On the Rationality of Cycling in the Theory of Moves Framework

(Extended Abstract)

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

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

Categories and Subject Descriptors

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

Keywords

Theory of Moves, Cyclic behavior

1. INTRODUCTION

Steven J. Brams's 1994 book Theory of Moves provides a complete information game framework where play commences at a particular state and subsequent moves are determined from a finite lookahead in conjunction with backward induction analysis, resulting in convergence to *Non-Myopic Equilibrium*(NME) [1]. For certain 2×2 games, the initial state is the NME. This can happen by players realizing they have no benefit to move (for example, if the initial state is mutually preferred), or realizing that making an initial move results in a cycle back to the initial state. Though agents in the standard TOM framework do not move if doing so will result in a cycle, average payoff over a cycle may be of higher utility for agents than being stuck in any particular state. Consider a price war between businesses over a commonly offered good. While the businesses might cycle around eventually offering the same initial price, during each price level, products continue to be sold, and the overall profit for a business is a function of the time spent at each price point. While certain states might prove lucrative for certain agents (a price hike), others might be disadvantageous. Is it rational to engage in the cycle expecting that gains in desired states will offset the loss in others?

If we modify rationality rules of TOM and incorporate strategizing for time spent in each state and indefinite game play, equilibria solutions must be analyzed to determine which solutions, if any, are stable in the long run. We study different time constraints on moves in a cycle for a player and show that only certain time choices can be rational. We further construct a meta-matrix with those limited time options to derive equilibria in terms of time spent at each state. Our analysis produces a complete specification of when to cycle and how much time to spend at each state where an agent can choose to move in TOM play.

Appears in: *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2013)*, Ito, Jonker, Gini, and Shehory (eds.), May, 6–10, 2013, Saint Paul, Minnesota, USA.

2. MODIFIED TOM FRAMEWORK

In TOM, play starts at an outcome determined by an initial strategy profile chosen by the players. Either player can unilaterally switch its strategy, thereby changing the initial state into a new state. Players take turns moving until one player declines and the game terminates in the corresponding state. We now present an augmentation of basic TOM play to account for time and dynamic utility where t is the current time step, t_r^k is player p_k's time remaining for the current cycle (initially set to T), U_t^k is p_k's utility at t, U_f^k is p_k's final utility, and t_c is time spent in the current state. t_c and t are initialized to 0, t_r^k to b (the maximum time allowance per iteration), and U_t^k and U_f^k are initialized to $v^k(S_0)$ (p_k's valuation for state S_0).

1. **Initial Move** If player p_k makes an initial move after t_0, then $t_r^k \leftarrow t_r^k - t_0$, $t \leftarrow t + t_0$. If neither player makes an initial move by b, game play terminates.

2. **Subsequent Moves** Given current state S_i, current player p_k, and current values t_c and t:

   - $t_c \geq t_r^k$: Play terminates and $\forall$ players p_m:
   $$U_f^m \leftarrow \left[\frac{t}{(t+1)}\right] U_t^m + \left[\frac{1}{(t+1)}\right] v^m(S_i);$$
   - $t_c < t_r^k$ and p_k decides to remain in S_i: $t_c \leftarrow t_c+1$, $t \leftarrow t+1$, and $\forall$ players p_m
   $$U_t^m \leftarrow \left[\frac{(t-1)}{t}\right] U_{t-1}^m + \left[\frac{1}{t}\right] v^m(S_i);$$
   - $t_c < t_r^k$ and p_k moves to state $S_j \neq S_0$: then $t_r^k \leftarrow t_r^k - t_c$, $t \leftarrow t+1$, $t_c \leftarrow 0$, and $\forall$ player p_m:
   $$U_t^m \leftarrow \left[\frac{(t-1)}{t}\right] U_{t-1}^m + \left[\frac{1}{t}\right] v^m(S_i);$$
   - $t_c < t_r^k$ and p_k moves to state $S_j = S_0$: a cycle occurs. $t \leftarrow t+1$, $\forall$ players p_m, $t_r^m = b$ and
   $$U_t^m \leftarrow \left[\frac{(t-1)}{t}\right] U_{t-1}^m + \left[\frac{1}{t}\right] v^m(S_i).$$

Notice if $b = \infty$, the algorithm provides no clear play termination and hence no calculation of final utility. In such games, $U_f^k = \lim_{t \to +\infty} (U_t^k)$. Resultantly, the revised paradigm supports characterization of noncyclic games.

3. OPTIMAL STRATEGY CONSTRUCTION

If time constraints are given on game play, an agent's rational strategy must not only take into account moving or staying, but how long it should stay before moving. For the i^{th} round of game play, a player p_k[1] can construct a time strategy $s_i = (t_1, t_2, ..., t_n)$ where t_j indicates how long it will stay in it's j^{th} decision state. It would like to find the strategy S_i optimizing its utility function $u^k(S_i)$.

3.1 Games with a min/max move time limits

We limit our discussion to 2×2, strictly ordinal games with payoffs (a,w), (b,x), (c,y), (d,z); p_k has two decision states per round and seeks a time strategy $s_i = (t_1^k, t_2^k)$ given a minimum time limit ϵ each decision state and maximum time limit M per cycle. WLOG, suppose p^k prefers its first decision state to its first.

Theorem 1. *If there is a minimum time limit ϵ each decision state, a player should never spend more than ϵ time in its least preferred decision state.*

We now wish to find the optimal value for it's least preferred decision state. Because $u^k(\mathcal{S})$ is continuous on the closed bounded interval $[\epsilon, M - \epsilon]$, we are guaranteed it will attain a maximum at some point on this interval and that this maximum will occur either on the endpoints of the interval or at a critical point corresponding to:

$$\frac{\partial u^k}{\partial \delta} = \frac{t_1^{\bar{k}}(b-a) + t_2^{\bar{k}}(d-a) + \epsilon(c-a)}{M - \delta + \epsilon + t_1^{\bar{k}} + t_2^{\bar{k}}} = 0$$

Three cases arise:

1. $t_1^{\bar{k}}(b-a) + t_2^{\bar{k}}(d-a) + \epsilon(c-a) > 0$
   $\Rightarrow$ increasing δ increases $u^k \Rightarrow (\epsilon, \epsilon)$ optimal
2. $t_1^{\bar{k}}(b-a) + t_2^{\bar{k}}(d-a) + \epsilon(c-a) < 0$
   $\Rightarrow$ increasing δ decreases $u^k \Rightarrow (M - \epsilon, \epsilon)$ optimal
3. $t_1^{\bar{k}}(b-a) + t_2^{\bar{k}}(d-a) + \epsilon(c-a) = 0$;
   δ has no impact and p_k is indifferent to cycling.

Meta Matrix Construction.

Because p_k's strategy about a cycle is dependent upon $p_{\bar{k}}$'s and since players may deviate from strategies each decision state, p_k must also predict the long term consequences of cycling on the dynamics of $p_{\bar{k}}$'s strategy selections before determining if cycling is rational.

Nash Equilibria (NE) solutions occur when no player has an incentive to deviate from their contributing strategy, and the previous analysis indicates the only viable candidates are strategies of the form $(M - \delta, \epsilon)$ with δ on either end the interval $[\epsilon, M - \epsilon]$. We can construct a "meta matrix" like the one pictured where action profiles for each player consist of these strategies and each square corresponds to the solution composed of the intersecting player strategies.[2] Payoffs are derived from the utility functions and describe the net difference in utility incurred for a cycle using the solution associated with that square. NE in a meta matrix indicates a stable strategy set which in turn corresponds to a stable game solution. When a meta matrix lacks NE, players will engage in a constant cascade of deviation to different extremes of δ, cycling about the squares of the meta matrix itself. An agent can use the following procedure then to determine if cycling is rational:
(1) Construct meta matrix and assign payoffs using Equations 1 and 2.
(2) If all payoffs in the meta matrix are positive for a player, it is rational for it to induce the cycle.
(3) If a NE exists and the payoffs for a player is positive, it is rational for it to induce the cycle. Conversely, if payoff is negative, cycling is irrational.
(4) If no NE exist, but the average payoff about the meta matrix for a player is positive (negative), it should (should not) cycle.

Meta Matrix

	ϵ	$M - \epsilon$
ϵ	Δ_1^k $\Delta_1^{\bar{k}}$	Δ_4^k $\Delta_4^{\bar{k}}$
$M - \epsilon$	Δ_2^k $\Delta_2^{\bar{k}}$	Δ_3^k $\Delta_3^{\bar{k}}$

$$\Delta u^k(\delta_1, \delta_2) = \begin{cases} \dfrac{\epsilon(b+c-2a) + (M-\delta_2)(d-a)}{2M + 2\epsilon - \delta_1 - \delta_2} & |z > x \\ \dfrac{\epsilon(c+d-2a) + (M-\delta_2)(b-a)}{2M + 2\epsilon - \delta_1 - \delta_2} & |x > z \end{cases} \tag{1}$$

$$\Delta u^{\bar{k}}(\delta_1, \delta_2) = \begin{cases} \dfrac{\epsilon(x+y-2w) + (M-\delta_2)(z-w)}{2M + 2\epsilon - \delta_1 - \delta_2} & |z > x \\ \dfrac{\epsilon(z+y-2w) + (M-\delta_2)(x-w)}{2M + 2\epsilon - \delta_1 - \delta_2} & |x > z \end{cases} \tag{2}$$

3.2 Games with min time limit but no max

Our analysis showed that while a player's most rational strategy minimizes its time in the least preferred state, it does not always want to maximize the time it spends in its most preferred moving state. As long as decision states have minimum time limits, are there games that have NE solution even when lacking a maximum time constraint per cycle? Due to limited space, we only present the following theorem, which is not restricted to 2×2 games, and omit the analysis.

Theorem 2. *Consider an n-cycle $(S_1, S_2, ..., S_n)$ with m players. Let $\mathcal{S}_\epsilon = (\epsilon, \epsilon, ..., \epsilon)$. If for each player p_i, $v^i(S_0) \leq U^i(\mathcal{S}_\epsilon) \Rightarrow \mathcal{S}_\epsilon$ is NE.*

In conclusion, for games with min but no max time limits, players need only check if each player's valuation of the cyclic state is less than their average valuation about the cycle. If it is, then not only is the minimum solution NE, but it is mutually beneficial for all players and the player should cycle. If this test fails, then there is no NE for the game.

4. REFERENCES

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

[1] p_k's opponent is henceforth denoted as $p_{\bar{k}}$

[2] Example: For a game where R, C prefer their first decision state, the top left outcome corresponds to both players selecting $\delta = \epsilon$, resulting in overall time strategy $(M - \epsilon, M - \epsilon, \epsilon, \epsilon)$.