cannot be said to have a cognitive component, as each agent represents a single rule. In contrast, our agents represent players of low or minimal rationality, who must make decisions on how much to offer and to whom. Finally, while characteristic function games have been investigated widely in terms of outcomes or solutions through both experimental and mathematical approaches (see Kahan and Rapoport, 1984; Uhlich, 1990, for overviews), less attention has been paid to the learning processes associated with such games.

In the next section of our paper, we describe the coalition formation game, and also present the experimental study of Kahan and Rapoport (1974). We then describe our model, and finally present simulation results. While we only explore three-person games with a certain characteristic function, we can hope our results generalize to games with larger numbers of players.

Coalition Formation

Description

In the coalition formation game with a set of n players P, any of the players can join together to form a *coalition* $S \subseteq P$. Such a coalition can attain a guaranteed payoff of $v(S)$, called the *coalition value*, where v is defined over all coalitions and is known as the *characteristic function* of the game; along with P, this defines the coalition game. While it generally is advantageous for a player to be included in some coalition,[3] it is up each member of the coalition to secure a portion of $v(S)$ for herself or himself. Once a coalition S forms an agreement of how to split $v(S)$, then this agreement is enforced. While some coalitions may have greater guaranteed payoffs than others, individual players should be drawn to those coalitions where they can attain the greatest individual reward.

To provide an example, imagine a situation where there are three researchers (players), A, B, and C, each of whom has some resource needed to run a study. No researcher can run the study alone (so that $v(A) =$

[1]Correspondence may be addressed to Alex Chavez.

[2]*Probability matching* holds that humans and other animals choose between different alternatives in ratios that matches those of their respective rewards. The rule has been used successfully to model dynamic learning in strategic contexts (Erev and Roth, 1998, e.g.,). However Sarin and Vahid (2001) have contested that *non-probabilistic selection*, where the agent always selects the action that it currently perceives will yield the highest payoff, can produce results as good as those obtained under the Erev and Roth (1998) model. The latter corresponds to our Greedy agents, and the former to Matching agents.

[3]If the condition $v(AB) > v(A) + v(B)$ holds for all players (where AB is shorthand for the coalition $\{A, B\}$), then v said to have the property of superadditivity. Such a set of relations specifies that players achieve a higher joint payoff in a coalition compared to the sum of their payoffs when acting alone, and represents comparative advantages. Superadditivity holds for all games studied here.

$v(B) = v(C) = 0$), but any two of them can collaborate to run the study. Assume that the combined resources of A and B permit them to run the "best" study (say $v(AB) = 95$); that A and C can run the next "best" one (let $v(AC) = 90$); and B and C the worst one (let $v(BC) = 65$). Assume further that the coalition ABC does not yield any value, so that $v(ABC) = 0$ (there is a limit on the number of researchers who can be involved). Thus, the characteristic function v has been defined completely, so the situation constitutes a coalition game.

Now, researcher A must ask the question, "With whom should I propose a coalition, and how should I propose to allocate the resources assigned by v?" For example, A might be greedy and propose the split $(65, 30, 0)$, where A receives 65 (say, in units of recognition), B receives 30, and C is not included in the *winning coalition* AB (note $65 + 30 = 95 = v(AB)$). But realizing that C would be better off receiving even a small payoff, B might then propose the allocation $(0, 40, 25)$ for the coalition BC, where both B and C do better than they would under A's proposal. The bargaining might continue, with C trying to increase her payoff by proposing the allocation $(60, 0, 30)$ (excluding B from the coalition). This last proposed allocation is special because neither player in the coalition (e.g. A) can make an offer to the excluded player (B) without the excluded player (B) being able to make a counteroffer to the remaining player (C) in which they both do better than under the original allocation; in this sense, the proposed split is *stable* (Aumann and Maschler, 1964). For example, say A was not satisfied with the proposed allocation $(60, 0, 30)$, and decides to propose a coalition with B with the split $(61, 34, 0)$. In this case, B could offer $(0, 35, 30)$ to C, which is another stable split. Along with $(0, 0, 0)$ (where no players form a coalition) and $(60, 35, 0)$, these stable allocations form the solution concept of the "bargaining set" of Aumann and Maschler (1964) (See also chapters 3 and 4 of Kahan and Rapoport, 1984). As Kahan and Rapoport (1974) point out, the bargaining set solution does not predict *which* of the four above allocations will emerge.

An Experimental Study

In their study, Kahan and Rapoport (1974) used human subjects in a computerized experiment designed to test behavior in situations similar to the one described above. 48 undergraduate male subjects were divided into groups of 16 and participated in three separate experiments. In the first experiment, messages were public, so that all players were aware of the others' offers. Subjects had to send messages publicly in a fixed order (as opposed to being able to speak at will). In the second experiment, messages could be private, but again were sent in order. In the last experiment, messages could be private, but were sent at will. The 16 players in each experiment were broken up into 4 quartets, each of which played five 3-person characteristic function games for 4 iterations. For each game, one member of the quartet would sit out as an observer – this procedure was employed to allow subjects to reflect upon the task, and to increase the validity of the assumption of independence between games. Order of play between and within games was randomized subject to the condition that no player would be observer in two consecutive rounds. Subjects were given an extended practice session. The 5 games are shown in Table 1.

	Game: I	II	III	IV	V
Char. Function					
$v(AB)$	95	115	95	106	118
$v(AC)$	90	90	88	86	84
$v(BC)$	65	85	81	66	50
Quota Values					
ω_A	60	60	51	63	76
ω_B	35	55	44	43	42
ω_C	30	30	37	23	8

Table 1: Characteristic Function and Quota Solutions by Game

Quota Values

The type of characteristic function game considered here is a special case known as the *quota games* (Kahan and Rapoport, 1974, 1984), where the conditions $v(ABC) = v(A) = v(B) = v(C) = 0$ and $v(AB), v(AC), v(BC) > 0$ hold. Such games have *quota solutions*, which are generally accepted by cooperative game theory, and are given by the following equation for player i:

$$\omega_i = .5 \sum I(j,k) * v(jk) \quad \forall j,k \in P, j \neq k \qquad (1)$$

where $I(j,k) = 1$ if $i = j$ or $i = k$, and equals -1 otherwise (note that $\omega_i + \omega_j = v(ij)$). So for example, player A's quota simply is calculated by $.5(v(AB) + v(AC) - v(BC))$. These quotas represent normative predictions; when players follow a set of weak rationality conditions (e.g. of the bargaining set (Aumann and Maschler, 1964)), they will arrive at the quota values. The quota solution is a solution concept for characteristic function form games, just as the Nash equilibrium is one for normal form games. Table 1 shows the quota values for players A, B, and C in all 5 games.

Kahan and Rapoport (1974) consider for each player *the mean reward as a member of the winning coalition* (MRAC). Because solution concepts such as the quota make predictions about how much players will get *given* that they are in a coalition, MRAC is the appropriate measure if one wishes to test the theory. Kahan and Rapoport (1974) report that human subjects' overall deviations from the quotas are not significantly different from zero, for each of the experimental conditions; this reinforces that idea that the quota is an important theoretical notion. Table 2 displays the human data in an aggregate form.

	I	II	III	IV	V
Experimental MRAC, Averaged across Experiment					
A	57.43	63.00	53.73	62.07	71.6
B	38.90	54.40	43.40	45.20	45.70
C	29.53	26.93	34.67	19.27	18.17
Experimental Frequency of Coalition Structures					
A,B,C	.0208	.0000	.0000	.0000	.0000
AB,C	.5625	.5208	.3542	.7292	.8750
AC,B	.3125	.1667	.3333	.1458	.1250
BC,A	.1050	.3125	.3125	.1250	.0000

Table 2: Human Data (Kahan and Rapoport, 1974)

While MRAC is the appropriate measure for comparing human performance with theoretical predictions, if one wishes simply to measure how well the subjects do in terms of wealth extracted from the game, then the mean reward (MR) is more appropriate. MRAC and MR can differ greatly; if a stubborn player s always refuses to accept any amount below his quota + 20, for example, then he may be included in the winning coalition once or twice out of a hundred trials, so that his MRAC would be $\omega_s + 20$, but his MR close to zero. The difference arises because in rounds when a player is not in the winning coalition, he or she receives a payoff of zero, which the MRAC ignores. We report both MRAC and MR below.

The Model

We wish to consider the behavior of agents using a simple learning rule in the context of the five coalition formation games above, and to see if these agents achieve outcomes close to those of humans in the experimental study of Kahan and Rapoport (1974). The model assumes that each player updates a belief about how much payoff it can expect from every other agent. We shall refer this value as player i's *aspiration level*, following Macy and Flache (2002), to player j at a given time t, or $A_i^j(t)$ for short. Aspiration levels are updated over time by adding a fraction of the difference between actual reward received from the environment, and the payoff level expected, as given by the equation (e.g., Sutton and Barto, 1998; Macy and Flache, 2002):

$$A_i^j(t) = A_i^j(t-1) + \alpha[r_i^j(t) - A_i^j(t-1)] \qquad (2)$$

where $\alpha \in (0,1)$ represents a recency constant, and $r_i^j(t)$ is the reward received by player i from being in a coalition with player j at time t, as specified by their agreement. For all results reported here, we set $\alpha = 0.2$, a typical value in the reinforcement learning literature. In simulations not reported here, α values between 0.1 and 0.4 yielded results similar to those in this paper. In addition to the updating rule given in Equation (2), we considered two methods for agents to decide whom to offer to. They are given in Table 3[4].

	Probability of i Offering to j
Greedy	1 if $j = \arg\max_{p \in P} A_i^p(t)$; 0 else
Matching	$A_i^j(t) / \sum_{p \in P} A_i^p(t)$

Table 3: Selection Rules

In addition, each agent makes offers *at* its aspiration levels, and not below. So for example, if agent A has aspiration levels of 50 to agent B and 65 to agent C, then it would offer $v(AC) - 65$ to agent C with probability 1 under the Greedy rule. Under the Matching rule, it would offer $v(AB) - 50$ to agent B with probability 65/115 and $v(AC) - 65$ to agent C with probability 50/115. While Sarin and Vahid (2001) note that Greedy (i.e. non-probabilistic) action selection is in line with the traditional economic precept of choice as maximization over beliefs, others have found that humans and other animals use probability matching to select between actions associated with a reward (e.g. Gallistel, 1990). Thus, we investigate the performance of both types of selection rules.

Also, our agents are myopic subjective maximizers in their offer behavior, in that they make offers based on the maximum they currently "think" they can get, without considering the possible ramifications of their offer behavior on the future state of the system. Such agents represent players who vastly simplify their objective environment, collapsing the available history of offer behavior for each other player into a single real value, A_i^j (Sarin and Vahid, 2001). Again, we wish to explore the behavior of agents representing players of minimal rationality, and to see if they can model the overall results of humans. The simple agents described in this section satisfy such a condition.

Finally, simulations were run as follows. At the start of a simulation, each agent i's initial aspiration level to j was initialized from the uniform distribution on $[0, v(ij)]$[5]. Next, for each episode, an agent was selected at random as the initial offerer. This agent made an offer to a receiving agent, according to the rules described above. The receiving agent accepted if the offer amount was greater than or equal to its current aspiration level to the offering agent. If it declined, then the receiving

[4] In simulations not reported here, other relatively exploitative selection methods drawn from reinforcement learning literature, such as ϵ-greedy and Softmax selection (Sutton and Barto, 1998), produced results similar to those reported below for the Greedy agent.

[5] We found that setting $A_i^j(0) = v(ij)/2$ – that is, to half of the coalition value between i and j – did not affect MRAC or MR in expectation (it resulted in less variance).

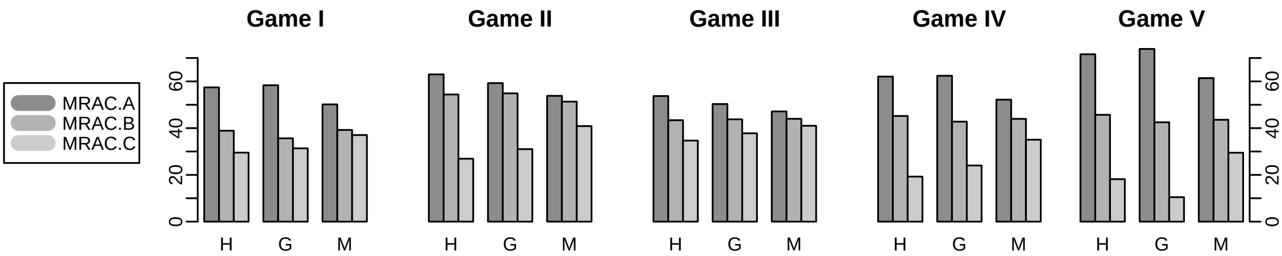

Figure 1: Mean Reward as Members of Winning Coalition for **H**uman Subjects, **G**reedy Agents, and **M**atching Agents

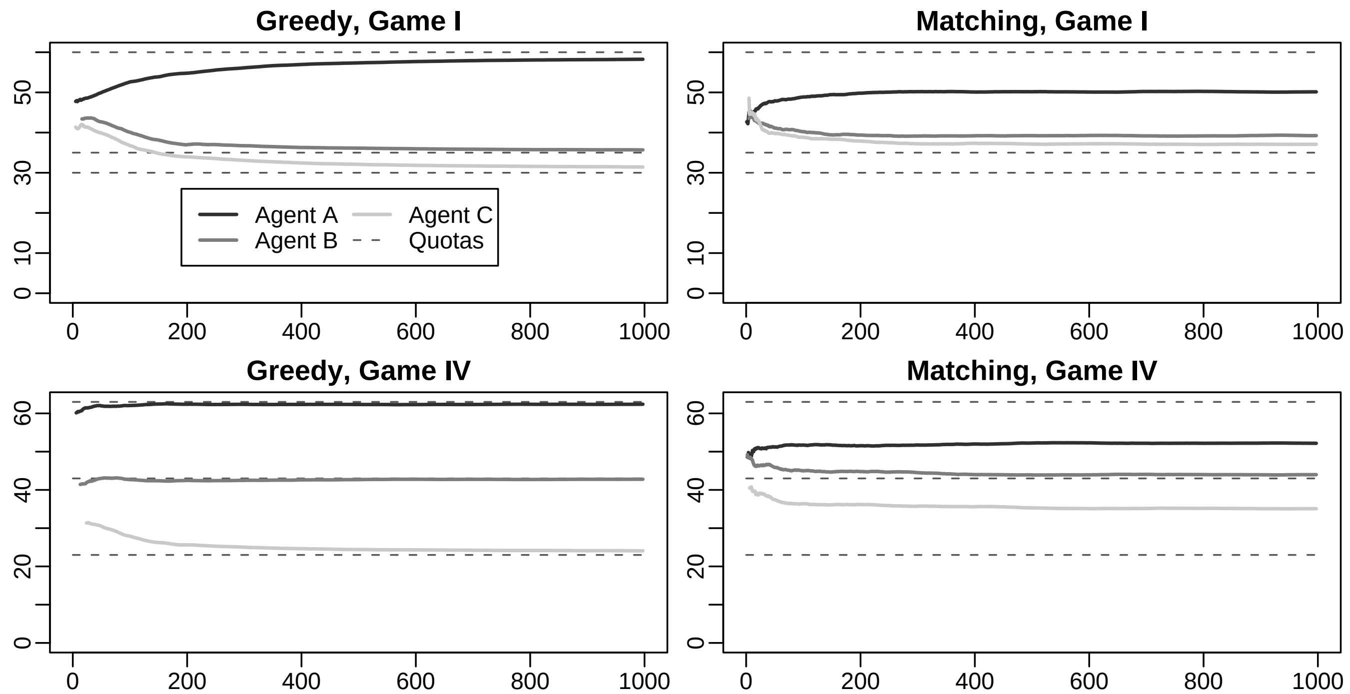

Figure 2: MRAC over Time in Games I and IV, for Greedy and Matching Agents

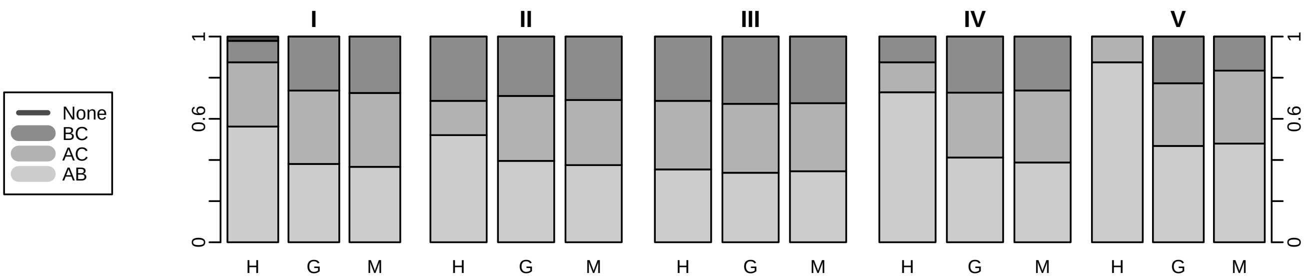

Figure 3: Frequency of Coalition Formation for **H**uman Subjects, **G**reedy Agents, and **M**atching Agents

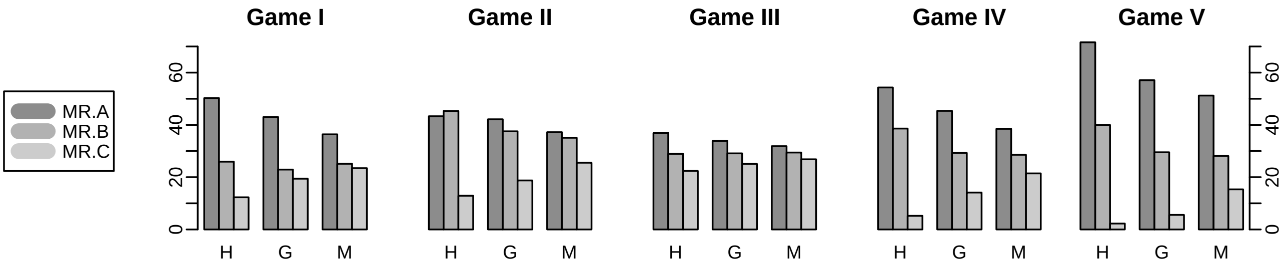

Figure 4: Mean Reward for **H**uman, **G**reedy, and **M**atching Agents

agent would become the new offering agent, and so on. The process continued until a receiving agent accepted an offer, which always happened within 6 rounds. When an episode ended, agents' aspiration levels were updated, and the next episode was started. This process continued until a maximum number of episodes was reached, whereupon the simulation would terminate.

Results

The MRAC values for the data from Kahan and Rapoport (1974) for human subjects are presented along with results for Greedy and Matching agents (20 simulations of 1,000 episodes each) in Figure 1. As Figure 1 shows, the MRAC values of the agents, especially for the Greedy type, closely fit those of human subjects ($r = .98$ for the Greedy agents, and $r = .96$ for the Matching agents). Figure 2 shows the rate of convergence of MRAC values for Greedy and Matching agents in Games I and IV (Games II, III, and V are not displayed for space reasons). As can partially be seen from the figure, MRAC values for Greedy agents rapidly approach the quota solutions in Games II, III, and IV, more slowly in Games I and V, and never for Matching agents (this is explained later). Figure 3 shows frequencies of coalition structures for human subjects and for simulations. The model does not predict some important differences in these frequencies. Namely, player A is included in the winning coalition more often than in the human data than in the simulations, and player C less often (this can be seen for player A by looking at the AC and AB blocks together, and for player C by looking at the BC and AC blocks together).

As discussed previously, to gauge the overall performance of players, it is useful to consider the measure of mean reward (MR) of players. MR values for human subjects, Greedy agents, and Matching agents are shown in Figure 4. The MR values for both agent types fit the data of human subjects well, with Greedy agents again doing better than the Matching agents ($r = .97$ vs. $r = .95$). However, there are some noticeable departures. For example, player A has lower MR values and players C higher, for agents vs. humans. This is attributable directly to the differences in the frequencies of coalition formation (Figure 3), where player A is in the winning coalition less often (and player C less more) for agents vs. humans.

We would like to make two statements here about our findings. The first is that systems of relatively exploitative agents (e.g. Greedy agents) converge to neighborhoods of quota solutions, and that this outcome is robust to variations of initial aspiration levels. The second is that the difference between an agent's initial aspiration level and its quota value is related to its speed of convergence in MRAC value (the smaller the difference, the faster the convergence). This may be important because although humans subjects arrived at the quota solution in only 4 iterations of play, their initial offers appear to have been very close to their quota values[6]. Whether this occurred as a result of practice sessions, transfer between games or from acting as the observer, or from deliberation is not clear. We merely point out that if the initial aspiration levels of the agents are close to their quota values, then their MRAC values converge almost immediately to the quotas (see e.g. agents A and B in Game IV of Figure 2).

Conclusion

There is now an extensive literature in behavioral game theory that aims at describing, modeling, and explaining human behavior in games (see Camerer, 2003; Kagel and Roth, 1995, for literature reviews)). This literature has discovered that a variety of simple learning models provide reasonably good predictions of human behavior in laboratory games. A number of authors have found that reinforcement learning models supplemented with additional heuristic rules (e.g., pertaining to the specifics of the game) achieve good predictive power. Moreover, other simple models of learning, e.g. using activation-based recall (Lebiere et al., 2003), have also provided good accounts of human behavior in games.

In the present work, we found that systems of agents using a simple learning rule can model human data reasonably well, and that their performance converges to the theoretical predictions of quota values in a class of 3-player, coalition formation games. The MRAC values for both Greedy and Matching agents were very close to those of humans, while their MR values showed some systematic differences from the human data. Although we do not claim the learning rule presented in Equation (2)

[6]Kahan and Rapoport (1974) report that first offers made to the winning coalition had an average deviation over games and coalitions of -2.95, e.g..

fully accounts for the way people play these games, we submit that it is a fruitful and practicable way of investigating strategic environments. While much remains to be done, it is intriguing that the agents' behavior so closely matches the human data, much more so than has been reported for simple reinforcement learning models in non-cooperative games.

The results of both agent and human bargaining results are interesting, but perhaps the most important questions regarding bargaining games pertain to their dynamics. When will (and when should) a player accept an offer? If a player chooses to reject, how much and to whom should he/she offer? The answers to these questions may depend not only on the coalition values and other initial parameters of the game, but also on the history of offers and round number. For example, if only 2 rounds are left before the game is ended by default, can player A (with the highest total coalition value) extract more reward by making an aggressive offer? or, perhaps he/she must accept to avoid forfeiting any positive reward for the game. To answer these questions, different approaches such as classifier systems are needed, for example to find bargaining rules such as (for Agent A): "If round is 4 and offeringAgent is B and amount is at least aspirationLevel$-0.1 * v(AB)$, then accept."

In sum, the prospects for mutual benefit between the cognitive modeling community and the behavioral game theory community are bright indeed.

Acknowledgements

The work reported here was supported in part by NSF grant number SES-9709548 and funding from the University Scholars Program at the University of Pennsylvania. We thank two anonymous reviewers, Frank Lee, and Dario Salvucci for helpful comments.

References

Aumann, R. J. and Maschler, M. (1964). The bargaining set for cooperative games. In Dresher, M., Shapley, L. S., and Tucker, A. W., editors, *Advanced in Game Theory*. Princeton University Press.

Camerer, C. and Ho, T. (1999). Experience-weighted attraction learning in normal form games. *Econometrica*, 67:827–874.

Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Russell Sage Foundation and Princeton University Press, New York, NY and Princeton, NJ.

Dworman, G. O., Kimbrough, S. O., and Laing, J. D. (1995a). Bargaining in a three-agent coalitions game: An application of genetic programming. In *Working Notes: AAAI–95 Fall Symposium Series, Genetic Programming*, pages 9–16, Boston, MA, November 10-12, 1995. AAAI. File: gpconf08.doc.

Dworman, G. O., Kimbrough, S. O., and Laing, J. D. (1995b). On automated discovery of models using genetic programming: Bargaining in a three-agent coalitions game. *Journal of Management Information Systems*, 12(3):97–125.

Dworman, G. O., Kimbrough, S. O., and Laing, J. D. (1995c). On automated discovery of models using genetic programming in game–theoretic contexts. In Nunamaker, Jr., J. F. and Sprague, Jr., R. H., editors, *Proceedings of the Twenty-Eighth Annual Hawaii International Conference on System Sciences, Volume III: Information Systems: Decision Support and Knowledge-Based Systems*, pages 428–438, Los Alamitos, CA. IEEE Computer Society Press.

Dworman, G. O., Kimbrough, S. O., and Laing, J. D. (1996). Bargaining by artificial agents in two coalition games: A study in genetic programming for electronic commerce. In Koza, J. R., Goldberg, D. E., Fogel, D. B., and Riolo, R. L., editors, *Genetic Programming 1996: Proceedings of the First Annual Genetic Programming Conference, July 28-31, 1996, Stanford University*, pages 54–62. The MIT Press.

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

Gallistel, C. R. (1990). *The Organization of Learning*. The MIT Press, Cambridge, MA.

Kagel, J. H. and Roth, A. E., editors (1995). *The Handbook of Experimental Economics*. Princeton University Press, Princeton, NJ.

Kahan, J. P. and Rapoport, A. (1974). Test of the bargaining set and kernel models in three-person games. In Rapoport, A., editor, *Game Theory as a Theory of Conflict Resolution*, pages 119–160. D. Reidel, Dordrecht, The Netherlands.

Kahan, J. P. and Rapoport, A. (1984). *Theories of Coalition Formation*. Lawrence Earlbaum Associates, Hillsdale, NJ.

Lebiere, C., Gray, R., Salvucci, D., and West, R. (2003). Choice and learning under uncertainty: a case study in baseball batting. In *Proceedings of the 25th Annual Conference of the Cognitive Science Society*.

Macy, M. W. and Flache, A. (2002). Learning dynamics in social dilemmas. *Proceedings of the National Academy of Science (PNAS)*, 99(suppl. 3):7229–7236.

Mookherjee, D. and Sopher, B. (1997). Learning and decision costs in experimental constant sum games. *Games and Economic Behavior*, 19(1):97–132.

Ritter, F. E. and Wallach, D. P. (1998). Models of two-person games in act-r and soar. In *Proceedings of the 2nd European Conference on Cognitive Modelling*.

Roth, A. E. and Erev, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior*, 8:164–212.

Sarin, R. and Vahid, F. (2001). Predicting how people play games. *Games and Economic Behavior*, 34:104–122.

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. The MIT Press, Cambridge, MA.

Uhlich, G. R. (1990). *Descriptive Theories of Bargaining*. Springer-Verlag, Berlin, Germany.