On the Rationality of Cycling in the Theory of Moves Framework

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

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

ABSTRACT

Theory of Moves (TOM) is a novel approach to game theory for determining rational strategies during the play of dynamic games [2]. While alternate models such as normal form games exist, players of these games are limited to single shot interactions with each other, but within TOM, sequences of moves and counter moves are allowed. As a consequence of this framework potential cyclic behavior may arise. Unfortunately, standard TOM framework suggests that agents do not move from the initial state if the possibility of cyclic behavior is detected. However, in a plethora of real life scenarios, cycling can benefit a player over time. We first extend the TOM framework by allowing agents to choose how much time to stay in each state while specifying time limits for moves. This generalization allows for cycling behavior in addition to normal, acyclic TOM play. We present additional rationality rules to handle the choice of move time and cyclic play and identify conditions for the existence of solutions that involve cycles. Moreover, if solutions do exist, equilibria are determined so a player can predict the rational outcome upon engaging a cycle. A variety of time constraints on move times are investigated and the effects of these contrasts on the solution space and equilibria are analyzed.

Categories and Subject Descriptors

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

General Terms

Theory, Economics, Algorithms

Keywords

Theory of Moves, Cyclic behavior

1. INTRODUCTION

Reasoning and learning mechanisms in single stage games continue to be an active research area in multi agent systems [1, 3, 5, 6, 8, 9, 7]. To handle situations in which agents engage in a natural move-counter move process or take decisions as a function of some initial state extensive form game frameworks have been studied with solution concepts including subgame perfect equilibrium [4]. 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 [7]. While this model does address scenarios where agents can alternate moves, it still does not adequately address scenarios 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.

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) [2]. 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.

The rest of this paper is organized as follows: Section 2 proposes revised rationality rules for TOM to account for dynamic utility, time constraints, and cycle constraints; Section 3 shows how non-cyclic games are still supported under the new framework, Section 4 explores games of a cyclic nature in which maximum time limits are specified while Section 5 explores cyclic games with no maximum time limit on play. Finally, Section 6 provides a brief conclusion and insight into future research. We note that our analysis is limited to games for which minimum time limits are *always*

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

Figure 1: Game 1

$$\begin{array}{llllll} R & C & & R & & C \\ (r_1, c_1) & (r_2, c_2) & \ldots & (r_b, c_b) & \ldots & (r_{n-1}, c_{n-1}) \mid (r_1, c_1) \\ [r_b, c_b] & [r_b, c_b] & \ldots & [r_b, c_b] & \ldots & [r_1, c_1] \end{array}$$

Figure 2: Non-cyclic Game

specified on play. This is because if no minimum time limits were specified, trivially players would continually deviate from any given solution by reducing time spent in decision states that were not its most preferred.

2. GAME DESCRIPTION

The TOM framework has the capacity to model move-counter-move processes between players, but disallows cycles, focusing on payoff received in terminal states. Consequentially how long players might spend on any given decision state is disregarded. In games of dynamically distributed utility, time restrictions may have a significant impact on optimal strategy selections, especially where cycling is concerned. Augmenting the TOM framework to allow for cyclic play necessitates an overhaul of game execution and such a revision is presented below.

2.1 Game Constraints

We can specify time constraints on a game G by use of an interval $I = [\epsilon, M]$, where M is the maximum time each cycle any player may spend, combined, in its decision states; and ϵ is the minimum time each player must spend in each decision state. For games with no time maximum time limit, we need only let $M \to \infty$.

2.2 Notational Conventions

For the c^{th} round of an m-player, n-state cycle in a game G, each player p_k has d_k decision states and associated strategy $s_k^c = (0, 0, ..., t_1^k, ..., t_2^k, ..., t_{d_k}^k, ..., 0)$, where t_i^k specifies how long p_k spends in its i^{th} decision state in the cycle. A *game solution* $\mathcal{S}^c$ is an n-tuple composed of all player strategies specifying entirely the time each player in the game will spend in each state of the cycle.

$$\mathcal{S}^c = \sum_{\forall p} (s_k^c) = (t_1, t_2, ..., t_n) \tag{1}$$

$\mathcal{T}$ is the sum of all time values in $\mathcal{S}^c$ (the total time spent in the cycle if $\mathcal{S}^c$ is used). p_k has valuation function $v^k(S_i)$ describing valuation for each state S_i in G, and utility function, $u^k(\mathcal{S}^c)$ describing its utility for round c if $\mathcal{S}^c$ is used.

$$u^k(\mathcal{S}^c) = \sum_{i=1}^{n} (\frac{t_i}{\mathcal{T}})(v^k(S_i) \tag{2}$$

Example of Notation:.

Consider Game 1 diagramed in Figure 1 for subgame originating at (3,4) with initial player R and I = (0,10]. R and C's decision states are (3,4) (1,1), and (2,2) (4,2); respectively. Suppose R and C, for the c^{th} round, have strategies $s_R^c = (2, 0, 1, 0)$ (indicating R will spend 2 seconds in (3,4) and 1 second in (1,1)), and $s_C^c = (0, 4, 0, 8)$ (indicates C will spend 4 seconds in (2,2) and 8 in (4,2)). Then the associated solution for that round is $\mathcal{S}^c = (2, 4, 1, 8)$, with $\mathcal{T} = 15$, and the utilities for each player are:

$$u^R(\mathcal{S}^c) = (\frac{2}{15})3 + (\frac{4}{15})2 + (\frac{1}{15})1 + (\frac{8}{15})4 = \frac{47}{15} \approx 3.13$$
$$u^C(\mathcal{S}^c) = (\frac{2}{15})4 + (\frac{4}{15})3 + (\frac{1}{15})1 + (\frac{8}{15})2 = \frac{37}{15} \approx 2.46$$

2.3 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 M (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 $M = \infty$, the algorithm provides no clear play termination and hence no calculation of final utility. In such games, for each player p_k

$$U_f^k = \lim_{t \to +\infty} (U_t^k) \tag{3}$$

a, w	d, z
b, x	c, y

Figure 3: Arbitrary Game

R	C	R	C	
(a,w)	(b,x)	(c,y)	(d,z)	| (a,w)
[a,w]	[a,w]	[a,w]	[a,w]	

Figure 4: Diagram of Backwards Induction for Arbitrary, Cyclic, 2 × 2 Subgame

3. NON-CYCLIC GAMES

Our rules are not limited to describing games of a cyclic nature. The revised paradigm can characterize a non-cyclic game such as the one in Figure 2[1] in the following way: define time constraints so that $M = \infty$, and consider a solution $\mathcal{S}^1 = (t_0, t_1, ..., t_{b-1}, \infty, -, ..., -)$, meaning R makes an initial move after t_0, C moves after t_1, etc. Then:

$U^R_{t_0} = r_1$

$$U^R_{(t_0+t_1)} = \left[\frac{t_0}{t_0+t_1}\right] U^R_{(t_0+t_1-1)} + \left[\frac{t_1}{t_0+t_1}\right](c_2) = \frac{r_1 t_0 + c_2 t_1}{t_0+t_1}$$

Let $t^* = \sum_{i=0}^{b-1}(t_i)$, then the utility for R at the b^{th} state is:

$$U^R_{t^*+1} = \frac{r_0 t_0 + c_1 t_1 + ... + c_{b-1} t_{b-1} + (1) r_b}{t^*+1}$$

R blocks at (r_b, c_b) and will chose not to move in any future time steps, so game play is "caught" in rule 2(b), and for all $t > b$, R's utility will be:

$$U^R_t = \left[\frac{t-1}{t}\right] U^R_{(t-1)} + \left[\frac{1}{t}\right](r_b) = \frac{r_0 t_0 + c_1 t_1 + ... + r_b(t - t^*)}{t}$$

By equation 3, $U^R_f = \left[\frac{t_0(r_0 - r_b) + t_1(c_1 - r_b) + ... + r_b t}{t}\right]$

$= \lim_{t \to +\infty} \frac{r_b}{t} = r_b$

Similar analysis proves the same for opponents: player's utilities converges to (r_b, c_b), in accordance with TOM.

4. CYCLIC GAMES UNDER MAX TIME RESTRAINTS

Both traditional TOM and the revised paradigm result in identical equilibria solutions for non-cyclic games, but our real goal is to study cyclic games so we can analyze when it is rational for players to engage cycles. We first turn our attention to games with a maximum time limit specified about each cycle. We begin by determining the optimal strategy for a player p_k if we assume it will cycle and then propose construction of a meta matrix to predict global long term game solutions that will result as a result of using these optimal strategies, or emergent unstable behavior to determine if cycling itself is rational.

We limit our discussion to 2×2, strictly ordinal games. p_k[2] has two decision states per round and seeks an optimal time strategy $s_i = (t^k_1, t^k_2)$ given time constraints ϵ, M. WLOG, suppose p^k prefers its first decision state.

[1] G is non-cyclic under TOM because R prefers (r_b, c_b) and will blocks here.

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

4.1 Rational Time Spent in Least Preferred Decision State

Because p_k only has two decision states each cycle, it would be useful if it had a clear decision on how long to stay in one of these states which was independent of the other. The following theorem provides such a result.

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.*

PROOF. For an arbitrary cycle of game G with initial player p_k, let p_k and $p_{\bar{k}}$ have feasible strategies $s^k = (\delta_1, \delta_2)$ and $s^{\bar{k}} = (t^{\bar{k}}_1, t^{\bar{k}}_2)$, respectively. The solution composed of these strategies, $\mathcal{S} = (\delta_1, t^{\bar{k}}_1, \delta_2, t^{\bar{k}}_2)$, with $\mathcal{T} = \delta_1 + t^{\bar{k}}_1 + \delta_2 + t^{\bar{k}}_2$, generates the following utility for p_k:

$$u^{p_k}(\mathcal{S}) = \frac{\delta_1}{\mathcal{T}} a + \frac{t^{\bar{k}}_1}{\mathcal{T}} b + \frac{\delta_2}{\mathcal{T}} c + \frac{t^{\bar{k}}_2}{\mathcal{T}} d$$

s^k feasible $\Rightarrow M - \epsilon \geq \delta_1, \delta_2 \geq \epsilon$, and G is cyclic so p_k prefers the second state the least out of the two. Assume $\delta_2 > \epsilon$. Now let $\delta_3 = \delta_2 - \epsilon$ and consider an alternate strategy for p_k as $s^{k*} = (\delta_1 + \delta_3, \delta_2 - \delta_3)$. Note that $\delta_1 + \delta_3 > \delta_1 \geq \epsilon$ and $\delta_2 - \delta_3 = \epsilon$ so s^{k*} is feasible. The solution composed of s^{k*} and $s^{\bar{k}}$, $\mathcal{S}^* = (\delta_1 + \delta_3, t^O_1, \delta_2 - \delta_3, t^O_2)$ with $\mathcal{T}^* = (\delta_1 + \delta_3) + t^{\bar{k}}_1 + (\delta_1 - \delta_3) + t^{\bar{k}}_2 = \delta_1 + t^{\bar{k}}_1 + \delta_2 + t^{\bar{k}}_2 = \mathcal{T}$ generates the following utility for p_k:

$u^k(\mathcal{S}^*) = (\frac{\delta_1 + \delta_3}{\mathcal{T}^*})a + (\frac{t^O_1}{\mathcal{T}^*})b + (\frac{\delta_2 - \delta 3}{\mathcal{T}^*})c + (\frac{t^O_2}{\mathcal{T}^*})d$

This game is cyclic $\Rightarrow a > c \Rightarrow ta - tc > 0$, so $\forall t > 0$

$$\begin{aligned} u^k(\mathcal{S}^*) &= (\frac{\delta_1 + \delta_3}{\mathcal{T}^*})a + (\frac{t^O_1}{\mathcal{T}^*})b + (\frac{\delta_2 - \delta 3}{\mathcal{T}^*})c + (\frac{t^O_2}{\mathcal{T}^*})d \\ &= (\frac{\delta_1 + \delta_3}{\mathcal{T}})a + (\frac{t^O_1}{\mathcal{T}})b + (\frac{\delta_2 - \delta 3}{\mathcal{T}})c + (\frac{t^O_2}{\mathcal{T}})d \\ &= (\frac{\delta_1}{\mathcal{T}})a + (\frac{t^O_1}{\mathcal{T}})b + (\frac{\delta_2}{\mathcal{T}})c + (\frac{t^O_2}{\mathcal{T}})d + \left[(\frac{\delta_3}{\mathcal{T}})a - (\frac{\delta_3}{\mathcal{T}})c\right] \\ &> (\frac{\delta_1}{\mathcal{T}})a + (\frac{t^O_1}{\mathcal{T}})b + (\frac{\delta_2}{\mathcal{T}})c + (\frac{t^O_2}{\mathcal{T}})d = u^k(\mathcal{S}) \end{aligned}$$

$\therefore$ if $\delta_2 > \epsilon$, R always has incentive to deviate from $\mathcal{S}$ by switching strategies from s^R to s^{R*}. Since the choice of δ_2 was arbitrary, and because the result did not depend on the order of game play, this means that for any player, it is not rational to spend more than ϵ time in the least preferred state. □

4.2 Rational Time Spent in Most Preferred Decision State

Because of the previous result, we might be naturally inclined to assume that p_k should spend all remaining time in its most preferred decision state. Coupled with Theorem 1 this would indicate that p_k's optimal strategy would always be $s = (M - \epsilon, \epsilon)$. We provide a counter example showing this is not true, but state this as a theorem so as to discuss the intuition behind why this is not generally the case.

THEOREM 2. *It is not always rational to spend the maximum time in the most preferred decision state*

PROOF. Compare two arbitrary strategies for p_k: one where the most preferred state is maximized and one where it is not: $s^k = (M - \epsilon, \epsilon)$ and (2) $s^{k*} = (M - \epsilon - \delta, \epsilon)$ where

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

Figure 5: Example 1

$M-\epsilon > M-\epsilon-\delta > \epsilon$, respectively. Assume $p_{\bar{k}}$ has feasible strategy $s^{\bar{k}} = (t_1^{\bar{k}}, t_2^{\bar{k}})$. The associaited solutions are $\mathcal{S} = (M-\epsilon, t_1^{\bar{k}}, \epsilon, t_2^{\bar{k}})$ and $\mathcal{S}^* = (M-\epsilon-\delta, t_1^{\bar{k}}, \epsilon, t_2^{\bar{k}})$, with $\mathcal{T} = M-\epsilon+t_1^{\bar{k}}+\epsilon+t_2^{\bar{k}} = M+t_1^{\bar{k}}+t_2^{\bar{k}}$, and the associated utilities for p_k are given by:

$$u^k(\mathcal{S}) = (\frac{M-\epsilon}{\mathcal{T}})a + (\frac{t_1^{\bar{k}}}{\mathcal{T}})b + (\frac{\epsilon}{\mathcal{T}})c + (\frac{t_2^{\bar{k}}}{\mathcal{T}})d$$

$$u^k(\mathcal{S}^*) = (\frac{M-\epsilon-\delta}{\mathcal{T}-\delta})a + (\frac{t_1^{\bar{k}}}{\mathcal{T}-\delta})b + (\frac{\epsilon}{\mathcal{T}-\delta})c + (\frac{t_2^{\bar{k}}}{\mathcal{T}-\delta})d$$

Consider the following sequence of equivalent statements:

$$T > (M-\epsilon)$$

$$\delta T > \delta(M-\epsilon)(\text{because } \delta > 0)$$

$$T(M-\epsilon) - \delta T < T(M-\epsilon) - \delta(M-\epsilon)$$

$$\frac{(M-\epsilon-\delta)}{(T-\delta)} < \frac{(M-\epsilon)}{T} \text{ (because } T>0,\ T-\delta>0)$$

$$\frac{(M-\epsilon-\delta)}{(T-\delta)}a < \frac{(M-\epsilon)}{T}a \text{ (assuming payoffs are positive)}$$

$$\frac{(M-\epsilon-\delta)}{(T-\delta)}a = \frac{(M-\epsilon)}{T}a - k_1 \text{ for some } k_1 > 0$$

$$(\frac{t_1^{\bar{k}}}{T-\delta})b > (\frac{t_1^{\bar{k}}}{T})b$$

$$(\frac{t_1^{\bar{k}}}{T-\delta})b = (\frac{t_1^{\bar{k}}}{T})b + k_2 \text{ for some } k_2 > 0$$

$$(\frac{\epsilon}{T-\delta})b = (\frac{\epsilon}{T})b + k_3 \text{ for some } k_3 > 0$$

$$(\frac{t_2^{\bar{k}}}{T-\delta})b = (\frac{t_2^{\bar{k}}}{T})b + k_4 \text{ for some } k_4 > 0$$

Then $u^k(\mathcal{S}^*) = u^k(\mathcal{S}) + k_2 + k_3 + k_4 - k_1$

$$= u^k(\mathcal{S}) + \left[\frac{t_1^{\bar{k}}\delta}{\mathcal{T}(\mathcal{T}-\delta)}\right]b + \left[\frac{\epsilon\delta}{\mathcal{T}(\mathcal{T}-\delta)}\right]c + \left[\frac{t_2^{\bar{k}}\delta}{\mathcal{T}(\mathcal{T}-\delta)}\right]d - \left[\frac{(\mathcal{T}-(M-\epsilon))\delta}{\mathcal{T}(\mathcal{T}-\delta)}\right]a$$

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

$$= u^k(\mathcal{S}) + \frac{\delta}{\mathcal{T}(\mathcal{T}-\delta)}\left[t_1^{\bar{k}}\delta b + \epsilon\delta c + t_2^{\bar{k}}\delta d - (\mathcal{T}-(M-\epsilon))\delta a\right]$$

$$= u^k(\mathcal{S}) + \frac{\delta}{\mathcal{T}(\mathcal{T}-\delta)}\left[t_1^{\bar{k}}(b-a) + t_2^{\bar{k}}(d-a) + \epsilon(c-a)\right]$$

$$\frac{\delta}{\mathcal{T}(\mathcal{T}-\delta)} > 0 \Rightarrow u^k(\mathcal{S}^*) > u^k(\mathcal{S}) \iff$$

$$t_1^{\bar{k}}(b-a) + t_2^{\bar{k}}(d-a) + \epsilon(c-a) > 0 \tag{4}$$

If eqn 4 is true, then p_k's rational choice is to select the strategy associated with $\mathcal{S}^*$, $s^{k*} = (M-\epsilon-\delta, \epsilon)$. $\square$

The proof hinges on the truth of eqn 4, but is such a situation feasible? Consider game in Figure 5 for the subgame with initial player R and initial state (2,3). Assume C has strategy $t^C = (\delta, \epsilon)$. Then $t_1^C(b-a) + t_2^C(d-a) + \epsilon(c-a) = 2\delta + \epsilon - \epsilon > 0$, and according to eqn 4 it is *not* rational for R to select the strategy that maximizes its time in the initial state, $s^R = (M-\epsilon, \epsilon)$, because there exists some δ, where $M-\epsilon > M-\delta > \epsilon$, such that the alternate strategy $s^{R*} = (M-\delta, \epsilon)$ will award a higher utility. Following the current example, if $\epsilon = 1$ and $M = 10$ and C's strategy is (ϵ, ϵ), then consider two solutions: $\mathcal{S} = (M-\epsilon, t_1^C, \epsilon, t_2^C) = (9,1,1,1)$ and $\mathcal{S}^* = (M-\epsilon-\delta, t_1^C, \epsilon, t_2^C) = (9-\delta,1,1,1)$ where $\delta = 1$. Then the utilities for R are $u^R(\mathcal{S}^*) = 2.16$ and $u^R(\mathcal{S}^*) = 2.18$. Clearly R should *not* maximize the time it spends in the initial state. This counter example provides evidence that it is *not* always optimal for a player to spend the maximum remaining time in their most preferred state.

4.2.1 Selecting δ

We showed p_k's most rational decision is to spend no more than ϵ time in its least preferred decision state but the choice of how much time to spent in the most preferred decision state is not always so trivial: there are games where there exists δ with $M-\epsilon > M-\delta > \epsilon$, such that spending $M-\delta$ time in the most preferred state will result in a higher utility as opposed to spending $M-\epsilon$ time there. If $p_{\bar{k}}$'s strategy is $(t_1^{\bar{k}}, t_2^{\bar{k}})$, p_k seeks to find δ such that the associated solution $\mathcal{S} = (M-\delta, t_1^{\bar{k}}, \epsilon, t_2^{\bar{k}})$ maximizes $u^k(\mathcal{S})$.

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 where $\frac{\partial u^k}{\partial \delta} = 0$

$$\frac{\partial u^k}{\partial \delta} = \frac{t_1^{\bar{k}}(b-a) + t_2^{\bar{k}}(d-a) + \epsilon(c-a)}{M-\delta+\epsilon+t_1^{\bar{k}}+t_2^{\bar{k}}} \tag{5}$$

$M-\delta+\epsilon+t_1^{\bar{k}}+t_2^{\bar{k}} > 0$ so a critical point occurs only when:

$$t_1^{\bar{k}}(b-a) + t_2^{\bar{k}}(d-a) + \epsilon(c-a) = 0 \tag{6}$$

We digress to notice if eqn 6 is true, δ has no impact, which is only possible if u^k is constant. We're inclined to discover this constant to determine if cycling is rational in this case. Solving for $t_1^{\bar{k}}$ and plugging in to the utility function reveals:

$$u^k(\mathcal{S}) = \frac{(M-\delta)}{\mathcal{T}}a + \frac{t_1}{\mathcal{T}}b + \frac{\epsilon}{\mathcal{T}}c + \frac{t_3}{\mathcal{T}}d$$

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

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

$$= \frac{(M-\delta)a + \left[\frac{-t_2^{\bar{k}}(d-a)-\epsilon(c-a)}{(b-a)}\right]b + \epsilon c + t_2^{\bar{k}}d}{M-\delta + \left[\frac{-t_2^{\bar{k}}(d-a)-\epsilon(c-a)}{(b-a)}\right] + \epsilon + t_2^{\bar{k}}}$$

$$= \frac{(b-a)(M-\delta)a + t_2^{\bar{k}}ab + \epsilon ab - \epsilon ac - t_2^{\bar{k}}ad}{(b-a)(M-\delta) - t_2^{\bar{k}}d - \epsilon c + \epsilon b + t_2^{\bar{k}}b}$$

$$= a\frac{\left[(b-a)(M-\delta) + t_2^{\bar{k}}b - t_2^{\bar{k}}d + \epsilon b - \epsilon c\right]}{\left[(b-a)(M-\delta) + t_2^{\bar{k}}b - t_2^{\bar{k}}d + \epsilon b - \epsilon c\right]}$$

$$= a$$

Because p_k prefers its first decision state, this shows that when eqn 6 is true, so long as p_k minimizes time spent in its least preferred decision state, it will fare no better or worse if it decides to cycle.

Continuing our search for optimal δ selection, notice now that if equation 6 is false there are no feasible $\delta > 0$ which give critical points, indicating the max and min values occur on the end points of I at either $\delta = \epsilon$ or $\delta = M - \epsilon$. To decide which end point maximizes u^k, we need only examine the sign of $\frac{\partial u^k}{\partial \delta}$. Three cases arise:

1. $t_1^{\bar{k}}(b-a) + t_2^{\bar{k}}(d-a) + \epsilon(c-a) > 0$
   Increasing δ results in increased u^k, so p_k's optimal strategy is (ϵ, ϵ). Notice this agrees with the analysis of inequality 4 which we used to construct a counter example as to why a player should not always maximize the time it spends in its most preferred state.

2. $t_1^{\bar{k}}(b-a) + t_2^{\bar{k}}(d-a) + \epsilon(c-a) < 0$
   Increasing δ results in decreased u^k, so p_k's optimal strategy is $(M-\epsilon, \epsilon)$

3. $t_1^{\bar{k}}(b-a) + t_2^{\bar{k}}(d-a) + \epsilon(c-a) = 0$
   Selection of δ has no impact on utility so p_k is prefers (ϵ, ϵ) and $(M-\epsilon, \epsilon)$ equally.

4.3 Meta Matrix Construction

In the preceding section, we saw that p_k is indifferent to cycling when case eqn 6 is satisfied. But what about cases 1 and 2? We were able to construct each player's rational strategy if it were to cycle, but how do we know when cycling is the rational decision? 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 predict the long term consequences of cycling on the dynamics of $p_{\bar{k}}$'s strategy selections before determining if cycling is rational.

Pure Nash Equilibria[3] 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.[4] 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 7 and 8.

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

$$\Delta u^k(\delta_1, \delta_2) = \begin{cases} \dfrac{\epsilon(b+c-2a) + (M-\delta_2)(d-a)}{2M + 2\epsilon - \delta_1 - \delta_2} & |z > x \\ \dfrac{\epsilon(c+d-2a) + (M-\delta_2)(b-a)}{2M + 2\epsilon - \delta_1 - \delta_2} & |x > z \end{cases} \tag{7}$$

$$\Delta u^{\bar{k}}(\delta_1, \delta_2) = \begin{cases} \dfrac{\epsilon(x+y-2w) + (M-\delta_2)(z-w)}{2M + 2\epsilon - \delta_1 - \delta_2} & |z > x \\ \dfrac{\epsilon(z+y-2w) + (M-\delta_2)(x-w)}{2M + 2\epsilon - \delta_1 - \delta_2} & |x > z \end{cases} \tag{8}$$

4.3.1 Game 1: NE solution exists and it is rational for R to induce the cycle

	ϵ	$(M-\epsilon)$
ϵ	$(\frac{1}{2}, -\frac{1}{2})$	$(\frac{2\epsilon}{M+2\epsilon}, -\frac{2\epsilon}{M+2\epsilon})$
$(M-\epsilon)$	$(\frac{M}{M+2\epsilon}, -\frac{M}{M+2\epsilon})$	$(\frac{1}{2}, -\frac{1}{2})$

Table 1: Unfilled meta matrix

Consider the game in Figure 5 and associated meta matrix in Table 1. If $2\epsilon = M$, $\frac{M}{M+2\epsilon} = \frac{1}{2} = \frac{2\epsilon}{M+2\epsilon}$ so all four states are NE. But no matter which strategy profile we chose the associated solution will be $(\epsilon, \epsilon, \epsilon, \epsilon)$. M $< 2\epsilon$ is not feasible, as ϵ is minimum time in each state and there are two states. If $M > 2\epsilon$ using table 4.3.1 we find one NE at $(M-\epsilon, M-\epsilon)$ with corresponding solution $(\epsilon, \epsilon, \epsilon, \epsilon)$. So upon inducing the cycle, players select $\delta = M - \epsilon$ coordinating to solution $\mathcal{S} = (\epsilon, \epsilon, \epsilon, \epsilon)$ and converging to utility

$$u^R(\mathcal{S}) = (\frac{1}{4})(2) + (\frac{1}{4})(4) + (\frac{1}{4})(1) + (\frac{1}{4})(3) = 2.5$$
$$u^C(\mathcal{S}) = (\frac{1}{4})(3) + (\frac{1}{4})(1) + (\frac{1}{4})(4) + (\frac{1}{4})(2) = 2.5$$

Clearly cycling with $\mathcal{S}$ is beneficial for R and therefore rational[5]. While detrimental for C, $\mathcal{S}$ minimizes the loss.

4.3.2 Game 2: NE Solutions exist but it is not rational for R to induce the cycle

Consider the game in Figure 6 with associated meta matrix shown in Table 3. If $M > 4\epsilon$, cycling under $(M -$

[3]For the remainder of the paper, NE refer to *pure Nash Equilibria*

[4]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 $(M-\epsilon, M-\epsilon, \epsilon, \epsilon)$.

[5]We also could have simply observed pure positive payoffs in the meta matrix for R to find cycling rational.

$M > 2\epsilon$	$M > 2\epsilon$
$M + (M) > 2\epsilon + (M)$	$M + (2\epsilon) > 2\epsilon + (2\epsilon)$
$2\frac{M}{M+2\epsilon} > 1$	$1 > \frac{4\epsilon}{M+2\epsilon}$
$\frac{M}{M+2\epsilon} > \frac{1}{2}$	$\frac{1}{2} > \frac{2\epsilon}{M+2\epsilon}$
$-\frac{M}{M+2\epsilon} < -\frac{1}{2}$	$-\frac{1}{2} < -\frac{2\epsilon}{M+2\epsilon}$

Table 2: Deduction for Example 1

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

Figure 6: Example 2 | R should not cycle

$\epsilon, \epsilon, \epsilon, M-\epsilon)$ and $(\epsilon, \epsilon, \epsilon, M-\epsilon)$ benefits R. We deduce table 4.3.2 and find C has a pure strategy for $s^C = (\epsilon, \epsilon)$ and the only NE is $\mathcal{S}_{NE} = (M-\epsilon, \epsilon, \epsilon, \epsilon)$ which is detrimental to both players. We test this with M=10 and $\epsilon = 1$. Then $\mathcal{S}_{NE} = (M-\epsilon, \epsilon, \epsilon, \epsilon) = (9,1,1,1)$, $\mathcal{T} = 12$.

$$\mathcal{S}_{NE} = (M-\epsilon, \epsilon, \epsilon, \epsilon) = (9,1,1,1) \Rightarrow \mathcal{T} = 12.$$
$$u^R(\mathcal{S}_{NE}) = (\frac{9}{12})(3) + (\frac{1}{12})(2) + (\frac{1}{12})(1) + (\frac{1}{12})(4) \approx 2.83$$
$$u^C(\mathcal{S}_{NE}) = (\frac{9}{12})(3) + (\frac{1}{12})(1) + (\frac{1}{12})(4) + (\frac{1}{12})(2) \approx 2.83$$

So while there are solutions that would benefit R, they do not occur as C would always have the incentive to deviate, so R should not chose to cycle.

4.3.3 Game 3: No NE but cycling is rational for R

Consider the game in Figure 7 with associated meta matrix in Table 5. Using Tables 4.3.2 and 4.3.1, we find there are no NE, so what behavior will emerge? Because there is no stable solution, during the play of the game players will continually switch their strategies, cycling about the solutions corresponding to the different squares in the meta matrix. For example, say players start with solution $\mathcal{S} = (M-\epsilon, M-\epsilon, \epsilon, \epsilon)$. In the next cycle R will deviate to minimize the time it spends in (2,3), in response C will deviate to minimize time in (3,1), then R will deviate to maximizing time in (2,3), finally C will deviate to maximize time in (3,1), bringing players back to $\mathcal{S}$. Over time, these values will normalize to
$\mathcal{S}_{av} = (\frac{M-\epsilon+\epsilon}{2}, \frac{M-\epsilon+\epsilon}{2}, \epsilon, \epsilon) = (\frac{M}{2}, \frac{M}{2}, \epsilon, \epsilon)$[6].

$$\Delta(u^R) = \frac{6M+8\epsilon}{2M+4\epsilon} - 2 = \frac{M}{M+2\epsilon}$$
$$\Delta(u^C) = \frac{5M+10\epsilon}{2M+4\epsilon} - 3 = \frac{-M}{2M+4\epsilon}$$

[6]Note that while cycling is detrimental for C, there are solutions for where both players mutually benefit from cycling. Consider M=15, $\epsilon = 0.5$

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

Figure 7: Example 3 | R should cycle

	ϵ	$(M-\epsilon)$
ϵ	$(\frac{M-4\epsilon}{2M}, -\frac{1}{2})$	$(\frac{-2\epsilon}{M+2\epsilon}, -\frac{2\epsilon}{M+2\epsilon})$
$(M-\epsilon)$	$(\frac{M-4\epsilon}{M+2\epsilon}, -(\frac{M}{M+2\epsilon}))$	$(-\frac{1}{2}, -\frac{1}{2})$

Table 3: Unfilled meta matrix

$M \geq 2\epsilon$	$M \geq 2\epsilon$
$-2\epsilon \geq -M$	$-2\epsilon \geq -M$
$-4\epsilon \geq -M - 2\epsilon$	$-M - 2\epsilon \geq -2M$
$2(-2\epsilon) \geq (-1)(M+2\epsilon)$	$(-1)(M+2\epsilon) \geq (2)(-M)$
$\frac{-2\epsilon}{M+2\epsilon} \geq \frac{-1}{2}$	$\frac{-1}{2} \geq \frac{-M}{M+2\epsilon}$

Table 4: Deduction for Example 2

Take the example where M=15, $\epsilon = 0.5$

$$\mathcal{S}_{av} = (7.5, 7.5, 0.5, 0.5) \Rightarrow \mathcal{T} = 16$$
$$u^R(\mathcal{S}_{av}) = (\frac{7.5}{16})(2) + (\frac{7.5}{16})(4) + (\frac{0.5}{16})(1) + (\frac{0.5}{16})(3) \approx 2.94$$
$$u^C(\mathcal{S}_{av}) = (\frac{7.5}{16})(3) + (\frac{7.5}{16})(2) + (\frac{0.5}{16})(4) + (\frac{0.5}{16})(1) = 2.5$$

5. CYCLIC GAMES WITH NO MAX TIME RESTRAINTS

Previous analysis showed that while p_k's rational strategy minimizes time spent in the least preferred state, it does not always maximize the remaining time spent in its most preferred decision state and there are games where the rational choice is to spend ϵ time in the most preferred state as well. This indicates that as long as a minimum time limit is specified, there is the possibility that it might be rational to engage a cycle even when no maximum time limits are defined. For this class of games a natural question arises: do any of these games have NE solutions? We address this question directly in with Theorem 3 and show the answer is not only yes, but that the NE which exists is unique; but before presenting the theorem build up the intuition behind it beginning with the following lemma.

5.1 Existence of Unique NE

LEMMA Consider an n-cycle $(S_1, S_2, ..., S_n)$ with m players. For some solution $\mathcal{S} = (t_1, t_2, ..., t_n)$ and player p_i with utility function u^i and decision state S_m:

1. *$v^i(S_m) < u^i(\mathcal{S}) \Rightarrow p_i$ should deviate from $\mathcal{S}$ by decreasing time in S_m*
2. *$v^i(S_m) > u^i(\mathcal{S}) \Rightarrow p_i$ shouuld deviate from $\mathcal{S}$ by increasing time in S_m*
3. *$v^i(S_m) = u^i(\mathcal{S}) \Rightarrow p_i$ has no incentive to alter time in S_m*

PROOF. **Part 1**
Consider an n-cyclic subgame $(s_1, s_2, ..., s_n)$ with $I = [\epsilon, \infty)$. Let $\mathcal{S} = (t_1, t_2, ..., t_n)$ be any arbitrary solution, with $\mathcal{T} = (t_1 + t_2 + ... + t_n)$. If for some player p_i with decision state S_m, $v^i(S_m) > u^i(\mathcal{S})$, then let $\mathcal{S}^* = (t_1, t_2, ., t_m + \delta, ..., t_n)$ and observe the following sequence of equivalent statements:

$v^i(S_m) > u^i(\mathcal{S})$

	ϵ	$(M-\epsilon)$
ϵ	$(\frac{2M-2\epsilon}{2M}, -\frac{1}{2})$	$(\frac{2\epsilon}{M+2\epsilon}, -\frac{2\epsilon}{M+2\epsilon})$
$(M-\epsilon)$	$(\frac{2M-2\epsilon}{M+2\epsilon}, -(\frac{M}{M+2\epsilon}))$	$(\frac{1}{2}, -\frac{1}{2})$

Table 5: Unfilled meta matrix

$$v^i(S_m) > \frac{t_1}{\mathcal{T}}v^i(S_1) + \frac{t_2}{\mathcal{T}}v^i(S_2) + ... + \frac{t_n}{\mathcal{T}}v^i(S_n)$$

$$\mathcal{T}v^i(S_m) > t_1v^i(S_1) + t_2v^i(S_2) + ... + t_nv^i(S_n)$$

$$\delta(\mathcal{T}v^i(S_m)) > \delta(t_1v^i(S_1) + t_2v^i(S_2) + ... + t_nv^i(S_n)) \ \forall \ \delta > 0$$

$$\delta(\mathcal{T}v^i(S_m)) + (\mathcal{T}t_1v^i(S_1) + \mathcal{T}t_2v^i(S_2) + ... + \mathcal{T}t_nv^i(S_n)) > \delta(t_1v^i(S_1)+t_2v^i(S_2)+...+t_nv^i(S_n))+(\mathcal{T}t_1v^i(S_1)+\mathcal{T}t_2v^i(S_2)+ ... + \mathcal{T}t_nv^i(S_n))$$

$$\mathcal{T}(t_m+\delta)v^i(S_m)+\mathcal{T}t_1v^i(S_1)+\mathcal{T}t_2v^i(S_2)+...+\mathcal{T}t_{m-1}v^i(S_{m-1})+ \mathcal{T}t_{m+1}v^i(S_{m+1}) + ... + \mathcal{T}t_nv^i(S_n)) > (\mathcal{T}+\delta)t_1v^i(S_1) + (\mathcal{T}+ \delta)t_2v^i(S_2) + ... + (\mathcal{T}+\delta)t_nv^i(S_n)$$

$$\frac{t_1v^i(S_1)}{\mathcal{T}+\delta} + \frac{t_2v^i(S_2)}{\mathcal{T}+\delta} + ... + \frac{t_{m-1}v^i(S_{m-1})}{\mathcal{T}+\delta} + \frac{(t_m+\delta)v^i(S_m)}{\mathcal{T}+\delta} + \frac{t_{m+1}v^i(S_{m+1})}{\mathcal{T}+\delta} + ... + \frac{t_nv^i(S_n)}{\mathcal{T}+\delta} > \frac{t_1v^i(S_1)}{\mathcal{T}} + \frac{t_2v^i(S_2)}{\mathcal{T}} + ... + \frac{t_nv^i(S_n)}{\mathcal{T}}$$

$$u^i(\mathcal{S}^*) > u^i(\mathcal{S})$$

$\therefore$ It is rational for p_i to deviates to $\mathcal{S}^*$ by increasing time in S_m as this increases its utility.

Part 2

Consider an n-cyclic subgame $(S_1, S_2, ..., S_n)$ with $I = [\epsilon, \infty)$. Let $\mathcal{S} = (t_1, t_2, ..., t_n)$ be any arbitrary solution, with $\mathcal{T} = (t_1 + t_2 + ... + t_n)$. If for some player p_i and decision state S_m, $v^i(S_m) < u^i(\mathcal{S})$, then let $\mathcal{S}^* = (t_1, t_2, ..., t_m - \delta, ..., t_n)$ and observe the following sequence of equivalent statements:

$$v^i(S_m) < u^i(\mathcal{S})$$

$$v^i(S_m) < \frac{t_1}{\mathcal{T}}v^i(S_1) + \frac{t_2}{\mathcal{T}}v^i(S_2) + ... + \frac{t_n}{\mathcal{T}}v^i(S_n)$$

$$\mathcal{T}v^i(S_m) < t_1v^i(S_1) + t_2v^i(S_2) + ... + t_nv^i(S_n)$$

$$-\delta(\mathcal{T}v^i(S_m)) > -\delta(t_1v^i(S_1) + t_2v^i(S_2) + ... + t_nv^i(S_n)) \ \forall \ \delta > 0$$

$$-\delta(\mathcal{T}v^i(S_m)) + (\mathcal{T}t_1v^i(S_1) + \mathcal{T}t_2v^i(S_2) + ... + \mathcal{T}t_nv^i(S_n)) > -\delta(t_1v^i(S_1)+t_2v^i(S_2)+...+t_nv^i(S_n))+(\mathcal{T}t_1v^i(S_1)+\mathcal{T}t_2v^i(S_2)+ ... + \mathcal{T}t_nv^i(S_n))$$

$$\mathcal{T}(t_m-\delta)v^i(S_m)+\mathcal{T}t_1v^i(S_1)+\mathcal{T}t_2v^i(S_2)+...+\mathcal{T}t_{m-1}v^i(S_{m-1})+ \mathcal{T}t_{m+1}v^i(S_{m+1}) + ... + \mathcal{T}t_nv^i(S_n)) > (\mathcal{T}-\delta)t_1v^i(S_1) + (\mathcal{T}- \delta)t_2v^i(S_2) + ... + (\mathcal{T}-\delta)t_nv^i(S_n)$$

$$\frac{t_1v^i(S_1)}{\mathcal{T}-\delta} + \frac{t_2v^i(S_2)}{\mathcal{T}-\delta} + ... + \frac{t_{m-1}v^i(S_{m-1})}{\mathcal{T}-\delta} + \frac{(t_m-\delta)v^i(S_m)}{\mathcal{T}-\delta} + \frac{t_{m+1}v^i(S_{m+1})}{\mathcal{T}-\delta} + ... + \frac{t_nv^i(S_n)}{\mathcal{T}-\delta} > \frac{t_1v^i(S_1)}{\mathcal{T}} + \frac{t_2v^i(S_2)}{\mathcal{T}} + ... + \frac{t_nv^i(S_n)}{\mathcal{T}}$$

$u^i(\mathcal{S}^*) > u^i(\mathcal{S})$ $\therefore$ it is rational for p_i to deviate to $\mathcal{S}^*$ by decreasing time in S_m

The proof for part 3 is trivial $\square$

The lemma shows that if a player's valuation of decision state S_m is greater than the utility it receives cycling with $\mathcal{S}$, it is rational to deviate from $\mathcal{S}$ by increasing time in S_m. But this begets the question: is there a specific amount of time it should deviate by so that this new solution is optimal? Alternatively, can it overshoot its deviation and end up wanting to backtrack and deduce time in S_m? The answer can be found in the following corollary:

COROLLARY *Consider an n-cycle $(S_1, S_2, ..., S_n)$ with m players. For some solution $\mathcal{S} = (t_1, t_2, ..., t_n)$, some state S_m and some player p_i, if $v^i(S_m) > U^i(\mathcal{S}) \Rightarrow \forall \xi > 0$, $\mathcal{S}* = (t_1, t_2, ..., t_{m-1}, \xi, t_{m+1}, ..., t_n)$ is not NE*

Proof. Suppose for p_i $v^i(S_m) > U^i(\mathcal{S})$ and p_i increases time in S_m by δ resulting in $\mathcal{S}^* = (t_1, t_2, ..., t_m + \delta, ..., t_n)$
$v^i(S_m) > u^i(\mathcal{S})$
$v^i(S_m) > \frac{t_1}{\mathcal{T}}v^i(S_1) + \frac{t_2}{\mathcal{T}}v^i(S_2) + ... + \frac{t_n}{\mathcal{T}}v^i(S_n)$
$\mathcal{T}v^i(S_m) > t_1v^i(S_1) + t_2v^i(S_2) + ... + t_nv^i(S_n)$
$\mathcal{T}v^i(S_m) + (\delta v^i(S_m)) > t_1v^i(S_1) + t_2v^i(S_2) + ... + t_nv^i(S_n) + (\delta v^i(S_m))$

$$v^i(S_m) > \frac{\left[t_1v^i(S_1) + t_2v^i(S_2) + ... + (t_m+\delta)v^i(S_m) + ... + t_nv^i(S_n)\right]}{\mathcal{T}+\delta}$$

$v^i(S_m) > u^i(\mathcal{S}^*)$

Since the selection of δ was arbitrary, p_i will always have the incentive to unilaterally deviate from $\mathcal{S}^*$ by increasing time in S_m[7] $\square$

What about the alternative, when a player wants to minimize the time it spends in a state? Players might wish to deviate by decreasing the time in a state, but with a lower bound ϵ the possibility for a NE solution exists: the one where players spend ϵ in each decision state. So as long as for each players and all the states s_m it moves from, if $s_m < u^p(\epsilon, \epsilon, ..., \epsilon)$ then $(\epsilon, \epsilon, \epsilon...., \epsilon)$ is NE because while the players have an incentive to deviate, it is impossible to do so. Furthermore, notice that by nature of the cycle, if this relationship holds for a player's cyclic states, then it holds for all its decision states. Our analysis can now be formalized with the following theorem:

Theorem 3. *Consider an n-cycle $(S_1, S_2, ..., S_n)$ with m players. Let $\mathcal{S}_\epsilon = (\epsilon, \epsilon, ..., \epsilon)$. If $\forall$ players p_i $v^i(S_0) \leq U^i(\mathcal{S}_\epsilon) \Rightarrow \mathcal{S}_\epsilon$ is the unique NE.*

In conclusion, if $I = [\epsilon, \infty)$, player p_k need only check if each of its opponent's valuation of the cyclic state is less than their valuation about the average of the cycle. If it is, then not only is the minimum solution $\mathcal{S}_\epsilon = (\epsilon, \epsilon, ..., \epsilon)$ a NE, but it is mutually beneficial for all players and p_k should cycle. If this test fails, then there is no NE for the game.

5.2 Example Game

Consider the game in Figure 8 with initial player R and state (2,3) and $I = [\epsilon, \infty)$. If we want to know if there is a mutually beneficial NE solution for the game, we check if

[7]Analysis shows the same situation for decreasing the amount it spends in a state, but is omitted due to space constraints.

$2, 3$	$4, 2$
$3, 2.5$	$1, 5$

Figure 8: Example 4 | Mutually beneficial to cycle

each player values the average payoff over the cycle as opposed to their valuation for the cyclic state. For R: $2 < \frac{2+3+1+4}{4} = 2.25$, and for C: $3 < \frac{3+2.5+5+2}{4} = 3.125$, so Theorem 2 is satisfied and we conclude $\mathcal{S}_\epsilon = (\epsilon, \epsilon, \epsilon, \epsilon)$ is NE. Try $\epsilon = 0.5$.

$$u_R(\mathcal{S}_\epsilon) = \frac{0.5}{2}2 + \frac{0.5}{2}3 + \frac{0.5}{2}1 + \frac{0.5}{2}4 = 2.5$$
$$u_C(\mathcal{S}_\epsilon) = \frac{0.5}{2}3 + \frac{0.5}{2}2.5 + \frac{0.5}{2}5 + \frac{0.5}{2}2 = 3.125$$

Test by first making R deviate by spending additional 0.2 time in its most preferred state then doing the same for C

$\mathcal{S}_\epsilon = (0.7, 0.5, 0.5, 0.5)$, $\mathcal{T} = 2.2\epsilon$
$$u_R(\mathcal{S}_\epsilon) = \frac{0.7}{2.2}2 + \frac{0.5}{2.2}3 + \frac{0.5}{2.2}1 + \frac{0.5}{2.2}4 = 2.45$$
$$u_C(\mathcal{S}_\epsilon) = \frac{0.7}{2.2}3 + \frac{0.5}{2.2}2.5 + \frac{0.5}{2.2}5 + \frac{0.5}{2.2}2 = 3.11$$
$\mathcal{S}_\epsilon = (0.5, 0.7, 0.5, 0.5)$, $\mathcal{T} = 2.2\epsilon$
$$u_R(\mathcal{S}_\epsilon) = \frac{0.5}{2.2}2 + \frac{0.7}{2.2}3 + \frac{0.5}{2.2}1 + \frac{0.5}{2.2}4 = 2.54$$
$$u_C(\mathcal{S}_\epsilon) = \frac{0.5}{2.2}3 + \frac{0.7}{2.2}2.5 + \frac{0.5}{2.2}5 + \frac{0.5}{2.2}2 = 3.06$$
Note in this case that R fares better if C deviates but C does worse, so it has no incentive to do so.

6. CONCLUSIONS AND FUTURE WORK

TOM is a novel approach to analyzing and playing dynamic games. Unfortunately, it neglects the influence of time on strategy selection restricting players to static, end game payoff distribution. Because of this, cycling, while shown to exist, is treated indistinguishably from games with pure strategies for staying in non-cyclic initial states, failing to supply players with a suitable rationality paradigm for optimizing their strategy in such situations. To determine if cycling is preferable over non-movement, a player must ask itself two questions: are there *any* solutions under which cycling proves beneficial, and are any of these feasible when considering infinite play?

We first revised the rationality rules of TOM to support an analysis of games where utility varies as a function of time and to incorporate time limitations on game play, and used this to characterize both cyclic and non-cyclic games. We then turned our attention to the analysis of indefinite cycling. We found in such games, each player's rational strategy is to minimize the time it spends in its *least* preferred state but that the converse is not necessarily true. For games with strict bounds on time, we found if equilibria exist they exist only as solutions corresponding to strategies on either of these bounds. We constructed a meta matrix for each game and theorized that NE in the meta matrix correspond to stable strategy sets that converge during indefinite game play. We find that even in the absence of equilibria solutions, where players constantly deviate to different strategies, cycling can be rational. For games limited only to minimum time limits, we found there are still scenarios where cycling is rational: those in which each player's valuation of the cyclic state is less than its average valuation about the cycle.

Although our analysis focused exclusively on the class of 2×2 games, we believe some of these results can be generalized to arbitrary n player, m state games. The difficulty arises from the fact that cycles of various lengths can occur, and so there is no straight forward way to determine the number of decision states each player makes, if any, in a cycle, nor what order player arises in. We have developed the geometric intuition behind these concepts, and are attempting to establish efficient means to characterize the solution space and equilibria of a game as a result. Furthermore, we plan to conduct simulations to obtain empirical results to supplement the theoretical ones discussed here.

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

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

[5] 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.

[6] 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.

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

[8] 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.

[9] 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.