# The Evolution of \Theory of Mind"

Daniel Monte
*Sao Paulo School of Economics (EESP-FGV)*

Nikolaus Robalino
*Simon Fraser University*

and

Arthur Robson
*Simon Fraser University*

February 19, 2012

Abstract. This paper provides an evolutionary foundation for our capacity to attribute mental states such as belief, desire, and intent to ourselves, and to others. This ability, referred to as "Theory of Mind", is intrinsic to game theory and is viewed by many as the capstone of social cognition. We argue here that theory of mind allows organisms to efficiently modify their behavior in strategic environments with a persistent element of novelty. We find that in such non-stationary environments it yields a sharp, unambiguous advantage over less sophisticated, behavioral approaches to strategic interaction.

## 1. Introduction

An individual with *theory of mind* (*ToM*) has the ability to conceive of himself, and of others, as having agency, and so to attribute to himself and others mental states such as belief, desire, knowledge, and intent. It is generally accepted in psychology that human beings beyond early infancy possess *ToM*. Further, it is conventional in game theory to make the crucial assumption, without much apology, that agents have *ToM*.

The present paper considers *ToM* in greater depth by addressing the question: *Why* and *how* might have such an ability evolved? In what types of environments would *ToM* yield a distinct advantage over alternative, less sophisticated, approaches to strategic interaction? In general terms, the answer we propose is that *ToM* is an evolutionary adaptation for dealing with strategic environments that have a persistent element of novelty.

The argument made here in favor of theory of mind is a substantial generalization and reformulation of the argument in Robson (2001) concerning the advantage of having

daniel.monte@fgv.br
nkasimat@sfu.ca
robson@sfu.ca

an own utility function in a non-strategic setting. In that paper, an own utility function permits an optimal response to novelty. Suppose an agent has experienced all of the possible outcomes, but has not experienced and does not know the probabilities with which these are combined. This latter element introduces the requisite novelty. If the agent has the biologically appropriate utility function, she can learn the correct gamble to take; if she acts correctly over a sufficiently rich set of gambles, she must possess, although perhaps only implicitly, the appropriate utility function.

We shift focus here to a dynamic model in which players repeatedly interact with one another but in which the set of games that they might face becomes larger and larger with time. We presume individuals have an appropriate own utility function. The focus is then on the advantage to an agent of conceiving of her opponents as rational actors—as having preferences, in particular, and understanding that they act optimally in the light of these. Having a template into which the preferences of an opponent can be fitted enables a player to deal with a higher rate of innovation than can a behavioral type of individual that adapts to each game as a distinct set of circumstances. In other words, the edge to *ToM* derives from a capacity to extrapolate to novel circumstances information that was learned about preferences in a specific case.

The distinction between the *ToM* and behavioral types might usefully be illustrated with reference to the following observations of vervet monkeys (Cheney and Seyfarth (1990), p. 213). If two groups are involved in a skirmish, sometimes a member of the losing side is observed to make a warning cry used by vervets to signal the approach of a leopard. All the vervets will then urgently disperse, saving the day for the losing combatants. The issue is: What is the genesis of this deceptive behavior? One possibility, corresponding to our theory of mind type, is that the deceptive vervet appreciates what the effect of such a cry would be on the others, understands that is, that they are averse to a leopard attack and exploits this aversion deliberately. The other polar extreme corresponds to our behavioral adaptive learners. Such a type has no model whatever of the other monkeys' preferences and beliefs. His alarm cry behavior conditions simply on the circumstance that he is losing a fight. By accident perhaps, he once made the leopard warning in such a circumstance and it had a favorable outcome. Subsequent reapplication of this strategy continued to be met with success, reinforcing the behavior.

Consider the argument in greater detail. We begin by fixing a two-stage extensive form with perfect information. In each period, each of a large number of player 1's is randomly matched to an opponent drawn from a large number of player 2's. In addition, the outcomes needed to complete the game are drawn randomly from some large but finite set. Each player has a strict ordering over the set of outcomes drawn for that player. Each player is fully aware of his own ordering but does not know the strict preference ordering of his opponent.

We compare two types of players – behavioral (or naive) and theory-of-mind (sophisticated) types. In the two-stage setting, this distinction is only important for player 1, since the optimal choice by the player 2's relies only on 2's preferences. The naive players behave in a fashion that is consistent with simple adaptive learning in psychology and with evolutionary game theory in economics. Each game is seen as a fresh problem, so naive learners must adaptively learn to play each such different game.

The *ToM* type of player 1, on the other hand, is disposed to learn the other agent's preferences. It is relevant now that the pattern of play is revealed to all players at the end of that period. Each time the player 1's see the player 2's being forced to make a choice, the player 1's learn how the player 2's rank the two outcomes. [1] For simplicity, we do not suppose the player 1's use the transitivity of the preference ordering of the player 2's. This assumption clearly loads the dice against the result we generate concerning the evolutionary advantage of the *ToM* type.

We now introduce innovation by periodically adding outcomes to the pool of existing ones. Suppose the arrival rate of such novelty is sufficiently low. Then in the limit, the *ToM* player 1's are exposed to almost all possible pairwise choices by the player 2's and hence they play appropriately in all but a vanishing fraction of the games they face. In this case, the fraction of possible games that naive players are exposed to will also tend to one in the limit, so they also play appropriately on all but a diminishing fraction of games. If the growth rate is sufficiently high, on the other hand, there will again be no apparent advantage to the *ToM* types, since both types will be informed about a vanishing fraction of cases they face.

The key observation is that, in an intermediate range of growth rates, the *ToM* types will be informed about the opponent's preferences with a probability converging to 1, while the naive types will be informed of the game to be played with a probability that converges to zero. In this simple, strong and robust sense, then, the *ToM* type outdoes the naive type. The key reason for the greater success of the *ToM* type is simply that there are vastly more possible games that can be generated from a given number of outcomes than there are outcome pairs. The sharpness of this result enables us to get away from calculating and comparing average payoffs.

The two-stage game considered so far apparently is rather special. The player 2's have no need of strategic sophistication at all, and the need for strategic sophistication on the part of the player 1's is limited to knowledge of player 2's preferences. How do our two-stage game results extend to a general setting? We then turn to a general $I$-stage game of perfect information to consider this question. There are now $I$ player types, where one type associated with each stage. Again, the last player does not need any strategic sophistication, and the advantage of *ToM* to the second-to-last player is similar to that obtained already for the player 1's in a two-stage game. However, the third-to-last player now apparently faces a more complex task. He must not only learn the preferences of the last two players but also learn that the second-to-last player knows the last players preferences as well. This last requirement is one of increasing depth of knowledge; the first requirement is one of increasing scope. In the present model,

[1] The assumption that there are a large number of player 2's means that a deviation by a single player 2 will not change the distribution observed by the player 1's. Assuming that the player 1's condition only on this distribution, there is no incentive then for the player 2's to choose strategically, or in any way contrary to the myopic payo .

however, the depth requirement is finessed. When the last player makes a choice that reveals his preferences between two outcomes, this choice is common knowledge to all the sophisticated players. Furthermore, the increasing scope of what the second-to-last player needs to know has no effect on the critical rate of innovation beyond which sophisticated players are inevitably unable to learn. On the other hand, the task faced by the naive players becomes more formidable, in terms of the critical rate of innovation, simply because there are now more possible games. The results from the two-stage game then generalize in a straightforward way.

## 2. The Model

We describe a model where the number of possible situations that agents might face grows over time and where, initially at least, agents are unsure about how others will behave in these different circumstances.

### *2.1. The Environment*

Consider $I$ equally "large" populations of infinitely lived players. Every player in population $i = 1, \ldots, I$ is endowed with the same preference ordering over a countably infinite set of objects, which we refer to as $Z$. Denote the preferences of population $i$ individuals as $\succsim^i \subset Z \times Z$. These preferences are complete and transitive. The members of population $i$ may be referred to as *player* $i$'s or as having *preference type* $i$. It assumed that each type's preferences are private information.

There are an infinite number of discrete periods. In each period, all the players are randomly matched up in $I$-sized groups, with one member of each matching from each preference type. The players of each matching in period $t$ then independently play a perfect information game, denoted by $\Gamma_t$. In any given period, a large number of identical games are played, each involving $I$ players. After playing the period $t$ game the groups dissolve. For simplicity, all possible $I$ member matchings of the $I$ distinct preference types are equally likely in every period.

We refer to $\Gamma_t$ as the period $t$ game. These games generally involve different outcomes. [2] More specifically, we assume the following throughout: (1) every $\Gamma_t$ is an $I$ stage game in which the player 1's move first, then the player 2's move, then the player 3's, and so on, and (2) there are exactly $A$ moves at every information set. We now elaborate further on how stage games are realized.

Fix a sequence of outcome sets $\{Z_t\}$ [3] and let $G_t$ be the set all perfect information games that could be generated from $Z_t$ that satisfy the two restrictions above. Assume each period game is drawn with equal probability from $G_t$. [4]

The sequence of available outcomes $\{Z_t\}$ comprises a dynamic environment with ever growing complexity. Although only a finite set of outcomes is available at each

[2] It is possible to allow the game forms to di er also, but this is precluded in the interests of simplicity.

[3] Each type's preference ordering over $Z_t$ is derived from the type's original ordering over $Z$.

[4] This is for simplicity. All that matters is that players have su cient exposure to all of their opponents' relevant choices.

date, as time goes by, there are more and more of these. Precisely, one new outcome is introduced at each of a sequence of arrival dates, denoted $t_1, ..., t_k, ...$, where $t_k$ is the arrival date of the $k$-th new outcome, so that $t_k \leq t_{k+1}$ for all $k = 1, 2, \ldots,$. Once introduced, outcomes are never removed from the pool of available objects. As a result, in periods $t_k$ to $t_{k+1} - 1$ the number of available outcomes is $|Z_0| + k$, where $Z_0$ is the initial set of objects.

This $k$-th object might be chosen at random from the set of remaining possibilities, but we will not spell out the details of how these drawings occur. Although there may be implications for the rank distribution of a newly introduced object, we assume the players do not avail themselves of any such information. Without loss of generality, then take $Z$ as the set of integers $1, 2, \ldots, .$ Further, take $Z_0 = \{1, 2, \ldots, N\}$, the first object introduced as $N+1$, the second as $N+2$, and so on. In periods $t = t_k, \ldots, t_{k+1}-1$ games are then drawn from the set of all the games that can be generated from the outcome set $\{1, 2, \ldots, N, \ldots, N + k\}.$ Thereby, the strategic environment is summarized as:

$$\mathscr{E} = \left(I, A, \left(\succsim^1, \ldots, \succsim^I\right), \{t_k\}_{k=1}^{\infty}\right).$$

### 2.2. Strategies and Histories

In each period, all players know the entire history of play. [5] Somewhat more formally, a history of length $t - 1$ is a sequence

$$H_{t-1} = \left\{\left(\Gamma_\tau, \tilde{}^1_\tau, \tilde{}^2_\tau, \ldots, \tilde{}^I_\tau\right)\right\}_{\tau=1}^{t-1},$$

where $\tilde{}^i_\tau$ is the aggregate distribution of period $t$ play, at nodes reached by $i$ preference types, in period $t$. Let $\mathscr{H}_{t-1}$ denote the set of all $t - 1$ length histories.

The focus of the analysis is on how players learn in the environments described above. This learning is facilitated since there are no supergame effects here. Individual players ignore the effects of their actions on the future behavior of their opponents and play myopically. Why couldn't players fool their opponents about their preferences? Recall that we assume a player reacts only to the *distribution* of choices made by the remaining players. Since any particular player has no effect on these distributions, and hence cannot affect the behavior of her future opponents, any such particular player must behave myopically. In the two-stage case, player 1's cannot, in any case, advantageously mislead the 2's because 2's do not react to 1's previous behavior. More interestingly, the type 1's disregard the choice of any particular type 2, since this choice cannot affect the distribution of player 2's choices, and so type 2's have no incentive to mislead the player 1's.

### 2.3. Cognitive Types

We are now ready to define the two cognitive types, theory of mind ("theory," for short, or "sophisticated") and behavioral (or "naive"), that are the subjects of the analysis.

---

[5] However, players never observe an opponent's *payoffs.*

Behavioral types are characterized by an inability to attribute mental states to other agents. They cannot then establish any relationship between the choices made by others in different contexts. Specifically, a *behavioral* (or *naive*) type is a player that adapts to each period game as a distinct set of circumstances, learning to play appropriately in a game only from repeated exposure to it. [6]

Theory types make inferences about preferences and beliefs using observed behavior. A *theory of mind* type knows that other agents are endowed with preferences and beliefs, and that they act optimally according to these. This is common knowledge among *ToM* types.

We adopt the following simplifying assumption.

A1: *Each preference type has a strict preference between any distinct pair of outcomes. The strictness of each preference ordering is common knowledge among* ToM *types.* [7] *Speci cally, it is commonly known by theory types that for each* $i$ $I$ *and* $z; z'$ $Z$, $z \sim^i z'$ *if and only if* $z = z'$:

When the players are theory types A1 implies there are histories that make commonly known some information about preferences. For any $z; z'$ $Z_t$ and $i$ $I$, where $z \neq z'$, say that $H_t$ *reveals* $z \succ^i z'$ if all theorists commonly know that $H_t$ could not have happened if $z' \succ^i z$. [8]

The section closes with illustrative examples of preference revelation. Assume A1 throughout.

Consider first "last mover" (the $I$ type) preference revelation. Suppose the subgame in Figure 1 is reached by a positive measure of player $I$'s. Suppose further, that $z \succ^I z'$.

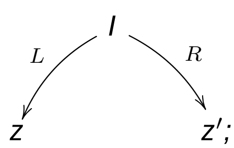


Figure 1: If $z \succ^I z'$ then this is revealed when this subgame is reached.

Since all players are rational, all player $i$'s who reach this subgame choose action "$L$". Let $H_t$ be the resulting history. Then $H_t$ reveals $z \succ^I z'$ since "$L$" could not have been chosen in the event that $z' \succ^I z$:

[6] Our behavioral agents approach a game in an essentially nontheoretic fashion. They merely associate games with acts according to some statistical procedure. They are reinforcement learners with only an ability to tell the di erence between distinct social situations. More rational behavioral types might respond optimally to beliefs about what happens at other players' information sets (Fudenberg and Kreps, 1995). These players are behavioral in the sense that these beliefs are updated according to experience in an adaptive fashion.

[7] Consider the following salient case. Suppose each $z \in Z_t$ is an $I$-tuple, $(x^1, \ldots, x^i, \ldots, x^I)$, where $x^i$ is the amount of money, say, allocated to player $i$. The present assumption then rules out independent preferences since it rules out that, in general, $(x^1, \ldots, x^i, \ldots, x^I) \sim^i (\hat{x}^1, \ldots, x^i, \ldots, \hat{x}^I)$. Nevertheless, the results can also be obtained with independent preferences if we assume that $(x^1, \ldots, x^i, \ldots, x^I) \sim^i (\hat{x}^1, \ldots, \hat{x}^i, \ldots, \hat{x}^I)$ if and only if $x^i = \hat{x}^i$ and make this common knowledge among *ToM* types.

[8] Without A1 each history would then be consistent with a wide range of preference based models.

What kind of histories reveal information about $\succsim^i$ when $i < I$? Consider the perspective of any *ToM* player who observes the subgame in Figure 2, where $\{z \succ^I x, z' \succ^I x'\}$ have been previously revealed and where $z \neq z'$. The observer knows that player $I - 1$

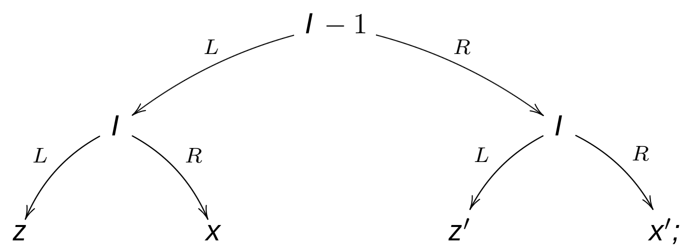


Figure 2: If $\{z \succ^I x, z' \succ^I x'\}$ have been revealed, $I-1$'s choice here reveals $I-1$'s preference between $z, z'$.

knows how player $I$ will choose. Hence, if the $I-1$ types choose "$L$", $z \succ^{I-1} z'$ becomes commonly known.

## 3. Results

Recall (from Section 2.1) that the environment is given by:

$$\mathscr{E} = \left(I, A, \left(\succsim^1, \ldots, \succsim^I\right), \{t_k\}_{k=1}^{\infty}\right).$$

The main results are about how different cognitive types cope in various environments—in particular, how they learn to play as we vary $I$, $A$, and, particularly, the arrival dates.

In very slowly and in very rapidly complexifying environments there are no benefits to *ToM*. There is however an intermediate, and significant, range in which *ToM* has a striking and unambiguous advantage over naivete. These results are expressed in Theorem 1 below. Before stating the result we need to first introduce some key notation.

Let $R_t^i$ denote the set of pairs in $Z_t \times Z_t$ on which $\succsim^i$ has been revealed. That is, $R_t^i = \left\{\{z, z'\} : \succsim^i \text{ revealed on } \{z, z'\}\right\}.$ Assume $\{z, z\} \in R_t^i$, for every $z \in Z_t$, and moreover that if $\{z, z'\} \in R_t^i$ then $\{z', z\} \in R_t^i$. A primary object of interest is the count of how much information has been revealed about $i$ type preferences. This is given by $K_t^i = |R_t^i|$, which belongs to the set $\{|Z_t|, |Z_t| + 2, \ldots, |Z_t|^2\}$. Let

$$L_t^i \equiv K_t^i / |Z_t|^2.$$

This is the fraction of $i$ type binary choices that have been revealed by period $t$. We will also keep track of the number of games players have been exposed to. The basic game tree has $A^I$ terminal nodes. Thus the number of possible games in period $t$ is

$|G_t| = |Z_t|^{A^I}$. Letting $K_t^N$ be the number of distinct games that have been realized along $H_t$; the fraction of period $t$ games to which players have been exposed is then

$$L_t^N \equiv K_t^N = |Z_t|^{A^I}:$$ [9]

The arrival date sequence is specified in (A2) below. The formulation chosen here parameterizes the rate at which the environment becomes increasingly complex, in a fashion that yields a straightforward connection between this rate and the advantages to theory of mind.

A2: *Let the arrival sequence* $\{t_k\}$ *be given as follows. Fix* $\quad \geq 0$. *For each* $k = 1; 2; : : : ;$ *let*

$$t_k = \lfloor [|Z_0| + k]^\alpha \rfloor : \tag{1}$$ [10]

We are now in a position to state the main results (Theorems 1 and 2). The proofs of these are given in the Appendix. First, for the simplest possible environment—

Theorem 1: *Assume (A1-2), and suppose* $I = 2$ *and* $A = 2$:

- *A) If* $\quad [0; 2)$ *then* $L_t^2 \longrightarrow 0$ *and* $L_t^N \longrightarrow 0$ *a.e. as* $t \longrightarrow \infty$. *That is, both the sophisticated and the naive type are overwhelmed by the rapid rate of arrival of new outcomes.*
- *B) If* $\quad (4; \infty)$, *then* $L_t^2 \longrightarrow 1$ *and* $L_t^N \longrightarrow 1$ *in probability, as* $t \longrightarrow \infty$. *That is, the rate of arrival of new outcomes is slow enough that theory types are able to essentially learn everything and naive types are exposed to all games.*
- *C) Finally, however, if* $\quad (2; 4)$, *then* $L_t^2 \longrightarrow 1$ *in probability but* $L_t^N \longrightarrow 0$ *a.e. as* $t \longrightarrow \infty$. *That is, for this intermediate range of arrival rates, the* ToM *type learns essentially everything, while the naive type learns essentially nothing.*

There are several aspects of the above result that bear emphasis. With the possible exception of the two isolated critical values of 2 and 4, the results are dramatic—either everything is learnt in the limit or nothing is. Indeed, it is not hard to see that, for either type, the range where nothing is learnt is inescapable in that the arrival rate of novelty outstrips there the maximum rate at which learning can occur. So the real contribution of the above is the much less obvious result that full learning occurs essentially whenever it is even possible that it could. These results make irrelevant the various simplifying assumptions that we invoked here.

In terms of the contest between the two types, there is then an interval over which the *ToM* type learns everything and the naive type learns nothing. The simplicity

[9] The count of the number of games that can be composed from $|Z_t|$ outcomes assumes that the order of the outcomes matters. If we allow outcomes after a given choice by a player to be permuted, and allow the subtrees after the choices by the rst player to be permuted, this reduces the number of games by a constant factor. However, the present results depend crucially only on the leading power to which $|Z_t|$ is raised.

[10] Here $\lfloor \cdot \rfloor$ denotes the oor function. It seems more plausible, perhaps, that these arrival dates would be random. This makes the analysis mathematically more complex, but does not seem to fundamentally change the results. The present assumption is then in the interests of simplicity.

of these results implies that we can finesse the issue of considering payoffs explicitly. Whatever these payoffs might be it is clear that the *ToM* type is outdoing the naive type in this intermediate range. [11]

### 3.1. General Environments

The two-stage game environment considered so far is special in a number of ways. The type 2's have no need of strategic sophistication at all, and the strategic sophistication of the player 1's is limited to knowledge of player 2's preferences. In fact, all that is required in order to play properly here is a first level theory of mind—some attribution by the player 1's of preferences to the player 2's.

We obtain, however, similar results for a general environment $\mathscr{E}$, with an arbitrary number of stages and moves. That is, the advantage of *ToM* is, in general, rather similar to that obtained already for the player 1's in a two-stage game. This is despite the need, in an $I$-stage game, for player 1, for example, to know what player 2's know about the preferences about the remaining players. In the current model of learning, however, such higher order sophistication comes for free. That is, when a player reveals her preferences all theory players simultaneously learn that player's preferences and that the other theory players know these preferences. The growth in complexity here does not then directly stem from higher and higher orders of belief, since when learning about preferences occurs here it is common knowledge. Any growth in complexity that there is stems from the more prosaic need for players moving near the start of the game to obtain preference information about more and more players. Remarkably, this greater complexity does not show up as a decreased ability to respond to novelty.

We argue then that evolution might proceed stage by stage in a general environment with $I$ preference types. First, the player $I-1$'s derive an advantage from *ToM* over naivete, regardless of the sophistication of the $i < I-1$ types for much the same reason as in the two-stage game. Once the $I-1$'s can be taken to be *ToM* it is then straightforward to demonstrate an edge to *ToM* over naivete for the player before him. The advantage to theory of mind is then established by induction.

We state the main result for environments with arbitrary (finite) $I$ and $A$—

Theorem 2: *Assume (A1-2). Let $T$ be the number of terminal nodes pertaining to the basic game tree. Here, the statements about the $L_t^i$s assume that all players $i+1;:::;I$ are theory types.*

*A) If $[0;2)$ then for each $i$ $I$, $L_t^i \longrightarrow 0$ and $L_t^N \longrightarrow 0$ a.e. as $t \longrightarrow \infty$. That is, both the sophisticated and the naive type are overwhelmed by the rapid rate of arrival of new outcomes.*

*B) If $(T;\infty)$, then for each $i$ $I$, $L_t^i \longrightarrow 1$ and $L_t^N \longrightarrow 1$ in probability as $t \longrightarrow \infty$. That is, the rate of arrival of new outcomes is slow enough that theory types are able to essentially learn everything and naive types are exposed to all games.*

[11] Suppose on the contrary that the *ToM* were right in the limit 5/8 of the time and the naive types were right 3/8 of the time. Without knowing more about the games in question, their payo s in particular, it does not follow automatically the the *ToM* type is superior.

C) *Finally, however, if $\quad (2; T)$, then $L_t^i \longrightarrow 1$ in probability but $L_t^N \longrightarrow 0$ as $t \longrightarrow \infty$. That is, for this intermediate range of arrival rates, the ToM type learns essentially everything, while the naive type learns essentially nothing.*

The general results here are again dramatic—except possibly at two points, everything is learnt in the limit or nothing is. Again, when nothing is learnt, it is because it is simply mechanically impossible to keep up with the rate of novelty, so that the key contribution of this theorem is to show that everything is learnt essentially whenever this is not mechanically ruled out.

The proof of Theorem 2 is somewhat intricate. Why is that? It is always clear that the rate of learning must be small if $L_t^i$, for example, is close to one. When $\quad > 2$, however, the proof involves showing that this is the *only* circumstance under which the rate of learning is small. There are two factors that complicate showing this. The first is that there are $i$-type subgames in which $i$'s player choice cannot reveal information about $\succsim^i$ because there is insufficient knowledge about the remaining players' choices.

The second factor is more awkward. It concerns the existence of $i$-type subgames with outcomes that are avoided by the remaining opponents, thus making it difficult to reveal information about $\succsim^i$. Such games may arise even as $t \longrightarrow \infty$. However, A1 implies that these problematic games are a vanishing small fraction of all games in the limit as $t \longrightarrow \infty$.

Again, there is then an interval over which the *ToM* type learns everything and the naive type learns nothing. Another point deserving emphasis is that the range in which theory types learn everything is independent of the complexity of the periodic interactions. The introduction of more and more stages (corresponding to more distinct preference types) and actions would seem bound to shift the transition point from no learning to full learning for the *ToM* types.

However, as the number of stages grows, although the speed of learning must surely be adversely affected, whenever $\quad > 2$ theory types can learn in any perfect information game that has a bounded number of terminal nodes. Consider a three stage environment, for example. As long as $\quad > 2$, player 1's will learn the player 3's preferences completely in the limit. In addition, at the same time that 3's choices reveal information about 3's preferences to player 1, they reveal the same information to the type 2's and the type 1's know this. But now, given this all this knowledge, and that $\quad > 2$, player 1 can also completely learn player 2's preferences. [12]

Naive types on the other hand do worse and worse as the period game becomes increasingly complicated. In that case, the naive types face a larger set of possible games, and hence can only keep up with a slower rate of novelty. Their disadvantage in this regard is particularly striking as the number of preference types increases. Recall that the number of terminal nodes, $T$, from Theorem 2 is $A^I$ in our model. Hence the range in which naive players are surpassed by theory types grows with each increase in $I$ and $A$:

[12] We can again  nesse the issue of considering payo  s explicitly. Whatever these payo  s might be, the *ToM* type is outdoing the naive type in the relevant intermediate range.

## 4. Extensions

What can be said about games with imperfect information? Normal form games, as a key case? For a general normal form game with multiple Nash equilibria, it is not clear how to disentangle the lack of knowledge of payoffs from the lack of information about which equilibrium is to be played, at least in the absence of strong assumptions. It is perhaps realistic to suppose that normal form games crop up along with the games of perfect information emphasized here, using outcomes drawn from the same set. Whether or not any learning can be accomplished on such normal form games, our approach shows that learning would arise based only on the games of perfect information. Given only that the naive types are not given some *ad hoc* advantage over the *ToM*'s when playing the normal form games, the results here will be robust.

How much do the current results depend on the particular model described here? Although the environment is rather particular, it is best seen merely as a test to discriminate between the underlying characteristics of *ToM*'s and of the naive players. That is, although other environments might show the difference in a less clear-cut fashion, it is hard to picture an even-handed test at which the naive players would outdo the *ToM* players.

For example, it is only for simplicity that we restrict attention to a fixed game tree. The tree itself could be random: it might involve a random number of moves or a random order of play, for example. Similarly, players could be allowed to move multiple times, sequentially, and so on.

Similarly, the assumption that individuals have a strict ranking over each distinct pair of outcomes is basically innocuous. If indifference is allowed, suppose, for example, that individuals randomize over each pair of indifferent outcomes. The indifference of player $i$ between $z$ and $z'$ would then be common knowledge to the *ToM* types, if ever player $i$ chose $z$ over $z'$ and $z'$ over $z$. This would lead a difference of detail rather than substance.

We assume here that *ToM* types do not apply transitivity in their deductions about the preferences of other players. This might make a substantive difference to the relevant ranges of the growth parameter . The new value of cannot exceed 2, since applying transitivity could not be disadvantageous. It is actually not hard to see that applying transitivity could not lower the critical value of below 1. In any case, there would continue to be the key intermediate range of values where *ToM* outperforms naivete; applying transitivity could only make this range larger.

Further, the results do not seem to be highly dependent on the exact specification of the two types. More sophisticated naive types could clearly do better than the ones we describe here. If naive types assign beliefs to subgames, rather than to entire games, for example, they would do as well as *ToM* types in the two stage game case. More generally, with three or more stages, such more sophisticated naive players would do better than the naive players considered here, but not as well as the sophisticated players.

## 5. The Evolution of *ToM* in Economics and Other Literatures

There are only a handful of papers within economics and game theory that attempt to provide an evolutionary foundation for *ToM.* Most of the existing work, in fact, has focused on the evolutionary basis of strategic choice rather than of *ToM per se.*

One such exception is Stahl (1993), who asks when evolutionary selection might favor strategic sophistication over naivete. In that paper, degrees of sophistication correspond to levels of of iteration of best responses. [13] The model yields a negative result pertaining to the emergence of higher order strategic intelligence. Naive types—in particular, those that luck upon the appropriate way to play—will not be eliminated from the population. Moreover, if higher orders of intelligence entail costs in terms of reproductive fitness then, and if aggregate play converges, all of the sophisticated types will die out in the long run. A crucial difference between Stahl (1993) and the present work is that the former focuses on a single symmetric normal form game. On the other hand, the non-stationarity of the environment, with a randomly chosen game in each period, generates our positive result in favor of sophistication. Another important difference between the two works lies in how we make the distinction between sophisticated and naive types. Here the *ToM* cognitive types are set apart by a disposition to view others as intentional rather than by an enhanced ability to iterate. Our naive types simply lack the ability to see others like this.

Another game theoretic work that addresses the evolutionary foundations of strategic intelligence is Mohlin (forthcoming, 2012). [14] Mohlin also obtains somewhat negative results concerning the emergence of sophisticated agents. Whether or not higher cognitive types are supported in the limiting distribution depends on the characteristics of the particular games in question (i.e., the iterated best replies). [15] In contrast, our results concerning *ToM* are not derived from the specific details of the periodic interactions but from the growing complexity of the environment, involving a widening variety of games.

Outside of economics and game theory, considerable effort has been devoted to identifying the ecological factors yielding selective pressures for social intelligence (see for instance Byrne and Whiten (1988) and Whiten and Byrne (1997)). [16] In this regard, the *Machiavellian Intelligence Hypothesis (MIH)* has emerged as the central explanatory theory. Although there are several discernible varieties of the MIH, a feature common to all of them is the idea that higher order social cognition is derived from social living and the resulting complexity (Byrne and Whiten, 1997). [17]

---

[13] A smart$_n$ cognitive type chooses a strategy that is $n$th-order rationalizable. The smart$_0$ types play in a predetermined way, i.e., a particular player of this type selects the same pure strategy in every period.

[14] Two di erent notions of strategic sophistication are put forward by Mohlin: a Level$_k$ cognitive hierarchy model (see, for instance, Stahl and Wilson (1995)) and a heterogeneous ctitious play model. In the Level$_k$ model, an anchor type, Level$_0$, plays a xed strategy, Level$_k$ best responds to Level$_{k-1}$, and so on. In heterogeneous ctitious play, the lowest cognitive types best responds to the empirical distribution of past play. In general, each higher type best responds to the type below him, and so on.

[15] The agents in the paper interact over a xed, nite set of normal form games.

[16] Social intelligence here is a suite of cognitive abilities encompassing *ToM,* in particular.

[17] For a game theoretic treatment of the MIH, describing the evolution of intelligence as the result of an arms race in memory capacity, see Robson (2003) .

An influential argument concerning the MIH is due to Dawkins and Krebs (1978) and Krebs and Dawkins (1984). They begin with the premise that natural selection will lead animals to manipulate conspecifics, with deceptive communication if necessary. [18] Presumably, an ability to deceive would often enhance the communicator's reproductive chances. However, it would also set in motion selective pressure for a strategy that anticipates and detects deception. Here we find a compelling rationale for the evolution of an ability to discriminate between others' states of mind. An individual with such an ability can extract true intent from possibly deceptive superficial behavior, use such information to better predict the protagonist's behavior, and then use these predictions to optimize his own choices. [19]

## 6. Conclusions | Complexity and the Hierarchy of Beliefs | Experimental Implementation

One interesting aspect of the model here deserves emphasis. This concerns how the problem facing players becomes more complex in a game of perfect information with a larger number of stages, $I$. In the conventional account, the greater complexity arises, for example, from the need for player 1's to know that player 2's know that ... player $I$'s payoffs. In the current model of learning about others' preferences, when a player makes a choice that reveals his preference over some pair of choices, this is common knowledge. The only greater complexity that does arise with more stages is that player 1, for example, must learn the preferences of players $2, \ldots, I$. Although this is clearly a harder and harder problem for player 1, as $I$ increases, this difficulty is not reflected in the value of . That is, if $> 2$, player 1 can learn player $I$'s binary preferences. Given these preferences, and still just with $> 2$, player 1 can learn player $I - 1$'s preferences. Given these two sets of preferences, and still just with $> 2$, player 1 can learn $I - 2$'s preferences, and so on. This model does not directly then reinforce a notion of complexity based on the level reached in the common knowledge hierarchy; rather the complexity is merely the more prosaic issue of identifying the preferences of the later players, perhaps by backwards recursion.

It would be of inherent interest to experimentally implement a version of the model here, perhaps simplified to have no innovation. That is, put a reasonably large number of subjects into each of $I$ pools—one for each role in the game. Induce the same preferences over a large set of outcomes for each of the player $i$'s for $i = 1, \ldots, I$ by using monetary payoffs. No player knows the other players' payoffs. Play the game otherwise as above. How fast would players learn other players' preferences? Would they be closer to the sophisticated *ToM* types described above or to the naive types? How would the number of stages $I$ affect matters?

---

[18] Previously, the accepted view was that communication among animals provided honest information | a position they attacked because it suggests that behavior could evolve because it bene ted others.

[19] An alternative | but not mutually exclusive | hypothesis is that intelligence derived from ecological complexity (Robson and Kaplan (2003), for example). To the extent that cognitive abilities suited to environmental complexity also enhance an individual's social intelligence, this ecological hypothesis could illuminate the emergence of *ToM*. Strategic intelligence applies a suite of cognitive skills that might well have rst arisen to solve non-social problems | memory, attention, capacities for counterfactual and causal reasoning, etc.

## 7. Appendix

The proofs of Theorems 1 and 2 are given in this appendix. We focus initially on the cases where the fraction of revealed information is negligible in the limit ($L_t^i; L_t^N \longrightarrow 0$). After this, the positive claims about learning are established (Section 7.2). In all of the following fix an environment $\mathscr{E} = (I; A; (\succsim^1; :::; \succsim^I); \{t_k\}_{k=1}^{\infty})$ with an arrival date sequence given by $t_k = \lfloor [|Z_0| + k]^{\alpha} \rfloor$ for some $\quad \mathbb{R}_+:$

### 7.1. No learning in the limit

It is shown here that if outcomes arrive at too fast a rate, learning cannot occur even when the greatest possible amount of information is revealed in every period.

Lemma 1: *In each of the following convergence is sure.*

*i) Suppose* $\quad [0; 2)$*: Then* $L_t^i \longrightarrow 0$ *for each preference type* $i = 1; :::; I:$

*ii) Suppose there are* $T$ *terminal nodes. If* $\quad [0; T)$*; then* $L_t^N \longrightarrow 0:$

*Proof.* Consider any environment in which the underlying game tree has $T$ terminal nodes. Clearly $K_t^i < t \cdot T$ everywhere. Similarly, since only one game is played in each period, $K_t^N \leq t$ surely. Thus, $L_t^i < T \cdot t = |Z_t|^2$ and $L_t^N \leq t = |Z_t|^T$ surely. Since $|Z_t| = |Z_0| + k$ whenever

$$\lfloor (|Z_0| + k)^{\alpha} \rfloor \leq t < \lfloor (|Z_0| + k + 1)^{\alpha} \rfloor;$$

it follows that $t < (|Z_t| + 1)^{\alpha}:$ Hence

$$L_t^j < T \cdot [|Z_t| + 1]^{\alpha} = |Z_t|^2 \quad \text{and} \quad L_t^N < [|Z_t| + 1]^{\alpha} = |Z_t|^T: \tag{2}$$

This establishes the claim since obviously whenever $\quad < 2;$ for instance, (2) implies $L_t^i \longrightarrow 0$ surely. ■

### 7.2. Results About Learning

Theorem 1 is just a special case of Theorem 2. Thus the remainder of this Appendix is devoted to proving the claims in Theorem 2 about full learning (the $L_t^i$ and $L_t^N$'s converging to one in probability). Attention is restricted to the *ToM* players, and so to the $L_t^i$s. As hypothesized in Theorem 2, when considering $L_t^i$, players $i+1; :::; I$ are taken to be *ToM*. The corresponding claim about the behavioral types goes through with minor changes to the analysis.

We will rely on the auxiliary claims—Propositions 1, 2, and 3—to be stated next. In the subsequent section (7.2.2) we use these to establish our claims about learning. The proofs of the auxiliary claims themselves are deferred to Sections 7.2.3-7.2.5.

### 7.2.1. Auxiliary Results

The first auxiliary result relates $L_t^i$, the fraction of what is known about $\succsim^i$ over pairs of outcomes, to what is commonly known about $\succsim^i$ in relation to $A$-tuples of outcomes. Let $\mathbf{Z}_t$ denote the set of $A$-tuples from $Z_t$ and let $\mathbf{U}_t^i \subseteq \mathbf{Z}_t$ be the $A$-tuples on which the $\succsim^i$-preferred outcome is not commonly known.

**Proposition 1:** *For all $i$ $I$ and $t$*

$$\frac{|\mathbf{U}_t^i|}{|\mathbf{Z}_t|} \geq \left[1 - L_t^i\right]^A - u_t^i;$$

*where $u_t^i \geq 0$ and $u_t^i \longrightarrow 0$; surely.*

The gist of the next result, Proposition 2, is the following. Suppose preference types $i; : : : ; I - 1$ are all *ToM* and that $L_t^{i+1}; : : : ; L_t^I$ each converge to one in probability. Then, in the limit, the probability of revealing new information about $\succsim^i$ is small *only if* the fraction of extant knowledge about preferences, $L_t^i$; is close to one. Indeed, although the probability of revealing new information about $\succsim^i$ is clearly small if $1 - L_t^i$ is small, the converse is not as obviously true.

The following proposition, however, provides an appropriate bound. It decomposes $E(K_{t+1}^i \mid H_t) - K_t^i$ into a factor of $1 - L_t^i$, which accounts for what is yet to be revealed about $\succsim^i$; and a residual term $\;_t^{i\varepsilon}$: The residual stems from two factors that complicate proving that the probability of revealing new information about $\succsim^i$ is small only if $L_t^i$; is close to one. These are (1) $i$-type subgames in which $i$'s choice cannot reveal information about $\succsim^i$ because there is insufficient knowledge about the remaining players' choices and (2) $i$-type subgames with outcomes that are avoided by the remaining opponents, so making it difficult to reveal information about $\succsim^i$.

Proposition 1 will be used in the proof of the following.

**Proposition 2:** *Suppose each of the random variables $L_t^{i+1}; : : : ; L_t^I$ converges to one in probability. Then for each " $[0; 1]$ there exists a random variable $\;_t^{i\varepsilon} \geq 0$; such that*

$$E(K_{t+1} \mid H_t) - K_t \geq {''}^{A^i} \cdot \left[\left[1 - L_t^i\right]^{A^i} - \;_t^{i\varepsilon}\right];$$

*where $\;_t^{i\varepsilon}$ converges in probability to a continuous function, $m^i : [0; 1] \longrightarrow [0; 1]$ such that $\lim_{\varepsilon \longrightarrow 0} m^i('') = 0$:*

The next result is simply that whenever $\;> 2$ the result of Proposition 2 ensures $L_t^i$ converges to one.

**Proposition 3:** *Fix a preference type $i$: Suppose for each " $[0; 1]$ there exists a random variable $\;_t^{i\varepsilon} \geq 0$; such that*

$$E(K_{t+1} \mid H_t) - K_t \geq {''}^{A^i} \cdot \left[\left[1 - L_t^i\right]^{A^i} - \;_t^{i\varepsilon}\right];$$

*where $\;^{i\varepsilon}_{t}$ converges in probability to a continuous function, $m^i : [0;1] \longrightarrow [0;1]$ such that $\lim_{\varepsilon \longrightarrow 0} m^i(") = 0$: Then, if the arrival rate of complexity, ; is greater than two, $L^i_t$ converges to one in probability.*

With Propositions 1-3 we are now ready to establish the Theorem 2 claims about $L^i_t$ converging to one when $> 2$.

### *7.2.2. Proof of the Theorem 2 Claim About $L^i_t \longrightarrow 1$*

The proof proceeds by induction. For the inductive step suppose $L^{i+1}_t; \ldots ; L^I_t$ all converge to one in probability. By Proposition 2, $L^i_t$ satisfies the hypothesis of Proposition 3. This implies $L^i_t \longrightarrow 1$ in probability whenever $> 2$: What is required then is to show that $L^I_t \longrightarrow 1$ in probability when $> 2$: To that end, consider the games where an $A$-tuple from $\mathbf{U}^I_t$ is assigned to each of the type $I$ information sets. The fraction of these games is—since the $I$ type has $A^{I-1}$ information sets—

$$\left[\frac{|\mathbf{U}^I_t|}{|\mathbf{Z}_t|}\right]^{A^{I-1}} :$$

Clearly, in the event such a game is drawn, $I$ type choice, at any of her information sets, reveals new information about her preferences. Hence, since games are uniformly drawn in each period,

$$E(K^I_{t+1} \mid H_t) - K^I_t \geq \left[\frac{|\mathbf{U}^I_t|}{|\mathbf{Z}_t|}\right]^{A^{I-1}} :$$

Proposition 1 then implies, where $u^I_t \geq 0$ converges to zero as $t \longrightarrow \infty$,

$$\begin{aligned} E(K^I_{t+1} \mid H_t) - K^I_t &\geq \left[\left[1 - L^I_t\right]^A - u^I_t\right]^{A^{I-1}} \\ &\geq [1 - L^I_t]^{A^I} - \sum_{s=0}^{A^{I-1}} \binom{A^{I-1}}{s} \left[u^I_t\right]^s : \end{aligned} \tag{3}$$

Hence $E(K^I_{t+1} \mid H_t) - K^I_t$ satisfies the hypothesis of Proposition 3 and therefore $L^I_t \longrightarrow 1$ in probability. This completes the proof.

The remaining sections (7.2.3-7.2.5) are dedicated to proving the auxiliary results.

### *7.2.3. Proof of Proposition 1*

Recall that $(z_1; \ldots ; z_A)$ $\mathbf{U}^i_t \subseteq \mathbf{Z}_t$ whenever the $i$ preferred outcome in $\{z_1; \ldots ; z_A\}$ is not commonly known. There are possibly many different values $|\mathbf{U}^i_t|$ can assume given $K^i_t$ revelations about $\succsim^i$ : To prove the claim we first solve, for any $K$, the problem

$$\min |\mathbf{U}^i_t|; \quad \text{s.t.} \quad K^i_t = K:$$

The bound is obtained by expressing $L_t^i$ and the minimized vale of $|\mathbf{U}_t^i|$ (conditional on $K_t^i = K$) in common terms.

Let us proceed to give a complete description of the above minimization problem. First, let $N_t(z) \in \{1, \ldots, |Z_t|\}$ be the number of outcomes in $Z_t$ that are commonly known to be $\succsim^i$-worse than $z$. We claim that

$$|\mathbf{Z}_t| - |\mathbf{U}_t^i| = \sum_{z \in Z_t} \left[ [N_t(z)]^A - [N_t(z) - 1]^A \right]. \tag{4}$$

To see this fix $z \in Z_t$. There are exactly $[N_t(z)]^A$ elements of $\mathbf{Z}_t$ that have all coordinates in the revealed $\succsim^i$-worse-than set of $z$. Subtract the number of $A$-tuples with every coordinate strictly worse than $z$ (thus $z$ is not itself present) to obtain

$$[N_t(z)]^A - [N_t(z) - 1]^A.$$

This is the number of $A$-tuples, $(z_1, \ldots, z_A)$, where at least one $z_r = z$ and moreover it is commonly known that $z \succsim^i z_s$ for each $s \in \{1, \ldots, A\}$. Summing these last indented expressions over $Z_t$ gives the count of $A$-tuples among whose coordinates, $\{z_1, \ldots, z_A\}$, the $i$ preferred outcome is commonly known. But this is $|\mathbf{Z}_t| - |\mathbf{U}_t^i|$ by definition, which establishes (4).

Now express $K_t^i$ in terms of the $N_t(z)$'s by setting $A = 2$ in (4); this yields

$$K_t^i = \sum_{z \in Z_t} \left[ [N_t(z)]^2 - [N_t(z) - 1]^2 \right] = 2 \cdot \sum_{z \in Z_t} [N_t(z) - 1] + |Z_t|. \tag{5}$$

In the following write $r_t(z)$ for the place of $z$ in the $\succsim^i$ ordering of $Z_t$—worst to best (i.e. $r_t(z) = |Z_t|$ for the $\succsim^i$-best outcome, and so on). Any solution to the above minimization problem must then solve the programming problem: for any $K \in \{|Z_t|, |Z_t| + 2, \ldots, |Z_t|^2\}$,

$$\max_{\{N_t(z)\}} \sum_{z \in Z_t} \left[ [N_t(z)]^A - [N_t(z) - 1]^A \right] \quad \text{subject to}$$

$$2 \cdot \sum_{z \in Z_t} [N_t(z) - 1] + |Z_t| = K, \tag{6}$$

$$N_t(z) \in \{1, 2, \ldots, r_t(z)\} \text{ for each } z \in Z_t.$$

In the remainder $\{N_t^K(z)\}_{z \in Z_t}$ will denote a solution to this problem.

Each summed term in the objective function is convex and strictly increasing in the $N_t(z)$'s. Hence, the solution to (6) must be at a corner in the sense that all but at most one $z \in Z_t$ has $N_t^K(z) \in \{1, r_t(z)\}$. Convexity also implies there is always a solution satisfying $N_t^K(z) \geq N_t^K(z')$ whenever $z \succ^i z'$. That is, we can proceed by filling the $N_t^K(z)$'s from the best outcome on downward, until exhausting the constraint $K_t^i = K$: Precisely, for any fixed $K$ one solution is given by

$$N_t^K(z) = \begin{cases} 1; & \text{if} \quad r_t(z) < r_K; \\ M_K; & \text{if} \quad r_t(z) = r_K; \\ r_t(z); & \text{if} \quad r_t(z) > r_K; \end{cases} \tag{7}$$

for some $r_K$ and $M_K$, where $r_K \quad \{1;:::;|Z_t|\}$ and $M_K \quad \{1;:::;r_K\}$: [20]

Suppose

$$K > 2 \cdot [|Z_t| - 1] + |Z_t|:$$

Using our solution from (7) in (4) and (5) (see Footnote 20 also), we obtain

$$|\mathbf{U}_t^i| \geq |\mathbf{Z}_t| - \left[ \sum_{s=r_K+1}^{|Z_t|} \left[ s^A - [s-1]^A \right] + [M_K]^A - [M_K - 1]^A + [r_K - 1] \right]$$

and

$$K = \sum_{r_K+1}^{|Z_t|} \left[ s^2 - [s-1]^2 \right] + [M_K]^2 - [M_K - 1]^2 + [r_K - 1]:$$

All but the first and last terms of the terms under the summation cancel, yielding,

$$|\mathbf{U}_t^i| \geq |\mathbf{Z}_t| - \left[ |Z_t|^A - r_K^A + [M_K]^A - [M_K - 1]^A + [r_K - 1] \right];$$

and

$$K = |Z_t|^2 - r_K^2 + [M_K]^2 - [M_K - 1]^2 + r_K - 1:$$

These last two expressions give, after some algebra,

$$\frac{|\mathbf{U}_t^i|}{|\mathbf{Z}_t|} \geq \left[ \frac{r_K}{|Z_t|} \right]^A - \left[ [M_K]^A - [M_K - 1]^A + r_K - 1 \right] = |\mathbf{Z}_t| \tag{8}$$

and

$$1 - L_t^i = \left[ \frac{r_K}{|Z_t|} \right]^2 - \left[ [M_K]^2 - [M_K - 1]^2 + r_K - 1 \right] = |Z_t|^2:$$

---

[20] For instance, if $K \leq 2 \cdot [|Z_t| - 1] + |Z_t|$, then $r_K = |Z_t|$, and $M_K$ solves $K = 2 \cdot [M_K - 1] + |Z_t|$. That is, $N_t^K(z) = 1$ for all but the $\succsim^i$-best outcome in $Z_t$, and $N_t^K(z') = M_K$ for $z'$, the $\succsim^i$-best outcome in $Z_t$. On the other hand, if $K > 2 \cdot [|Z_t| - 1] + |Z_t|$, then $r_K$ and $M_K$ are obtained from

$$2 \cdot \sum_{s=r_K+1}^{|Z_t|} (s-1) + 2 \cdot (M_K - 1) + |Z_t| = K.$$

In the latter case there is exactly one $r_K \in \{1, \ldots, |Z_t| - 1\}$ and a corresponding $M_K \in \{1, \ldots, r_K\}$ satisfying this equation.

Clearly now,

$$\left[\frac{r_K}{|Z_t|}\right]^2 \geq 1 - L_t^i;$$

and hence

$$\left[\frac{r_K}{|Z_t|}\right]^A > \left[1 - L_t^i\right]^A :$$

Using this in (8) gives

$$\frac{|\mathbf{U}_t^i|}{|\mathbf{Z}_t|} \geq \left[1 - L_t^i\right]^A - \left[[M_K]^A - [M_K - 1]^A + r_K - 1\right] = |\mathbf{Z}_t| : \tag{9}$$

Now consider the case in which

$$K \leq 2 \cdot [|Z_t| - 1] + |Z_t| :$$

Then, the solution described in (7) has $r_K = |Z_t|$. Using this in (4) and (5) yields, after some algebra (again refer to Footnote 20),

$$\frac{|\mathbf{U}_t^i|}{|\mathbf{Z}_t|} \geq 1 - \left[[M_K]^A - [M_K - 1]^A + [|Z_t| - 1]\right] = |\mathbf{Z}_t|$$

and

$$1 - L_t^i = 1 - \left[[M_K]^2 - [M_K - 1]^2 + |Z_t| - 1\right] = |Z_t|^2 :$$

Then

$$\frac{|\mathbf{U}_t^i|}{|\mathbf{Z}_t|} - \left[1 - L_t^i\right]^A \geq \frac{|\mathbf{U}_t^i|}{|\mathbf{Z}_t|} - [1 - L_t^i]$$

$$\geq \frac{[M_K]^2 - [M_K - 1]^2 + |Z_t| - 1}{|Z_t|^2} - \frac{[M_K]^A - [M_K - 1]^A + |Z_t| - 1}{|Z_t|^A} \tag{10}$$

$$\geq - \frac{[M_K]^A - [M_K - 1]^A + |Z_t| - 1}{|Z_t|^A} :$$

Since $M_K \leq |Z_t|$ this means

$$\frac{|\mathbf{U}_t^i|}{|\mathbf{Z}_t|} \geq \left[1 - L_t^i\right]^A - \frac{|Z_t|^A - [|Z_t| - 1]^A + |Z_t| - 1}{|\mathbf{Z}_t|} :$$

Setting $u_t^i$ from the statement of the proposition to

$$\left[|Z_t|^A - [|Z_t| - 1]^A + r_K - 1\right] = |\mathbf{Z}_t|$$

establishes the desired result since $u_t^i$ defined this way converges to zero surely as $|Z_t|$ tends to infinity.

### 7.3. Proof of Proposition 2

In the following fix a preference type $i$. We forgo writing the $i$ superscripts whenever no confusion will result.

Let $P_t^{i+1}(z)$ be the fraction of $i+1$ subgames for which it is common knowledge that outcome $z$ obtains whenever play reaches : We first establish

$$E(K_{t+1}\,|\,H_t) - K_t \geq \left[\sum_{\mathbf{z}\,\epsilon\,\mathbf{U}_t^i} \left[P_t^{i+1}(z_1)\cdot P_t^{i+1}(z_2)\right]\cdot\ \cdots\ \cdot P_t^{i+1}(z_A)\right]^{A^{i-1}} : \qquad (11)$$

To verify this, first recall that $\mathbf{U}_t^i \subseteq \mathbf{Z}_t$ are the $A$-tuples on which the $\succsim^i$-preferred outcome is not commonly known. Then observe that

$$P_t^{i+1}(z_1)\cdot P_t^{i+1}(z_2)\cdot\ \cdots\ \cdot P_t^{i+1}(z_A) \qquad (12)$$

is the fraction of $i$ subgames in which the type $i$'s can uniquely and correctly predict (and this is common knowledge) the $A$ continuation outcomes, $(z_1; : : : ; z_A);$ corresponding to each of their $A$ moves. In the event a positive mass of $i$ types reach such a subgame, and if additionally $(z_1; : : : ; z_A)$ $\mathbf{U}_t^i$, then $i$'s choice there reveals new information about $\succsim^i$ : Thus summing (12) over $\mathbf{U}_t^i$ gives the fraction of $i$ player subgames in which $i$ choice yields a revelation about $\succsim^i$. Since the $i$ types have $A^{i-1}$ information sets and the subgames are uniformly and independently drawn, the probability of drawing a period game where new information is revealed by the $i$'s is at least

$$\left[\sum_{\mathbf{z}\,\epsilon\,\mathbf{U}_t^i} \left[P_t^{i+1}(z_1)\cdot P_t^{i+1}(z_2)\cdot\ \cdots\ \cdot P_t^{i+1}(z_A)\right]\right]^{A^{i-1}} :$$

This establishes (11).

Next, define for each $" > 0;$

$$\mathbf{S}_t^{\varepsilon} = \left\{\mathbf{z}\ \ \mathbf{Z}_t : P_t^{i+1}(z_1)\cdots P_t^{i+1}(z_A) < "^A = |\mathbf{Z}_t|\right\} :$$

Using this in (11), we obtain

$$E(K_{t+1}\,|\,H_t) - K_t$$

$$\geq \left[\sum_{\mathbf{z}\,\epsilon\,\mathbf{U}_t^i\setminus\mathbf{S}_t^{\varepsilon}} \left[P_t^{i+1}(z_1)\cdots P_t^{i+1}(z_A)\right] + \sum_{\mathbf{z}\,\epsilon\,\mathbf{S}_t^{\varepsilon}} \left[P_t^{i+1}(z_1)\cdots P_t^{i+1}(z_A)\right]\right]^{A^{i-1}}$$

$$\geq \left[\sum_{\mathbf{z}\,\epsilon\,\mathbf{U}_t^i\setminus\mathbf{S}_t^{\varepsilon}} \left[P_t^{i+1}(z_1)\cdots P_t^{i+1}(z_A)\right]\right]^{A^{i-1}}$$

$$\geq \left[ \frac{"A}{|\mathbf{Z}_t|} \cdot |\mathbf{U}_t^i \setminus \mathbf{S}_t^{\varepsilon}| \right]^{A^{i-1}} \tag{13}$$

$$\geq \ "A^i \cdot \left[ \frac{|\mathbf{U}_t^i|}{|\mathbf{Z}_t|} - \frac{|\mathbf{S}_t^{\varepsilon}|}{|\mathbf{Z}_t|} \right]^{A^{i-1}} :$$

A binomial expansion of the last above line yields

$$E(K_{t+1} \,|\, H_t) - K_t \geq "A^i \left[ \left[ \frac{|\mathbf{U}_t^i|}{|\mathbf{Z}_t|} \right]^{A^{i-1}} - \sum_{r=1}^{A^{i-1}} \binom{A^{i-1}}{r} \left[ \frac{|\mathbf{S}_t^{\varepsilon}|}{|\mathbf{Z}_t|} \right]^r \right] :$$ [21]

Proposition 3 then implies

$$\left[ \frac{|\mathbf{U}_t^i|}{|\mathbf{Z}_t|} \right]^{A^{i-1}} \geq \left[ [1 - L_t]^A - u_t^i \right]^{A^{i-1}} ;$$

where $u_t^i \geq 0$ and $u_t^i \longrightarrow 0$: Another application of the binomial theorem gives

$$\left[ \frac{|\mathbf{U}_t^i|}{|\mathbf{Z}_t|} \right]^{A^{i-1}} \geq [1 - L_t]^{A^i} - \tilde{u}_t^i;$$

for some $\tilde{u}_t \geq 0$, where again $\tilde{u}_t \longrightarrow 0$: Thus, for all $" > 0$

$$E(K_{t+1} \,|\, H_t) - K_t \ \geq \ "A^i \cdot \left[ [1 - L_t]^{A^i} - \tilde{u}_t - \sum_{r=1}^{A^{i-1}} \binom{A^{i-1}}{r} \left[ \frac{|\mathbf{S}_t^{\varepsilon}|}{|\mathbf{Z}_t|} \right]^r \right] ; \tag{14}$$

where $\tilde{u}_t \longrightarrow 0$ surely.

What remains is to show that the $|\mathbf{S}_t^{\varepsilon}|=|\mathbf{Z}_t|$ terms can be made small (in probability) with the appropriate choice of $"$: We show this in two parts, where Part 1 is used in the proof of Part 2.

**Part 1:** *Let $N_t^j(z)$ be the number of outcomes that have been revealed $\succsim^j$-worse* [22] *than z. Then*

$$P_t^{i+1}(z) \geq \frac{1}{|Z_t|} \left[ \frac{N_t^{i+1}(z)}{|Z_t|} \times \cdots \times \frac{N_t^I(z)}{|Z_t|} \right]^{A^I} : \tag{15}$$

[21] Recall, $(y-x)^N = \sum_{r=0}^{N} \binom{N}{r} y^{N-r} \cdot [-x]^r$. Thus, whenever $0 \leq y \leq 1$, it follows that $(y-x)^N \geq y^N - \sum_{r=1}^{N} \binom{N}{r} x^r$.

[22] *Note that $z \succsim^i z$, in particular.*

**Proof of Part 1:** We will show that the right-hand-side of (15) is a lower bound on the fraction of $i+1$ subgames where it is common knowledge that all the remaining players have dominant acts resulting in outcome $z:$

To that end, consider an $i+1$ player subgame constructed in the following way. Fix, for the $i+1$ types, an action $a^{i+1} \ A$ and assign outcomes that are commonly known to be $\succsim^{i+1}$-worse than $z$ to any terminal node reachable whenever these players choose $a \neq a^{i+1}:$ Since there are $T^{i+1}$ such terminal nodes, there are $[N_t^{i+1}(z)]^{T^{i+1}}$ ways to do this. Then, fix an action for the $i+2$ players, $a^{i+2} \ A;$ and assign outcomes that have been revealed $\succsim^{i+2}$-worse than $z$ to any terminal node attainable when the $i+1$ types choose $a^{i+1}$ and the $i+2$ types subsequently choose $a \neq a^{i+2}:$ With $T^{i+2}$ of these end nodes, there are $[N_t^{i+2}(z)]^{T^{i+2}}$ ways of doing this. Proceed this way, obtaining a sequence of moves $(a^{i+1};:::;a^{I-1});$ until reaching player $I:$ There, fix an action $a^I:$ Assign $z$ to the terminal node reached by the sequence $(a^{i+1};:::;a^{I-1};a^I)$ and $A-1$ outcomes that are known to be $\succsim^{I}$-worse than $z$ to the remaining $T^I = A-1$ end nodes. There are $[N_t^I(z)]^{T_I}$ ways to do this.

In the event that $i+1$ types reach any subgame constructed in this way, every *ToM* player must assign probability one to the $z$ outcome obtaining—irrespective of the period game in which the subgame is embedded. Given the fixed sequence of actions, $a^{i+1}\cdots a^I;$ the fraction of $i+1$ subgames that can be constructed in this way is

$$\frac{1}{|Z_t|}\left[\frac{N_t^{i+1}(z)}{|Z_t|}\right]^{T^{i+1}}\left[\frac{N_t^{i+2}(z)}{|Z_t|}\right]^{T^{i+2}}\cdots\left[\frac{N_t^{I}(z)}{|Z_t|}\right]^{T^{I}}:$$

Observe next that there are $A^{I-(j+1)}$ terminal nodes in a $j+1$ subgame. Therefore $T^j < A^I$ and the last indented expression is greater than

$$\frac{1}{|Z_t|}\left[\frac{N_t^{i+1}(z)}{|Z_t|}\times\cdots\times\frac{N_t^{I}(z)}{|Z_t|}\right]^{A^I}:$$

This establishes what was claimed since whenever play reaches one of the above described subgames it is common knowledge that the sequence of actions $a^{i+1}\cdots a^I$ obtains, which by construction yields outcome $z:$

**Part 2:** *Suppose each of the random variables, $L_t^{i+1};L_t^{i+2};:::;L_t^I;$ converges to one in probability. Then for each $" \ [0;1]$ there exists a random variable $^{\varepsilon}_{t}$ such that*

$$\frac{|\mathbf{S}_t^{\varepsilon}|}{|\mathbf{Z}_t|} < \ ^{\varepsilon}_{t}; \tag{16}$$

*where $^{\varepsilon}_{t}$ converges in probability to a continuous function, $f:[0;1] \longrightarrow [0;1];$ with $\lim_{\varepsilon\longrightarrow 0} f(") = 0:$*

**Proof of Part 2:** First, we express $|\mathbf{S}_t^\varepsilon|/|\mathbf{Z}_t|$ in terms of the simpler sets

$$S_t^\varepsilon = \left\{z \in Z_t : P_t^{i+1}(z) < \varepsilon/|Z_t|\right\}.$$

To that end, observe that the fraction of $A$-tuples, $(z_1, \ldots, z_a, \ldots, z_A),$ having the property, $P_t^{i+1}(z_a) \geq \varepsilon/|Z_t|$ for every $s = 1, \ldots, A,$ is

$$\left[1 - \frac{|S_t^\varepsilon|}{|Z_t|}\right]^A.$$

Then, since every $(z_1, \ldots, z_A)$ in $\mathbf{S}_t^\varepsilon$ has at least one coordinate, $z_a,$ with $P_t^{i+1}(z_a) < \varepsilon/|Z_t|,$ it follows that

$$\frac{|\mathbf{S}_t^\varepsilon|}{|\mathbf{Z}_t|} \leq 1 - \left[1 - \frac{|S_t^\varepsilon|}{|Z_t|}\right]^A. \tag{17}$$

Next, let $r_t^j(z)$ denote the place of $z$ in the $\succsim^j$ ranking of $Z_t$ (worst to best, i.e., $r_t^j(z)$ is the actual number of outcomes in $Z_t$ that are $\succsim^j$-worse than $z$). Recall also that $N_t^j(z)$ is the number of outcomes in $Z_t$ that have been revealed $\succsim^j$-worse than $z$. Writing

$$\frac{N_t^j(z)}{|Z_t|} = \frac{r_t^j(z)}{|Z_t|} \cdot \frac{N_t^j(z)}{r_t^j(z)},$$

the result of Part 1, (15), means that $z \in S_t^\varepsilon$ implies

$$\begin{aligned} &\frac{N_t^j(z)}{r_t^j(z)} < \varepsilon^{1/2A^I} \quad \text{for some } j \in \{i+1, \ldots, I,\} \quad \text{or} \\ &\frac{r_t^j(z)}{|Z_t|} < \varepsilon^{1/2A^I} \quad \text{for some } j \in \{i+1, \ldots, I\}. \end{aligned} \tag{18}$$

In other words,

$$\begin{aligned} S_t^\varepsilon \subseteq &\left[\bigcup_{j=i+1}^{I} \left\{z \in Z_t : N_t^j(z)/r_t^j(z) < \varepsilon^{1/2A^I}\right\}\right] \\ &\bigcup \left[\bigcup_{j=i+1}^{I} \left\{z \in Z_t : r_t^j(z)/|Z_t| < \varepsilon^{1/2A^I}\right\}\right] \end{aligned}$$

and therefore,

$$\begin{aligned} \frac{|S_t^\varepsilon|}{|Zt|} \leq &\frac{1}{|Z_t|} \sum_{j=i+1}^{I} \left|\left\{z \in Z_t : N_t^j(z)/r_t^j(z) < \varepsilon^{1/2A^I}\right\}\right| \\ &+ \frac{1}{|Z_t|} \sum_{j=i+1}^{I} \left|\left\{z \in Z_t : r_t^j(z)/|Z_t| < \varepsilon^{1/2A^I}\right\}\right|. \end{aligned} \tag{19}$$

We now show that under the hypotheses of Part 2,

$$\frac{1}{|Z_t|}\sum_{j=i+1}^{I}\left|\left\{z \quad Z_t \,:\, N_t^j(z){=}r_t^j(z) < {}^{\prime\prime 1/2A^I}\right\}\right| \longrightarrow 0$$

in probability. It suffices to show that for each $j = i+1; : : : ; I$ and for all $\quad > 0$

$$\frac{1}{|Z_t|}\left|\left\{z \quad Z_t \,:\, N_t^j(z){=}r_t^j(z) < 1 - \quad \right\}\right| \longrightarrow 0$$

in probability whenever $L_t^j$ converges to 1 in probability. With that in mind, recall that $N_t^j(z)$ is the number of outcomes that have been revealed $\succsim^j$-worse than $z$. Observe that since $K_t^j$ is the number of pairs on which $\succsim^j$ is commonly known,

$$K_t^j = 2 \cdot \sum_{z \in Z_t} N_t^j(z) - |Z_t|: \tag{20}$$

Now, let $Z_t^{j\eta} \subseteq Z_t$ be the set of outcomes for which $N_t^j(z){=}r_t^j(z) < 1 - \quad :$ [23] Using (20) we have

$$\begin{aligned} K_t^j &= 2 \cdot \left[\sum_{z \in Z_t \setminus Z_t^{j\eta}} N_t^j(z) + \sum_{z \in Z_t^{j\eta}} N_t^j(z)\right] - |Z_t| \\ &< 2 \cdot \left[\sum_{z \in Z_t \setminus Z_t^{j\eta}} r_t^j(z) + (1 - \quad) \sum_{z \in Z_t^{j\eta}} r_t^j(z)\right] - |Z_t| \\ &= 2 \cdot \left[\sum_{z \in Z_t} r_t^j(z) - \quad \cdot \sum_{z \in Z_t^{j\eta}} r_t^j(z)\right] - |Z_t| \end{aligned} \tag{21}$$

By assumption (A1),

$$\sum_{z \in Z_t} r_t^j(z) = \sum_{s=1}^{|Z_t|} s = |Z_t| \cdot [|Z_t| - 1]{=}2:$$

[23] Observe that for each $j, z$ and $t$, $\frac{N_t^j(z)}{r_t^j(z)} \geq \frac{1}{r_t^j(z)} \geq \frac{1}{|Z_t|}$. If $1 - \eta < 1/|Z_t|$ for some $t$, then each $Z_t^{j\eta}$ will be empty. However for su ciently large $\tau$, for each $t \geq \tau$, it must be the case that every $Z_t^{j\eta}$ is empty only if $N_t^j(z) = r_t^j(z)$.

Using this in the last line of (21) it follows that

$$K_t^j < |Z_t|^2 - 2 \cdot \quad \cdot \sum_{z \epsilon Z_t^{j\eta}} r_t^j(z):$$

Then, the following bound is obtained by putting in the above expression the $\succsim^j$-worst outcomes (lowest $r_t^j(z)$) in $Z_t^{j\eta}$:

$$K_t^j < |Z_t|^2 - 2 \cdot \quad \cdot \sum_{s=1}^{|Z_t^{j\eta}|} s = |Z_t|^2 - \quad \cdot |Z_t^{j\eta}| \cdot [|Z_t^{j\eta}| - 1]:$$

Divide both sides of this last expression by $|Z_t|^2$ to obtain

$$L_t^j < 1 - \quad \cdot \frac{|Z_t^{j\eta}|\{|Z_t^{j\eta}| - 1\}}{|Z_t|^2}:$$

Clearly now, if $L_t^j$ converges to one in probability then $|Z_t^{j\eta}|=|Z_t|$ must converge to zero in probability, for all $\quad > 0$: We have thus established that for all sufficiently small $" > 0$

$$\frac{1}{|Z_t|} \sum_{j=i+1}^{I} \left| \left\{ z \quad Z_t : N_t^j(z)=r_t^j(z) < "^{1/2A^I} \right\} \right| \longrightarrow 0$$

in probability, i.e., choose $\quad$ so that $1 - \quad = "^{1/2A^I}$, the summands in the above expression are then just $|Z_t^{j\eta}|=|Z_t|$.

Next (refer to (19)), we have for each $j = i + 1; : : : ; I$

$$\left| \left\{ z \quad Z_t : r_t^j(z)=|Z_t| < "^{1/2A^I} \right\} \right| < |Z_t| \cdot "^{1/2A^I}:$$ [24]

Obviously then,

$$\frac{1}{|Z_t|} \sum_{j=i+1}^{I} \left| \left\{ z \quad Z_t : r_t^j(z)=|Z_t| < "^{1/2A^I} \right\} \right| < [I - i] \cdot "^{1/2A^I}:$$

In sum, letting

$$\hat{}^{\varepsilon}_t = \frac{1}{|Z_t|} \sum_{j=i+1}^{I} \left| \left\{ z \quad Z_t : N_t^j(z)=r_t^j(z) < "^{1/2A^I} \right\} \right| + [I - i] \cdot "^{1/2A^I}; \qquad (22)$$

[24] We have used here that $r_t^j(z) = r_t^j(z')$ if and only if $z = z'$, which implies there are at most $\lfloor |Z_t| \cdot \varepsilon^{1/2A^I} \rfloor$ outcomes in $Z_t$ with $r_t^j(z)/|Z_t| < \varepsilon^{1/2A^I}$.

we have established, using (17) and (19), that

$$\frac{|\mathbf{S}_t^\varepsilon|}{|\mathbf{Z}_t|} \leq 1 - \left[1 - \frac{|S_t^\varepsilon|}{|Z_t|}\right]^A < 1 - \left[1 - \hat{}_t^\varepsilon\right]^A;$$

where $\hat{}_t^\varepsilon \longrightarrow [I - i] \cdot {}''^{1/2A^I}$ in probability. The proof of Part 2 is completed by setting

$$_t^\varepsilon = 1 - \left[1 - \hat{}_t^\varepsilon\right]^A$$

and by defining

$$f('') = 1 - \left[1 - [I - i] \cdot {}''^{1/2A^I}\right]^A$$

Recalling (14), in particular, Parts 1 and 2 yield the proof of Proposition 2. Specifically, let $_t^{i\varepsilon}$ and $m^i$ from the statement of Proposition 2 be

$$_t^{i\varepsilon} = \tilde{u}_t + \sum_{r=1}^{A^{i-1}} \binom{A^{i-1}}{r} [\,_t^\varepsilon]^r$$

and

$$m^i('') = \sum_{r=1}^{A^{i-1}} \binom{A^{i-1}}{r} [f('')]^r;$$

where $_t^\varepsilon$ and $f$ come from the proof of Part 2 above.

### 7.3.1. Proof of Proposition 3

The proof is given in three parts. The first two parts together show that the $L_t$ processes converge in probability to some random variable $L$: In the third part it is shown that $L$ must equal one a.e. In order to establish the convergence of the $L_t$'s we show they belong to a class of generalized martingales with the martingale convergence property. We make use of the following definition and result in this connection [(Egghe, 1984), Definition VIII.1.3 and Theorem VIII.1.22].

w-submil Convergence: *The adapted process $(L_t; H_t)$ is a* ***weak sub-martingale in the limit (w-submil)*** *if almost surely, for each $\;> 0$, there is a $T$ such that $\;\geq t \geq T$ implies $P\{E(L_\tau \mid H_t) - L_t \geq -\;\} > 1 -\;$: If $L_t$ is a w-submil, then there exists a random variable $L$ such that $L_t \longrightarrow L$ in probability.* [25]

For the remainder of this section fix a preference type $i$ and assume the hypotheses of Proposition 2, and also that $\;> 2$:

**Part 1:** *For each $'' > 0$ there exists a random variable $_t^\varepsilon$ such that the following is true. For each pair of consecutive arrival dates $t^* <\; ^*$,*

$$E(L_{\tau^*} \mid H_{t^*}) - L_{t^*} < 0 \implies E(L_{\tau^*} \mid H_{t^*}) > \;_{t^*}^\varepsilon;$$

[25] *We suppress the $i$'s from now on.*

*where* ${}^{\varepsilon}_{t} \longrightarrow 1 - m(\varepsilon)^{\frac{1}{A^i}}$ *in probability, and* $m = m^i$*; is the function described in the statement of Proposition 3.*

**Proof of Part 1:** Write $Q_t = |Z_t|^2$ so that $L_t = K_t / Q_t$. Fix a pair of consecutive arrival dates $t^*$ and $\tau^*$. Then

$$
\begin{aligned}
L_{\tau^*} - L_{t^*} &= \frac{K_{\tau^*} - K_{t^*}}{Q_{\tau^*}} - \frac{Q_{\tau^*} - Q_{t^*}}{Q_{\tau^*}} \cdot L_{t^*} \\
&\geq \sum_{t=t^*}^{\tau^*-1} \left[K_{t+1} - K_t\right] / Q_{\tau^*} - \frac{Q_{\tau^*} - Q_{t^*}}{Q_{\tau^*}}.
\end{aligned}
\tag{23}
$$

Hence,

$$
\begin{aligned}
& E(L_{\tau^*} \mid H_{t^*}) - L_{t^*} < 0 \implies \\
& \sum_{t=t^*}^{\tau^*-1} E(K_{t+1} - K_t \mid H_{t^*}) < Q_{\tau^*} - Q_{t^*}.
\end{aligned}
\tag{24}
$$

By the hypothesis of Proposition 2, for all $\varepsilon > 0$,

$$
\sum_{t=t^*}^{\tau^*-1} E(K_{t+1} - K_t \mid H_{t^*}) \geq \varepsilon^{A^i} \cdot \sum_{t=t^*}^{\tau^*-1} E([1 - L_{t^*}]^{A^i} - {}^{\varepsilon}_{t} \mid H_{t^*}).
$$

Then, since $L_t$ is non-decreasing between arrival dates,

$$
\begin{aligned}
& \sum_{t=t^*}^{\tau^*-1} E(K_{t+1} - K_t \mid H_{t^*}) \\
& > \varepsilon^{A^i} \cdot \left[ [\tau^* - t^*] \cdot E([1 - L_{\tau^*-1}]^{A^i} \mid H_{t^*}) - \sum_{t=t^*}^{\tau^*-1} E({}^{\varepsilon}_{t} \mid H_{t^*}) \right].
\end{aligned}
\tag{25}
$$

Using this in equation (24), and invoking Jensen's inequality, yields after some algebra:

$$
\begin{aligned}
& \text{If} \quad E(L_{\tau^*} \mid H_{t^*}) - L_{t^*} < 0, \quad \text{then} \\
& E(L_{\tau^*-1} \mid H_{t^*}) > 1 - \left[ \left[\frac{1}{\varepsilon}\right]^{A^i} \cdot \frac{Q_{\tau^*} - Q_{t^*}}{\tau^* - t^*} + \frac{1}{\tau^* - t^*} \sum_{t=t^*}^{\tau^*-1} E({}^{\varepsilon}_{t} \mid H_{t^*}) \right]^{\frac{1}{A^i}}.
\end{aligned}
\tag{26}
$$

Since $L_{\tau^*} \geq L_{\tau^*-1} \cdot Q_{t^*} / Q_{\tau^*}$ surely, multiplying the second line of (26) by $Q_{t^*} / Q_{\tau^*}$ gives:

$$
\begin{aligned}
& \text{If} \quad E(L_{\tau^*} \mid H_{t^*}) - L_{t^*} < 0, \quad \text{then} \\
& E(L_{\tau^*} \mid H_{t^*}) > \\
& \frac{Q_{t^*}}{Q_{\tau^*}} \cdot \left[ 1 - \left[ \left[\frac{1}{\varepsilon}\right]^{A^i} \cdot \frac{Q_{\tau^*} - Q_{t^*}}{\tau^* - t^*} + \frac{1}{\tau^* - t^*} \sum_{t=t^*}^{\tau^*-1} E({}^{\varepsilon}_{t} \mid H_{t^*}) \right]^{\frac{1}{A^i}} \right].
\end{aligned}
\tag{27}
$$

Define $^{\varepsilon}_{t^*}$ as the last line in (27).

To complete the proof of Part 1 we show that

$$^{\varepsilon}_{t^*} \longrightarrow 1 - m(")^{\frac{1}{A^i}}$$

in probability. To verify this limit, first note that if $t^*$ is the arrival date of the $k$-th new outcome then

$$\frac{Q_{\tau^*} - Q_{t^*}}{{}^* - t^*} = \frac{[n+k+1]^2 - [n+k]^2}{\lfloor (n+k+1)^\alpha \rfloor - \lfloor (n+k)^\alpha \rfloor};$$

which converges to zero when $> 2$: Obviously $Q_{t^*}=Q_{\tau^*} \longrightarrow 1$; and finally

$$\frac{1}{{}^* - t^*} \sum_{t=t^*}^{\star\tau - 1} E( {}^{\varepsilon}_{t} \mid H_{t^*}) \longrightarrow m(")$$

in probability as $t^* \longrightarrow \infty$; since by assumption $^{\varepsilon}_{t}$ converges to $m(")$ in probability and is finite valued everywhere. In view of (27) this completes the proof of Part 1.

We use Part 1 to establish the following.

**Part 2:** *$L_t$ converges in probability to some random variable $L$.*

**Proof of Part 2:** As a notational convenience, let hatted variables denote variables sampled at arrival dates, e.g., $\hat{L}_k = L_{t_k}$. We show first that the arrival date subsequence $\{\hat{L}_k\}$ is a w-submil and thus converges in probability. Specifically, the following will be established. For each $> 0$ there exists an $M$ such that for all arrival dates $t_m$, $t_n$, where $M \leq m < n$,

$$P\left\{E(\hat{L}_n \mid \hat{H}_m) - \hat{L}_m \geq - \right\} > 1 - : \tag{28}$$

To that end, suppose

$$E(\hat{L}_n \mid \hat{H}_m) - \hat{L}_m < 0:$$

Then, since

$$E(\hat{L}_n \mid \hat{H}_m) - \hat{L}_m = \sum_{k=m}^{n-1} E(E(\hat{L}_{k+1} \mid \hat{H}_k) - \hat{L}_k \mid \hat{H}_m);$$

there is at least one arrival date, $t_k$, in this range for which

$$E(\hat{L}_{k+1} \mid \hat{H}_m) - E(\hat{L}_k \mid \hat{H}_m) < 0:$$

Let $r$ be the last arrival in $\{m; ; : : : ; n-1\}$ at which this is true. Using the $^{\varepsilon}_{t}$ defined in Part 1 gives

$$E(\hat{L}_{r+1} \mid \hat{H}_m) \;>\; E(\hat{\ }^{\varepsilon}_{r} \mid \hat{H}_m);$$

and for each $k = r + 1; : : : ; n - 1;$

$$E(\hat{L}_{k+1} \mid \hat{H}_m) - E(\hat{L}_k \mid \hat{H}_m) \;\geq\; 0:$$

Hence,

$$E(\hat{L}_n \mid \hat{H}_m) - \hat{L}_m \;>\; E(\hat{\ }^{\varepsilon}_{r} \mid \hat{H}_m) - \hat{L}_m \;\geq\; E(\hat{\ }^{\varepsilon}_{r} \mid \hat{H}_m) - 1: \tag{29}$$

Now observe that since $\hat{\ }^{\varepsilon}_{k}$ converges to $1 - m('')^{\frac{1}{A^i}}$ in probability, we can choose an arrival $M$ large enough so that

$$P\left\{E(\hat{\ }^{\varepsilon}_{k} \mid \hat{H}_m) - 1 > -2 \cdot m('')^{\frac{1}{A^i}}\right\} > 1 - 2 \cdot m('')^{\frac{1}{A^i}}$$

for all $k$ and $m$ with $k \geq m \geq M:$ Equation (29) then implies that for $k \geq m \geq M$,

$$P\left\{E(\hat{L}_n \mid \hat{H}_m) - \hat{L}_m > -2 \cdot m('')^{\frac{1}{A^i}}\right\} > 1 - 2 \cdot m('')^{\frac{1}{A^i}}:$$

To see that $\{\hat{L}_k\}$ is a w-submil obtain (28) by choosing $''$ so that $m('')^{\frac{1}{A^i}} < \;=2:$

Having established that $\hat{L}_k$ is a w-submil, we proceed to show that $L_t$ is also a w-submil. Consider any dates $t$ and  where $t < \;:$ Then

$$L_\tau - L_t \geq L_{\tau^*} - L_{t^*}\frac{Q_{t^*}}{Q_{\tau^*}}$$

everywhere, when $t^*$ is the first arrival date after $t$ and $^*$ is the greatest arrival date less than or equal to $:$ Since $\{\hat{L}_k\}$ is a w-submil, bounded above by 1, $L_{\tau^*} - L_{t^*} \longrightarrow 0$ in probability, as $t^* \longrightarrow \infty$. Furthermore, $Q_t{=}Q_{t+1} \longrightarrow 1$. Hence the right hand side of the last indented expression converges to zero in probability which establishes that $\{L_t\}$ is a w-submil. In light of the w-submil convergence result stated above, this establishes the claim made in Part 2.

**Part 3:** *$L$; the limiting function of $L_t$, is equal to one a.e.*

**Proof of Part 3:** Let $^{\varepsilon}_{t}$ and $m = m^i$ be as described in the statement of Proposition 3. Then, for all $'' > 0$

$$E\left(L_\tau\right) = \sum_{t=0}^{\tau-1} E\left(K^i_{t+1} - K^i_t\right)/Q_\tau$$

$$\geq ''^{A^i} \cdot \frac{}{Q_\tau} \cdot \left[\frac{1}{}\sum_{t=1}^{\tau-1} \cdot E\left[\left[1 - L^i_t\right]^{A^i} - {}^{\varepsilon}_{t}\right]\right]: \tag{30}$$

Since $=Q_\tau \longrightarrow \infty$ whenever $> 2$ [26] and $L_t$ is everywhere finite, it must be the case that

$$\limsup \left( \frac{1}{} \cdot \sum_{t=1}^{\tau-1} E\left[ \left[1 - L_t^i\right]^{A^i} - \ _t^{\varepsilon} \right] \right) \leq 0; \qquad (31)$$

for each $" > 0$: It is straightforward to show that

$$\lim_{\tau \longrightarrow \infty} \left[ \frac{1}{} \cdot \sum_{t=1}^{\tau-1} E\left[ \left[1 - L_t^i\right]^{A^i} - \ _t^{\varepsilon} \right] \right] = E\left[ [1 - L]^{A^i} - m(") \right]:$$

Hence, (31) yields

$$E\left[ [1 - L]^{A^i} - m(") \right] \leq 0; \text{ for all } " > 0:$$

Since $m(")$ is continuous and tends to zero as $" \longrightarrow 0$; this last indented expression implies

$$E\left[ [1 - L]^{A^i} \right] = 0:$$

Finally, $L \leq 1$ surely, yields $L = 1$ a.e.

*D. Monte*
*email:* daniel_monte@sfu.ca

*N. Robalino*
*email:* nkasimat@sfu.ca

*A. Robson*
*email:* robson@sfu.ca

## References

[1] Byrne, R.W. and A. Whiten (1997): "Machiavellian Intelligence," in *Machiavellian Intelligence II: Extensions and Evaluations*, Whiten, A. and R.W. Byrne (eds.), pages: 1-23. Cambridge University Press: Cambridge, UK.

[2] Byrne, R.W. and Whiten, A. (eds) (1988): *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes and Humans*. Blackwell: Oxford, UK.

[3] Cheney, D.L. and R.M. Seyfarth (1990): *How Monkeys See the World: Inside the Mind of Another Species*. University of Chicago Press: Chicago.

[4] Dawkins, R. and J.R. Krebs (1978): "Animal Signals: Mind Reading and Manipulation," in *Behavioral Ecology: An Evolutionary Approach*, Krebs, J.R. and N.B. Davies (eds.), pages: 282-308. Basil Blackwell: Oxford, UK.

[5] Egghe, L. (1984): *Stopping Time Techniques for Analysts and Probabilists*. Cambridge University Press: Cambridge, UK.

[26] Recall that $|Z_t| = n + k$ whenever $\lfloor (n + k)^\alpha \rfloor \leq t < \lfloor (n + k + 1)^\alpha \rfloor$ therefore $t/Q_t \geq |Z_t|^\alpha / |Z_t|^2$.

[6] Fudenberg, D. and D.M. Kreps (1995): "Learning in Extensive Form Games I. Self-Confirming Equilibria," *Games and Economic Behavior*, **8**, 20–55.

[7] Krebs, J.R. and R. Dawkins (1984): "Animal Signals: Mind Reading and Manipulation," in *Behavioral Ecology: An Evolutionary Approach*, Krebs, J.R. and N.B. Davies (eds.), pages: 380-401. Basil Blackwell: Oxford, UK.

[8] Mohlin, E. (2012): "Evolution of Theories of Mind," Forthcoming in *Games and Economic Behavior*.

[9] Robson, A.J. (2001): "Why Would Nature Give Individuals Utility Functions?," *Journal of Political Economy*, **109**, 900–914.

[10] Robson, A.J. (2003): "The Evolution of Rationality and the Red Queen," *Journal of Economic Theory*, **15**, 177–202.

[11] Robson, A.J. and H. Kaplan (2003): "The Evolution of Human Longevity and Intelligence in Hunter-Gatherer Economies," *American Economic Review*, **93**, 150–169.

[12] Stahl, D.O. (1993): "Evolution of Smart$_n$ Players," *Games and Economic Behavior* , **5**, 604–617.

[13] Stahl, D.O. and P.W. Wilson (1995): "On Players' Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, **10**, 218–254.

[14] Whiten, A. and R.W. Byrne (eds.) (1997): *Machiavellian Intelligence II: Extensions and Evaluations*. Cambridge University Press: Cambridge, UK.