An Adversarial Environment Model for Bounded Rational Agents in Zero-Sum Interactions

Inon Zuckerman[1], Sarit Kraus[1], Jeffrey S. Rosenschein[2], Gal Kaminka[1]

[1]Department of Computer Science
Bar-Ilan University
Ramat-Gan, Israel
{zukermi,sarit,galk}@cs.biu.ac.il

[2]The School of Engineering
and Computer Science
Hebrew University, Jerusalem, Israel
jeff@cs.huji.ac.il

ABSTRACT

Multiagent environments are often not cooperative nor collaborative; in many cases, agents have conflicting interests, leading to adversarial interactions. This paper presents a formal *Adversarial Environment* model for bounded rational agents operating in a zero-sum environment. In such environments, attempts to use classical utility-based search methods can raise a variety of difficulties (e.g., implicitly modeling the opponent as an omniscient utility maximizer, rather than leveraging a more nuanced, explicit opponent model).

We define an Adversarial Environment by describing the mental states of an agent in such an environment. We then present behavioral axioms that are intended to serve as design principles for building such adversarial agents. We explore the application of our approach by analyzing log files of completed Connect-Four games, and present an empirical analysis of the axioms' appropriateness.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Intelligent agents,Multiagent Systems*;
I.2.4 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods —*Modal logic*

General Terms

Design, Theory

Keywords

Agents, Multiagent Systems, Modal Logic

1. INTRODUCTION

Early research in multiagent systems (MAS) considered cooperative groups of agents; because individual agents had limited resources, or limited access to information (e.g., limited processing power, limited sensor coverage), they worked together by design to solve problems that individually they could not solve, or at least could not solve as efficiently.

MAS research, however, soon began to consider interacting agents with individuated interests, as representatives of different humans or organizations with non-identical interests. When interactions are guided by diverse interests, participants may have to overcome disagreements, uncooperative interactions, and even intentional attempts to damage one another. When these types of interactions occur, environments require appropriate behavior from the agents situated in them. We call these environments *Adversarial Environments*, and call the clashing agents *Adversaries*.

Models of cooperation and teamwork have been extensively explored in MAS through the axiomatization of mental states (e.g., [8, 4, 5]). However, none of this research dealt with adversarial domains and their implications for agent behavior. Our paper addresses this issue by providing a formal, axiomatized mental state model for a subset of adversarial domains, namely simple zero-sum adversarial environments.

Simple zero-sum encounters exist of course in various two-player games (e.g., *Chess*, *Checkers*), but they also exist in n-player games (e.g., *Risk*, *Diplomacy*), auctions for a single good, and elsewhere. In these latter environments especially, using a utility-based adversarial search (such as the Min-Max algorithm) does not always provide an adequate solution; the payoff function might be quite complex or difficult to quantify, and there are natural computational limitations on bounded rational agents. In addition, traditional search methods (like Min-Max) do not make use of a model of the opponent, which has proven to be a valuable addition to adversarial planning [9, 3, 11].

In this paper, we develop a formal, axiomatized model for bounded rational agents that are situated in a zero-sum adversarial environment. The model uses different modality operators, and its main foundations are the *SharedPlans* [4] model for collaborative behavior. We explore environment properties and the mental states of agents to derive behavioral axioms; these behavioral axioms constitute a formal model that serves as a specification and design guideline for agent design in such settings.

We then investigate the behavior of our model empirically using the *Connect-Four* board game. We show that this game conforms to our environment definition, and analyze players' behavior using a large set of completed match log

AAMAS'07 May 14–18 2007, Honolulu, Hawai'i, USA.

files. In addition, we use the results presented in [9] to discuss the importance of opponent modeling in our *Connect-Four* adversarial domain.

The paper proceeds as follows. Section 2 presents the model's formalization. Section 3 presents the empirical analysis and its results. We discuss related work in Section 4, and conclude and present future directions in Section 5.

2. ADVERSARIAL ENVIRONMENTS

The adversarial environment model (denoted as `AE`) is intended to guide the design of agents by providing a specification of the capabilities and mental attitudes of an agent in an adversarial environment. We focus here on specific types of adversarial environments, specified as follows:

1. **Zero-Sum Interactions:** positive and negative utilities of all agents sum to zero;
2. **Simple AEs:** all agents in the environment are adversarial agents;
3. **Bilateral AEs:** AE's with exactly two agents;
4. **Multilateral AEs':** AE's of three or more agents.

We will work on both *bilateral* and *multilateral* instantiations of *zero-sum* and *simple* environments. In particular, our adversarial environment model will deal with interactions that consist of N agents ($N \geq 2$), where all agents are adversaries, and only one agent can succeed. Examples of such environments range from board games (e.g., *Chess*, *Connect-Four*, and *Diplomacy*) to certain economic environments (e.g., N-bidder auctions over a single good).

2.1 Model Overview

Our approach is to formalize the mental attitudes and behaviors of a single adversarial agent; we consider how a single agent perceives the AE. The following list specifies the conditions and mental states of an agent in a simple, zero-sum AE:

1. The agent has an individual intention that its own goal will be completed;
2. The agent has an individual belief that it and its adversaries are pursuing *full conflicting goals* (defined below) — there can be only one winner;
3. The agent has an individual belief that each adversary has an intention to complete its own *full conflicting goal*;
4. The agent has an individual belief in the (partial) profile of its adversaries.

Item 3 is required, since it might be the case that some agent has a *full conflicting goal*, and is currently considering adopting the intention to complete it, but is, as of yet, not committed to achieving it. This might occur because the agent has not yet deliberated about the effects that adopting that intention might have on the other intentions it is currently holding. In such cases, it might not consider itself to even be in an adversarial environment.

Item 4 states that the agent should hold some belief about the profiles of its adversaries. The profile represents all the knowledge the agent has about its adversary: its weaknesses, strategic capabilities, goals, intentions, trustworthiness, and more. It can be given explicitly or can be learned from observations of past encounters.

2.2 Model Definitions for Mental States

We use Grosz and Kraus's definitions of the modal operators, predicates, and meta-predicates, as defined in their *SharedPlan* formalization [4]. We recall here some of the predicates and operators that are used in that formalization: $Int.To(A_i, \alpha, T_n, T_\alpha, C)$ represents A_i's intentions at time T_n to do an action α at time T_α in the context of C. $Int.Th(A_i, prop, T_n, T_{prop}, C)$ represents A_i's intentions at time T_n that a certain proposition *prop* holds at time T_{prop} in the context of C. The potential intention operators, $Pot.Int.To(...)$ and $Pot.Int.Th(...)$, are used to represent the mental state when an agent considers adopting an intention, but has not deliberated about the interaction of the other intentions it holds. The operator $Bel(A_i, f, T_f)$ represents agent A_i believing in the statement expressed in formula f, at time T_f. $MB(A, f, T_f)$ represents mutual belief for a group of agents A.

A snapshot of the system finds our environment to be in some state $e \in E$ of environmental variable states, and each adversary in any $L_{A_i} \in L$ of possible local states. At any given time step, the system will be in some world w of the set of all possible worlds $w \in W$, where $w = E \times L_{A_1} \times L_{A_2} \times ...L_{A_n}$, and n is the number of adversaries. For example, in a Texas Hold'em poker game, an agent's local state might be its own set of cards (which is unknown to its adversary) while the environment will consist of the betting pot and the community cards (which are visible to both players).

A *utility* function under this formalization is defined as a mapping from a possible world $w \in W$ to an element in $\Re$, which expresses the desirability of the world, from a single agent perspective. We usually normalize the range to [0,1], where 0 represents the least desirable possible world, and 1 is the most desirable world. The implementation of the utility function is dependent on the domain in question.

The following list specifies new predicates, functions, variables, and constants used in conjunction with the original definitions for the adversarial environment formalization:

1. ϕ is a null action (the agent does not do anything).
2. G_{A_i} is the set of agent A_i's goals. Each goal is a set of predicates whose satisfaction makes the goal complete (we use $G^*_{A_i} \in G_{A_i}$ to represent an arbitrary goal of agent A_i).
3. g_{A_i} is the set of agent A_i's subgoals. Subgoals are predicates whose satisfaction represents an important milestone toward achievement of the full goal. $g_{G^*_{A_i}} \subseteq g_{A_i}$ is the set of subgoals that are important to the completion of goal $G^*_{A_i}$ (we will use $g^*_{G^*_{A_i}} \in g_{G^*_{A_i}}$ to represent an arbitrary subgoal).
4. $P^{A_j}_{A_i}$ is the profile object agent A_i holds about agent A_j.
5. C_A is a general set of actions for all agents in A which are derived from the environment's constraints. $C_{A_i} \subseteq C_A$ is the set of agent A_i's possible actions.
6. $Do(A_i, \alpha, T_\alpha, w)$ holds when A_i performs action α over time interval T_α in world w.
7. $Achieve(G^*_{A_i}, \alpha, w)$ is *true* when goal $G^*_{A_i}$ is achieved following the completion of action α in world $w \in W$, where $\alpha \in C_{A_i}$.
8. $Profile(A_i, P^{A_i}_{A_j})$ is *true* when agent A_i holds an object profile for agent A_j.

Definition 1. **Full conflict** (*FulConf*) describes a zero-sum interaction where only a single goal of the goals in conflict can be completed.

$FulConf(G^*_{A_i}, G^*_{A_j}) \Rightarrow (\exists \alpha \in C_{A_i}, \forall w, \beta \in C_{A_j})$
$(Achieve(G^*_{A_i}, \alpha, w) \Rightarrow \neg Achieve(G^*_{A_j}, \beta, w)) \vee$
$(\exists \beta \in C_{A_j}, \forall w, \alpha \in C_{A_i})(Achieve(G^*_{A_j}, \beta, w) \Rightarrow$
$\neg Achieve(G^*_{A_i}, \alpha, w))$

Definition 2. **Adversarial Knowledge** (*AdvKnow*) is a function returning a value which represents the amount of

knowledge agent A_i has on the profile of agent A_j, at time T_n. The higher the value, the more knowledge agent A_i has.
$AdvKnow : P_{A_i}^{A_j} \times T_n \rightarrow \Re$

Definition 3. **Eval** — This evaluation function returns an estimated expected utility value for an agent in A, after completing an action from C_A in some world state w.
$Eval : A \times C_A \times w \rightarrow \Re$

Definition 4. **TrH** — (Threshold) is a numerical constant in the [0,1] range that represents an evaluation function (*Eval*) threshold value. An action that yields an estimated utility evaluation above the TrH is regarded as a *highly beneficial action*.

The *Eval* value is an estimation and not the real utility function, which is usually unknown. Using the real utility value for a rational agent would easily yield the best outcome for that agent. However, agents usually do not have the real utility functions, but rather a heuristic estimate of it.

There are two important properties that should hold for the evaluation function:

Property 1. The evaluation function should state that the most desirable world state is one in which the goal is achieved. Therefore, after the goal has been satisfied, there can be no future action that can put the agent in a world state with higher *Eval* value.
$(\forall A_i, G^*_{A_i}, \alpha, \beta \in C_{A_i}, w \in W)$
$Achieve(G^*_{A_i}, \alpha, w) \Rightarrow Eval(A_i, \alpha, w) \geq Eval(A_i, \beta, w)$

Property 2. The evaluation function should project an action that causes a completion of a goal or a subgoal to a value which is greater than TrH (a highly beneficial action).
$(\forall A_i, G^*_{A_i} \in G_{A_i}, \alpha \in C_{A_i}, w \in W, g^*_{G_{A_i}} \in g_{G_{A_i}})$
$Achieve(G^*_{A_i}, \alpha, w) \vee Achieve(g^*_{G_{A_i}}, \alpha, w) \Rightarrow$
$Eval(A_i, \alpha, w) \geq TrH.$

Definition 5. **SetAction** We define a set action (*SetAction*) as a set of action operations (either complex or basic actions) from some action sets C_{A_i} and C_{A_j} which, according to agent A_i's belief, are attached together by a temporal and consequential relationship, forming a chain of events (action, and its following consequent action).
$(\forall \alpha^1, \ldots, \alpha^u \in C_{A_i}, \beta^1, \ldots, \beta^v \in C_{A_j}, w \in W)$
$SetAction(\alpha^1, \ldots, \alpha^u, \beta^1, \ldots, \beta^v, w) \Rightarrow$
$((Do(A_i, \alpha^1, T_{\alpha^1}, w) \Rightarrow Do(A_j, \beta^1, T_{\beta^1}, w)) \Rightarrow$
$Do(A_i, \alpha^2, T_{\alpha^2}, w) \Rightarrow \ldots \Rightarrow Do(A_i, \alpha^u, T_{\alpha^u}, w))$
The consequential relation might exist due to various environmental constraints (when one action forces the adversary to respond with a specific action) or due to the agent's knowledge about the profile of its adversary.

Property 3. As the knowledge we have about our adversary increases we will have additional beliefs about its behavior in different situations which in turn creates new set actions. Formally, if our $AdvKnow$ at time T_{n+1} is greater than $AdvKnow$ at time T_n, then every $SetAction$ known at time T_n is also known at time T_{n+1}.
$AdvKnow(P_{A_i}^{A_j}, T_{n+1}) > AdvKnow(P_{A_i}^{A_j}, T_n) \Rightarrow$
$(\forall \alpha^1, \ldots, \alpha^u \in C_{A_i}, \beta^1, \ldots, \beta^v \in C_{A_j})$
$Bel(A_{ag}, SetAction(\alpha^1, \ldots, \alpha^u, \beta^1, \ldots, \beta^v), T_n) \Rightarrow$
$Bel(A_{ag}, SetAction(\alpha^1, \ldots, \alpha^u, \beta^1, \ldots, \beta^v), T_{n+1})$

2.3 The Environment Formulation

The following axioms provide the formal definition for a simple, zero-sum Adversarial Environment (AE). Satisfaction of these axioms means that the agent is situated in such an environment. It provides specifications for agent A_{ag} to interact with its set of adversaries A with respect to goals $G^*_{A_{ag}}$ and G^*_A at time T_{Co} at some world state w.
$AE(A_{ag}, A, G^*_{A_{ag}}, A_1, \ldots, A_k, G^*_{A_1}, \ldots, G^*_{A_k}, T_n, w)$
1. A_{ag} has an $Int.Th$ his goal would be completed:
$(\exists \alpha \in C_{A_{ag}}, T_\alpha)$
$Int.Th(A_{ag}, Achieve(G^*_{A_{ag}}, \alpha), T_n, T_\alpha, AE)$
2. A_{ag} believes that it and each of its adversaries A_o are pursuing full conflicting goals:
$(\forall A_o \in \{A_1, \ldots, A_k\})$
$Bel(A_{ag}, FulConf(G^*_{A_{ag}}, G^*_{A_o}), T_n)$
3. A_{ag} believes that each of his adversaries in A_o has the $Int.Th$ his conflicting goal $G^*_{A_{o_i}}$ will be completed:
$(\forall A_o \in \{A_1, \ldots, A_k\})(\exists \beta \in C_{A_o}, T_\beta)$
$Bel(A_{ag}, Int.Th(A_o, Achieve(G^*_{A_o}, \beta), T_{Co}, T_\beta, AE), T_n)$
4. A_{ag} has beliefs about the (partial) profiles of its adversaries
$(\forall A_o \in \{A_1, \ldots, A_k\})$
$(\exists P^{A_o}_{A_{ag}} \in P_{A_{ag}}) Bel(A_{ag}, Profile(A_o, P^{A_o}_{A_{ag}}), T_n)$

To build an agent that will be able to operate successfully within such an AE, we must specify behavioral guidelines for its interactions. Using a naive *Eval* maximization strategy to a certain search depth will not always yield satisfactory results for several reasons: (1) the search *horizon problem* when searching for a fixed depth; (2) the strong assumption of an optimally rational, unbounded resources adversary; (3) using an *estimated* evaluation function which will not give optimal results in all world states, and can be exploited [9].

The following axioms specify the behavioral principles that can be used to differentiate between successful and less successful agents in the above Adversarial Environment. Those axioms should be used as specification principles when designing and implementing agents that should be able to perform well in such Adversarial Environments. The behavioral axioms represent situations in which the agent will adopt potential intentions to (*Pot.Int.To(...)*) perform an action, which will typically require some means-end reasoning to select a possible course of action. This reasoning will lead to the adoption of an *Int.To(...)* (see [4]).

A1. Goal Achieving Axiom. The first axiom is the simplest case; when the agent A_{ag} believes that it is one action (α) away from achieving his conflicting goal $G^*_{A_{ag}}$, it should adopt the potential intention to do α and complete its goal.
$(\forall A_{ag}, \alpha \in C_{A_{ag}}, T_n, T_\alpha, w \in W)$
$(Bel(A_{ag}, Do(A_{ag}, \alpha, T_\alpha, w) \Rightarrow Achieve(G^*_{A_{ag}}, \alpha, w))$
$\Rightarrow Pot.Int.To(A_{ag}, \alpha, T_n, T_\alpha, w)$

This somewhat trivial behavior is the first and strongest axiom. In any situation, when the agent is an action away from completing the goal, it should complete the action. Any fair *Eval* function would naturally classify α as the maximal value action (property 1). However, without explicit axiomatization of such behavior there might be situations where the agent will decide on taking another action for various reasons, due to its bounded decision resources.

A2. Preventive Act Axiom. Being in an adversarial situation, agent A_{ag} might decide to take actions that will damage one of its adversary's plans to complete its goal, even if those actions do not explicitly advance A_{ag} towards its conflicting goal $G^*_{A_{ag}}$. Such preventive action will take place when agent A_{ag} has a belief about the possibility of its adversary A_o doing an action β that will give it a high

utility evaluation value ($> TrH$). Believing that taking action α will prevent the opponent from doing its β, it will adopt a potential intention to do α.
$(\forall A_{ag}, A_o \in A, \alpha \in C_{A_{ag}}, \beta \in C_{A_o}, T_n, T_\beta, w \in W)$
$(Bel(A_{ag}, Do(A_o, \beta, T_\beta, w) \wedge Eval(A_o, \beta, w) > TrH, T_n) \wedge$
$Bel(A_{ag}, Do(A_{ag}, \alpha, T_\alpha, w) \Rightarrow \neg Do(A_o, \beta, T_\beta, w), T_n)$
$\Rightarrow Pot.Int.To(A_{ag}, \alpha, T_n, T_\alpha, w)$

This axiom is a basic component of any adversarial environment. For example, looking at a *Chess* board game, a player could realize that it is about to be checkmated by its opponent, thus making a preventive move. Another example is a *Connect Four* game: when a player has a row of three chips, its opponent must block it, or lose.

A specific instance of A1 occurs when the adversary is one action away from achieving its goal, and immediate preventive action needs to be taken by the agent. Formally, we have the same beliefs as stated above, with a changed belief that doing action β will cause agent A_o to achieve its goal.

Proposition 1: Prevent or lose case.
$(\forall A_{ag}, A_o \in A, \alpha \in C_{A_{ag}}, \beta \in C_{A_o}, G^*_{A_o}, T_n, T_\alpha, T_\beta, w \in W)$
$Bel(A_{ag}, Do(A_o, \beta, T_\beta, w) \Rightarrow Achieve(G^*_{A_o}, \beta, w), T_n) \wedge$
$Bel(A_{ag}, Do(A_{ag}, \alpha, T_\alpha, w) \Rightarrow \neg Do(A_o, \beta, T_\beta, w))$
$\Rightarrow Pot.Int.To(A_{ag}, \alpha, T_n, T_\alpha, w)$

Sketch of proof: Proposition 1 can be easily derived from axiom A1 and the property 2 of the *Eval* function, which states that any action that causes a completion of a goal is *a highly beneficial* action.

The preventive act behavior will occur implicitly when the *Eval* function is equal to the real world utility function. However, being bounded rational agents and dealing with an estimated evaluation function we need to explicitly axiomatize such behavior, for it will not always occur implicitly from the evaluation function.

A3. Suboptimal Tactical Move Axiom. In many scenarios a situation may occur where an agent will decide not to take the current most beneficial action it can take (the action with the maximal utility evaluation value), because it believes that taking another action (with lower utility evaluation value) might yield (depending on the adversary's response) a future possibility for a highly beneficial action. This will occur most often when the *Eval* function is inaccurate and differs by a large extent from the *Utility* function. Put formally, agent A_{ag} believes in a certain $SetAction$ that will evolve according to its initial action and will yield a high beneficial value ($> TrH$) solely for it.
$(\forall A_{ag}, A_o \in A, T_n, w \in W)$
$(\exists \alpha^1, \ldots, \alpha^u \in C_{A_i}, \beta^1, \ldots, \beta^v \in C_{A_j}, T_{\alpha^1})$
$Bel(A_{ag}, SetAction(\alpha^1, \ldots, \alpha^u, \beta^1, \ldots, \beta^v), T_n) \wedge$
$Bel(A_{ag}, Eval(A_o, \beta^v, w) < TrH < Eval(A_{ag}, \alpha^u, w), T_n)$
$\Rightarrow Pot.Int.To(A_{ag}, \alpha^1, T_n, T_{\alpha^1}, w)$

An agent might believe that a chain of events will occur for various reasons due to the inevitable nature of the domain. For example, in *Chess*, we often observe the following: a move causes a check position, which in turn limits the opponent's moves to avoiding the check, to which the first player might react with another check, and so on. The agent might also believe in a chain of events based on its knowledge of its adversary's profile, which allows it to foresee the adversary's movements with high accuracy.

A4. Profile Detection Axiom. The agent can adjust its adversary's profiles by observations and pattern study (specifically, if there are repeated encounters with the same adversary). However, instead of waiting for profile information to be revealed, an agent can also initiate actions that will force its adversary to react in a way that will reveal profile knowledge about it. Formally, the axiom states that if all actions (γ) are not *highly beneficial actions* ($< TrH$), the agent can do action α in time T_α if it believes that it will result in a non-highly beneficial action β from its adversary, which in turn teaches it about the adversary's profile, i.e., gives a higher $AdvKnow(P^{A_j}_{A_i}, T_\beta)$.
$(\forall A_{ag}, A_o \in A, \alpha \in C_{A_{ag}}, \beta \in C_{A_o}, T_n, T_\alpha, T_\beta, w \in W)$
$Bel(A_{ag}, (\forall \gamma \in C_{A_{ag}}) Eval(A_{ag}, \gamma, w) < TrH, T_n) \wedge$
$Bel(A_{ag}, Do(A_{ag}, \alpha, T_\alpha, w) \Rightarrow Do(A_o, \beta, T_\beta, w), T_n) \wedge$
$Bel(A_{ag}, Eval(A_o, \beta, w) < TrH) \wedge$
$Bel(A_{ag}, AdvKnow(P^{A_j}_{A_i}, T_\beta) > AdvKnow(P^{A_j}_{A_i}, T_n), T_n) \Rightarrow$
$Pot.Int.To(A_{ag}, \alpha, T_n, T_\alpha, w)$

For example, going back to the *Chess* board game scenario, consider starting a game versus an opponent about whom we know nothing, not even if it is a human or a computerized opponent. We might start playing a strategy that will be suitable versus an average opponent, and adjust our game according to its level of play.

A5. Alliance Formation Axiom The following behavioral axiom is relevant only in a multilateral instantiation of the adversarial environment (obviously, an alliance cannot be formed in a bilateral, zero-sum encounter). In different situations during a multilateral interaction, a group of agents might believe that it is in their best interests to form a temporary *alliance*. Such an alliance is an agreement that constrains its members' behavior, but is believed by its members to enable them to achieve a higher utility value than the one achievable outside of the alliance.

As an example, we can look at the classical *Risk* board game, where each player has an individual goal of being the sole conquerer of the world, a zero-sum game. However, in order to achieve this goal, it might be strategically wise to make short-term ceasefire agreements with other players, or to join forces and attack an opponent who is stronger than the rest.

An alliance's terms defines the way its members should act. It is a set of predicates, denoted as *Terms*, that is agreed upon by the alliance members, and should remain true for the duration of the alliance. For example, the set *Terms* in the *Risk* scenario, could contain the following predicates:
1. Alliance members will not attack each other on territories X, Y and Z;
2. Alliance members will contribute C units per turn for attacking adversary A_o;
3. Members are obligated to stay as part of the alliance until time T_k or until adversary's A_o army is smaller than Q.

The set *Terms* specifies inter-group constraints on each of the alliance member's ($\forall A^{al}_i \in A^{al} \subseteq A$) set of actions $C^{al}_i \subseteq C$.

Definition 6. **Al_val** — the total evaluation value that agent A_i will achieve while being part of A^{al} is the sum of $Eval_i$ (*Eval* for A_i) of each of A^{al}_j Eval values after taking their own α actions (via the agent(α) predicate):
$Al\_val(A_i, C^{al}, A^{al}, w) = \sum_{\alpha \in C^{al}} Eval_i(A^{al}_j, agent(\alpha), w)$

Definition 7. **Al_TrH** — is a number representing an Al_val

threshold; above it, the alliance can be said to be a *highly beneficial alliance*.

The value of Al_TrH will be calculated dynamically according to the progress of the interaction, as can be seen in [7]. After an alliance is formed, its members are now working in their normal adversarial environment, as well as according to the mental states and axioms required for their interactions as part of the alliance. The following *Alliance* model (AL) specifies the conditions under which the group A^{al} can be said to be in an alliance and working with a new and constrained set of actions C^{al}, at time T_n.
$AL(A^{al}, C^{al}, w, T_n)$
1. A^{al} has a MB that all members are part of A^{al}:
$MB(A^{al}, (\forall A_i^{al} \in A^{al}) member(A_i^{al}, A^{al}), T_n)$
2. A^{al} has a MB that the group be maintained:
$MB(A^{al}, (\forall A_i^{al} \in A^{al}) Int.Th$
$(A_i, member(A_i, A^{al}), T_n, T_{n+1}, C_o), T_n)$
3. A^{al} has a MB that being members gives them high utility value:
$MB(A^{al}, (\forall A_i^{al} \in A^{al}) Al\_val(A_i^{al}, C^{al}, A^{al}, w) \geq Al\_TrH, T_n)$

Members' profiles are a crucial part of successful alliances. We assume that agents that have more accurate profiles of their adversaries will be more successful in such environments. Such agents will be able to predict when a member is about to breach the alliance's contract (item 2 in the above model), and take counter measures (when item 3 will falsify). The robustness of the alliance is in part a function of its members' trustfulness measure, objective position estimation, and other profile properties. We should note that an agent can simultaneously be part of more than one alliance.

Such a temporary alliance, where the group members do not have a joint goal but act collaboratively for the interest of their own individual goals, is classified as a *Treatment Group* by modern psychologists [12] (in contrast to a *Task Group*, where its members have a joint goal). The *Shared Activity* model as presented in [5] modeled *Treatment Group* behavior using the same *SharedPlans* formalization.

When comparing both definitions of an *alliance* and a *Treatment Group* we found an unsurprising resemblance between both models: the environment model's definitions are almost identical (see SA's definitions in [5]), and their *Selfish-Act* and *Cooperative Act* axioms conform to our adversarial agent's behavior. The main distinction between both models is the integration of a *Helpful-behavior act* axiom, in the *Shared Activity* which cannot be part of ours. This axiom states that an agent will consider taking action that will lower its Eval value (to a certain lower bound), if it believes that a group partner will gain a significant benefit. Such behavior cannot occur in a pure *adversarial environment* (as a zero-sum game is), where the alliance members are constantly on watch to manipulate their alliance to their own advantage.

A6. Evaluation Maximization Axiom. In a case when all other axioms are inapplicable, we will proceed with the action that maximizes the heuristic value as computed in the *Eval* function.
$(\forall A_{ag}, A_o \in A, \alpha \in C_{ag}, T_n, w \in W)$
$Bel(A_{ag}, (\forall \gamma \in C_{ag}) Eval(A_{ag}, \alpha, w) \geq Eval(A_{ag}, \gamma, w), T_n)$
$\Rightarrow Pot.Int.To(A_{ag}, \alpha, T_n, T_\alpha, w)$

T1. Optimality on *Eval* = *Utility* The above axiomatic model handles situations where the *Utility* is unknown and the agents are bounded rational agents. The following theorem shows that in bilateral interactions, where the agents have the real *Utility* function (i.e., *Eval* = *Utility*) and are rational agents, the axioms provide the same optimal result as classic adversarial search (e.g., *Min-Max*).

THEOREM 1. *Let A^e_{ag} be an unbounded rational* AE *agent using the* Eval *heuristic evaluation function, A^u_{ag} be the same agent using the true* Utility *function, and A_o be a sole unbounded utility-based rational adversary. Given that* Eval = Utility*:*
$(\forall \alpha \in C_{A^u_{ag}}, \alpha` \in C_{A^e_{ag}}, T_n, w \in W)$
$Pot.Int.To(A^u_{ag}, \alpha, T_n, T_\alpha, w) \rightarrow$
$Pot.Int.To(A^e_{ag}, \alpha`, T_n, T_\alpha, w) \wedge$
$((\alpha = \alpha`) \vee (Utility(A^u_{ag}, \alpha, w) = Eval(A^e_{ag}, \alpha`, w)))$

Sketch of proof — Given that A^u_{ag} has the real utility function and unbounded resources, it can generate the full game tree and run the optimal *MinMax* algorithm to choose the highest utility value action, which we denote by, α. The proof will show that A^e_{ag}, using the AE axioms, will select the same or equal utility $\alpha`$ (when there is more than one action with the same max utility) when *Eval* = *Utility*.
($A1$) Goal achieving axiom — suppose there is an α such that its completion will achieve A^u_{ag}'s goal. It will obtain the highest utility by *Min-Max* for A^u_{ag}. The A^e_{ag} agent will select α or another action with the same utility value via *A1*. If such α does not exist, A^e_{ag} cannot apply this axiom, and proceeds to *A2*.
($A2$) Preventive act axiom — (1) Looking at the basic case (see *Prop1*), if there is a β which leads A_o to achieve its goal, then a preventive action α will yield the highest utility for A^u_{ag}. A^u_{ag} will choose it through the utility, while A^e_{ag} will choose it through *A2*. (2) In the general case, β is a *highly beneficial action* for A_o, thus yields low utility for A^u_{ag}, which will guide it to select an α that will prevent β, while A^e_{ag} will choose it through *A2*.[1] If such β does not exist for A_o, then *A2* is not applicable, and A^e_{ag} can proceed to *A3*.
($A3$) Suboptimal tactical move axiom — When using a heuristic *Eval* function, A^e_{ag} has a partial belief in the profile of its adversary (item 4 in AE model), which may lead it to believe in *SetActions* (*Prop1*). In our case, A^e_{ag} is holding a full profile on its optimal adversary and knows that A_o will behave optimally according to the real utility values on the complete search tree, therefore, any belief about suboptimal *SetAction* cannot exist, yielding this axiom inapplicable. A^e_{ag} will proceed to *A4*.
($A4$) Profile detection axiom — Given that A^e_{ag} has the full profile of A_o, none of A^e_{ag}'s actions can increase its knowledge. That axiom will not be applied, and the agent will proceed with *A6* (*A5* will be disregarded because the interaction is bilateral).
($A6$) Evaluation maximization axiom — This axiom will select the max *Eval* for A^e_{ag}. Given that *Eval* = *Utility*, the same α that was selected by A^u_{ag} will be selected.

3. EVALUATION

The main purpose of our experimental analysis is to evaluate the model's behavior and performance in a real adversarial environment. This section investigates whether bounded

[1]A case where following the completion of β there exists a γ which gives high utility for Agent A^u_{ag}, cannot occur because A_o uses the same utility, and γ's existence will cause it to classify β as a low utility action.

rational agents situated in such adversarial environments will be better off applying our suggested behavioral axioms.

3.1 The Domain

To explore the use of the above model and its behavioral axioms, we decided to use the *Connect-Four* game as our adversarial environment. *Connect-Four* is a 2-player, zero-sum game which is played using a 6x7 matrix-like board. Each turn, a player drops a disc into one of the 7 columns (the set of 21 discs is usually colored yellow for player 1 and red for player 2; we will use White and Black respectively to avoid confusion). The winner is the first player to complete a horizontal, vertical, or diagonal set of four discs with its color. On very rare occasions, the game might end in a tie if all the empty grids are filled, but no player managed to create a 4-disc set.

The *Connect-Four* game was solved in [1], where it is shown that the first player (playing with the white discs) can force a win by starting in the middle column (column 4) and playing optimally However, the optimal strategy is very complex, and difficult to follow even for complex bounded rational agents, such as human players.

Before we can proceed checking agent behavior, we must first verify that the domain conforms to the adversarial environment's definition as given above (which the behavioral axioms are based on). First, when playing a *Connect-Four* game, the agent has an intention to win the game (item 1). Second (item 2), our agent believes that in *Connect-Four* there can only be one winner (or no winner at all in the rare occurrence of a tie). In addition, our agent believes that its opponent to the game will try to win (item 3), and we hope it has some partial knowledge (item 4) about its adversary (this knowledge can vary from nothing, through simple facts such as age, to strategies and weaknesses).

Of course, not all *Connect-Four* encounters are adversarial. For example, when a parent is playing the game with its child, the following situation might occur: the child, having a strong incentive to win, treats the environment as adversarial (it intends to win, understands that there can only be one winner, and believes that its parent is trying to beat him). However, the parent's point of view might see the environment as an educational one, where its goal is not to win the game, but to cause enjoyment or practice strategic reasoning. In such an educational environment, a new set of behavioral axioms might be more beneficial to the parent's goals than our suggested adversarial behavioral axioms.

3.2 Axiom Analysis

After showing that the *Connect-Four* game is indeed a zero-sum, bilateral adversarial environment, the next step is to look at players' behaviors during the game and check whether behaving according to our model does improve performance. To do so we have collected log files from completed *Connect-Four* games that were played by human players over the Internet. Our collected log file data came from Play by eMail (PBeM) sites. These are web sites that host email games, where each move is taken by an email exchange between the server and the players. Many such sites' archives contain real competitive interactions, and also maintain a ranking system for their members. Most of the data we used can be found in [6].

As can be learned from [1], *Connect-Four* has an optimal strategy and a considerable advantage for the player who starts the game (which we call the *White* player). We will concentrate in our analysis on the second player's moves (to be called *Black*). The White player, being the first to act, has the so-called *initiative advantage*. Having the advantage and a good strategy will keep the Black player busy reacting to its moves, instead of initiating *threats*. A threat is a combination of three discs of the same color, with an empty spot for the fourth winning disk. An *open threat* is a threat that can be realized in the opponent's next move. In order for the Black player to win, it must somehow turn the tide, take the advantage and start presenting threats to the White player. We will explore Black players' behavior and their conformance to our axioms.

To do so, we built an application that reads log files and analyzes the Black player's moves. The application contains two main components: (1) a *Min-Max* algorithm for evaluation of moves; (2) *open threats detector* for the discovering of open threats. The *Min-Max* algorithm will work to a given depth, d and for each move α will output the heuristic value for the next action taken by the player as written in the log file, $h(\alpha)$, alongside the maximum heuristic value, $max_h(\alpha)$, that could be achieved prior to taking the move (obviously, if $h(\alpha) \neq max_h(\alpha)$, then the player did not do the optimal move heuristically). The threat detector's job is to notify if some action was taken in order to block an open threat (not blocking an open threat will probably cause the player to lose in the opponent's next move).

The heuristic function used by *Min-Max* to evaluate the player's utility is the following function, which is simple to compute, yet provides a reasonable challenge to human opponents:

Definition 8. Let $Group$ be an adjacent set of four squares that are horizontal, vertical, or diagonal. $Group_b^n$ ($Group_w^n$) be a $Group$ with n pieces of the *black* (*white*) color and $4-n$ empty squares.

$$h = ((Group_b^1 * \alpha) + (Group_b^2 * \beta) + (Group_b^3 * \gamma) + (Group_b^4 * \infty)) - ((Group_w^1 * \alpha) + (Group_w^2 * \beta) + (Group_w^3 * \gamma) + (Group_w^4 * \infty))$$

The values of α, β and δ can vary to form any desired linear combination; however, it is important to value them with the $\alpha < \beta < \delta$ ordering in mind (we used 1, 4, and 8 as their respective values). Groups of 4 discs of the same color means victory, thus discovery of such a group will result in ∞ to ensure an extreme value.

We now use our estimated evaluation function to evaluate the Black player's actions during the *Connect-Four* adversarial interaction. Each game from the log file was input into the application, which processed and output a reformatted log file containing the h value of the current move, the max_h value that could be achieved, and a notification if an open threat was detected. A total of 123 games were analyzed (57 with White winning, and 66 with Black winning). A few additional games were manually ignored in the experiment, due to these problems: a player abandoning the game while the outcome is not final, or a blunt irrational move in the early stages of the game (e.g., not blocking an obvious winning group in the first opening moves). In addition, a single tie game was also removed. The simulator was run to a search depth of 3 moves. We now proceed to analyze the games with respect to each behavioral axiom.

Table 1: Average heuristic difference analysis

	Black losses	Black Won
Avg' min_h	-17.62	-12.02
Avg' 3 lowest h moves (min_h^3)	-13.20	-8.70

Table 2: Black's winnings percentages

	% of games
$min_h^3 < -11.5$	12%
$min_h^3 > -4$	5%
$-11.5 \leq min_h^3 \leq -4$	83%

3.2.1 Affirming the Suboptimal tactical move axiom

The following section presents the heuristic evaluations of the Min-Max algorithm for each action, and checks the amount and extent of suboptimal tactical actions and their implications on performance.

Table 1 shows results and insights from the games' heuristic analysis, when search depth equals 3 (this search depth was selected for the results to be comparable to [9], see Section 3.2.3). The table's heuristic data is the difference between the present maximal heuristic value and the heuristic value of the action that was eventually taken by the player (i.e., the closer the number is to 0, the closer the action was to the maximum heuristic action).

The first row presents the difference values of the action that had the maximal difference value among all the Black player's actions in a given game, as averaged over all Black's winning and losing games (see respective columns). In games in which the Black player loses, its average difference value was -17.62, while in games in which the Black player won, its average was -12.02. The second row expands the analysis by considering the 3 highest heuristic difference actions, and averaging them. In that case, we notice an average heuristic difference of 5 points between games which the Black player loses and games in which it wins. Nevertheless, the importance of those numbers is that they allowed us to take an educated guess on a threshold number of 11.5, as the value of the *TrH* constant, which differentiates between normal actions and *highly beneficial* ones.

After finding an approximated *TrH* constant, we can proceed with an analysis of the importance of *suboptimal* moves. To do so we took the subset of games in which the minimum heuristic difference value for Black's actions was 11.5. As presented in Table 2, we can see the different min_h^3 average of the 3 largest ranges and the respective percentage of games won. The first row shows that the Black player won only 12% of the games in which the average of its 3 highest heuristically difference actions (min_h^3) was smaller than the suggested threshold, $TrH = 11.5$.

The second row shows a surprising result: it seems that when $min_h^3 > -4$ the Black player rarely wins. Intuition would suggest that games in which the action evaluation values were closer to the maximal values will result in more winning games for Black. However, it seems that in the *Connect-Four* domain, merely responding with somewhat easily expected actions, without initiating a few surprising and suboptimal moves, does not yield good results. The last row sums up the main insights from the analysis; most of Black's wins (83%) came when its min_h^3 was in the range of -11.5 to -4. A close inspection of those Black winning games shows the following pattern behind the numbers: after standard opening moves, Black suddenly drops a disc into an isolated column, which seems a waste of a move. White continues to build its *threats*, while usually disregarding Black's last move, which in turn uses the isolated disc as an anchor for a future winning *threat*.

The results show that it was beneficial for the Black player to take suboptimal actions and not give the current highest possible heuristic value, but will not be too harmful for its position (i.e., will not give *high beneficial value* to its adversary). As it turned out, learning the threshold is an important aspect of success: taking wildly risky moves ($min_h^3 < -11.5$) or trying to avoid them ($min_h^3 > -4$) reduces the Black player's winning chances by a large margin.

3.2.2 Affirming the Profile Monitoring Axiom

In the task of showing the importance of monitoring one's adversaries' profiles, our log files could not be used because they did not contain repeated interactions between players, which are needed to infer the players' knowledge about their adversaries. However, the importance of opponent modeling and its use in attaining tactical advantages was already studied in various domains ([3, 9] are good examples).

In a recent paper, Markovitch and Reger [9] explored the notion of learning and exploitation of opponent weakness in competitive interactions. They apply simple learning strategies by analyzing examples from past interactions in a specific domain. They also used the *Connect-Four* adversarial domain, which can now be used to understand the importance of monitoring the adversary's profile.

Following the presentation of their theoretical model, they describe an extensive empirical study and check the agent's performance after learning the weakness model with past examples. One of the domains used as a competitive environment was the same *Connect-Four* game (*Checkers* was the second domain). Their heuristic function was identical to ours with three different variations (H1, H2, and H3) that are distinguished from one another in their linear combination coefficient values. The search depth for the players was 3 (as in our analysis). Their extensive experiments check and compare various learning strategies, risk factors, predefined feature sets and usage methods. The bottom line is that the *Connect-Four* domain shows an improvement from a 0.556 winning rate before modeling to a 0.69 after modeling (page 22). Their conclusions, showing improved performance when holding and using the adversary's model, justify the effort to monitor the adversary profile for continuous and repeated interactions.

An additional point that came up in their experiments is the following: after the opponent weakness model has been learned, the authors describe different methods of integrating the opponent weakness model into the agent's decision strategy. Nevertheless, regardless of the specific method they chose to work with, all integration methods might cause the agent to take suboptimal decisions; it might cause the agent to prefer actions that are *suboptimal* at the present decision junction, but which might cause the opponent to react in accordance with its weakness model (as represented by our agent) which in turn will be beneficial for us in the future. The agent's behavior, as demonstrated in [9] further confirms and strengthens our *Suboptimal Tactical Axiom* as discussed in the previous section.

3.2.3 Additional Insights

The need for the *Goal Achieving*, *Preventive Act*, and *Evaluation Maximization* axioms are obvious, and need no further verification. However, even with respect to those axioms, a few interesting insights came up in the log analysis. The *Goal achieving* and *Preventive Act* axioms, though theoretically trivial, seem to provide some challenge to a human player. In the initial inspection of the logs, we encountered few games[2] where a player, for inexplicable reasons, did not block the other from winning or failed to execute its own winning move. We can blame those faults on the human's lack of attention, or a typing error in its move reply; nevertheless, those errors might occur in bounded rational agents, and the appropriate behavior needs to be axiomatized.

A typical *Connect-Four* game revolves around generating threats and blocking them. In our analysis we looked for explicit preventive actions, i.e., moves that block a group of 3 discs, or that remove a future threat (in our limited search horizon). We found that in 83% of the total games there was at least one preventive action taken by the Black player. It was also found that Black averaged 2.8 preventive actions per game on the games in which it lost, while averaging 1.5 preventive actions per game when winning. It seems that Black requires 1 or 2 preventive actions to build its initial taking position, before starting to present threats. If it did not manage to win, it will usually prevent an extra threat or two before succumbing to White.

4. RELATED WORK

Much research deals with the axiomatization of teamwork and mental states of individuals: some models use knowledge and belief [10], others have models of goals and intentions [8, 4]. However, all these formal theories deal with agent teamwork and cooperation. As far as we know, our model is the first to provide a formalized model for explicit adversarial environments and agents' behavior in it.

The classical *Min-Max* adversarial search algorithm was the first attempt to integrate the opponent into the search space with a weak assumption of an optimally playing opponent. Since then, much effort has gone into integrating the opponent model into the decision procedure to predict future behavior. The $M*$ algorithm presented by Carmel and Markovitch [2] showed a method of incorporating opponent models into adversary search, while in [3] they used learning to provide a more accurate opponent model in a 2-player repeated game environment, where agents' strategies were modeled as finite automata. Additional *Adversarial planning* work was done by Willmott et al. [13], which provided an adversarial planning approach to the game of *GO*.

The research mentioned above dealt with adversarial search and the integration of opponent models into classical utility-based search methods. That work shows the importance of opponent modeling and the ability to exploit it to an agent's advantage. However, the basic limitations of those search methods still apply; our model tries to overcome those limitations by presenting a formal model for a new, mental state-based adversarial specification.

5. CONCLUSIONS

We presented an *Adversarial Environment* model for a bounded rational agent that is situated in an N-player, zero-sum environment. We used the *SharedPlans* formalization to define the model and the axioms that agents can apply as behavioral guidelines.

The model is meant to be used as a guideline for designing agents that need to operate in such adversarial environments. We presented empirical results, based on *Connect-Four* log file analysis, that exemplify the model and the axioms for a bilateral instance of the environment.

The results we presented are a first step towards an expanded model that will cover all types of adversarial environments, for example, environments that are non-zero-sum, and environments that contain *natural* agents that are not part of the direct conflict. Those challenges and more will be dealt with in future research.

[2]These were later removed from the final analysis.

6. ACKNOWLEDGMENT

This research was supported in part by Israel Science Foundation grants #1211/04 and #898/05.

7. REFERENCES

[1] L. V. Allis. A knowledge-based approach of Connect-Four — the game is solved: White wins. Master's thesis, Free University, Amsterdam, The Netherlands, 1988.

[2] D. Carmel and S. Markovitch. Incorporating opponent models into adversary search. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, pages 120–125, Portland, OR, 1996.

[3] D. Carmel and S. Markovitch. Opponent modeling in multi–agent systems. In G. Weiß and S. Sen, editors, *Adaptation and Learning in Multi–Agent Systems*, pages 40–52. Springer-Verlag, 1996.

[4] B. J. Grosz and S. Kraus. Collaborative plans for complex group action. *Artificial Intelligence*, 86(2):269–357, 1996.

[5] M. Hadad, G. Kaminka, G. Armon, and S. Kraus. Supporting collaborative activity. In *Proc. of AAAI-2005*, pages 83–88, Pittsburgh, 2005.

[6] http://www.gamerz.net/~pbmserv/.

[7] S. Kraus and D. Lehmann. Designing and building a negotiating automated agent. *Computational Intelligence*, 11:132–171, 1995.

[8] H. J. Levesque, P. R. Cohen, and J. H. T. Nunes. On acting together. In *Proc. of AAAI-90*, pages 94–99, Boston, MA, 1990.

[9] S. Markovitch and R. Reger. Learning and exploiting relative weaknesses of opponent agents. *Autonomous Agents and Multi-Agent Systems*, 10(2):103–130, 2005.

[10] Y. M. Ronald Fagin, Joseph Y. Halpern and M. Y. Vardi. *Reasoning about knowledge*. MIT Press, Cambridge, Mass., 1995.

[11] P. Thagard. Adversarial problem solving: Modeling an oponent using explanatory coherence. *Cognitive Science*, 16(1):123–149, 1992.

[12] R. W. Toseland and R. F. Rivas. *An Introduction to Group Work Practice*. Prentice Hall, Englewood Cliffs, NJ, 2nd edition edition, 1995.

[13] S. Willmott, J. Richardson, A. Bundy, and J. Levine. An adversarial planning approach to Go. *Lecture Notes in Computer Science*, 1558:93–112, 1999.