Discovery, Utilization, and Analysis of Credible Threats for 2 X 2 Incomplete Information Games in TOM

Jolie Olsen
University of Tulsa
Tulsa, OK
jolie.d.olsen@gmail.com

Sandip Sen
University of Tulsa
Tulsa, OK
sandip@utulsa.edu

ABSTRACT

Steven Brams's Theory of Moves (TOM) is an alternative to traditional game theoretic treatment of real-life interaction in which players choose strategies based on analysis of future moves and counter-moves that arise if game-play commences at a specified start state and either player can choose to move first. In repeated play, players using TOM rationality arrive at non-myopic equilibrium[2]. One advantage of TOM is its ability to model scenarios in which power asymmetries exist between players. In particular, threat power, i.e., the ability of an agent to threaten and sustain immediately disadvantageous outcomes to force a desirable result long term, can be utilized to induce Pareto optimal states in games such as Prisoner's Dilemma which result in Pareto dominated outcomes using traditional methods. Unfortunately, prior work on TOM is limited by an assumption of complete information. This paper presents a mechanism that can be used by an agent to utilize threat power when playing a strict, ordinal 2×2 game under incomplete information. We also analyze the benefits of threat power and support this analysis with empirical evidence.

Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: Multiagent Systems

General Terms

Design, Management, Theory

Keywords

Theory of Moves, Threats, Games of incomplete information

1. INTRODUCTION

Reasoning and learning mechanisms in single-stage games continue to be an active research area in multiagent systems [1, 4, 7, 8, 11, 12]. Unfortunately, many approaches presuppose reasoning in which agents act simultaneously and payoffs are immediately distributed after a single interaction. The limitation of such an approach is that it fails to adequately represent an abundance of real-life scenarios in which agents engage in a natural move-counter move process or take decisions as a function of some initial state. In analyzing such dynamic games, concepts such as single-shot stage games and Nash equilibria (NE) [5, 6] are ill-equipped and more elaborate techniques are required to understand realistic outcomes. Hence, extensive form game frameworks have been studied with solution concepts including subgame perfect equilibrium [5]. In these games, both the start state and the initial player is specified and a finite game tree, without cycles, is analyzed to derive equilibrium strategies [10]. While this model does address scenarios where agents can alternate moves, it still does not adequately attend to settings where two or more agents are considering their options from a state of the world, where any of the agents can be the first mover and where there is a possibility of cycling between world states.

In Steven J. Brams's 1994 book Theory of Moves [3], a framework is provided where play commences at a particular state and subsequent moves are determined from a finite lookahead in conjunction with backward induction analysis. As a result, either player has the opportunity to initiate play. The basic TOM framework considers complete information games and analyzes convergence to *Non-Myopic Equilibrium* (NME). Brams's work also studies games in which there exist inherent asymmetries between players and establishes the concepts of threat power (the ability to sustain a negative outcome in the short term in order to ultimately arrive at a more preferred state, moving power (the ability to sustain game cycling), and order power (the ability to force a game to proceed in a specified order).

TOM's reliance on complete information results in obvious limitations. Ghosh and Sen proposed a probabilistic learning algorithm to circumvent this requirement in which agents operating with incomplete information engage in repeated play and converge to a primitive form of NME [9], which we refer to as *basic NME*. If an agent wishes to utilize threat power, however, knowledge of an opponent's payoff structure is also essential. So while TOM learners suffice for simple games between symmetric players, an enhanced technique, which allows inference of threat states, is required if one player wields greater authority than its opponent.

Unfortunately, in the domain of 2×2 games, knowledge of one's payoff matrix coupled with basic NME convergence does not always suffice to deduce complete information of an opponent's payoff structure. In fact such inference is possible only in a handful of games. However, we show that the *complete* knowledge of an opponent's payoff is not necessary to infer threat power, and so long as the range of possible payoffs is constrained in a certain way, threat power can be utilized.

We consider all possible 2×2 games with strict, ordinal preferences (players' preferences for a total order of the 4 possible outcomes). We introduce an equivalence relation over these games which preserves threat states in 97% of

partitions, along with a mechanism for inferring the equivalence class of a game being played. Additionally, even if a player finds itself playing a game in one of the of inconsistent partitions, we show that, for certain categories of threats, it will not lose utility by utilizing a threat from any game within the partition. This result allows TOM agents to recognize and utilize threat powers in 2×2 incomplete information games with strict, ordinal preferences.

Our analysis of threat power shows that at times it might be detrimental to utilize threat power. We then characterize threats by their relative benefits and risks and support our conclusions with experimental results. To the best of our knowledge, this is the first concerted effort in characterizing and utilizing of threat power to the incomplete information setting.

The rest of this paper is organized follows: section 2 covers relevant background information including an overview of TOM, NME, and threat power. Section 3 explains definitions and notations used. Section 4 presents our mechanism for discovering threat states in incomplete information settings. Section 5 analyzes the effectiveness of different threats with section 6 providing simulation results which match this analysis. Finally, section 7 provides a brief conclusion and closing remarks about foreseeable future research goals that follow from this research.

2. BACKGROUND

This section reviews the essential aspects of the Theory of Moves framework and describes threat power entirely.

2.1 Rationality Rules

TOM is a dynamic approach to game theory in which players engage in a move-counter-move process rather than being confined to a single shot interaction. Unlike its predecessors such as normal form games, TOM requires the initial state of a game to be determined as a function of initial strategies selected by players. Once an initial state has been established, the following **basic rules** dictate play of the game [2, p. 23]:

• Play starts at an outcome, called the initial state, determined by an initial strategy profile chosen by the players.
• Either player can unilaterally switch its strategy, and thereby change the initial state into a new state, in the same row or column as the initial state. The player who switches is called Player 1 (P1).
• Player 2 (P2) can respond by unilaterally switching its strategy, thereby moving the game to a new state.
• The alternating responses continue until the player whose turn it is to move next chooses not to switch its strategy. The game terminates in a final state, which is the outcome of the game.

The following are **supplemental rules**, as listed in [2]:

• If, at any state in the move-countermove process, a player whose turn it is to move next receives its best payoff, it will not move from this state.
• If it is rational for one player to move and the other player not to move from the initial state, then the player who moves takes precedence: its move overrides the player who stays, so the outcome will be that induced by the player who moves.
• If one player, say, C can induce a better state for itself by moving than by staying, but R by moving can induce a state Pareto-superior to C's induced state - then R will move, even if it otherwise would prefer to stay, to effect a better outcome.

The authors propose a final supplemental rule in addition to those that Brams proposed:

• If a player, say R, can induce a better state for itself by moving but its opponent, C, by moving, can induce a state Pareto superior to R's induced state, and C has a pure strategy for moving, then R will not move from the initial state so that C's move will take precedence

	S	D
S	$3,3$	$2,4$
D	$4,2$	$1,1$

Figure 1: Game 57 | Chicken

R	C	R	C	
(3,3)	(4,2)	(1,1)	(2,4)	(3,3)
[3,3]	[2,4]	[2,4]	[2,4]	

Figure 2: Game 57 | Chicken - Subgame with Initial State: 0, Initial Player: R

2.2 NME and backwards induction

To select an initial strategy and make rational decisions about moving, a player must think non-myopically by conducting analysis not only on its opponent's current strategy but what it believes their future strategies will be. This analysis is called *backwards induction* and outcomes induced as a result of it are termed *non-myopic equilibria* or *NMEs*. For the purpose of this paper, we define NMEs which conform only to the basic rules as *basic* NMEs and those which conform to all of the rules as *cooperative* NMEs. NMEs guarantee Nash pay off in the worst case, and often are Pareto optimal states that manifest in games such as *Prisoner's Dilemma* which are unattainable under Nash conventions. Furthermore, not all games contain a pure strategy NE but every 2 × 2 game has at least one pure NME [2, p. 33].

Let us analyse Fig. 1. Two pure NE exist in this game: (2,4) and (4,2), excluding the perceived "fairest" state: (3,3). Fortunately, three NME exist in this game, one of which is (3,3). To identify NME, backwards induction is computed on all subgames, with a subgame defined as a hypothetical move counter-move process originating from a specified initial state and designated first player.

Examine the subgame with start state (3,3) and initial player row[1] given in Fig. 2. The backwards induction process begins by examining the last position in the rotation, (2,4): if play makes it to this point, C has two options: stay at (2,4) or switch strategies resulting in the payoff distribution (3,3). Clearly, (2,4) is preferred so C chooses to stay. Thus if game play arrives at (2,4) with C playing, (2,4) will be induced; it is dubbed a *survivor* and indicated by square brackets. This is known as *blockage* because play will not

[1]For the remainder of the paper, row player will be designated as R and column player designated as C

	W	M
B	$3,3$	$1,4$
A	$2,2$	$4,1$

Figure 3: Game 30 | The Cuban Missile Crisis - row player *US* and column player *SU*

commence beyond this point. At (1,1), R has two options: stay at (1,1) or switch rows to (2,4). R prefers (2,4) so it becomes the survivor at (1,1). At (4,2), it might appear C will stay rather than move to (1,1) but C moves. It does so because it realizes based on the preceding analysis that if play proceeds to (1,1), it is rational for R to move, and play will eventually terminate at (2,4): C's most preferred state. Therefore (2,4) is the survivor at (4,2). Finally at initial state (3,3), R has the decision to stay and receive 3 or move and receive 2. R prefers (3,3) so it becomes the survivor, indicating blockage and no initial move from R. Therefore, (3,3) is the basic NME for this subgame.[2]

Similar backwards induction on the subgame with same initial state but C as the initial player reveals its basic NME is likewise (3,3). Therefore, for subgames of Chicken with initial state (3,3), the NME induced is (3,3) regardless of initial player. We then call (3,3) the *unique NME* for subgames games of Chicken originating at (3,3).

2.3 Threat Power

An essential justification for TOM is the implicit assumption in traditional game theory that players have matched capabilities. In real life natural inequalities often arise between opponents. Many political and foreign policy events involving such asymmetric "players" can appear perplexing when analyzed exclusively within the context of traditional game theory, as the equilibira predicted often fail to match outcomes that appear repeatedly in real life.

Let us examine Fig. 3. R nor C have a dominant strategy and no pure NE exists; traditional theory fails to offer a rational choice of moves. Let us assume one player, C, can endure a less than desirable outcome in the short time if this results in a long term benefit. If C locates a Pareto inferior state it can, by threatening that state, induce R to move. Assume players are currently at state (3,3) and R's move is next. C can threaten: "if you switch from row 2 to row 1 with the hopes of inducing (4,1), then I will refuse to move from (2,2)." If C follows through on this threat, over time, R will learn to terminate play upon arrival at (3,3) because doing otherwise results in a lower payoff. This is an example of a *compellent threat*, the other type being a *deterrent threat*. We formalizes these concepts now.

Assume a player p_1 makes a threat against opponent p_2 in an attempt to induce outcome (i, j) which has a payoff for each player represented as (x^1_{ij},x^2_{ij}).

• p_1's *breakdown state* is the Pareto inferior outcome it's threatening. The associated strategy is p_1's *breakdown strategy*

• p_1's *threat state* is the Pareto dominated outcome it's trying to induce by using a threat. The associated strategy is p_1's *threat strategy*

• p_1's threat is *real* iff when carried out, it worsens the payoff for p_2.

• p_1's threat is *rational*, when successful in deterring p_2, it improves p_1's own payoff over what it would be if p_2 moved from (i, j).

• A threat is *credible* if it is both real and rational.

[2]In this paper, basic NME are indicated by boldfaced payoff in brackets beneath the initial state.

$3,3$	$4,1$
$2,2$	$1,4$

(a): Game 23

$3,3$	$2,2$
$1,4$	$4,1$

(b): Game 23 after player transposition

Figure 4: Game 23 as a Structurally Unique Game

Consider threat state (x^1_{ij},x^2_{ij}) and the existence of a Pareto inferior breakdown state (x^1_{mn},x^2_{mn}) from the perspective of R. There are two possibilities for R to carry out a threat:

• *R stays: threat and breakdown strategy are the same.* In this case, $m = i$, and the threat is called a **compellent threat**

• *R moves: threat and breakdown strategy are different.* In this case, m $\neq$i, and the threat is called a **deterrent threat**.

3. DEFINITIONS AND NOTATION

3.1 Domain

We limit our study to two-player, two-action games of total conflict and total order. Literature has traditionally proposed 57 such games exist by claiming two games are equivalent if a transposition of their players results in the same set of outcome. That is clearly not the case when dealing with threat power. To see why, refer to Fig. 4 which presents Game 23 on the left and the result of player transposition on the right. While these games are structurally distinct in the traditional sense, (a) contains a t_c for R while (b) does not. Because of this, we consider 108 games: the 57 structurally distinct games and their player inversions[3].

3.2 e values

For any subgame the *computing effort*, or e-value, is defined as the number of unilateral strategy switches necessary to induce its NME. If the subgame induces a cycle, the e-value is equal to the number of states in the rotation. For example, in the subgame represented in Fig. 1, the e-value is 0; and any cyclical 2×2 subgame has an e-value of 4. On first glance it might not be apparent the distinction between NMEs and e-values, and in games of a non-cyclic nature a player *can* infer one set of values from the other. However, in cyclic games e-values provide more information and NME is not sufficient alone to deduce them. This distinction provides motivation for development of the enhanced TOM Learners Algorithm in Section 4.1 so we formalize it below.

LEMMA 1. *Knowledge of* e *value* $\Rightarrow$ *knowledge of NME.*

PROOF. e value, by definition, specifies NME as its position in game rotation □

The converse is not necessarily true. Consider a subgame where the initial state is the NME. There is no way of knowing if the NME is induced by cycling or stagnant players so the e value cannot be deduced

[3]6 games are identical when inverted regardless of switching the row and column player so these games are not included twice in the domain

3.3 Game Notation

We represent a game between a player p_k and its opponent, denoted $p_{\bar{k}}$ as a set $g_i = \{ P_i^k, P_i^{\bar{k}}, e_i\}$ where P_i^k and $P_i^{\bar{k}}$ are p_k and $p_{\bar{k}}$'s payoff matricies, respectively. Consequently, we can extrapolate four outcomes $\{O_{1,1}, O_{1,2}, O_{2,1}, O_{2,2}\}$ where $O_{i,j} = (P_{i,j}^k, P_{i,j}^{\bar{k}})$. If ξ represents a given outcome, then ξ^D is the outcome diaganol to ξ, ξ_P^{ND} is the outcome resulting by a unilateral strategy switch by p_k (p_k's direct neighbor), and ξ_P^{NI} is the outcome resulting by a unilateral strategy switch by $p_{\bar{k}}$ (p_k's indirect neighbor). e_i is a vector matrix which holds a vector for each outcome, each of which contains two e-values, one for each subgame originating there. Each player p_k has a strategy profile $S_k = \{s_k^a, s_k^b\}$. An outcome in the game O_k is composed by a combination of player strategies. g_i thus is defined explicitly by the players' four unique strategy combinations.

3.4 Knowledge set

We represent p_k's knowledge concerning g_i with the variable $K_i^k \subset g_i$. The following demonstrate the potential knowledge sets p_k can possess:

(i) $K_i^k = \{P_i^k\}$; (ii) $K_i^k = \{P_i^k, P_i^{\bar{k}}\}$; (iii) $K_i^k = \{P_i^k, e_i\}$; (iv) $K_i^k = \{P_i^k, P_i^{\bar{k}}, e_i\}$

First note that ii is equivalent to iv.[4], so we may limit our discussion to sets i through iii. We define i as **incomplete**, ii as **complete**, and iii as **partial**.

If an agent has a complete knowledge set, it can trivially identify and subsequently utilize threat states. But with a partial knowledge set, this is not always the case. Before continuing, we present the following theorem which indicates a partial knowledge set is not sufficient to deduce a complete one, a result which will provide essential justification for development of the mechanism proposed in Section 4. Due to space constraints, the proof for this theorem has been omitted, but is available upon request of the authors.

THEOREM 2. *Let p_k be a player of game g_i with opponent $p_{\bar{k}}$. Assume p_k has a partial knowledge set. It is impossible to guarantee inference of a complete knowledge set.*

3.5 Equivalence Relation on the 2 × 2 games

Consider the following binary relation $\sim_k$:

$$\text{Let } g_i = \{ P_i^k, P_i^{\bar{k}}, e_i\},\ g_j = \{ P_j^k, P_j^{\bar{k}}, e_j\}$$
$$g_i \sim_k g_j \iff P_i^k = P_j^{\bar{k}} \wedge e_i = e_j$$

A similar relation can be defined for $p_{\bar{k}}$. WLOG we discuss $\sim_k$ and denote it simply as $\sim$. 57 equivalence classes are derived from $\sim$. We utilize this partitioning as it preserves threat states nicely, providing a strategy for discovering threat states later on when limited to partial knowledge.

LEMMA 3. *At most one t_c can occur in any equivalence class constructed by $\sim$.*

THEOREM 4. *Let $g_i \sim g_j$. If g_i contains some t_c, p_k receives no worse payoff using this threat in subgames of g_j.*

[4]If a player's knowledge set is ii it need only perform backwards induction to obtain e values.

$$P_k^{s,p}(S_i) = \sum_{l=i+1}^{4} \left(\prod_{n=i+1}^{l-1} P_k^{s,p}(S_n) \right) (1 - P_k^{s,p}(S_l)) f(O_l^k, O_i^k)$$

$$f(x,y) = \begin{cases} 1 & : x > y \\ 0.5 & : x = 0 \\ -1 & : x < y \end{cases}$$

Figure 5: Formula for Computing P values

4. MECHANISM FOR UTILIZING THREAT POWER IN GAMES OF INCOMPLETE INFORMATION

When an agent's knowledge set for a game is complete, discovery of credible threats is trivial but for limited knowledge sets threat states are not always as obvious. The goal then is to develop a protocol for agents to identify credible threats even when limited by incomplete or partial knowledge sets. We present herein such a mechanism.

4.1 Part I: Learning Phase

In the proceeding section, we show that if an agent has a partial knowledge set, it can identify the equivalence class of the game it is playing, and Theorem 4 indicates that if an agent knows this equivalence class then threat power can be utilized. Therefore, we wish to first employ a method by which a player can infer a partial knowledge set from an incomplete one. Ghosh and Sen proposed a learning algorithm in which players with incomplete knowledge sets, through repeated play, converge to a single outcome [9]. While most often these learners converge to an NME, the algorithm provides no insight for games which are cyclic and hence e-values cannot be obtained. An enhanced version of the algorithm to procure e-values is presented with the following procedure:

• Each agent selects an initial strategy, thus determining the subgame played

• Each agent calculates $P_j^{s,p}(S_i)$, the probability of player p_j moving from state S_i in subgame with initial state/player combination (s,p). For a player p_k, two distinct P values can be calculated: (1) The probability that p_k's *opponent*, denoted $p_{\bar{k}}$, will move from S_i; calculated as the ratio of times $p_{\bar{k}}$ has moved from S_i rather than stayed, initialized to 1.0 to provide for initial exploration, and (2) the probability p_k itself will move, calculated with the following formula:

$$P_k^{s,p}(S_i) = \sum_{l=i+1}^{4} \left(\prod_{n=i+1}^{l-1} P_k^{s,p}(S_n) \right) (1 - P_k^{s,p}(S_l)) f(O_l^k, O_i^k)$$

where O_l^k and O_i^k are p_k's payoff at outcomes O_l and O_i, respectively; and f is a function constructed to reflect p_k's preference between two outcomes.

• Play terminates when a cycle is reached or both players choose not to move.

4.2 Part II: Inference Phase

Upon completion of the Learning Phase, p_k has acquired a partially complete knowledge set. The task then becomes to identify the equivalence class of the game it is playing. An obvious brute force approach can be employed: an agent exhaustively generates all possible opponent payoff matrices,

computes backwards induction on all subgames to generate e-values, checks each games for equivalence using its partial knowledge disregarding games that are not equivalent and checking this list at each step against all equivalence classes until it finds a match.

As an alternative, the authors propose the following method: p_k generates 4 *belief variables* α, β, γ, and δ representing its belief of $p_{\bar{k}}$'s preference for each state in the game. An agent trivially knows:

$\alpha \in \{1,2,3,4\}$, $\beta \in \{1,2,3,4\}$, $\gamma \in \{1,2,3,4\}$, $\delta \in \{1,2,3,4\}$

Simple *prediction rules* coupled with partial knowledge are used to eliminate values from these sets. Any further elimination necessary reduces to solving a simple constraint satisfaction problem. Before describing the CSP, the prediction rules are listed with the following convention used: $e_i(\alpha_k)$ is the e-value for subgame with initial player/state combination (α,p_k).[5]

1. $e(\alpha) = <4,4> \iff \alpha$ is the mutually most preferred.
   Proof: Assume $e_i(\alpha) = <4,4>$ for some α. $\Rightarrow \alpha_R > \alpha_R^D$ and $\alpha_C > \alpha_C^{ND}$ and $\alpha_C > \alpha_C^{NI}$ (from the subgame with R initializing). $\Rightarrow \alpha_R \geq 2$ and $\alpha_C \geq 3$. Also, $\alpha_C > \alpha_C^D$ and $\alpha_R > \alpha_R^{ND}$ and $\alpha_R > \alpha_R^{NI}$ (from the subgame with C initializing). $\Rightarrow \alpha_R = 4$ and $\alpha_C = 4$. $\therefore e_i(\alpha) = <4,4> \Rightarrow$ a mutual best state.

   Now, assume α is a mutual best state. $\Rightarrow$ regardless of which player begins, if we do backwards induction, no state can block, and α will cycle. $\therefore \alpha$ is a mutual best state $\Rightarrow e_i(\alpha) = <4,4>$

   Thus, by (i) and (ii), $e_i(\alpha) = <4,4> \iff \alpha$ is a mutual best state.

2. $e(\alpha_k) \in \{0,4\} \iff \alpha$ is not p_k's least preferred state.

3. $e(\alpha_k) \in \{1,2,3\} \Rightarrow \alpha$ is not p_k's most preferred state.

4. If α, β, γ are distinct outcomes and $e(\alpha_k) = e(\beta_k) = e(\gamma_k) = 0 \Rightarrow$ the remaining state is p_k's least preferred state.

5. If $e(\alpha_k) = 0$ and $e(\alpha_k) = 4$ and for all remaining states β, $e(\beta_k) \in \{1,2,3\}$, $\Rightarrow \alpha$ is p_k's most preferred state.

6. $e(\alpha_k) = 4$ or $e(\alpha_k^{NI}) = 3 \Rightarrow \alpha$ is one of p_k's two most preferred states.

7. If $e(\alpha_k) = 0$ and $P^k(\alpha) = 2 \Rightarrow$ the game does not contain a mutually least preferred state.

8. If α is $p_{\bar{k}}$'s most preferred state, and $P^k(\alpha^D) < P^k(\alpha)$ $\Rightarrow e(\alpha) = <4,0>$

9. If $P^k(\alpha^D) < P^k(\alpha)$ and $e(\alpha) = <4,0> \Rightarrow \alpha$ is one of $p_{\bar{k}}$'s two most preferred states. Furthermore, if α is $p_{\bar{k}}$'s second most preferred state $\Rightarrow \alpha^D$ is $p_{\bar{k}}$'s most preferred state.

10. If $P^k(\alpha^D) < P^k(\alpha)$ and $e(\alpha_k) = <4,2> \Rightarrow \alpha$ is $p_{\bar{k}}$'s most preferred state.

[5]a proof is given for the first proposition, but due to space constraints subsequent proofs, which can be developed in a similar manner, are omitted.

R	C	R	C	
(a,w)	(b,x)	(c,y)	(d,z)	(a,w)
$P(\alpha)$	$P(\beta)$	$P(\gamma)$	$P(\delta)$	

Figure 6: General Game: Generation of P Values - Example 1

	C	D
C	$3,4$	$4,2$
D	$2,3$	$1,1$

Figure 7: Game 1 | Pure Superfluous Compellent Threat for C

After application of rules if an exact opponent payoff has not been generated, inequalities can be derived from the agent's knowledge of e-values to solve a CSP. To establish inequalities, an agent regenerates the P values discussed in the previous section. This is done as follows: examine Fig. 6 which models an arbitrary game g_i with start state α, initial player R, and clockwise rotation α, β, γ, δ. If $m = C_i^R(\alpha)$ and ϕ is the outcome at the m^{th} position in the rotation $\Rightarrow P(\phi) = 0$ because blockage occurs at ϕ. Then for all outcomes θ such that θ occurs earlier then ϕ in the rotation, $P(\theta) = 1$. An example of how this is done is given:

First let us assume that after application of prediction rules an agent has constrained his belief variables to the following sets:

$\alpha \in \{1,2,3\}$, $\beta \in \{1,2\}$, $\gamma \in \{4\}$, $\delta \in \{2,3\}$

Assume in a subgame with rotation mentioned $m = C_i^R(\alpha)$ and $m = 4 \Rightarrow \alpha_C > \delta_C \wedge \alpha_C > \beta_C$. Simple constraint satisfaction eliminates the following: 1 and 2 from α, 3 from δ, and 2 from β to arrive at an exact payoff vector {3,1,4,2}

The application of prediction rules followed by constraint satisfaction is sufficient to predict 103 of 108 classes, and for games in the remaining 5 classes, the inefficient brute force approach can be applied as a last resort.

.

5. FORMAL ANALYSIS OF THREAT POWER

While on surface level all threats might appear the same, scrutiny reveals certain threats might be more effective than others. We propose the following categories for classifying threats, and term threats which do not fall in to one of the two categories as *useful*.

5.1 Superfluous Threats

There are a number of threats that while credible, are superfluous. We can consider two separate types of superfluous threats: those which are superfluous regardless of how an initial strategy selection method, and those which are superfluous because we could obtain the threat state by employing learning to select initial strategies.

5.1.1 Pure Superfluous Threats

Consider a game in which there exists one unique NME. If a threatener's threat state is the NME, then threat power is rendered ineffective. Certainly, the agent need not implement threat power because regardless of initial state, the NME, and hence the threat state, will be induced.

Game 1, in Fig. 7 presents an example of such a super-

	C	D
C	2, 2	4, 1
D	1, 4	3, 3

Figure 8: Game 32 | Prisoner's Dilemma - Superfluous Credible Threat when Learning for C and R

$$U_p(g, p_k) = \sum_{o \in O} P(g, o) p_g^{p_k}(o)$$
$$p_g^{p_k}(o) = \sum_{o_p \in O} \sum_{p_m \in N} P(o_p, o, p_m) P(p_m) v_k(o_p)$$

Figure 9: Function for Computing Potential Utility

fluous threat. In Game 1, there exists one unique NME: (3,4) which is also a deterrent threat for R. However, no matter which subgame is played, if agents use TOM rationality, (3,4) will be induced. Even if initial states are chosen at random, game play converges to the threatener's desired result. Threat power is superfluous in this case.

5.1.2 Learning Induced Superfluous Threats

Consider a game in which there exists two unique NMEs, one of which Pareto dominates the other. Suppose the Pareto dominant NME is a credible threat for one agent. *Prisoner's Dilemma* is an example of one such game and is shown in Fig. 8. In PD, (3,3) is a deterrent threat for both R and C. Furthermore, (3,3) is the NME for subgames with initial state (3,3), (4,1), and (1,4), while (2,2) is the NME for subgame with initial state (2,2). This means that if players are using TOM rules, the outcome depends on the initial state. Recall that an outcome is a specification of strategies from both players. Because this credible threat is the Pareto dominant of the two NMEs, players can simply learn to select (3,3) as the initial state by choosing as a rule the initial strategy that has fared best for them in the past.

5.2 Detrimental Threats

5.2.1 Potential Utility of a Threat

If agents have the ability to select their initial strategies in order to influence where game play initiates, then it is natural to consider the potential utility of a game. We define agent p_k's *potential utility* $U_p(g, p_k)$ of a game g by way of the formula given in Fig. 9 which is the sum over all outcomes o of g of the utility of o as a starting state, $p_g^a(o)$, times the probability of g initiating at o, $P(g, o)$. The second equation describes $p_g^{p_k}(o)$: p_k's utility of o as a start state where $P(O_p,o,p_m)$ the probability of inducing outcome o_p given start state o and initial player p_m, $P(p_m)$ the probability of p_m being selected as the initial player for that iteration, and $v_k(O_p)$ is p_k's valuation (payoff) of outcome o_p.

by the sum of the probability of each outcome being the end result of all subgames initiating at o, times p_k's valuation for that outcome, v_k. Note that if an outcome o converges to a unique NME, $p_g^{p_k}(o) = v_k(o)$.

If we consider a vector $\vec{t_g} =< P(g, o_1), P(g, o_2), \ldots, P(g, o_m) >$ to hold the probabilities of starting at each outcome and write $p_g^{p_k}(o)$ as a vector, then an alternate definition for potential utility can be given as $U_p(g, p_k) = \vec{t_g} \cdot p_g^{p_k}$.

	S	D
S	3, 3	2, 1
D	4, 2	1, 4

Figure 10: Game 47 | Detrimental Threat

5.2.2 Definition of Detrimental Threat

If $U_p^l(g, p_k)$ is a lower bound on potential utility, $U_p^u(g, p_k)$ an upper bound, and $U_p^{u*}(g, p_k)$ an upper bound when threat power is being used with credible threat o, and if the following inequality can be established:

$$U_p^{u*}(g, p_k) \leq U_p^l(g, p_k) < U_p^u(g, p_k),$$

then we can consider threat state o to be detrimental. The justification for this is simple: if the upper bound of potential utility is strictly greater than the upper bound on potential utility when threat power is being considered, then there exists an NME which p_k values more that is not a threat state. If the upper bound when using threat power is less than or equal to the lower bound without threat power, then the worst p_k can do without threat power is the same as when threat power is used, thus p_k cannot lose any utility by choosing the higher NME outcome as its dominant strategy, and it stands to lose out on this NME outcome as the effectiveness of threat power increases.

Considering the game in figure 10 with two unique NMEs: (3,3) for games commencing at (3,3) and (1,4), and (4,2) for games commencing at (2,1) and (4,2). R has a compellent threat at (3,3). Consider a game where the start state is selected randomly, then the probability vector $\vec{t_{47}} = (0.25, 0.25, 0.25, 0.25)$. In this case, $U_p(47, R) = 0.25 \times 3 + 0.25 \times 4 + 0.25 \times 3 + 0.25 \times 4 = 3.5$. The worst we can ever do is to start at some exclusive combination of (3,3) or (1,4) which would lead to a potential utility of 3. Now observe that as threat power becomes more effective, the probability vector $(1, 0, 0, 0) \rightarrow U_p^{u*}(47, R) = 3$. So as threat power becomes more effective, R's potential utility decreases. Hence it is never in R's best interest to utilize its compellent threat.

6. EMPIRICAL ANALYSIS OF THREAT POWER

The goal of our simulations was to observe the benefits of using threat power. We present empirical results that agree with the analysis presented in Section 5.

6.1 Simulation Setup

Two distinct sets of simulations were performed: ones where threat power was used and ones where it was disregarded. In each run of the simulations, a game g_i was chosen at random. For the threat power runs, one player is deemed threatener and its opponent threatenee. Two phases then occur: a learning/inference phase (which lasts 1,000 iterations or until convergence, whichever occurs first) in which the threatener employs the mechanism detailed in Section 4 to identify existing credible threats, and a threat phase (which lasts 1000 iterations) in which the threatener utilizes said threats against its opponent. For the runs without threat power, the simulation consists of the same number of iterations, but with both agents employing only the Enhanced TOM Learners learning algorithm. Iterations are defined as both agents playing a subgame of g_i until one

$$u(s_k^i,p_k)=\sum_{O_t\in s_k^i}\sum_{o_p\in O}\sum_{p_m\in N}(P(o_p,o_t,p_m)P(p_m)v_k(O_p))$$

Figure 11: Formula for Calculation of Utility of an Initial Strategy

player decides not to move (in which case the induced outcome is the blocked state) or a cycle occurs (in which case the induced outcome is the initial state). The subgame played each iteration is determined by initial strategy selections of each agent, and the initial player is chosen deterministically. If the first player declines to make an initial move, the second player is given the opportunity to move.

6.2 Utility of an Initial Strategy

For game g, p_k's i^{th} strategy s_k^i, is associated with a set of outcomes, termed an *action profile*: $s_R^i = \{O_{i,1}, O_{i,2}\}$ for R, or $s_C^i = \{O_{1,i}, O_{2,i}\}$ for C. Consider Fig. 8. R has two strategies: C and D. C is action profile {(C,C),(C,D)} and D is action profile {(D,C),(D,D)}. p_k can estimate the utility of s_k^i as shown in Fig. 11, where O is the set of possible outcomes in g, N the set of players, $P(O_p,O_t,p_m)$ the probability of inducing outcome o_p given start state o_t and initial player p_m, $P(p_m)$ the probability of p_m being selected as the initial player for that iteration, and $v_k(O_p)$ is p_k's payoff at outcome O_p.

6.3 Methods for Selecting Initial Strategies

In TOM the subgame played is determined by players' initial strategy choices. As a rule, during threat phases, if a credible threat exists the threatener selects its threat strategy initially. If its threat state differs from the initial state, it then implements its breakdown strategy to punish its opponent for deviation. For learning phases and for the threatenee, the following methods for strategy selection can be implemented:

6.3.1 Random Strategy Selection

With random strategy selection (R), an agent chooses an initial strategy randomly. If p represents the probability of each outcome in the game being selected as an initial state, then p has a discrete uniform distribution during learning phases if both agents use this method and a discrete uniform distribution over all outcomes in the threatener's threat strategy during threat phases.

6.3.2 Exploratory Learning Strategy Selection

In the Exploratory Learning method (EL), an agent gathers data upon termination of each iteration to learn the behavior of its opponent. To determine which initial strategy to select for an iteration, an agent first calculates the utility of all strategies in its strategy profile then selects the strategy with the highest, with exploration. The function for calculating the probability of player p_k selecting s_k^i as its initial strategy at time t is given in Fig. 12, with I_p the number of iterations remaining in the current phase. The function allows for exploration as the initial strategy is chosen randomly on the first iteration, and subsequent strategies are selected by highest utility with a monotonic increasing probability. This probability is a sigmoid function dependent upon how many iterations the agents are set to play.

$$P(s_k^i,p_k,t)=\begin{cases}\frac{1}{1+e^{-2t/I_p}} & : s_k^i=\arg\max u(s,p_k)\\ \frac{1}{|S|} & : s_k^i\neq\arg\max u(s,p_k)\end{cases}$$

Figure 12: Function for Probability of Selecting Initial Strategy

	H +T.P.			
	Type	all C.T.	t_c	t_d
Row	A	+4.08%	+3.43%	+10.13%
	S	+4.06%	+04.57%	-00.26%
	U	+4.36%	+02.75%	+19.86%
	D	-00.68%	-00.68%	00.00%
Col	A	-11.53%	-12.46%	-03.80%
	S	+00.67%	+00.77%	-00.17%
	U	-22.48%	-24.49%	-06.31%
	D	+31.44%	+31.44%	00.00%

Table 1: Simulation Results of using Hybrid Random Exploration Strategy Plus threat power

6.3.3 Hybrid Strategy Selection

In the Hybrid Strategy Selection method (H), R is used during the learning phase, and EL during the threat phase.

6.4 Results

Each simulation was performed on 10,000 randomly generated 2 × 2 cardinal matrices. No-conflict games and partial order games were discarded. In threat power runs, R was designated as threatener and C as threatenee.

Simulation results are listed in Tables 1, 2, and 3. The uppermost row shows the strategy selection method used. Note that R is not afforded an individual analysis; because agents employing this method select strategies at random, they never learn from past interactions, rendering both threat power and learning inert. + T.P. indicates threat power was used. The column headers following "Type" indicate the group of threats being analyzed: t_c refers to compellent threats, t_d deterrent, and ∀ C.T. indicates all credible threats were analyzed. The row headers underneath "Type" indicate the classification of threats being considered: A corresponds to all threats, S superfluous, D detrimental, and U useful. The table is divided in half: the top half detailing results for R and the bottom half for C. The number in each individual cell represents the average percent utility gain

	EL + T.P.			
	Type	all C.T.	t_c	t_d
Row	A	+04.75%	+04.13%	+10.35%
	S	+05.97%	+05.78%	07.68%
	U	+04.48%	+03.52%	+12.67%
	D	-06.68%	-06.68%	00.00%
Col	A	-10.93%	-11.62%	-05.28%
	S	+02.15%	+02.25%	+01.26%
	U	-23.53%	-25.41%	-09.59%
	D	+37.77%	+37.77%	00.00%

Table 2: Simulation Results of using Learning/Exploration Strategy Plus Threat Power

		H		
	Type	all C.T.	t_c	t_d
Row	A	+02.66%	+01.92%	+08.93%
	S	+05.02%	+04.42%	+09.92%
	U	+00.49%	-00.40%	+08.02%
	D	+04.31%	+04.31%	00.00%
Col	A	+03.12%	+03.17%	+02.70%
	S	+04.49%	+04.32%	+06.01%
	U	+02.15%	+02.38%	+00.38%
	D	+00.30%	+00.30%	00.00%

Table 3: Simulation Results of Using Learning/Exploration Strategy without Threat Power

which occurred in the threat phase as opposed to the learning phase for games which possess the category of threats represented by the intersection of the cell's row and column header. For example, a number at the intersection of S and t_c in the bottom half represents the percent utility increase incurred by C for all superfluous, compellent threats.

6.5 Analysis of Results

In this section we provide an interpretation of the results in terms the threat classification given in Section 5.

Superfluous Threats As predicted, superfluous threats showed no added improvements for agents using TP as opposed to EL. This can be seen by comparing the results of all superfluous threats for R in Table 2 against Table 3. We see that while threat power *did* help R incur an increase in utility, the same utility increase (and more) occurred when R used a learning strategy without threat power.

Detrimental Threats As predicted, detrimental threats resulted in utility decrease when utilized by R. A surprising result was the greatest single increase in utility in the simulation was for C when R tried to use detrimental threats against it. Examination of Table 3 implies this effect is a direct result of TP because in games without TP, R experiences no utility decrease nor does C experience an increase.

Useful Threats As expected, useful threats performed well during threat phases. When using H, threateners experienced on average a 20% increase in utility, more than twofold the utility increase experienced in simulations where threat TP was not involved but learning was.

One marked difference between using EL versus TP is in the impact on an agent's opponent. While EL increases the utility of *both* agents, TP has a negative impact on the threatenee, resulting in a 25% utility decrease for certain games. This drastically affects the social welfare. However for superfluous threats we don't see a marked increase in performance, nor do we see a decrease in performance for our threatenee. As predicted, this trend is preserved for games in which threat power was not used.

7. CONCLUSIONS

Theory of Moves is a novel approach to analyzing games of a dynamic and asymmetric nature which is limited by a restrictive reliance on complete information. We present a mechanism which extends the application of TOM and threat power to incomplete information games by equipping agents with the ability to identify credible threats from experience. We analyzed the effectiveness of threats in terms of individual and social welfare and prescribed three classifications: superfluous, detrimental, and useful. Empirical results on simulations of agents with and without threat power reinforced this classification. We explored the impact initial strategy selection has on agent utility and showed that often an exploratory learning strategy produces the same effect that threat power achieves.

As far as we are aware, the application of TOM outside the domain of 2×2 strictly ordered games is largely unchartered territory. Though the inference mechanism detailed in Section 4 is tailored to 2×2 games, we believe a similar mechanism can be developed for games with more states, and hope to develop results which generalize to arbitrary N $\times$ M games, though the complexity of such a mechanism could be NP complete. Furthermore, we plan to develop additional rules which will classify the 2×2 games entirely, thus eliminating any need for solving a CSP. Cooperation strategies in partial order games are also of particular interest. We would like to perform a thorough analysis on moving and order power and determine if adaptations of the mechanism provided herein equips agents with incomplete knowledge sets the ability to utilize these powers in addition to threat power.

8. REFERENCES

[1] E. Alonso, M. d'Inverno, D. Kudenko, M. Luck, and J.Noble. Learning in multi-agent systems. *Knowledge Engineering Review*, 16(3), 2001.

[2] S. J. Brams. *Theory of Moves*. Cambridge University Press, 1994.

[3] S. J. Brams. *Theory of Moves*. Cambridge University Press, Cambridge: UK, 1994.

[4] M. Littman and P. Stone. A Polynomial-time Nash Equilibrium algorithm for repeated games. *Decision Support Systems*, 39:55–66, 2005.

[5] R. B. Myerson. *Game Theory: Analysis of Conflict*. Harvard University Press, 1991.

[6] J. F. Nash. Non-cooperative games. *Annals of Mathematics*, 54:286 – 295, 1951.

[7] L. Panait and S. Luke. Cooperative multi-agent learning: The state of the art. *Autonomous Agents and Multi-Agent Systems*, 11(3):387–434, 2005.

[8] R. Powers, Y. Shoham, and T. Grenager. A general criterion and an algorithmic framework for learning in multi-agent systems. *Machine Learning*, 67:45–76, 2007.

[9] S. Sen and A. Ghosh. Theory of moves learners: Towards non-myopic equilibrium. *AAMAS*, 2005.

[10] Y. Shoham and K. Leyton-Brown. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, 2008.

[11] Y. Shoham, R. Powers, and T. Grenager. If multi-agent learning is the answer, what is the question? *Artificial Intelligence*, 171(7):365–377, 2007. Special issue on Foundations of Multi-Agent Learning.

[12] P. Stone and M. Veloso. Collaborative and adversarial learning: A case study in robotic soccer. In S. Sen, editor, *Working Notes for the AAAI Symposium on Adaptation, Co-evolution and Learning in Multiagent Systems*, pages 88–92, Stanford University, CA, Mar. 1996.