THE EVOLUTION OF 'THEORY OF MIND': THEORY AND EXPERIMENTS

By

Erik O. Kimbrough, Nikolau Robalino and Arthur J. Robson

September 2013
Revised January 2014

COWLES FOUNDATION DISCUSSION PAPER NO. 1907R

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

The Evolution of "Theory of Mind": Theory and Experiments*

Erik O. Kimbrough

Nikolaus Robalino

Arthur J. Robson

January 24, 2014

Abstract. This paper investigates the evolutionary foundation for our capacity to attribute preferences to others. This ability is intrinsic to game theory, and is a key component of "Theory of Mind", perhaps the capstone of social cognition. We argue here that this component of theory of mind allows organisms to efficiently modify their behavior in strategic environments with a persistent element of novelty. Our notion of "Theory of Mind" (*ToM*) yields a sharp, unambiguous advantage over less sophisticated approaches to strategic interaction because agents with *ToM* extrapolate to novel circumstances information about opponents' preferences that was learned previously. We then report on experiments investigating *ToM* in a simpler version of the theoretical model. We find highly significant learning of opponents' preferences, providing strong evidence for the presence of *ToM* as in our model in the subjects. Moreover, scores on standard measures of autism-spectrum behaviors are significant determinants of individual speeds of learning, so our notion of *ToM* is significantly correlated with theory of mind as in psychology.

---

*Affiliation and mailing address for all three authors—Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada. Email addresses—ekimbrough@gmail.com, nkasimat@sfu.ca, robson@sfu.ca, respectively. We thank Bertille Antoine, Ben Greiner, Matt Jackson, Leanna Mitchell, Daniel Monte, Chris Muris, Krishna Pendakur, Luis Rayo, Phil Reny, and Larry Samuelson for helpful discussions. The paper also benefited from comments by participants in seminars at Caltech, Stanford, UCR, UCSB, and Cal Poly, and at conferences sponsored by the Becker-Friedman Institute at the University of Chicago, by the Toulouse School of Economics, by the Max Planck Institute at Ringberg Castle, and by the Economic Science Association in Santa Cruz, CA. Kimbrough and Robson acknowledge support from the SSHRC SFU Small Grants Program; Kimbrough thanks the SSHRC Insight Development Grants Program; Robalino and Robson thank the Human Evolutionary Studies Program at SFU; Robson thanks the Guggenheim Foundation, the Canada Research Chair Program and the SSHRC Insight Grants Program. Some of the analysis was performed using the open source statistical software R (R Development Research Team, 2013).

1. Introduction

An individual with *theory of mind* 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 theory of mind.[1] More specifically, it is conventional in game theory to make the crucial assumption, without much apology, that agents have theory of mind in the sense of imputing preferences to others.

The present paper considers theory of mind in greater depth by addressing the question: *Why* and *how* might this ability to impute preferences to others have evolved? In what types of environments would this ability yield a distinct advantage over alternative, less sophisticated, approaches to strategic interaction? In general terms, the answer we propose is that this aspect of theory of mind 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 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; conversely, if she acts correctly over a sufficiently rich set of gambles, she must possess, although perhaps only implicitly, the appropriate utility function.

We shift attention here to a dynamic model in which players repeatedly interact with one another but in which novelty is repeatedly introduced. More

---

[1] The classic experiment that suggests children have theory of mind is the "Sally-Ann" test described in Baron-Cohen, Leslie, and Frith (1985). According to this test, young children begin to realize that others may have false beliefs shortly after age four. This test relies on children's verbal facility. Onishi and Baillargeon (2005) push the age back to 15 months using a non-verbal technique. Infants are taken to express that their expectations have been violated by lengthening the duration of their gaze. The presence of this capacity in such young individuals increases the likelihood that it is, to some degree at least, innate.

precisely, although the game tree is fixed, the outcomes needed to complete the game are randomly drawn in each period from an outcome set that grows over time. We presume individuals have an appropriate own utility function, but do not know the utility functions of their opponents. The focus is then on the advantage to an agent of conceiving of her opponents as also being agents—in particular, understanding that they act optimally in the light of their preferences and so endeavoring to learn these. Having a template into which the preferences of an opponent can be fitted enables a player to better deal with the innovation that arises from new outcomes than can a "naive type" that adapts to each game as a distinct set of circumstances. In other words, the edge to theory of mind derives from a capacity to extrapolate to novel circumstances information that was learned about others' preferences in a previous situation.

This outlines our dynamic interpretation of the aspect of theory of mind concerning the preferences of others. Our interpretation is in the spirit of revealed preference in that the implications of knowledge of others' preferences are observable. This interpretation exploits the concept of theory of mind more fully and fruitfully than might a static interpretation. Throughout the paper, we refer to our dynamic interpretation of theory of mind, for simplicity, just as *ToM*.

The distinction between the *ToM* and naive types might 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 *ToM* 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 naive reinforcement 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 strategem continued to be met with success, reinforcing the behavior.[2]

Consider the argument in greater detail. We begin by fixing a game tree with perfect information, with stages $i = 1, \ldots, I$. There are I equally large populations, one for each of the associated "player roles." In each period, a large number of random matches are made, with each match having one player in each role $i = 1, \ldots, I$. The outcomes needed to complete the game are drawn randomly and uniformly in each period from the finite outcome set that is available then. Players have preference orderings over the set of outcomes that are ever possible, and so preferences over the finite subset of these that is available in each period. Each player is fully aware of his own ordering but does not directly know the preference ordering of his opponents.

Occasionally, a new outcome is added to the set of potential outcomes, where each new outcome is drawn independently from a given distribution. The number of outcomes available grows to infinity at a parametric rate. The crucial aspect of this model is the introduction of novelty, rather than the growing complexity that is also generated. That is, in a model in which outcomes were also dropped, so the outcome set remained of constant size, similar results obtain, but in a slightly more awkward fashion. We view our strategic environment as a convenient test-bed on which we can derive the speeds with which the various types can learn. The basic results do not seem specific to this particular environment, so these differences in relative learning speeds would likely be manifested in many alternative models.

All players see the complete history of the games played—the outcomes that were chosen to complete the game, the choices that were made by all player roles, but not the payoffs of others. The types of players here differ with respect to the extent and the manner of utilization of this information. We compare two main categories of types of players—naive and theory of mind (ToM) types. The naive types' behavior is inspired by reinforcement learning, as implicit in evolutionary game theory, where they treat each new game as an unfamiliar set of circumstances. The ToM types are disposed to learn others' preferences.

---

[2] It is plausibly the capacity for such that naive learning that is subject to natural selection rather than the precise strategem itself.

They apply the information provided by the history available in each period to build up a detailed picture of the preferences of the other roles. All types are assumed to avail themselves of a dominant choice, whenever this is available. This assumption is in the spirit of focussing on the learning the preferences of others rather than considering the implications of knowing one's own preferences.

The crucial feature of naive types is that they make a (possibly mixed) choice that is the same for all new games. (This assumption can be relaxed as long as naive types behave inappropriately in some *positive fraction* of new games.) This characterization of naive types is in line with "evolutionary game theory," which was inspired, in turn, by the psychological theory of reinforcement learning. It is not crucial otherwise how naive players behave. Indeed, even if the naive types apply a fully Bayesian rational strategy the second time a game is played, they will still lose the evolutionary race here to the *SPE-ToM* type. More reasonable assumptions on the rate of learning for the naive types would only strengthen our results. Furthermore, the results favoring the *SPE-ToM* type hold even if the *ToM* types have a sufficiently small extra fixed cost.

The crucial aspect of *ToM* behavior is that, in the long run, once the history of the game has revealed the preferences of all subsequent players, *ToM* types map these preferences to an action. There is a particular *ToM* type, the *SPE-ToM* type, that maps these preferences to the *SPE* choice for the subgame, when this is unique. This *SPE-ToM* type is shown to evolutionarily dominate the population, in the long run. In the short run, the *ToM* types understand enough about the game that they can learn the preferences of other player roles. For example, it is common knowledge among all *ToM* types that all players use dominant actions, if available. It is not crucial otherwise how the *ToM* types behave—they could even *minimize* their payoffs according to a fully accurate posterior distribution over all the relevant aspects of the game, when the preferences of all subsequent players are not known.

We do not assume that the *ToM* types use the transitivity of opponents' preferences. (Indeed, the results here would apply even if preferences were not transitive.) The *ToM* types build up a description of others' preferences only by observing all the pairwise choices. Generalizing this assumption would only strengthen our results by increasing *ToM* types' learning speed.

Theorem 1 is the basic theoretical result here—in an intermediate range of growth rates of the outcome set, the *ToM* types will learn opponents' preferences with a probability that converges to one, while the naive types see a familiar game with a probability that converges to zero. The greater adaptation of the *ToM* type simply reflects that there are vastly more possible games that can be generated from a given number of outcomes than there are outcome pairs.

There are various ways the *ToM* types might exploit this greater knowledge at the expense of the naive types. We have set up the model to favor a simple and salient possibility, as expressed in the main conceptual result—Theorem 2—that eventually a unique *SPE* is attained, with the *SPE-ToM* type ultimately evolutionarily dominant, over all other *ToM* types, as well as over all the naive types.

A key result is then that it is better to be "smart"—a *ToM*—than it is to be a naive player. Indeed, our results hold if the *ToM*s incur a fixed cost, as long as this cost is small enough. This is important since the previous literature has tended to find an advantage to (lucky and) less smart players over smarter players—see, as a key example, Stahl (1993). The underlying reason for the reverse (and more plausible) result here is that we force individuals to address novel games. More particularly, the assumption that naive players use the same strategy in any novel game disallows a full range of naive types that adopt a full range of strategies conditional on every game. If a full spectrum of such strategies existed, that is, a suitable naive but lucky type would be unbeatable in the long run, and would beat the *ToM*s too, if these involved a cost. But the existence of such a full range of strategies covering novel games does not seem plausible.

After stating the theoretical results, we present experiments on theory of mind that buttress the current approach by allowing us to observe 1) the presence and extent of our revealed preference version of theory of mind in human subjects and 2) the degree to which this dynamic revealed preference interpretation of *ToM* corresponds to theory of mind as it is understood by psychologists. We construct an environment similar to that in the model, but simpler, in which *ToM* yields a distinct strategic advantage, and observe the extent to which our subjects exploit this advantage.

In the experiments, subjects play a sequence of two-player extensive form games where each player role has two moves at each decision node. In each repetition, a game is constructed by drawing outcomes without replacement from a finite set. All players in a given role had the same (induced) preferences, but these players knew only their own payoff at each outcome and not that of their opponent, as is the crucial feature of the theoretical model. We randomly and anonymously paired subjects in each of 90 repetitions to observe the ability of players 1 to learn (and to exploit their knowledge of) the preferences of players 2. As reflects the theoretical model, many games in later periods that would appear novel to a naive reinforcement learner could be understood by an agent with *ToM* who had observed previous choices in the subgames. The rate at which subjects achieve subgame perfect equilibrium outcomes measures the extent to which individuals exhibit *ToM* by learning their opponents' preferences.

At the end of each experimental session, we collected two measures of theory of mind that are commonly used in psychology. Specifically, we asked the students to complete two short Likert scale surveys measuring the extent of autism spectrum behaviors. One was the Autism-Spectrum Quotient (AQ) survey due to Baron-Cohen et al. (2001); the other was the Broad Autism Phenotype Questionnaire (BAP), due to Hurley et al. (2007).

There were two striking results of the experiments that corroborate the present approach. First, we observed highly significant learning of player 2's preferences by players 1, but no such significant learning of specific games. That is, iron-clad support for the formal model of the paper is expressed in real-world behavior. Individuals do behave as if they ascribe preferences to opponents and endeavor to learn these, given that it is advantageous. Not surprisingly, this ability is present in real-world individuals to varying degrees. Second, there is strong evidence that this attribute is an aspect of theory of mind, as this term is understood in psychology: player 1's who report fewer autism-spectrum behaviors (i.e. have lower AQ and BAP scores) have a statistically significant tendency to learn player 2's preferences faster.

2. The Theoretical Model

2.1. **The Environment.**

We begin by defining the underlying games. The extensive game form is a fixed tree with perfect information and a finite number of stages, $I \geqslant 2$ and actions, A, at each decision node.[3]

There is one "player role" for each such stage, $i = 1, \ldots, I$, in the game. (In a reversal of the usual convention, the first player role to move is I and the last to move is 1. This simplifies the notation used in the proof.) Each player role is represented by an equally large population of agents. These agents will have different "types", that differ in their choice of strategy, but not in their payoff function. These types will be described precisely below, but they will be grouped into two broad "categories"—*ToM* and naive.

Independently in each period, all players are randomly and uniformly matched with exactly one player for each role in each of the resulting large number of games.[4]

All that is left to complete the description of the basic game, is the payoff for each player role—the mapping from outcomes to expected offspring. There is a fixed overall set of outcomes, each with consequences for the reproductive success of the I player roles. Player role $i = 1, \ldots, I$ is then characterized by a function mapping outcomes to expected numbers of offspring. A fundamental novelty is that, although each player role knows its own payoff at each outcome, it does not know the payoff for the other player roles.

For notational simplicity, however, we finesse consideration of explicit outcomes and payoff functions from outcomes to expected offspring. Given a fixed tree structure with T terminal nodes, we instead simply identify each outcome with a payoff vector and each game with a particular set of such payoff vectors assigned to the terminal nodes. We assume that all expected offspring payoffs lie in the compact interval $[m, M]$, for $M > m > 0$. The upper bound M is merely a technical convenience; the lower bound m ensures that no type would go extinct if it is temporarily outdone by another type.

---

[3] The restriction that each node induce the same number of actions, A, can readily be relaxed by allowing equivalent moves, in which case A can be interpreted as the *maximum* number of actions available at any node in the entire tree. Indeed, it is possible to allow the game tree to be randomly chosen. This would not fundamentally change the nature of our results but would considerably add to the notation required.

[4] Uniform matching is not crucial to our results but chosen in the interest of simplicity.

A1: *The set of all games is represented by* $Q = [m, M]^{TI}$, *for* $M > m > 0$. *That is, each outcome is a payoff vector in* $Z = [m, M]^{I}$, *with one component for each player role, and there are* T *such outcomes comprising each game.*

Let $t = 1, 2, \ldots,$ denote successive time periods. At date t, there is available a set of outcomes $Z_t \subset Z$, determined in the following way. There is an initial finite set of outcomes $Z_1 \subset Z$ where each of these outcomes is drawn independently from Z according to a cumulative distribution function F as follows.[5]

A2: *The cdf over outcomes* F *has a continuous probability density* f *that is strictly positive on* Z.

There is then a subsequence of time periods $\{t_k\}_{k=1}^{\infty}$. At date t_k, $k = 1, 2, \ldots,$ a k-th outcome is added to the existing ones by drawing it independently from Z according to F.[6] In between arrival dates the set of outcomes is fixed, and once an outcome is introduced it is available thereafter. The available set of outcomes in period t is then $Z_1 \bigcup \{z_1, \ldots, z_k\}$, whenever $t_k \leqslant t < t_{k+1}$, where $z_k \in Z$ denotes the introduced outcome at arrival date t_k. Figure 1 is a schematic representation of the game.

We parameterize 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.

---

[5] The assumption that the initial set is drawn from F can readily be relaxed.

[6] This abbreviated way of modeling outcomes introduces the apparent complication that the same payoff for role i might be associated with multiple possible payoffs for the remaining players. Knowing your own payoff does not then imply knowing the outcome. This issue could be addressed by supposing that there is a unique label attached to each payoff vector, and that each player role observes this label, as well as his payoff. However, with the current set-up, when the cdf F is continuous, the probability of any role's payoff arising more than once is zero. Each player i can then safely assume that a given payoff is associated to a unique outcome and a unique vector of other roles' payoffs. We then adopt this simpler set-up.
We do not consider how *ToM* types might update beliefs about opponents' payoffs in the light of their own observed payoff. All that we rely on is that, if history establishes another player role's preference between two outcomes for sure, then the *ToM* types learn. All that we rely on concerning the naive types is that they can only learn from repeated exposure to a given game.

A3: *Fix $\alpha \geqslant 0$. The arrival date sequence $\{t_k\}$ satisfies, for each $k = 1, 2, \ldots$, that $t_k = \lfloor |Z_{t_k}|^\alpha \rfloor = \lfloor (|Z_1| + k)^\alpha \rfloor$.*[7]

If the parameter α is low, the spacing between successive t_k's is low and the rate of arrival of novelty is high; if α is high, on the other hand, the rate of arrival of novelty is low. More particularly, if $\alpha < 2$, we will show that the rate of arrival of novel outcomes is too fast for the *ToM* types to keep up, given that they must see each opponent make a choice between each pair of outcomes. If $\alpha < T$, on the other hand, we will show that the rate of arrival of novel outcomes is too fast for the naive types to keep up, given they must see each new game at least once.[8]

Consider now a convenient formal description of the set of games available at each date.

Definition 1: *At date t, the* empirical *cdf based on sampling, with equal probabilities, from the outcomes that are actually available at date t, is denoted by the random function $F_t(z)$ where $z \in [m, M]^I$. The set of games at date t is the T-times product of Z_t. This is denoted Q_t. The empirical cdf of games at date t derives from T-fold independent sampling of outcomes according to F_t and is denoted by $G_t(q)$, where $q \in Q = [m, M]^{IT}$.*[9]

We suppose that, at each date t, an extensive form game denoted q_t is drawn according to G_t independently of history. The players in each match then play q_t. Players of each strategic type within a given player role are constrained to use the same strategy. For simplicity, indeed, the *ToM* types are ultimately constrained to use pure strategies.[10]

---

[7] Here $\lfloor \cdot \rfloor$ denotes the floor 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.

[8] Even if the *ToM* types used transitivity of opponents' preferences, $\alpha < 1$ is certainly too fast an arrival rate for the *ToM* types to keep up, even under the most favorable sequence of pairings of the new outcome with the old outcomes.

[9] Note that F_t and G_t are random variables measurable with respect to the information available at date t, in particular the set of available outcomes Z_t.

[10] That is, the *ToM* types use pure strategies when they know the preferences of all the subsequent players. This is a harmless simplification, since the *ToM* type that will prevail in the long run is a pure strategy in these circumstances. Naive types are assumed to mix uniformly when the game is new.

The cdf's F_t and G_t are well-behaved in the limit. This result is elegant and so warrants inclusion here. First note that the distribution of games implied by the cdf on outcomes, F, is given by G, say, which is the cdf on the payoff space $[m, M]^{IT}$ generated by T independent choices of outcomes distributed according to F. Clearly, G also has a continuous pdf g that is strictly positive on $[m, M]^{IT}$. These cdf's are then the limits of the cdf's F_t and G_t—

LEMMA 1: *It follows that* $F_t(z) \to F(z)$ *and* $G_t(q) \to G(q)$ *with probability one, and uniformly in* $z \in [m, M]^I$, *or in* $q \in [m, M]^{IT}$, *respectively.*

Proof. This follows directly from the Glivenko-Cantelli Theorem. (See Billingsley 1968, p. 275, and Elker, Pollard and Stute 1979, p. 825, for its extension to many dimensions). ■

The evolutionary bottom line is then as follows—each I-tuple playing each game generate children according to the outcome obtained. The current generation then dies and their offspring become the next generation of players. The offspring of each type of i player become i players of the same type in the following period. We normalize the number of children born to each type of i player by dividing this number by the total number of offspring produced by all players in role i.[11]

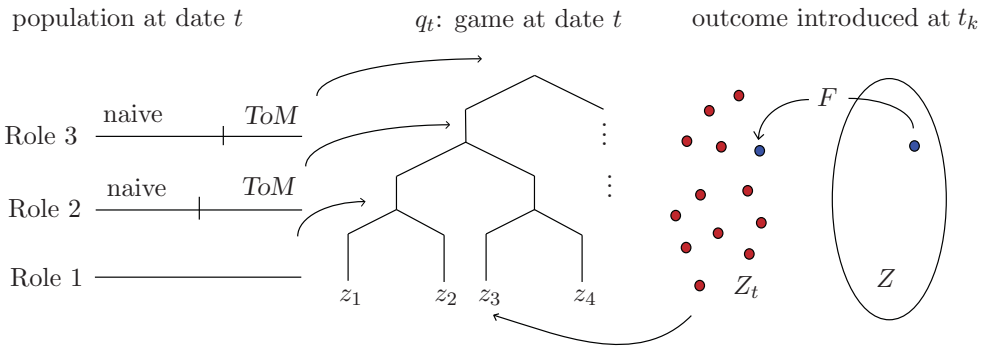

FIGURE 1: **A Schematic Representation of the Key Elements of the Model.**

[11] The assumption that each generation plays the game just once can be straightforwardly generalized so that individual dies and reproduces in each period, with a constant probability for each possibility. The expected number of times each individual plays the game could then be arbitrary.

We turn now to the specification of the "strategic types" within each player role.

2.2. **Strategic Types.**

We allow a finite number of different "strategic types" within each role, $i = 1, \ldots, I$. When making a choice at date t every player of any type is informed of a publicly observed history $H_t = \{Z_t, (q_1, \pi_1), \ldots, (q_{t-1}, \pi_{t-1})\}$, and the game q_t drawn in the current period. The history records the outcomes available at date t, the randomly drawn games up to the previous period, and the empirical distributions of choices made by previous generations. Although each player observes the outcomes assigned to each terminal node, as revealed by the payoff she is assigned at that node, it should be emphasized that she does not observe other roles' payoffs directly. In particular, for each player role i decision-node h that is reached by a positive fraction of players in period τ, $\pi_\tau(h) \in \Delta(A)$ records the aggregate behavior of date τ i player roles at h. Let $\boldsymbol{H}_t$ be the set of date t histories, and $\boldsymbol{H} = \bigcup_{t \geqslant 1} \boldsymbol{H}_t$. Recall that in each period t, every extensive form in Q_t shares the same underlying game tree. Then, let Σ_i denote the set of strategies available to the player role i's of any given date.

We partition each player role population into strategic types. Specifically, for each $i = 1, \ldots, I$, there is a finite set of functions $C_i \subset \{c : \boldsymbol{H} \times Q \longrightarrow \Sigma_i\}$. These are the i player *strategic types*. Each i player is associated with a $c \in C_i$, which determines his choice of strategy.[12] Moreover, we assume these types are inheritable. Specifically, an individual in period t with strategic type c chooses the strategy $c(H_t, q_t)$ in game q_t, his children choose $c(H_{t+1}, q_{t+1})$ in q_{t+1}, his grandchildren choose $c(H_{t+2}, q_{t+2})$, and so on. Variation in strategic types allows for different levels of sophistication within each player role. Some of these types are players who see others as having agency; other types do not see this.

As part of the specification of the map c, we assume that all individuals choose a strictly dominant action in the subgame they initiate, whenever such an action is available. For example, the player at the last stage of the game always chooses the outcome that she strictly prefers. This general assumption

---

[12] It is not required that *ToM* types remember the entire history. What is needed is that they update their beliefs about other players' preferences using the aggregate choices made in each period. It is not important whether naive players remember the entire history or not.

is in the spirit of focussing upon the implications of other players' payoffs rather than the implications of one's own payoffs. This assumption incorporates an element of sequential rationality, since such a dominant strategy is conditional upon having reached the node in question, that is, conditional on the previous history of the game.[13]

To be more precise, the assumption is—

A4: *Consider any i player role, and an i player subgame q. The action a at q is dominant for i if for every action $a' \neq a$, for every outcome z available in the continuation game after i's choice of a in q, and every outcome z' available in the continuation game after i's choice of a' in q, $z_i > z'_i$. For each $i = 1, \ldots, I$, every strategic type in C_i always chooses any such dominant action. When indifferent between several such dominant actions, a player mixes evenly between these actions.*

It is useful to summarize the taxonomy of agents here. Each player role corresponds to a stage $i = 1, \ldots, I$ in the game of perfect information. There is an equal and large number of players within each role, which population is divided into a finite number of types, where each type has the same payoff function, but differs in strategy. The final element of the taxonomy is that there are two main "categories" of types of players—

2.2.1. **Naive Players.** We adopt a relaxed concept of naivete, which serves to make the ultimate results stronger—

Definition 2: *Each map c for a naive type requires that she must choose a fixed arbitrary strategy whenever the game is novel. For specificity, suppose she then mixes uniformly over all available actions. Naive players also choose any strictly dominant action in the remaining subgame, as in A4.*

The assumption that the strategy used in novel games is fixed is in the spirit that naive players start as blank slates in such situations. However, our results would still hold, despite some extra complication, if, in any novel game, the

[13] Given suitable noise, this element of sequential rationality is assured, and this property can be made a result rather than an assumption. That is, a strategy that did not use a dominant choice would be driven to extinction under any plausible evolutionary dynamic. We omit this proof for conciseness.

naive players chose a best response to opponents' strategies that mix uniformly over all their available actions. Such a choice by the naive players would be Bayesian optimal initially, if all players' payoffs were independently distributed. This choice would be the actual true *SPE* choice for a positive fraction of new games, but not for all of them. The key property of any relaxed version of the assumption is that it should ensure that the naive players make inappropriate choices in a positive fraction of new games in the long run.

If the game is not new, the strategy of each type of naive player is unconstrained. Although it makes an implausible combination, the naive players could then be fully Bayesian rational with respect to all of the relevant characteristics of the game—not merely updating the full distribution of opponents' payoffs, but updating the distribution for all opponents' types. Nevertheless, the sophisticated *ToM* players will out-compete them, given only the naive players' inability to adapt fully to a new game. To the extent that naive players fail to attain such Bayesian rationality in games that are not new, our results would simply be strengthened.

2.2.2. **Theory of Mind Players.** Consider now a category of theory of mind strategic types. Intuitively, these types conceive of opponents as making choices according to well defined preferences and beliefs. All of the *ToM* types know there are some preferences influencing player role j's choices in every period, and they learn what these preference are.

Definition 3: *The important* long run *aspect of the behavior of ToM types is that, if the history of the game has revealed the preferences of all subsequent players, these ToM types map these preferences into an action. In particular, in every role, there is a positive fraction of a special type of ToM called SPE-ToM which plays a subgame perfect equilibrium action, given these known preferences of subsequent players and that the SPE is unique. Recall that, as part of the map c, ToM players choose any strictly dominant action in the remaining subgame, as in A4. In the* short run*, all the ToM players know that all* other *players also use dominant actions if available, as in A4; further, this is common knowledge among the ToM players. The presence of some ToM players in every role is also common knowledge among all the ToM types.*

The assumptions here on the *ToM* types seem reasonable. In particular, the assumption that *ToM* types have common knowledge that all types choose a dominant action is in the spirit of focussing here on the implications of the preferences of others, while presuming full use of one's own preferences. Note also that it is merely for expositional clarity that we describe the short run learning behavior of the *ToM* types in terms of common knowledge. The entire description can be recast in pure "revealed preference" terms. How this can be done is discussed after the statement of Theorem 1.

It should be emphasized that we place only weak restrictions on all types—naive or *ToM*, so the results are thereby strengthened. As long as the naive types make the same mixed choice in all novel games, their behavior is otherwise unrestricted and might be highly sophisticated. Similarly, when the *ToM* types have not ascertained the preferences of all subsequent players, their behavior is arbitrary and might be highly suboptimal. For example, they might *minimize* their expected payoff given a fully Bayesian view of their situation. It is not the case then that the *ToM*s strategically dominate the naive types—naive players might do much better than the *ToM* players in the short-run. In the long run, however, the assumption that there is a *SPE-ToM* ensures this type must evolutionarily dominate all the naive types and for that matter all the other *ToM* types.

Figure 1 is a schematic representation of the model.

2.3. **The Theoretical Results.**

There are two main theoretical results. The first shows that the *ToM* types learn the preferences of other roles, so these become common knowledge among all *ToM* types in all roles. The second shows how the *ToM*s might exploit this knowledge by playing the *SPE* of the game.

DEFINITION 4: *Suppose A4 holds. The history H_t reveals players in role i strictly prefer z to z' if and only if, whenever H_t occurs, it becomes common knowledge among ToMs that $z_i > z'_i$.*

It is established in the course of the proof of Theorem 1 that any such strict preference for player i can be revealed by some suitable possible history.

Now, for each $i = 1, \ldots, I$, let L_{it} denote the fraction of pairs $(z, z') \in Z_t \times Z_t$ where H_t reveals i's favored outcome between $\{z, z'\}$.[14] To evaluate the performance of the naive players, let γ_t be the fraction of games (of those in Q_t) that have been played previously at date t. Let $T \geqslant 4$ be the number of terminal nodes in the fixed game tree.[15] We then have the following key theoretical result that sets the stage for establishing the evolutionary dominance of the *SPE-ToM*s over all other players—

THEOREM 1: *Suppose assumptions A1-A4 all hold. If $\alpha < 2$, then L_{it} surely converges to zero, $i = 1, \ldots, I$; if $\alpha > 2$, however, then L_{it} converges to 1 in probability. That is, if the rate of arrival of novelty is sufficiently high, then the fraction of pairs of outcomes for which i's preferences have been revealed tends to 0; otherwise this fraction tends to 1, in probability. Similarly, if $\alpha < T$, then γ_t surely converges to zero; if $\alpha > T$, then γ_t converges to 1 in probability. That is, if the rate of arrival of novelty is sufficiently high, then the fraction of games that have been played before tends to 0; otherwise this fraction tends to 1, in probability.*[16]

This is proved in the Appendix. This result says that if $\alpha > 2$, and, in particular, if all types adopt strictly dominant acts, whenever these are available, where the *ToM*s have common knowledge that this is true, then all preferences are revealed in the limit to the *ToM*s. This is the crucial result here, since if, at the same time, $\alpha < T$, all the naive players see new games essentially always and mix uniformly, in a way that is generally inappropriate, with the most important exception of games in which they have a dominant action.

An intuitive description of how the *ToM* types learn preferences is useful. Consider a *ToM* type in a particular player role $j > 1$. The argument that this type can obtain the preferences of subsequent player roles proceeds by backwards induction on these subsequent roles. Players in the last role choose a preferred

---

[14] For simplicity, assume that players mix whenever indifferent and that this too is common knowledge among the *ToM* types. The large population in each player role i means that such indifference would then be evident to *ToM* players in other roles. Although such indifference does arise given that each finite outcome set is chosen with replacement, the probability of such indifference tends to 0 in the limit.

[15] If the I player roles have A actions each, then $T = A^I$.

[16] It is difficult to analyze the case that $\alpha = 2$ or $\alpha = T$, but these are non-generic.

action and this is revealed in the choices that $j > 1$ sees. A player in role j also knows that all *ToM* types in all roles now know this as well. Eventually a complete picture of player *1*'s preference can be built up as common knowledge among all the *ToM* types. As the induction hypothesis, suppose the preferences of $i-2, \ldots, 1$ for $i \leqslant j$ have been established as common knowledge among the *ToM* types. We need to show that j can similarly obtain the preferences of $i-1$. Suppose then that a game is drawn in which player role $i-1$ in fact has a dominant action, a, say, after which $i-2$ has a dominant action, after which $i-3$ has a dominant action, after which... We refer to this as the subgame starting with $i-1$ as being "forward dominance solvable." Furthermore, there is another action, a', say, that $i-1$ could take, after which again $i-2$ has a dominant action, after which... Player j knows the situation faced by $i-2, ..., 1$. Since, in fact, players in role $i-1$ have a dominant action, all types take this. Player j can see that all $i-1$'s have made the same choice, so that the *ToM*s there who made this choice must then prefer the outcome induced by a to the outcome induced by a'. Eventually, *ToM* $j \geqslant i$ can build up a complete picture of the preferences of the role $i-1$.

This description of learning shows how the common knowledge assumptions concerning the *ToM* types can be stripped to their bare revealed preference essentials. It is unimportant, that is, what or whether the *ToM* types think, in any literal sense. All that matters is that it is *as if* the *ToM*s in roles $i, ..., I$ add to their knowledge of role $i-1$'s preferences in the circumstances considered above. Once a *ToM* type in role i, for example, has experienced all of role $i-1$ binary choices being put to the test like this, given that this is already true for roles $i-2, ..., 1$, this role i *ToM* type can map the preferences for subsequent players to an action.

All that remains then, to complete the argument, is to show that the *ToM* types will do better than the naive types by exploiting their knowledge of all other players' preferences, while the naive types are overwhelmed by novel games. This will be true in a variety of circumstances; for simplicity, we focus on assumptions that yield the *SPE*.[17]

---

[17] This *SPE* is unique with probability that converges to *1*.

For simplicity, we impose the following restriction on the alternative *ToM* types—

A5: *Every ToM alternative to the SPE-ToM differs from the SPE-ToM at every reached decision node in a set of games that arises with positive probability under the distribution F.*

We now have the main conceptual result—

THEOREM 2: *Suppose assumptions A1-A4 and A5 all hold. Suppose that there are a finite number of types—naive and ToM, one of which is the SPE-ToM. If $\alpha \in (2, T)$, then the proportion of SPE-ToM in role i, R_{it}, say, tends to 1 in probability, $i = 2, ..., I$.*

The proof of this is also relegated to the appendix.

We focus here on the case that $\alpha \in (2, T)$. If $\alpha > T$, so the rate of introduction of novelty is slow, the relative performance of the two types depends on the detailed long run behavior of the naive players. If the naive players play a Bayesian rational strategy the second time they encounter a given game, they would tie the *ToM*s. There are less stringent conditions under which this would remain true. It is, in any case, not intuitively surprising that a clear advantage to *ToM* relies upon there being at least a minimum rate of introduction of novelty. In the case that $\alpha < 2$, the *ToM* players are overwhelmed with novelty, as are the naive players. The outcome then hinges on the short run behavior of the various types. As long as the naive players are not given a more sophisticated short run strategy than the *ToM*s, the naive types can, at best, match the *ToM*s. For example, if the naive types mix uniformly over all their choices in any new game, and the *ToM*s do this whenever they do not know subsequent players' preferences, the naive types cannot beat the *ToM*s.

We close this subsection with a number of additional remarks.

1) The key issue here is how *ToM* deals with *novelty*—the arrival of new outcomes—rather than with *complexity*—the unbounded growth of the outcome set. Indeed, the model could be recast to display this as follows. Suppose that a randomly chosen outcome is dropped whenever a new outcome is added, so the size of the outcome set is fixed, despite such updating events. There will then be a critical value such that, if the interval between successive updating events

is less than this critical value, the naive types will be mechanically unable to keep up with the flow of new games. There will also be an analogous but lower critical value for the *ToM* types. If the fixed interval between updating events is chosen to lie between these two critical values, the naive types will usually be faced with novel games; the *ToM* types will do better, with a stochastic but usually positive fraction of games in which the choices of role 2 players can all be predicted. This provides a version of the current results, although one that is noisier than the current approach.[18]

2) It is straightforward to show that the ascendancy of the *SPE-ToM* type is robust to the introduction of a sufficiently small fixed cost for all *ToM* types, so these results stand in sharp contrast to Stahl (1993), for example. It is not unreasonable that there should be a higher fixed cost of *ToM*, since, for example, it might require the maintenance of a more complex brain. However, this is not the only potential source of cost, and in fact, the *memory* demands of the naive types here are certainly greater than the memory demands of *ToM*. The naive types need to remember each game; the *ToM*s need only remember preferences over each pairwise choice for opponents, and if memory is costly then these costs would be lower for the *ToM*s in any case. In this sense, consideration of all costs might well reinforce the advantage of the *ToM*s.

3) Suppose, hypothetically, that the naive types have all been eliminated. The eventual ascendancy of each *SPE-ToM* type over the other *ToM* types is not a matter of strategic dominance but relies on the previous ascendancy of *SPE-ToM* types at all subsequent stages. That is, given a particular pattern of subsequent *ToM* roles, there may be a *ToM* that outdoes the *SPE-ToM*. It is only once *SPE* behavior has been established for subsequent players, by backwards induction, that the *SPE* choices become optimal.[19]

4) The ascendancy of *SPE-ToM* at each stage relies on the assumption that there is a large population in the corresponding role. Even though a non-*SPE*

---

[18] The need in the current model for the interval between updating events to grow with time is a reflection of the fact that each new outcome produces a larger number of novel games when there is already a larger number of outcomes.

[19] This is perhaps analogous to the difficulty that the Connecticut Yankee has at King Arthur's Court, according to Mark Twain. That is, to his consternation, the choice made by Twain's hero often fails to be optimal because the choice by his opponents is non-optimal.

choice might benefit the player role in question since it could advantageously modify the optimal choice of previous roles, this benefit is analogous to a public good. That is, the optimal choice by a small measure of players in the role in question must be sequentially rational.

5) Consideration of a long run equilibrium, as in the above two results, is simpler analytically than direct consideration of the speed of learning of the various types. More importantly, it also permits the use of weak restrictions on the naive and *ToM* types, as is desirable in this evolutionary context. As a related matter, learning by the *ToM* types relies on rather improbable events and so seems likely to be slow. That is, it might seem that this method of proof would produce weaker results than actually hold. However, the current method suffices to show that complete learning by the *ToM*s occurs whenever $\alpha > 2$. Since it is mechanically impossible to learn others' preferences when $\alpha < 2$, a more sophisticated method of proof cannot significantly improve the result.

6) Our results show how an increase in the rate of introduction of novelty might precipitate a transition from a regime in which there is no advantage to theory of mind to one in which a clear such advantage is evident. This is consistent with theory and evidence from other disciplines concerning the evolution of intelligence. For example, it is argued that the increase in human intelligence was in part due to the increasing novelty of the savannah environment into which they were thrust after we exited our previous arboreal niche. (For a discussion of the intense demands of a terrestrial hunter-gatherer lifestyle, see, for example, Robson and Kaplan, 2003.)

2.4. **Related Theoretical Literature.**

We outline here a few related theoretical papers in economics. The most abstract and general perspective on theory of mind involves a hierarchy of preferences, beliefs about others' preferences, beliefs about others' beliefs about beliefs about preferences, and so on. (Robalino and Robson, 2012, provide a summary of this approach.) Harsanyi (1967/68) provides the classic solution that short circuits the full generality of the hierarchical description.

A strand of literature is concerned to model individuals' beliefs in a more realistic fashion than does the general abstract approach. The first paper in this

strand is Stahl (1993) who considers a hierarchy of more and more sophisticated strategies analogous to iterated rationalizability. A smart$_n$ player understands that no smart$_{n-1}$ player would use a strategy that is not $(n-1)$-level rationalizable. A key aim of Stahl is to examine the evolution of intelligence in this framework. He obtains negative results—the smart$_0$ players who are right in their choice of strategy cannot be driven out by smarter players in a wide variety of plausible circumstances. Mohlin (2012) provides a recent substantial generalization of the closely related level-k approach that allows for multiple games, learning, and partial observability of type. Nevertheless, it remains true that lower types coexist with higher types in the long-run. This is not to deny that the level-k approach might work well in fitting observations. For example, Crawford and Iriberri (2007) provide an explanation for anomalies in private-value auctions based on this approach.

Our model fits only loosely in this context. Our setup sidesteps nontrivial higher order beliefs by making revelation of preferences common knowledge among the sophisticated players. A player role that needs to learn the preferences of a larger number of subsequent player roles then faces a problem only of greater breadth rather than one of greater depth.[20] At the same time, our approach demonstrates how apparently rather weaker assumptions than common knowledge of rationality and preferences suffice to generate the full revealed preference predictions for a game with perfect information. That is, play here evolves towards subgame perfect equilibrium in each of a sequence of different games, despite the continual introduction of novelty.

The line that we draw between smarter and less smart players separates the naive players who learn to play each game separately (as in evolutionary game theory) and the *ToM* players who infer others' preferences from their choices and eventually use these inferred preferences to choose optimally in novel games. In contrast to the level-k approach, we obtain a positive result concerning the evolution of intelligence. We consider a large and growing set of games, but, more particularly, the reason for the difference is that naive players are constrained to use the same strategy in every novel game. There is no type in our

[20] Such greater breadth might nevertheless rapidly overwhelm real-world players, as the number of stages increases; indeed, this is an important question for future research.

framework that is minimally smart but lucky enough to use the right strategy in every game. Indeed, the existence of such a type in our framework seems far-fetched.

There is by now a fairly large literature that examines varieties of, and alternatives to, adaptive learning. Camerer, Ho and Chong (2002), for example, extend a model of adaptive, experience-weighted learning (EWA) to allow for best-responding to predictions of others' behavior, and even for farsighted behavior that involves teaching other players. They show this generalized model outperforms the basic EWA model empirically, a result that is broadly consistent with our experimental findings. Bhatt and Camerer (2005) find neural correlates of choices, beliefs, and 2nd-order beliefs (what you think that others think that you will do). These correlates are suggestive of the need to transcend simple adaptive learning. Finally, Knoepfle, Camerer and Wang (2009) apply eye-tracking technology to infer what individuals pay attention to before choosing. Since individuals actually examine others' payoffs carefully, this too casts doubt on any simple model of adaptive learning. We show experimentally that people go further, actively seeking to learn others' payoffs when these are initially hidden.

3. Experiments on Theory of Mind

3.1. **Experimental Design.**

We report here the results of experiments that are simplified versions of the theoretical model. These test the ability of individuals to learn the preferences of others through repeated interaction and to use that information strategically to their advantage. The game tree is a two-stage extensive form where each player has two choices at each decision node.

There are then two player roles, 1 and 2.[21] Player roles differ in their position in the game tree and their (induced) preferences, but all players of a given role have identical preferences. In each period, each role 1 participant is randomly and anonymously matched with a single role 2 participant to play a two-stage extensive form game, as depicted in Figure C1, in Appendix C. We employ this

[21] Here we revert to the usual convention that role 1 moves before role 2.

matching scheme to at least diminish the likelihood of supergame effects. In each game, role 1 players always move first, choosing one of two intermediate nodes (displayed in the figure as blue circles), and then based on that decision, the role 2 player chooses a terminal node that determines payoffs for each participant (displayed in the figure as a pair of boxes).

When making their decisions, participants observe only their own payoff at each outcome and are originally uninformed of the payoff for the other participant.[22] Instead, they know only that payoff *pairs* are consistent over time. That is, whenever the payoff for role 1 is X, the payoff to role 2 will always be the same number Y. In Figure C1, which is shown from the perspective of a role 1 participant, his own payoff at each terminal node is shown in the orange box, while his counterpart's payoff is displayed as a "?" in the blue box. Similarly, when role 2 players make their decisions, they only observe their own payoffs and see a "?" for their counterpart (see figure C2).

In each period, the payoffs at each terminal node are drawn *without replacement* randomly from a finite set of V payoff pairs.[23] Each element in each pair of payoffs is unique, guaranteeing a strict preference ordering over outcomes. This set is fixed in the experiments in contrast to its growth in the theoretical model. We do not then attempt to study the theoretical long run in the experiments, but content ourselves with observing the rate of learning of opponents' preferences. Allowing for the strategic equivalence of games in which the two payoff pairs at a given terminal node are presented in reverse order, there are $\binom{V}{2}\binom{V-2}{2}/2$ strategically distinct games that can be generated from V payoff pairs, each of which has a unique subgame perfect equilibrium.

Thus, as in the theoretical model, despite their initial ignorance of their counterpart's preferences, role 1 players can learn about these preferences over time,

---

[22] Note that payoff privacy has the added benefit of mitigating the effects of non-standard preferences on individual choice; since individuals are unaware of exactly how their choices impact others' payoffs, altruistic and reciprocal actions, which may depend on the relative effect on own and other's payoffs (as in Charness and Rabin, 2002, for example), will be controlled. Indeed, it has long been known that payoff privacy encourages the achievement of equilibrium outcomes in market settings (Smith 1982).

[23] We sample with replacement in the theoretical model, although this assumption is merely a minor convenience. We do not allow replacement here to make the most of our experimental resources of time and money.

by observing how role 2 players respond to various choices presented to them. If role 1 players correctly learn role 2 players' preferences, they can increase their own payoff by choosing the *SPE* action. On the face of it, role 1 players will have then developed a theory of a role 2 player's mind.

This suggests investigating whether role 1 players choose in a manner that is increasingly consistent with the *SPE*. Initial pilot sessions revealed two issues with this strategy: 1) many of the randomly generated games include dominant strategies for player 1, which are not informative for inferring capacity to learn the preferences of others, as indeed reflected in the theoretical model, and 2) more subtly, there is a simple "highest mean" rule of thumb that also often generates *SPE* play. Consider a player 1 who is initially uncertain about player 2's preferences. From the point of view of player 1, given independence of player 2's preferences, player 2 is equally likely to choose each terminal node, given player 1's choice. The expected payoff maximizing strategy is to choose the intermediate node at which the average of potential terminal payoffs is highest. Indeed, our pilot sessions suggested that many participants followed this strategy, which was relatively successful.

For these reasons, we used a 3x1 within-subjects experimental design that, over the course of an experimental session, pares down the game set to exclude the games in which choice is too simple to be informative. Specifically, each session included games drawn from 7 payoff pairs (so there are 105 possible games). In eighteen of our sessions, payoff possibilities for each participant consisted of integers between 1 and 7, and in two sessions the set was {1,2,3,4,8,9,10}. This variation was intended to reduce noise by more strongly discouraging player 2 from choosing a dominated option, but observed player 2 choices in these sessions are comparable to those in other sessions, so we pool the data for analysis below. Each session lasted for 90 periods in which, in the first 15 periods, the game set included 15 randomly chosen games from the set of possible games, $\bar{Q}$, say. Finally, starting in the 16th period, we eliminate all games in which player 1 has a dominant strategy, and the next 15 periods consist of games randomly drawn from this subset of $\bar{Q}$. Finally, starting in the 31st period, we also eliminate all games in which the optimal strategy under the "highest mean" rule of thumb corresponds to the *SPE* of the game, and our final 60 periods consist

of randomly drawn games from this smaller subset. Thus, our final 60 periods make it harder for player 1 to achieve high payoffs, since the only effective strategy is to learn the preferences of the role 2 players.

Learning by role 1 players here would be disrupted by the presence of any role 2 player who fails to choose his dominant action. For this reason, we considered automating the role 2 player. However, on reflection, this design choice seems untenable. In the instructions, we would need to explain that algorithmic players 2 maximize their payoffs in each stage, which would finesse much of the inference problem faced by player 1—in essence the instructions would be providing a key part of the theory of mind. It is also conceivable that individuals would behave differently towards a computer program than they would towards a human agent.

A second potential issue is that foregone payoffs (due to role 1 player's choice) may lead to non-myopic behavior by some player 2s. Such behavior involves role 2 players solving a difficult inference problem. A spiteful (or altruistic) player 2, who wanted to punish (or reward) player 1 on the basis of player 2's foregone payoffs, first must infer that player 1 has learned player 2's preferences and then infer player 1's own preferences on the basis of this assumption. Player 2 could then, given his options, choose the higher or lower of the two payoffs for player 1 as either punishment or reward. However, players 2 chose their dominant action roughly 90% of the time, which suggests that these sources of error were not a prominent feature of our experiment.

We relate our results directly to theory of mind as in psychology, as measured by two short survey instruments. At the conclusion of the experiment, participants completed the Autism-Spectrum Quotient (AQ) survey designed by Baron-Cohen et al. (2001), since autism spectrum reflects varying degrees of inability to "read" others' minds. This short survey has been shown to correlate with clinical diagnoses of autism spectrum disorders, but it is not used for clinical purposes. The instrument was designed for use on adults of normal intelligence to identify the extent of autism spectrum behaviors in that population. Participants also completed the Broad Autism Phenotype Questionnaire (BAP) due to Hurley et al. (2007), which provides a similar measure of autism spectrum behavior and is highly correlated with the AQ. With this additional data we will be able to evaluate how each participant's ability to perform as player 1

in our experiments correlates with two other well-known *ToM* metrics.[24] Copies of the questionnaires are available in Appendices D and E.[25]

We report data from 20 experimental sessions with a total of 174 participants (87 in each role). Each experimental session consisted of 6, 8 or 10 participants, recruited from the students of Simon Fraser University between April and October 2013. Participants entered the lab and were seated at visually isolated computer terminals where they privately read self-paced instructions. A researcher was available to privately answer any questions about the instructions. After reading the instructions, if there were no additional questions, the experiment began. Instructions are available in Appendix B.

Each experimental session took between 90 and 120 minutes. At the conclusion of each session, participants were paid privately in cash equal to their payoffs from two randomly chosen periods. We use this protocol to increase the salience of each individual decision, thereby inducing participants to treat each game as payoff-relevant. For each chosen period, the payoff from that period was multiplied by 2 or 3 (depending on the session) and converted to CAD. Average salient experimental earnings were $25.00, with a maximum of $42.00 and a minimum of $6.00. In addition to their earnings from the two randomly chosen periods, participants also received $7 for arriving to the experiment on time. Upon receiving payment, participants were dismissed.

3.2. **Experimental Results.**

---

[24] One might be concerned that any differences we observe in behavior that are correlated with AQ are actually driven by differences in intelligence. Indeed, it is well-known that extreme autistics tend to have low IQs. Crucially, however, within the normal range of AQ scores (those surveyed who had not been diagnosed with an autism spectrum disorder), the survey measure is uncorrelated with intelligence (Baron-Cohen et al., 2001). Our sample consists of undergraduates none of whom (to our knowledge) are diagnosed with any autism spectrum disorder. Thus any relationship we observe between AQ and performance is unlikely to be due to differences in intelligence.

[25] In the first wave of these experiments performed in April and June 2013 (76 subjects total), we conducted the AQ questionnaire with a 5-point Likert scale that allowed for indifference rather than the standard 4-point scale which requires participants to either agree or disagree with each statement. The AQ questionnaire is scored by assigning 1 or 0 to each response and summing. In our data analysis below, we assign indifferent responses a score of 0.5.

Since the decision problem is trivial for player 2, our analysis focuses entirely on decisions by player 1. We focus on the probability with which player 1 chooses an action consistent with the *SPE* of the game. For a fixed game, and with repeated play with fixed matching and private information about individual payoffs, pairs frequently converge to non-cooperative equilibrium outcomes over time (McCabe et al., 1998).[26] This is not surprising since, in their environment an individual merely need learn her counterpart's preferences over two pairwise comparisons. However, since these experiments employ static repetition of the same game, the data do not clearly distinguish theory of mind from reinforcement learning. Our experiment is the first (that we know of) to test theory of mind capacity in a dynamic setting in which inferences drawn from the play of one game may be employed to predict play in novel, future games. In this sense our setting is more strategically complex than those previously studied, and hence we are able to both distinguish *ToM* from reinforcement learning *and* observe heterogeneity in *ToM* capabilities, which we can exploit in our data analysis.

First, we describe overall learning trends, and we show that individuals' performance as players 1 depends on how much information they have acquired about the preferences of players 2. This suggests that our players 1 exhibit *ToM* in the sense of the theoretical model. Finally, we compare our measure of *ToM* to measures from psychology and show that the learning speed of players 1 is significantly correlated with survey responses, suggesting that our theoretical concept of *ToM* corresponds, at least to some extent, with theory of mind as understood by psychologists.

3.2.1. **Learning Others' Preferences.**

Figure 2 displays a time series of the probability that player 1 chose an action consistent with knowledge of player 2's preferences (i.e. consistent with *SPE*) over the 90 periods of the experiment. After 15 periods, the game set no longer included instances where player 1 had a dominant strategy. After 30 periods, the game set no longer included instances where player 1 would choose correctly by following the "highest mean" rule of thumb. At period 31, when subjects

[26] See also Fouraker and Siegel (1963) and Oechssler and Schipper (2003).

enter the NoDominant/NoHeuristic treatment, there is a significant downtick in player 1's performance, but afterwards there is a notable upward trend in the probability of player 1 choosing optimally.[27] Despite the fact that individuals tend to learn player 2's preferences over time on average, we observe substantial heterogeneity in rates of learning, which we exploit in the next section.

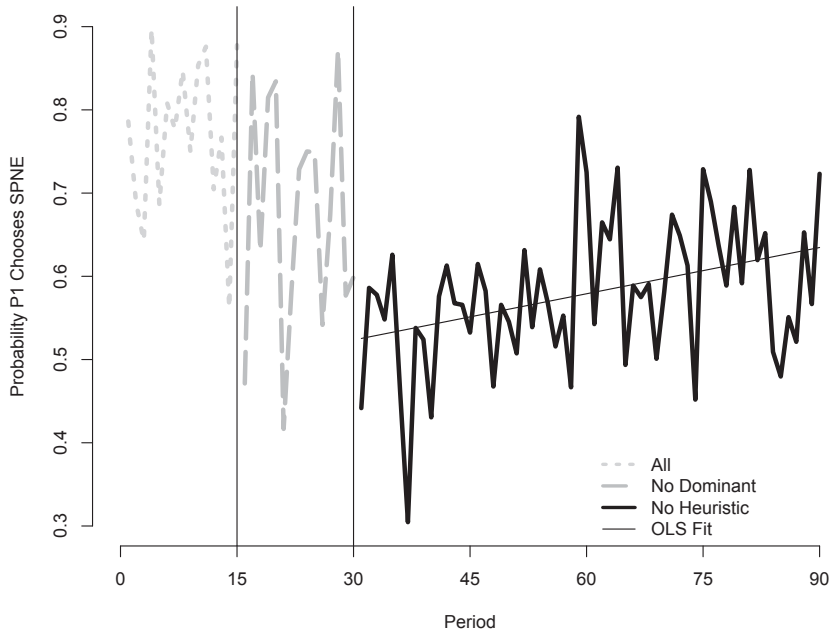

FIGURE 2: **Time Series of Learning Opponent's Preferences.**

To provide statistical support for these observations, Table 1 reports logistic regressions where the dependent variable takes a value of 1 if player 1 chooses an action consistent with the *SPE* of the game and 0 otherwise. We include treatment dummies for periods 1-15 and periods 16-30 to control for the game set. To identify the impact of feedback quality from player 2 choices on the likelihood of *SPE* choices, column (2) also includes two variables that control for the proportion of dominant choices made by players 2 in previous periods. Specifically, let $W_{i,t}$ be an indicator variable that takes a value of 1 when player i's partner chose the dominant action in the randomly chosen game q_t. Then

[27] Table F1 in Appendix F also reports summary statistics for each experimental session. Figure F1 displays a histogram of the individual rates of SPE consistent choices over all informative games—those in which the rule of thumb did not lead to SPE-consistent choice.

we compute the lagged proportion of dominant choices observed by player i as $\frac{\sum_{s=1}^{t-1} W_{is}}{t-1}$. Observing dominant choices by *all* players 2 is also informative, so we compute a second measure for each period of each session that measures the lagged proportion of dominant choices made by all players 2. To test for naive reinforcement learning, as in the theoretical model, column (3) also includes a variable that counts the number of times participants have played the randomly chosen game at time t in the past. Finally to test for *ToM* learning, as in the model, column (4) includes two additional variables that measure the amount of information player 1 has at a given time about player 2's preferences. Specifically, let q_t be a feasible subgame in period t and $I(q_t)$ be an indicator function that takes a value of 1 if any player 2 is observed making a choice in that subgame in period t (or in the mirror image subgame).[28] In a given period, there are two feasible subgames q_t^1 and q_t^2, say. We then measure the previous exposure to player 2's preferences in game q_t by computing: $\min\{\sum_{s=1}^{t-1} I(q_s^1), \sum_{s=1}^{t-1} I(q_s^2)\}$. This provides a rough measure of what player 1's should know about player 2's preferences. It is a function of the total number of times that player 2 has chosen between each of the two relevant outcome pairs. These two totals are then aggregated using the function min for simplicity. As with the variables we introduced in column (2), we also construct an analogous measure that includes only those choices made by the person with whom player 1 was paired. We also include both session and individual fixed effects.

The positive and significant estimated coefficient on Period in column (1) indicates that participants are increasingly likely to choose optimally over time. This is consistent with the evidence in Figure 2. In column (2), when we include two variables measuring the fraction of previous dominant choices by players 2, we find a significant effect only of dominant choices made by partnered player 2s, but not by all players 2. Player 1s who have observed more a greater share of dominant actions by their partners are more likely to choose optimally in later periods.[29] Column (3) tests for naive reinforcement learning as in the

---

[28] Recall that players 1 receive aggregated information about the choices of all players 1 and 2 in their session at the end of each period. See Figure C3 in the appendix.

[29] If player 1s were Bayesian rational, they would treat the observations on all player 2's as equally informative. It is plausible psychologically, however, that they pay particular attention to their partnered player 2, especially since this partner's behavior affects the current payoff

P1 Chose SPE	(1)	(2)	(3)	(4)
Period	0.009***	0.009***	0.009***	0.002
	(0.002)	(0.002)	(0.002)	(0.003)
No Dominant Options	0.811***	0.808***	0.805***	0.842***
	(0.096)	(0.096)	(0.098)	(0.098)
All Treatments	1.415***	1.430***	1.425***	1.444***
	(0.119)	(0.120)	(0.124)	(0.124)
Cumulative Fraction My Partner Chose Dominant$_{t-1}$		1.226***	1.226***	1.237***
		(0.440)	(0.440)	(0.441)
Cumulative Fraction All P2s Chose Dominant$_{t-1}$		-0.265	-0.264	-0.187
		(0.819)	(0.819)	(0.822)
# of Times Played Previously			0.009	0.013
			(0.051)	(0.051)
# of Previous Choices Observed My Partner				-0.010
				(0.022)
# of Previous Choices Observed All P2s				0.080***
				(0.019)
Constant	-0.497**	-1.293*	-1.286*	-1.329**
	(0.241)	(0.660)	(0.661)	(0.663)
Observations	7830	7743	7743	7743
Session Fixed Effects	Y	Y	Y	Y
Individual Fixed Effects	Y	Y	Y	Y

Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 1: **Logistic Regression Analysis of Learning.**

theoretical model. We find no evidence of a significant effect of repetition of the same game on the probability of choosing correctly. This is driven in part by the fact that our games were drawn from a relatively large set, which reduces the potential for repetition. Finally, column (4) includes our measures of the amount of information players 1 had about the preferences of player 2. A highly significant and positive coefficient on the variable measuring the information that could be gleaned from all previous choices by other players 2 implies that players 1 improve their performance by applying what they have learned about the preferences of players 2 in the past. That is, they exhibit *ToM* in the sense of our model. In contrast to our findings from column (2), we find that the total previous number of choices made by *all* players 2 is a better determinant

for player 1. Indeed, this finding is consistent with evidence that individuals overweight private information; see, for example, Goeree et al., (2007).

of learning than those made by their partner, as would be Bayesian rational. Importantly, when we include these variables, the coefficient on Period is no longer statistically significant, suggesting that the significant estimated trend in columns (1) - (3) was actually capturing the effects of *ToM*. Thus, even in this complex setting, individuals are able to learn the preferences of others. We summarize these observations below:

Finding 1: On average, there is a significant increase in understanding of others' preferences over time, despite individual variation.

Finding 2: The increase is driven by observation of player 2's preferences (*ToM*) rather than naive reinforcement learning, as is consistent with the theoretical results.

3.2.2. **Comparing Measures of ToM.**

Table 1 provides evidence that increases in the rate of *SPE* choices result from *ToM*. However, our data reveal clear heterogeneity across individuals. Thus, we exploit this heterogeneity to ask whether our measure of *ToM* correlates with previous survey measures of theory of mind from psychology. Specifically, we examine correlations between subjects' AQ and BAP scores and the rate at which players *learn* the preferences of others.

We estimate learning rates separately for each player 1 with logistic regressions where the dependent variable takes a value of 1 when the player chose a node consistent with *SPE* and 0 otherwise, and the independent variables are our measure of the information available about player 2's preferences from previous choices (as described above), the lagged proportion of dominant choices made by their partners, and a constant term. The β coefficient on the first independent variable provides an estimate of each individual's rate of learning.[30] In both computations, we restrict attention only to choices that are informative for

[30] The data reported here exclude one extreme outlier from our 20th session who chose the *SPE*-consistent action in 87/90 periods and whose estimated β was more than 12 times greater (18.64) than the next fastest-learning subject (1.46).

inferences about *ToM* by excluding games with dominant strategies for players 1 and games in which the "highest mean" heuristic corresponds to the *SPE*.[31]

We then compute simple correlation coefficients between estimated learning rates and measures of theory of mind from the AQ and BAP questionnaires. Recall that on both instruments, a higher score indicates increased presence of autism spectrum behaviors. Thus, negative correlations will indicate that our concept of *ToM* is analogous to the information in the AQ and BAP surveys, while the absence of correlation or positive correlations will indicate otherwise.

	Learning Rate
BAP	-0.22**
BAP_Rigid	-0.02
BAP_Aloof	-0.27***
BAP_Prag	-0.17*
AQ	-0.28***
AQ_Social	-0.28***
AQ_Switch	-0.14*
AQ_Detail	0.02
AQ_Commun	-0.23**
AQ_Imagin	-0.14*
*** p < 0.01, ** p < 0.05, * p < 0.1.	

TABLE 2: **Correlations between Autism Spectrum Measures and Individual Learning Rates.** BAP and AQ are overall scores from each instrument. Other variables are individual scores on subscales of each instrument. BAP_Rigid = Rigidity, BAP_Aloof = Aloofness, BAP_Prag = Pragmatic Language Deficit, AQ_Social = Social Skills, AQ_Switch = Attention Switching, AQ_Detail = Attention to Detail, AQ_Commun = Communication Skills, and AQ_Imagin = Imagination.

Table 2 report these simple correlations between measures from our experiment and survey measures of autism spectrum intensity.[32] From the table, we can see that learning rates are significantly negatively correlated with both the AQ and BAP scores as well as most of the subscales.[33] Taken together this provides solid evidence that our games measure theory of mind as it is conceived by psychologists. Reading through the questionnaires, this correlation

[31] Note that this regression equation is derived from the findings from column 4 of Table 1 in that we include only those independent variables that were statistically significant.
[32] Figure F2 displays these correlations for the AQ and BAP scores.
[33] Following convention, the BAP score is the mean of the three BAP subscale scores, and the AQ score is the sum of the five AQ subscale scores.

agrees with intuition. For example, consider the finding that our measure of *ToM* is highly significantly correlated with the two subscales that emphasize *social skills*: AQ_Social and BAP_Aloof. We highlight these subscales because they are explicitly designed to measure capacity for and enjoyment of social interaction, which is particularly reliant on theory of mind. One particularly telling item on the AQ_Social subscale asks individuals how strongly they agree with the statement:

"I find it difficult to work out people's intentions."

This is consistent with our notion of *ToM* in a strategic setting. We also observe that learning is correlated with the AQ_Commun and AQ_Imagin subscales. The latter measures "imagination" by asking respondents to what degree they enjoy/understand fiction and fictional characters. One question asks about the ability to impute motives to fictional characters, which suggests some overlap with the AQ_Social subscale.

Interestingly, the one AQ subscale that exhibits a non-negative correlation (AQ_Detail) emphasizes precision in individual habits and attention to detail. In a strategic setting such as ours, these traits might be expected to partly counteract the negative effects of other typical theory of mind deficits, perhaps accounting for the lack of significant correlation.

Importantly, our survey data exhibit scores in the normal range. Thus, differences in the strategic aspects of theory of mind vary significantly across individuals in the normal range of social intelligence.[34]

Finding 3: Our dynamic measure of *ToM* based on observed learning is significantly correlated with survey measures of theory of mind.

4. Conclusions

This paper presents a theoretical model of the evolution of theory of mind. The model demonstrates the advantages to learning opponents' behavior in simple games of perfect information. A departure from standard game theory is to

[34] Figure F3 in the online appendix displays histograms of our participants' AQ and BAP scores over the range of feasible scores.

allow the outcomes used in the game to be randomly selected from a growing outcome set. We show how sophisticated individuals who recognize agency in others can build up a picture of others' preferences while naive players who react only to the complete game remain in the dark. We impose plausible conditions under which sophisticated individuals who choose the *SPE* action will dominate all other types of individual, sophisticated or naive, in the long run.

We then perform experiments measuring the ability of real-world individuals to learn the preferences of others in a strategic setting. The experiments implement a simplified version of the theoretical model, using a two-stage game where each decision node involves two choices. We find 1) evidence of highly significant learning of opponents' preferences over time, but not of complete games, and 2) significant correlations between behavior in these experiments and responses to two well-known survey instruments measuring theory of mind from psychology. This validates the use of the term "theory of mind" in the present context. Indeed, the experiments here raise the interesting possibility of developing a test for autism that is behavioral rather than purely verbal.

In economics, theory of mind is implicated, in particular, as driving behavior in social settings involving reciprocity and mutualistic gains from exchange (see, for example, McCabe et al., 2003, and Izuma et al. 2011). Theory of mind is crucial here because individuals are thought to condition their behavior on others' beliefs and intentions, and presumes imputing preferences to those others.

We show that the essential capacity to attribute preferences to others is theoretically evolutionarily plausible and actually present in the population to a varying degree. Other social phenomena that assume the presence of theory of mind then gain firmer footing, and so an indirect contribution of our work is to set the stage for future research on such phenomena.

References

[1] **Simon Baron-Cohen, Alan M. Leslie & Uta Frith** (**1985**): Does the Autistic Child Have a 'Theory of Mind?'. *Cognition*, 21(1), 37–46.

[2] **Simon Baron-Cohen, Sally Wheelwright, Richard Skinner, Joanne Martin & Emma Clubley** (**2001**): The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Males and Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders*, 31(1), 5–17.

[3] **Meghana Bhatt & Colin F. Camerer** (**2005**): Self-referential Thinking and Equilibrium as States of Mind in Games: fMRI Evidence. *Games and Economic Behavior*, 52(2), 424–459.

[4] **Patrick Billingsley** (**1968**): *Convergence of Probability Measures*. 2nd edition, Chicago: John Wiley and Sons.

[5] **Colin F. Camerer, Teck-Hua Ho & Juin-Kuan Chong** (**2002**): Sophisticated Experience-Weighted Attraction Learning and Strategic Teaching in Repeated Games. *Journal of Economic Theory*, 104(1), 137–188.

[6] **Gary B. Charness & Matthew Rabin** (**2002**): Understanding Social Preferences With Simple Tests. *Quarterly Journal of Economics*, 117(3), 817–869.

[7] **Dorothy L. Cheney & Robert M. Seyfarth** (**1990**): *How Monkeys See the World: Inside the Mind of Another Species*. Chicago: University of Chicago Press.

[8] **Vincent P. Crawford & Nagore Iriberri** (**2007**): Level-k Auctions: Can a Non-Equilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions? *Econometrica*, 75(6), 1721–1770.

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

[10] **Johann Elker, David Pollard & Winfried Stute** (**1979**): Glivenko-Cantelli Theorems for Classes of Convex Sets. *Advances in Applied Probability*, 11(4), 820–833.

[11] **Lawrence E. Fouraker & Sidney Siegel** (**1963**): *Bargaining Behavior*. McGraw Hill.

[12] **Jacob K. Goeree, Thomas R. Palfrey, Brian W. Rogers & Richard D. McKelvey** (**2007**): Self-Correcting Information Cascades. *The Review of Economic Studies*, 74(3), 733–762, URL `http://restud.oxfordjournals.org/content/74/3/733.abstract`.

[13] **John C. Harsanyi** (**1967-68**): Games with Incomplete Information Played by 'Bayesian' Players, I-III. *Management Science*, 14, 159–182, 320–334, 486–502.

[14] **Robert S.E. Hurley, Molly Losh, Morgan Parlier, Stephen J. Reznick & Joseph Piven** (**2007**): The Broad Autism Phenotype Questionnaire. *Journal of Autism and Developmental Disorders*, 37(9), 1679–1690.

[15] **Keise Izuma, Kenji Matsumoto, Colin F. Camerer & Ralph Adolphs** (**2011**): Insensitivity to Social Reputation in Autism. *Proceedings of the National Academy of Sciences*, 108(42), 17302–17307, URL `http://www.pnas.org/content/108/42/17302.abstract`.

[16] **Daniel T. Knoepfle, Colin F. Camerer & Joseph T. Wang** (**2009**): Studying Learning in Games Using Eye-tracking. *Journal of the European Economic Association*, 7(2-3), 388–398.

[17] **Kevin A. McCabe, Stephen J. Rassenti & Vernon L. Smith** (**1998**): Reciprocity, Trust, and Payoff Privacy in Extensive Form Bargaining. *Games and Economic Behavior*, 24(1), 10–24.

[18] **Kevin A. McCabe, Mary L. Rigdon & Vernon L. Smith** (**2003**): Positive Reciprocity and Intentions in Trust Games. *Journal of Economic Behavior and Organization*, 52(2), 267–275.

[19] **Erik Mohlin** (**2012**): Evolution of Theories of Mind. *Games and Economic Behavior*, 75(1), 299–318.

[20] **Jörg Oechssler & Burkhard Schipper** (**2003**): Can You Guess the Game You are Playing? *Games and Economic Behavior*, 43(1), 137–152.

[21] **Kristine H. Onishi & René Baillargeon** (**2005**): Do 15-Month-Old Infants Understand False Beliefs? *Science*, 308(8), 255–258.

[22] **Benedikt M. Pötscher & Ingmar R. Prucha** (**1994**): Generic Uniform Convergence and Equicontinuity Concepts for Random Functions. *Journal of Econometrics*, 60(1), 23–63.

[23] **R Development Core Team** (**2013**): *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, URL `http://www.R-project.org/`, ISBN 3-900051-07-0.

[24] **Nikolaus Robalino & Arthur J. Robson** (**2012**): The Economic Approach to "Theory of Mind". *Philosophical Transactions of the Royal Society, Biological Sciences*, 367, 2224–2233.

[25] **Arthur J. Robson** (**2001**): Why Would Nature Give Individuals Utility Functions? *Journal of Political Economy*, 109(4), 900–914.

[26] **Arthur J. Robson & Hillard Kaplan** (**2003**): The Evolution of Human Longevity and Intelligence in Hunter-Gatherer Economies. *American Economic Review*, 93(1), 150–169.

[27] **Vernon L. Smith** (**1982**): Microeconomic Systems as an Experimental Science. *American Economic Review*, 72(5), 923–955.

[28] **Dale O. Stahl** (**1993**): Evolution of Smart$_n$ Players. *Games and Economic Behavior*, 5(4), 604–617.

Appendices (for online publication)

A. Proofs of the Theorems

A1-A4 are assumed throughout this appendix.

A.1. **Proof of Theorem 1.**

The first result here is Lemma 2, which verifies the claims of Theorem 1 regarding cases where either L_{it} or γ_t converge to zero.

Lemma 2: *Each of the following is true.*

i) Suppose $\alpha \in [0, 2)$. *Then* $L_{it} \longrightarrow 0$ *surely for each player role* $i = 1, \ldots, I$.

ii) Suppose the extensive form has T *terminal nodes. If* $\alpha \in [0, T)$, *then* $\gamma_t \longrightarrow 0$ *surely.*

Proof. Clearly $L_{it} \leqslant (|Z_t| + t \cdot 2T)/|Z_t|^2$ everywhere, since the maximal number of binary preference orderings that can be revealed for any player at any date is bounded above by $2T$. Similarly, since only one game is played in each period, $\gamma_t \leqslant t/|Z_t|^T$ surely. Since $|Z_t| = |Z_1| + k$ whenever $\lfloor (|Z_1| + k)^\alpha \rfloor \leqslant t < \lfloor (|Z_1| + k + 1)^\alpha \rfloor$, it follows that $t < (|Z_t| + 1)^\alpha$. Hence,

$$L_{it} < 1/|Z_t| + 2T \cdot \lfloor |Z_t| + 1 \rfloor^\alpha / |Z_t|^2 \quad \text{and} \quad \gamma_t < \lfloor |Z_t| + 1 \rfloor^\alpha / |Z_t|^T.$$

Surely $|Z_t| \longrightarrow \infty$. This completes the proof as the previous indented expression then implies that if $\alpha < 2$, then $L_{it} \longrightarrow 0$ surely, for instance. ■

In order to complete the proof of Theorem 1 it will next be proved that if $\alpha > 2$, then all players' preferences are revealed in the limit, i.e., $L_{it} \longrightarrow 1$, for each $i = 1, \ldots, I$. (The proof that γ_t converges to one when $\alpha > T$ proceeds along similar lines, and is thus omitted.) First a required notation—

Definition 5: *Let the random variable* K_{it} *denote the number of pairs of outcomes* $(z, z') \in Z_t \times Z_t$ *such that* i*'s preferences over* $\{z_i, z_i'\}$ *have been revealed to the ToMs along* H_t. *Let* K_{it} *range from zero to* $|Z_t|^2$, *the total number of pairs of outcomes available at date* t. *Hence* $L_{it} = K_{it}/|Z_t|^2$.

Note that, given $|Z_t|$ outcomes, there are $|Z_t|(|Z_t| - 1|)/2$ pairs of distinct outcomes available. This is the actual number of binary choices to be learned

for each $i \in I$. For convenience, however, we define K_{it} to count *all* of the $|Z_t|^2$ possible pairs. This generates a more concise expression for L_{it}, with $|Z_t|^2$ in the divisor, rather than $|Z_t|(|Z_t| - 1|)/2$. It is harmless provided we assume—as we will throughout—that K_{it} automatically includes all elements along the diagonal of $Z_t \times Z_t$, and that each revelation of a role i preference increases the count K_{it} by two (since the mirror image of each pair $(z, z') \in Z_t \times Z_t$ is also in $Z_t \times Z_t$).

The desired result is established by induction using the following two results.

LEMMA 3: *Suppose $\alpha > 2$. Consider the player role $i \geqslant 1$. In case $i > 1$ suppose that L_{jt} converges to one in probability for each $j < i$. Then, for each $\xi \in [0, 1]$ there is a sequence of random variables $\{\theta_{it}(\xi)\}$, non-increasing between arrival dates, such that*

$$E(K_{it+1} \mid H_t) - K_{it} \geqslant [\xi \cdot (1 - L_{it} - \theta_{it}(\xi))]^{A^{I-i}}, \text{ where } 1 - L_{it} - \theta_{it}(\xi) \geqslant 0,$$

for all $t \geqslant 1$. Furthermore, there is a non-random function $\theta_i(\xi) \geqslant 0$, which converges to zero as ξ tends to zero, such that $P\{\theta_{it}(\xi) - \theta_i(\xi) > \epsilon\} \longrightarrow 0$ as t tends to infinity, for each $\xi \in [0, 1]$ and each $\epsilon > 0$.

That is, in the limit, the probability of revealing new information about role *1* preferences is small *only if* the fraction of extant knowledge about *1* preferences, L_{1t}, is close to one. Similarly for role $i > 1$, provided L_{jt} converges to one for each $j < i \leqslant I$.[35]

LEMMA 4: *Consider the $i \in I$ player role. Suppose $\alpha > 2$. Suppose further that for each $\xi \in [0, 1]$ there is a sequence of random variables $\{\theta_{it}(\xi)\}$ such that*

$$E(K_{it+1} \mid H_t) - K_{it} \geqslant [\xi \cdot (1 - L_{it} - \theta_{it}(\xi))]^{A^{I-i}}$$

for all $t \geqslant 1$, where each random process $\{\theta_{it}(\xi)\}$ is as is stated in Lemma 3. Then, L_{it} converges in probability to one.

---

[35] Indeed, the probability of revealing new information about i is clearly small whenever $1 - L_{it}$ is small. The converse is not as obviously true. Lemma 3, however, provides an appropriate bound. It decomposes $E(K_{it+1} \mid H_t) - K_{it}$ into a factor of $1 - L_{it}$, which accounts for information yet to be revealed about i preferences, and a residual $\theta_{it}(\xi)$. The residual arises, for example, from i-type subgames in which i player choice does not reveal information because it is unclear what i players believe about the remaining players' choices.

Proofs of Lemmas 3 and 4 are given below (in sections A.1.1 and A.1.2, respectively).

To see now that Lemmas 3 and 4 deliver the desired result proceed by induction. Recall the manner in which the player roles here are enumerated. The I players move first, then the $I-1$ role players, and so on, with the 1 role players moving last. Suppose $\alpha > 2$. Fix an $i > 1$, and suppose $L_{jt} \longrightarrow 1$ in probability, for each $j < i$. Lemmas 3 and 4 together imply that $L_{it} \longrightarrow 1$. The result then follows by reapplying the same two lemmas in the case $i = 1$ to give that L_{1t} converges to one in probability.

A.1.1. **Proof of Lemma 3.**

The proof here is given for the cases in which $i > 1$. The proof of the result for $i = 1$ follows with minor adjustments to the notation used here and will thus be omitted.

Fix a player role $i > 1$. In establishing Lemma 3 we will keep track of some of the forward dominance-solvable i player subgames. Recall that each player role i subgame shares the same underlying tree, one in which i has A moves at his information set, $i-1$ has A moves at each of his information sets, and so on. With that in mind, fix two end-nodes of this i role subtree and consider subgames in which particular outcomes are available at these nodes. In particular, enumerate the terminal nodes of the i subtree as follows. Fix distinct actions $a_1, a_2 \in A$. Name "one" an end-node reachable after i chooses a_1, and label "two" one of the end-nodes reachable after i chooses a_2. Enumerate the remaining end-nodes "3" through "A^i" in some arbitrary way. This enumeration of the terminal nodes of the i role subtree will be implicit throughout this section.

DEFINITION 6: *Let $Q_i(z, z')$ be the set of i player subgames, in the full space of i role subgames Z^{A^i}, with outcome z at end-node "one", and z' at end-node "two" that satisfy additionally the following. For each game in $Q_i(z, z')$ the $i-1$ player subgames following i's choice of a_1 and a_2 are forward dominance-solvable (uniquely) resulting in z, and z', respectively, and moreover one of the actions a_1, a_2 is uniquely dominant for the i players themselves.* [36]

---

[36] In proving Lemma 3 for $i = 1$, define $Q_1(z, z')$ as the 1 player subgames with z available at a_1 and z' at a_2, with player 1's are not indifferent between z and z'.

DEFINITION 7: *Let Q_{it} denote all the i player subgames possible at date t,*[37] *and denote by $Q^*_{it} \subseteq Q_{it}$ the subgames for which all the relevant $1, \ldots, i-1$ player preferences have been revealed along H_t.*

DEFINITION 8: *Let N_{it} denote the pairs of outcomes (z, z') available in period t such that i's favored outcome among z and z' has not been revealed along H_t.*

The following describes events that are sure to deliver revelations about i player preferences. Suppose a subgame $q \in Q_i(z, z') \cap Q^*_{it}$ is reached where $(z, z') \in N_{it}$. All *ToM*s will then make the same prediction regarding the outcomes obtained in the subgames of q after i players choose either a_1 or a_2. Every *ToM* knows this, knows that every *ToM* knows, and so on. Suppose, for the sake of argument, that $z_i > z'_i$. By A4 every i player reaching q will there choose a_1. *ToM* players will then infer that $z_i > z'_i$, since if it were the case that $z_i \leqslant z'_i$, a positive fraction of the *ToM*s in role i would have chosen some $a \neq a_1$ rather than a_1.

By the above discussion it follows that the fraction of i subgames, among those in Q_{it}, at which i players are sure to reveal new information about their preferences is bounded below by $\sum_{N_{it}} |Q_i(z, z') \cap Q^*_{it}|/|Q_{it}|$. The set of games at date t is just the T-times product of Z_t, and each game is drawn uniformly from this set. The empirical distribution over games realized at date t can then be replicated by drawing A^{I-i} i player subgames uniformly and independently from Q_{it}. Therefore,

$$E(K_{it+1} - K_{it} \,|\, H_t) \geqslant \left(\sum_{(z,z') \in N_{it}} \frac{|Q_i(z, z') \cap Q^*_{it}|}{|Q_{it}|} \right)^{A^{I-i}} . \tag{1}$$

The bound is conservative, obtained by considering the case in which i players reveal new information at every i information set.

Consider now some additional required notation.

DEFINITION 9: *Let $\boldsymbol{Z}$, and $\boldsymbol{Z}_t$ denote the $[A^i - 2]$-times products of Z, and Z_t, respectively. Write $E_t(z, z') = |Q_i(z, z') \cap Q_{it}|/|\boldsymbol{Z}_t|$. This is the date t probability of drawing from Q_{it} an i role subgame that belongs to $Q_i(z, z')$, conditional on the subgame having z at end-node "one" and z' at end-node "two". Denote*

[37] Recall that Q_{it} can be identified with the A^i times product of Z_t.

by $E(z,z')$ the probability of drawing from Z^{A^i}—according to the distribution f^{A^i}—an i player subgame that belongs to $Q_i(z,z')$, conditional on the subgame having z at "one" and z' at "two".

The following result will be needed.

CLAIM 1: *$E_t(z,z')$ almost surely converges uniformly to $E(z,z')$. That is, almost surely* $\sup_{(z,z')\in Z\times Z}|E_t(z,z')-E(z,z')|\longrightarrow 0$.

Proof. Theorem 3.1 from Potscher and Prucha (1994) states that if E_t almost surely converges pointwise to E on $Z\times Z$, if E is continuous, and if the sequence E_t is almost surely asymptotically uniformly equicontinuous on $Z\times Z$, then E_t almost surely converges uniformly to E on $Z\times Z$. The equicontinuity condition (from Potscher and Prucha 1994) is

$$\limsup_{t\longrightarrow\infty}\sup_{(z,z')\in Z\times Z}\left\{\sup_{(x,x')\in B(z,z',\eta)}|E_t(z,z')-E_t(x,x')|\right\}\longrightarrow 0 \text{ a.s. as } \eta\longrightarrow 0,$$

where $B(z,z',\eta)$ denotes the open ball with radius η centered at (z,z'). Lemma 1 delivers the required pointwise convergence of $E_t(z,z')$ to $E(z,z')$. A2 gives that $E(z,z')$ is continuous. The proof is then completed by verifying the above equicontinuity condition.

Let $N(z,z',\eta)$ denote the set of outcome profiles $\boldsymbol{x}\in\boldsymbol{Z}$ such that for every coordinate, $y=(y_1,\ldots,y_i)$, of $\boldsymbol{x}$, and for each $j\leqslant i$, $|y_j-z_j|>\eta$ and $|y_j-z'_j|>\eta$. Consider $(x,x')\in B(z,z',\eta)$ and $\boldsymbol{x}\in N(z,z',\eta)$. For each coordinate y of $\boldsymbol{x}$, role $j\leqslant i$ prefers z_j to y_j if and only if he prefers x_j to y_j (similarly for the z' and x' outcomes). It follows that the subgame $(z,z',\boldsymbol{x})$ satisfies the forward dominance-solvability conditions characterizing subgames in $Q_i(z,z')$ if and only if $(x,x',\boldsymbol{x})$ satisfies the dominance-solvability conditions of subgames in $Q_i(x,x')$. (We use the fact that every subgame in $Q_i(z,z')$ can be uniquely identified with a profile of payoffs, $(z,z',\boldsymbol{x})$, for some $\boldsymbol{x}\in\boldsymbol{Z}$.) It then follows that $|Q_i(z,z')\cap N(z,z',\eta)|=|Q_i(x,x')\cap N(z,z',\eta)|$ whenever $(x,x')\in B(z,z',\eta)$. Now recall that $E_t(z,z')=|Q_i(z,z')\cap Q_{it}|/|\boldsymbol{Z}_t|$. Then, writing $Q_i(z,z')$ in this expression as the union of the disjoint sets $Q_i(z,z')\cap N(z,z',\eta)$ and $Q_i(z,z')\setminus N(z,z',\eta)$[38]

yields the following bound for each $(z, z') \in Z \times Z$, and $\eta > 0$,

$$\sup_{(x,x') \in B(z,z',\eta)} |E_t(z, z') - E_t(x, x')| \leqslant |\boldsymbol{Z}_t \setminus N(z, z', \eta)|/|\boldsymbol{Z}_t|.$$

Lemma 1 and A2 deliver the desired equicontinuity since together they imply that $|\boldsymbol{Z}_t \setminus N(z, z', \eta)|/|\boldsymbol{Z}_t|$ almost surely converges to zero as η tends to zero, uniformly in (z, z'). ■

Now recall the lower bound obtained in (1)—in particular the summand in the parentheses there. Keeping in mind the terms defined in Definition 7 we obtain,

$$\begin{aligned} &\frac{|Q_i(z, z') \cap Q^*_{it}|}{|\boldsymbol{Z}_t|} = \\ &\frac{|Q_i(z, z') \cap Q_{it}|}{|\boldsymbol{Z}_t|} - \frac{|Q_i(z, z') \cap Q_{it} \setminus Q^*_{it}|}{|\boldsymbol{Z}_t|} = E(z, z') - \phi_t(z, z'), \end{aligned} \tag{2}$$

where we have defined,

$$\phi_t(z, z') \equiv E(z, z') - E_t(z, z') + \frac{|Q_i(z, z') \cap Q_{it} \setminus Q^*_{it}|}{|\boldsymbol{Z}_t|}.$$

Write $\boldsymbol{S}(\xi) = \{(z, z') \in Z \times Z : E(z, z') < \xi\}$. Equation (2) gives—we use here the fact that $|Q_{it}| = |Z_t|^2 \cdot |\boldsymbol{Z}_t|$, and also that $E(z, z') - \phi_t(z, z') \geqslant 0$—

$$\begin{aligned} \sum_{(z,z') \in N_{it}} \frac{|Q_i(z, z') \cap Q^*_{it}|}{|Q_{it}|} &\geqslant \frac{1}{|Z_t|^2} \sum_{N_{it} \setminus \boldsymbol{S}(\xi)} \big(E(z, z') - \phi_t(z, z')\big) \\ &\geqslant \xi \cdot \left(\frac{|N_{it} \setminus \boldsymbol{S}(\xi)|}{|Z_t|^2} - \frac{1}{\xi} \times \sum_{N_{it} \setminus \boldsymbol{S}(\xi)} \frac{\phi_t(z, z')}{|Z_t|^2} \right) \\ &= \xi \cdot \left(\frac{|N_{it}|}{|Z_t|^2} - \frac{|N_{it} \cap \boldsymbol{S}(\xi)|}{|Z_t|^2} - \frac{1}{\xi} \times \sum_{N_{it} \setminus \boldsymbol{S}(\xi)} \frac{\phi_t(z, z')}{|Z_t|^2} \right). \end{aligned} \tag{3}$$

Define $\theta_{it}(\xi)$ as in the statement of the result as

$$\theta_{it}(\xi) = \frac{|N_{it} \cap \boldsymbol{S}(\xi)|}{|Z_t|^2} + \frac{1}{\xi} \times \sum_{N_{it} \setminus \boldsymbol{S}(\xi)} \frac{\phi_t(z, z')}{|Z_t|^2}.$$

---

[38] and, for $E_t(x, x')$, expressing $Q_i(x, x')$ as the union of $Q_i(x, x') \cap N(z, z', \eta)$ and $Q_i(x, x') \setminus N(z, z', \eta))$

Next, observe that $|N_{it}|/|Z_t|^2 = 1 - L_{it}$. Going back then to equation (1), and using (3), delivers

$$E(K_{i,t+1} \mid H_t) - K_{it} \geqslant \left[\xi \cdot (1 - L_{it} - \theta_{it}(\xi))\right]^{A^{I-i}}.$$

Note here that $1 - L_{it} - \theta_{it}(\xi) \geqslant 0$ as the expression is obtained by summing the terms $E(z, z') - \phi_t(z, z')$, which are non-negative, over the set $N_{it} \setminus \boldsymbol{S}(\xi)$.

It remains now to show that $\theta_{it}(\xi)$ satisfies the properties given in the statement of the lemma. With that in mind, first observe that $E(z, z')$ and $E_t(z, z')$ are constant in between arrival dates, and that the set $Q_{it} \setminus Q^*_{it}$ is non-increasing between arrival dates. For each (z, z'), $\phi_t(z, z')$ is then non-increasing between arrival dates. The set N_{it} is also non-increasing at these dates. It follows then that $\theta_{it}(\xi)$ is non-increasing between arrival dates as required.

Next, observe that if $L_{jt} \longrightarrow 1$ in probability for each $j < i$, then Claim 1 implies that $\theta_{it}(\xi) - |N_{it} \cap \boldsymbol{S}(\xi)|/|Z_t|^2$ converges in probability to 0.[39] Let $\theta_i(\xi)$ from the statement of the lemma be $\int_{\boldsymbol{S}(\xi)} f(z)f(z')dzdz'$. Clearly, $\theta_i(\xi) \geqslant 0$ for each $\xi \in [0, 1]$. A2 gives that $\theta_i(\xi)$ converges to zero as ξ tends to zero. In order to see that $P\{\theta_{it}(\xi) - \theta_i(\xi) > \epsilon\}$ tends to zero for each $\epsilon > 0$, as required, note that $|N_{it} \cap \boldsymbol{S}(\xi)| \leqslant |\{Z_t \times Z_t\} \cap \boldsymbol{S}(\xi)|$, and that Lemma 1 implies $|\{Z_t \times Z_t\} \cap \boldsymbol{S}(\xi)|/|Z_t|^2$ almost surely converges to $\theta_i(\xi)$.

A.1.2. **Proof of Lemma 4.**

Fix a player role $i \in I$ and assume the hypotheses of Lemma 4. In the remainder we suppress the i subscripts whenever it is possible to do so without confusion.

The proof is given in two parts. The first shows that L_t converges in probability to a random variable L. The second establishes that L equals one a.e. In order to prove the convergence of L_t we show that when $\alpha > 2$ these processes belongs to a class of generalized sub-martingales with the sub-martingale convergence property. In particular, we use the following definition and result in

---

[39] That is, we have $\left|\sum_{N_{it} \setminus \boldsymbol{S}(\xi)} \frac{\phi_t(z,z')}{|Z_t|^2}\right| \leqslant \sum_{Z_t \times Z_t} \frac{|\phi_t(z,z')|}{|Z_t|^2} \leqslant \sum_{Z_t \times Z_t} \frac{|E_t(z,z') - E(z,z')|}{|Z_t|^2} + \frac{|Q_{it} \setminus Q^*_{it}|}{|Q_{it}|}$. The summation on the RHS converges to zero by Claim 1. The second term on the RHS converges to zero in probability as long as $L_{jt} \longrightarrow 1$ in probability for each $j < i$.

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 submartingale in the limit (w-submil)*** *if almost surely, for each* $\eta > 0$*, there is a* n *such that* $\tau \geqslant t \geqslant n$ *implies* $P\{E(L_\tau \mid H_t) - L_t \geqslant -\eta\} > 1 - \eta$*. If* L_t *is an integrable w-submil, then there exists a random variable* L *such that* $L_t \longrightarrow L$ *in probability.*

Part 1: L_t *converges in probability to a random variable* L*.* In view of the w-submil convergence result, as a first step, we prove that the arrival date subsequence $\{L_{t_k}\}$ is a w-submil under the hypotheses of the lemma. Toward that end, consider consecutive arrival dates t^*, τ^*, with $\tau^* > t^*$. By the definition of K_t (Definition 5),

$$L_{\tau^*} - L_{t^*} = \frac{1}{|Z_{\tau^*}|^2} \sum_{t=t^*}^{\tau^*-1} [K_{t+1} - K_t] - \frac{|Z_{\tau^*}|^2 - |Z_{t^*}|^2}{|Z_{\tau^*}|^2} \cdot L_{t^*}. \tag{4}$$

Then, by the hypotheses of the claim being proved, for each $\xi \in [0, 1]$,

$$\begin{aligned} \sum_{t=t^*}^{\tau^*-1} E(K_{t+1} - K_t \mid H_{t^*}) &\geqslant \sum_{t=t^*}^{\tau^*-1} [\xi \cdot E(1 - L_t - \theta_{it}(\xi) \mid H_{t^*})]^{A^{I-i}} \\ &> [\tau^* - t^*] \cdot [\xi \cdot E(1 - L_{\tau^*-1} - \theta_{it^*}(\xi) \mid H_{t^*})]^{A^{I-i}}. \end{aligned} \tag{5}$$

The first line uses Jensen's inequality. The second uses the fact that L_t is non-decreasing between arrival dates and that $\theta_{it}(\xi)$ is non-increasing between arrival dates (see the definition of $\theta_{it}(\xi)$ in the statement of Lemma 3).

Combining equations (4) and (5) (after taking the conditional expectation in (4)) yields

$$\begin{aligned} &E(L_{\tau^*} \mid H_{t^*}) - L_{t^*} < 0 \implies \\ &[\xi \cdot E(1 - L_{\tau^*-1} - \theta_{it^*}(\xi) \mid H_{t^*})]^{A^{I-i}} < \frac{|Z_{\tau^*}|^2}{\tau^* - t^*} \cdot \frac{|Z_{\tau^*}|^2 - |Z_{t^*}|^2}{|Z_{\tau^*}|^2} \cdot L_{t^*}. \end{aligned}$$

Solving for $E(L_{\tau^*-1} \mid H_{t^*})$ in the last indented inequality, and then using the fact that surely $E(L_{\tau^*} \mid H_{t^*}) \geqslant E(L_{\tau^*-1} \mid H_{t^*}) \cdot |Z_{t^*}|^2/|Z_{\tau^*}|^2$ gives—

$$E(L_{\tau^*} \mid H_{t^*}) - L_{t^*} < 0 \implies$$
$$E(L_{\tau^*} \mid H_{t^*}) > \frac{|Z_{t^*}|^2}{|Z_{\tau^*}|^2}\left[1 - \theta_{it^*}(\xi) - \frac{1}{\xi}\cdot\left(\frac{|Z_{\tau^*}|^2 - |Z_{t^*}|^2}{\tau^* - t^*}\right)^{\frac{1}{A^{I-i}}}\right] \equiv 1 - A_{t^*}(\xi). \tag{6}$$

Note here that we have implicitly defined the new variable $A_t(\xi)$.

In the remainder hatted variables will be used to denote variables sampled at arrival dates, e.g., $\hat{L}_k = L_{t_k}$.

We next use (6) to establish the w-submil condition: For each $\eta > 0$ there exists an N such that for all arrival dates t_m, t_n, such that $n > m \geqslant N$,

$$P\{E(\hat{L}_n \mid \hat{H}_m) - \hat{L}_m \geqslant -\eta\} > 1 - \eta. \tag{7}$$

To that end, suppose $E(\hat{L}_n \mid \hat{H}_m) < \hat{L}_m$ for some integers m and n such that $n > m$. 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 k, with $m \leqslant k < n - 1$, such that $E(\hat{L}_{k+1} \mid \hat{H}_m) < E(\hat{L}_k \mid \hat{H}_m)$. Let r be the largest integer in $\{m, , \ldots, n-1\}$ for which this is the case, i.e., $E(\hat{L}_{k+1} \mid \hat{H}_m) \geqslant E(\hat{L}_k \mid \hat{H}_m)$, for each $k = r+1, \ldots, n-1$. According to (6), $E(\hat{L}_{r+1} \mid \hat{H}_m) > 1 - E(\hat{A}_r(\xi) \mid \hat{H}_m)$. Hence, since L_t is everywhere bounded above by one it follows that

$$E(\hat{L}_n \mid \hat{H}_m) - \hat{L}_m > -E(\hat{A}_r(\xi) \mid \hat{H}_m).$$

Recall now the hypothesis, $P\{\theta_{it}(\xi) - \theta_i(\xi) > \epsilon\} \longrightarrow 0$, for all $\xi \in [0, 1]$ and $\epsilon > 0$, where $\theta_i(\xi)$ is as in the statement of the lemma. Consider this in the following form: $P\{-\theta_{it}(\xi) \geqslant -\theta_i(\xi) - \epsilon\} \longrightarrow 1$, for each $\epsilon > 0$. We can replace the random variable $\theta_{it}(\xi)$ with $A_t(\xi)$ in this limit since $\theta_{it}(\xi) - A_t(\xi)$ converges surely to zero for each $\xi \in [0, 1]$. To see this (referring to (6) where $A_t(\xi)$ is defined) first note that $|Z_{t^*}|^2/|Z_{\tau^*}|^2$ converges surely to one. Then, recall that t^* and τ^* are consecutive arrival dates, and observe that

$$\frac{|\hat{Z}_{k+1}|^2 - |\hat{Z}_k|^2}{t_{k+1} - t_k} = \frac{(|Z_1| + k + 1)^2 - (|Z_1| + k)^2}{\lfloor(|Z_1| + k + 1)^\alpha\rfloor - \lfloor(|Z_1| + k)^\alpha\rfloor},$$

which converges surely to zero as k tends to infinity whenever $\alpha > 2$.

Now, given that $P\{-A_t(\xi) \geqslant -\theta_i(\xi) - \epsilon\} \longrightarrow 1$, for each $\epsilon > 0$, we can choose an arrival $N(\xi)$ large enough so that

$$P\{-E(\hat{A}_k(\xi) \,|\, \hat{H}_m) \geqslant -2 \cdot \theta_i(\xi)\} > 1 - 2 \cdot \theta_i(\xi)$$

for all k and m with $k > m \geqslant N(\xi)$, and thus for $k > m \geqslant N(\xi)$,

$$P\{E(\hat{L}_n \,|\, \hat{H}_m) - \hat{L}_m > -2 \cdot \theta_i(\xi)\} > 1 - 2 \cdot \theta_i(\xi).$$

By assumption, $\theta_i(\xi)$ converges to zero as ξ approaches zero. Hence (7) can be obtained by choosing ξ in the previous indented equation so that $\theta_i(\xi) < \eta/2$ establishing that the sequence $\{\hat{L}_k\}$ is a w-submil.

Having shown that $\{\hat{L}_k\}$ is a w-submil, it remains to verify that $\{L_t\}$ is also a w-submil. With that in mind, consider any dates t and τ where $t < \tau$. Let t^* denote the first arrival date after t and let τ^* denote the greatest arrival date less than or equal to τ. Then, $L_t \leqslant L_{t^*} \cdot [|Z_t| + 1]^2/|Z_t|^2$ everywhere, and thus

$$L_\tau - L_t \geqslant L_{\tau^*} - L_{t^*} \cdot [|Z_t| + 1]^2/|Z_t|^2$$

everywhere. The w-submil convergence result implies $L_{\tau^*} - L_{t^*} \longrightarrow 0$ in probability. Furthermore, $[|Z_t| + 1]^2/|Z_t|^2 \longrightarrow 1$ surely. Hence the right-hand side of the last indented expression converges to zero in probability establishing that $\{L_t\}$ is a w-submil.

Part 2: L_t converges to one in probability. Let L denote the limit, in probability, of L_t. By the hypotheses of Lemma 4 (defining here $K_o \equiv 0$, and invoking Jensen's inequality),

$$\begin{aligned} E(L_t) &= \sum_{t=0}^{\tau-1} E\left(K_{t+1} - K_t\right)/|Z_\tau|^2 \\ &\geqslant \frac{\xi^{A^{I-i}} \cdot \tau}{|Z_\tau|^2} \cdot \frac{1}{\tau} \sum_{t=1}^{\tau-1} \cdot E\left(1 - L_{it} - \theta_{it}(\xi)\right)^{A^{I-i}}. \end{aligned} \tag{8}$$

Recall that $\tau/|Z_\tau|^2 \longrightarrow \infty$ whenever $\alpha > 2$.[40] Since L_t is everywhere bounded by one, and since $1 - L_{it} - \theta_{it}(\xi) \geqslant 0$, equation (8) implies that for each $\xi \in (0, 1]$,

$$\lim_{\tau \longrightarrow \infty} \frac{1}{\tau} \cdot \sum_{t=1}^{\tau-1} E\left(1 - L_{it} - \theta_{it}(\xi)\right)^{A^{I-i}} = 0.$$

By the hypotheses of the lemma being proved each $E(1-L_{it}-\theta_{it}(\xi))$ term can be made arbitrarily close to $E(1-L_{it})$ by choosing $\xi > 0$ appropriately. But $E(1-L_{it})$ converges to $E(1-L)$. Thus the limit in the previous indented equation implies that $E(1-L) = 0$, and hence L must equal one almost everywhere.

A.2. **Proof of Theorem 2.**

Recall A2 describing the cdf F on the payoff space $[m, M]^I$ and the implied cdf for games given by G, on the payoff space $[m, M]^{IT}$.

DEFINITION 10: *Let μ denote the measure on games induced by F. In particular, for each measurable $S \subseteq Q$, $\mu(S) = \int_{q \in S} dG(q)$. Let μ_t denote the corresponding empirical measure. That is, $\mu_t(S) = |S \cap Q_t|/|Q_t|$.*

We establish the result of Theorem 2 by showing that the ratio of the population of any alternative type to that of the *SPE-ToM* type tends to zero in probability.[41] If the alternative type is a *ToM* type that differs from *SPE-ToM* only a set of μ measure zero, it should simply be identified with *SPE-ToM*. It also follows that $\mu(\bar{S}) > 0$ where $\bar{S}$ is the set of games for which player role i has no dominant choice at any node.[42] However, the set of games for which player role i has a dominant choice at some but not all nodes also has positive μ measure. For simplicity, we then rule out the possibility that the alternative *ToM* type differs from *SPE-ToM* with positive probability *only* on this set and agrees with it with probability one on $\bar{S}$.

We recall a key hypothesis of Theorem 2—

A6: *For each $i > 1$, every alternative i ToM type differs from the SPE-ToM at every i decision node in a set of games S with positive μ measure.*

---

[40] Recall that $|Z_t| = |Z_1| + k$ whenever $\lfloor(|Z_1| + k)^\alpha\rfloor \leqslant t < \lfloor(|Z_1| + k + 1)^\alpha\rfloor$ therefore $t/|Z_t|^2 \geqslant |Z_t|^\alpha/|Z_t|^2$.

[41] Recall there is a finite number of types.

[42] This follows since any game with a dominant choice at some node for i can be mapped to a game for which this is not true by swapping an outcome in the dominant set of outcomes with an outcome that is not in this set.

That is, in the limit, the alternative type will differ from the *SPE-ToM* on a set of games that occur with positive probability. What about the naive alternative types? Any such naive type differs from the *SPE-ToM* type on $\bar{S}$ *given* that the game is new. That the game is new will be assured with probability that tends to one, so we effectively assign $S = \bar{S}$ in this conditional sense.

For the remainder fix a player role $i < 1$ and fix one alternative type to the *SPE-ToM* in role i.

DEFINITION 11: *Let the random variable R_{it} be the fraction of the population in player role i that is SPE-ToM.*

The proof of Theorem 2 is by induction on i. It follows from A4 that $R_{1t} = 1$ is satisfied vacuously. The result is then established by proving that if $R_{jt} \longrightarrow 1$, in probability, $j = 1, ..., i-1$, then R_{it} converges in probability to one. Assume then in what follows that $R_{jt} \longrightarrow 1$, in probability, $j = 1, ..., i-1$.

Consider some prerequisites.

DEFINITION 12: *The random variable $I_t(\delta) \in \{0, 1\}$ is such that $I_t(\delta) = 1$ if and only if the game drawn at date t belongs to the set Q_δ, where Q_δ is the set of games where the minimum absolute payoff difference for any pair of outcomes, for any player is greater than $\delta \geqslant 0$.*

DEFINITION 13: *Define the random variable $D_t \in \{0, 1\}$ such that it satisfies the following. If the alternative is a naive type, then $D_t = 1$ if and only if the game drawn at date t is new, and is such that the game has no dominant strategy at any i decision node. If the alternative is a ToM type, then $D_t = 1$ if and only if at date t the alternative type behaves differently from the SPE-ToM at every i role information set.*

The restrictions that define Q_δ are measurable, so Q_δ itself is measurable. It is an immediate consequence of Lemma 1 that $P\{I_t(\delta) = 1 \mid H_t\}$ almost surely converges to $\mu(Q_\delta)$. Similarly, Lemma 1 implies if the alternative type is a *ToM* type, then $P\{D_t = 1 \mid H_t\}$ almost surely converges to $\mu(S)$.

DEFINITION 14: *The random variable $J_t(\varepsilon) \in \{0, 1\}$ is such that $J_t(\varepsilon) = 1$ if and only if 1) all $1, \ldots, i-1$ player preferences in the game drawn at t have been revealed to the ToM types; and 2) at each role $i-1$ decision node that can*

be reached by role i, the fraction of resulting play that reaches an SPE outcome in that subgame is at least $1-\varepsilon$.

As a key ingredient in the proof consider the following result.

CLAIM 2: *For each sufficiently small $\delta > 0$ and $\varepsilon > 0$ the following results hold given that $J_t(\varepsilon) = 1$ throughout. i) If the alternative type is a ToM type, given $I_t(\delta) = 1$ and $D_t = 1$ as well, then the ratio of the expected payoff of the alternative type to that of the SPE-ToM is at most $1 + \frac{\varepsilon}{1-\varepsilon}\frac{M}{m} - \frac{\delta}{M}$. ii) If the alternative type is naive, given $I_t(\delta) = 1$ and $D_t = 1$ as well, then the ratio of the expected payoff of the alternative type to that of the SPE-ToM is at most $1 + \frac{\varepsilon}{1-\varepsilon}\frac{M}{m} - \left(1 - \frac{1}{A}\right)\frac{\delta}{M}$. iii) Whenever $I_t(0) = 1$, the ratio of the expected payoff of the alternative type—ToM or naive—to that of the SPE-ToM is at most $1 + \frac{\varepsilon}{1-\varepsilon}\frac{M}{m}$.*

Proof. Fix a date t. Assume $I_t(0) = 1$, since this is required in each of the three claims. Let $z(h)$ then be the unique *SPE* payoff in the continuation game defined by the i role information set h, at date t. Let $m(h)$ be the measure of players that reach the i role information set h at date t.

Consider i). Since $J_t(\varepsilon) = 1$, at most a fraction ε of any i player cognitive type is matched with remaining players that do not behave as in the unique pure *SPE*. When matched with these non-*SPE* remaining players, the alternative type's expected payoff is at most M. Since $D_t = 1$, by assumption, the alternative type chooses differently from the *SPE-ToM* at every i information set. The ratio of the expected payoff of the alternative type to that of the *SPE-ToM* is then at most

$$\left((1-\varepsilon)\sum_h m(h)\left(z(h)-\delta\right) + \varepsilon \cdot M\right) \Big/ \left((1-\varepsilon)\sum_h m(h) z(h)\right) .$$

Since $z(h) \in [m, M]$, i) follows. The proof of ii) relies on a similar argument, the factor $1 - 1/A$ arising from naive mixed choice. To establish iii) observe that an alternative type cannot do better than the *SPE-ToM* when matched with remaining players that act as in the unique *SPE*—i.e., set δ in the expression above to zero. ■

From this point on, we focus on the case that the alternative type is *ToM* . (The detailed argument for a naive alternative is nearly identical.) Consider the

realized one period growth rate of the alternative *ToM* type relative to that of the *SPE-ToM* at date t. In view of Claim 2, this rate is bounded above by

$$\begin{aligned} & I_t(\delta)J_t(\varepsilon)D_t \cdot \ln\left(1 + \frac{\varepsilon}{1-\varepsilon}\frac{M}{m} - \frac{\delta}{M}\right) \\ & + I_t(0)(1 - I_t(\delta))J_t(\varepsilon) \cdot \ln\left(1 + \frac{\varepsilon}{1-\varepsilon}\frac{M}{m}\right) \\ & + (1 - I_t(0) + 1 - J_t(\varepsilon)) \cdot \ln(M/m). \end{aligned} \tag{9}$$

To see that the indicator functions here exhaust all possible cases, note first that the first two terms of the expression apply for every case in which $J_t(\varepsilon) = 1$, and $I_t(0) = 1$, x in the light of Claim 2. Then observe that the last term covers cases when either $J_t(\varepsilon) = 0$ or $I_t(0) = 0$. The $\ln(M/m)$ factor arising in the cases not covered by Claim 2 yields an upper bound given that the maximum ratio of expected offspring for any two types is $M/m < \infty$.

By Claim 2, (9) holds for each sufficiently small $\delta > 0$, and $\varepsilon > 0$. For the reminder, fix these numbers so that $\frac{\varepsilon}{1-\varepsilon}\frac{M}{m} - \frac{\delta}{M} < 0$.

For any indicator functions A, B, and C, $ABC \geqslant A + B + C - 2$. Moreover, $I_t(0)J_t(\varepsilon) \leqslant 1$. Thus, the quantity expressed in (9) is bounded above by

$$\begin{aligned} & \Delta_t(\delta, \varepsilon) \equiv (D_t + J_t(\varepsilon) - 1) \cdot \ln\left(1 + \frac{\varepsilon}{1-\varepsilon}\frac{M}{m} - \frac{\delta}{M}\right) \\ & + (1 - I_t(\delta)) \cdot \ln\left(\left[1 + \frac{\varepsilon}{1-\varepsilon}\frac{M}{m}\right] \Big/ \left[1 + \frac{\varepsilon}{1-\varepsilon}\frac{M}{m} - \frac{\delta}{M}\right]\right) \\ & + (1 - I_t(0) + 1 - J_t(\varepsilon)) \cdot \ln(M/m). \end{aligned} \tag{10}$$

The ratio of the population of the alternative type to that of the *SPE-ToM* at date τ is then bounded above by the random variable $r_0 \cdot r_\tau$, where $\ln r_\tau \leqslant \sum_{t=1}^{\tau} \Delta_t(\delta, \varepsilon)$. The result is established by showing that $\sum_{t=1}^{\tau} \Delta_t(\delta, \varepsilon)/\tau$ converges in probability to a negative constant for suitably chosen $\delta > 0$ and $\varepsilon > 0$ satisfying $\frac{\varepsilon}{1-\varepsilon}\frac{M}{m} - \frac{\delta}{M} < 0$.

We rely on the following claims.

Claim 3: *Suppose $\alpha > 2$. If $R_{jt} \longrightarrow 1$ in probability, $j = 1, ..., i-1$, then $\frac{1}{\tau}\sum_{t=1}^{\tau} J_t(\varepsilon)$ converges in probability to one, for each $\varepsilon > 0$.*

Proof. Fix $\varepsilon \in (0, 1]$. If $\alpha > 2$, then Theorem 1 applies. Under the hypotheses of the claim $J_t(\varepsilon) \longrightarrow 1$ in probability. Thus $E(J_t(\varepsilon))$ tends to one. Then $E(\frac{1}{\tau}\sum_{t=1}^{\tau} J_t(\varepsilon)) \longrightarrow 1$. Since $J_t(\varepsilon) \leqslant 1$ everywhere it follows that $\frac{1}{\tau}\sum_{t=1}^{\tau} J_t(\varepsilon)$ converges to one in probability. ■

Claim 4: *i)* $\frac{1}{\tau}\sum_{t=1}^{\tau} I_t(\delta)$ *almost surely converges to* $\mu(Q_\delta)$. *ii) Assume A6. Given the alternative is a ToM, then* $\frac{1}{\tau}\sum_{t=1}^{\tau} D_t$ *almost surely converges to* $\mu(S) > 0$, *where* S *is as described in A6.*

Proof. Consider i). Lemma 1 implies $E(I_t(\delta)|Z_\infty)$ converges to $\mu(Q_\delta)$ for almost every realized outcome set Z_∞. Hence, $\frac{1}{\tau}\sum_{t=1}^{\tau} E(I_t(\delta)|Z_\infty) \longrightarrow \mu(Q_\delta)$ almost surely and for almost every Z_∞. The random variables $(I_t(\delta)|Z_\infty)$ are independent, and the sequence satisfies Kolmogorov's criterion. The strong law of large numbers implies $\frac{1}{\tau}\sum_{t=1}^{\tau}[(I_t(\delta)|Z_\infty) - E(I_t(\delta)|Z_\infty)] \longrightarrow 0$, almost surely, for almost every Z_∞. Hence $\frac{1}{\tau}\sum_{t=1}^{\tau}(I_t(\delta)|Z_\infty) \longrightarrow \mu(Q_\delta)$, almost surely and for almost every Z_∞, so that $\frac{1}{\tau}\sum_{t=1}^{\tau} I_t(\delta) \longrightarrow \mu(Q_\delta)$, almost surely. This completes the proof of i). A similar proof establishes ii).[43] ■

Claims 3 and 4 (using also (10) and the fact that $\mu(Q_\delta)$ converges to one as δ tends to zero) then give that

---

[43] If the alternative is naive, and $\alpha < T$, then $\frac{1}{\tau}\sum_{t=1}^{\tau} D_t$ converges in probability to μ^*, where μ^* is the measure of games where i has no dominant action at any node. The proof is worth sketching since it diverges from the proof for the *ToM* alternative. Define the random variable A_t such that $A_t = 1$ if the game drawn at date t is new, and let $A_t = 0$ otherwise. Let B_t equal one if the game realized at t has no dominant strategy for role i at any i information set; let B_t be one otherwise. For any indicator functions A and B, $A + B - 1 \leqslant AB \leqslant B$. Hence, surely

$$\frac{1}{\tau}\sum_{t=1}^{\tau}(A_t - 1) + \frac{1}{\tau}\sum_{t=1}^{\tau} B_t \leqslant \frac{1}{\tau}\sum_{t=1}^{\tau} D_t \leqslant \frac{1}{\tau}\sum_{t=1}^{\tau} B_t. \tag{11}$$

$E(A_t)$ is just $E(1 - \gamma_t)$, where γ_t is the fraction of games played previously among those available at date t. Whenever $\alpha < T$, γ_t surely converges to zero (Theorem 1). Clearly then, $E(\frac{1}{\tau}\sum_{t=1}^{\tau} A_t) \longrightarrow 1$, whenever $\alpha < T$. Since A_t is surely bounded above by one, it follows that $\frac{1}{\tau}\sum_{t=1}^{\tau} A_t \longrightarrow 1$, in probability, whenever $\alpha < T$. In light of the above indented equation it then suffices to show that $\frac{1}{\tau}\sum_{t=1}^{\tau} B_t$ tends to μ^*. To see this is in fact the case note first that Lemma 1 implies that $E(B_t|Z_\infty)$ converges to μ^* almost surely and for almost every sequence of realized outcome sets Z_∞. Then note that the random variables $(B_t|Z_\infty)$ are independent, for each Z_∞, and that the sequence satisfies Kolmogorov's criterion. Thus, $\frac{1}{\tau}\sum_{t=1}^{\tau}[(B_t|Z_\infty) - E(B_t|Z_\infty)] \longrightarrow 0$, almost surely and for almost every Z_∞, so that $\frac{1}{\tau}\sum_{t=1}^{\tau}(B_t|Z_\infty) \longrightarrow \mu^*$, so $\frac{1}{\tau}\sum_{t=1}^{\tau} B_t \longrightarrow \mu^*$, almost surely.

$$\frac{1}{\tau}\sum_{t=1}^{\tau}\Delta_t(\delta,\varepsilon) \longrightarrow \mu(S)\cdot\ln\left(1+\frac{\varepsilon}{1-\varepsilon}\frac{M}{m}-\frac{\delta}{M}\right) + (1-\mu(Q_\delta))\cdot\ln\left(\left[1+\frac{\varepsilon}{1-\varepsilon}\frac{M}{m}\right]\Big/\left[1+\frac{\varepsilon}{1-\varepsilon}\frac{M}{m}-\frac{\delta}{M}\right]\right), \tag{12}$$

in probability.

We can choose $\varepsilon > 0$ and $\delta > 0$ so that $\frac{\varepsilon}{1-\varepsilon}\frac{M}{m}-\frac{\delta}{M} < 0$, and simultaneously the limiting value in (12) is negative. That is, choose a δ such that $\mu(S)\cdot\ln(1-\frac{\delta}{M})+(1-\mu(Q_\delta))\cdot\ln(1/[1-\frac{\delta}{M}]) < 0$, and then choose a sufficiently small but positive ε. This completes the proof of Theorem 2 since for such ε and δ, we have shown $\frac{\ln r_\tau}{\tau}$ is bounded above, in the limit, in probability, by a negative constant. Hence $r_\tau \longrightarrow 0$, in probability.

B. Experiment Instructions (for online publication)

Page 1

In this experiment you will participate in a series of two person decision problems. The experiment will last for a number of rounds. Each round you will be randomly paired with another individual. The joint decisions made by you and the other person will determine how much money you will earn in that round.

Your earnings will be paid to you in cash at the end of the experiment. We will not tell anyone else your earnings. We ask that you do not discuss your earnings with anyone else.

If you have a question at any time, please raise your hand.

Page 2

You will see a diagram similar to one on your screen at the beginning of the experiment. You and another person will participate in a decision problem shown in the diagram.

One of you will be Person 1 (orange). The other person will be Person 2 (blue). In the upper left corner, you will see whether you are Person 1 or Person 2.

You will be either a Person 1 or a Person 2 for the entire experiment.

Page 3

Notice the four pairs of squares with numbers in them; each pair consists of two earnings boxes. The earnings boxes show the different earnings you and the other person will make, denoted in Experimental Dollars. There are two numbers, Person 1 will earn what is in the orange box, and Person 2 will earn what is in the blue box if that decision is reached.

In this experiment, you can only see the earnings in your own box. That is, if you are Person 1 you will only see the earnings in the orange boxes, and if you are Person 2 you will only see the earnings in the blue boxes. Both boxes will be visible, but the number in the other person's box will be replaced with a "?".

However, for each amount that you earn, the amount the other person earns is fixed. In other words, for each amount that Person 1 sees, there is a corresponding, unique amount that will always be shown to Person 2.

For example, suppose Person 1 sees an earnings box containing "12" in round 1. In the same pair, suppose Person 2 sees "7". Then, at any later round, anytime Person 1 sees "12", Person 2 will see "7".

Together, you and the other person will choose a path through the diagram to an earnings box. We will describe how you make choices next.

Page 4

A node, displayed as a circle and identified by a letter, is a point at which a person makes a decision. Notice that the nodes are color coded to indicate whether Person 1 or Person 2 will be making that decision. You will always have two options.

If you are you Person 1 you will always choose either "Right" or "Down", which will select a node at which Person 2 will make a decision.

If you are Person 2 you will also choose either "Right" or "Down" which will select a pair of earnings boxes for you and Person 1.

Once a pair of earnings boxes is chosen, the round ends, and each of you will be able to review the decisions made in that round.

Page 5

In each round all pairs will choose a path through the same set of nodes and earnings boxes. This is important because at the end of each round, in addition to your own outcome, you will be able to see how many pairs ended up at each other possible outcome.

While you review your own results from a round, a miniature figure showing all possible paths through nodes and to earnings boxes will be displayed on the right hand side of the screen.

The figure will show how many pairs chose a path to each set of earnings boxes.

The Payoff History table will update to display your payoff from the current period.

Page 6

We have provided you with a pencil and a piece of paper on which you may write down any information you deem relevant for your decisions. At the end of the experiment, please return the paper and pencil to the experimenter.

At the end of the experiment, we will randomly choose 2 rounds for payment, and your earnings from those rounds will be summed and converted to $CAD at a rate of 1 Experimental Dollar = $2.

Important points:

- You will be either a Person 1 or a Person 2 for the entire experiment.
- Each round you will be randomly paired with another person for that round.
- Person 1 always makes the first decision in a round.
- Person 1's payoff is in the orange earnings box and Person 2's in the blue earnings box.
- Each person will only be able to see the numbers in their own earnings box.
- Earnings always come in unique pairs so that for each amount observed by Person 1, the number observed by Person 2 will be fixed.

In a given round, all pairs will choose a path through the same set of nodes and earnings boxes.

After each round you will be able to see how many pairs ended up at each outcome.

We will choose 2 randomly selected periods for payment at the end of the experiment.

Any questions?

C. Screenshots (for online publication)

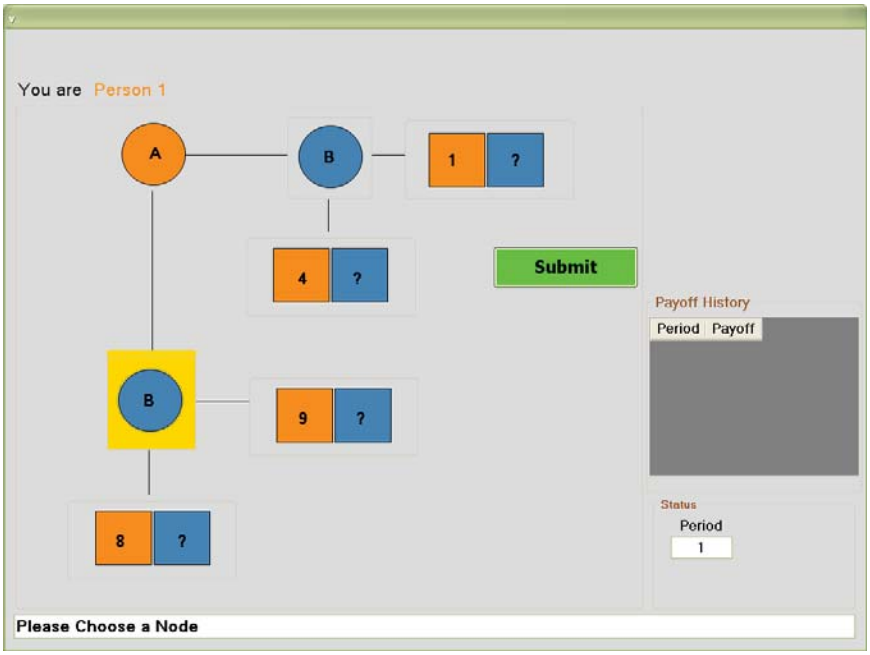

Figure C1: **Screenshot for Player 1.** This figure shows the screen as player 1 sees it prior to submitting his choice of action. The yellow highlighted node indicates that player 1 has provisionally chosen the corresponding action, but the decision is not final until the submit button is clicked. While waiting for player 1 to choose, player 2 sees the same screen except that she is unable to make a decision, provisional choices by player 1 are not observable, and the "Submit" button is invisible.

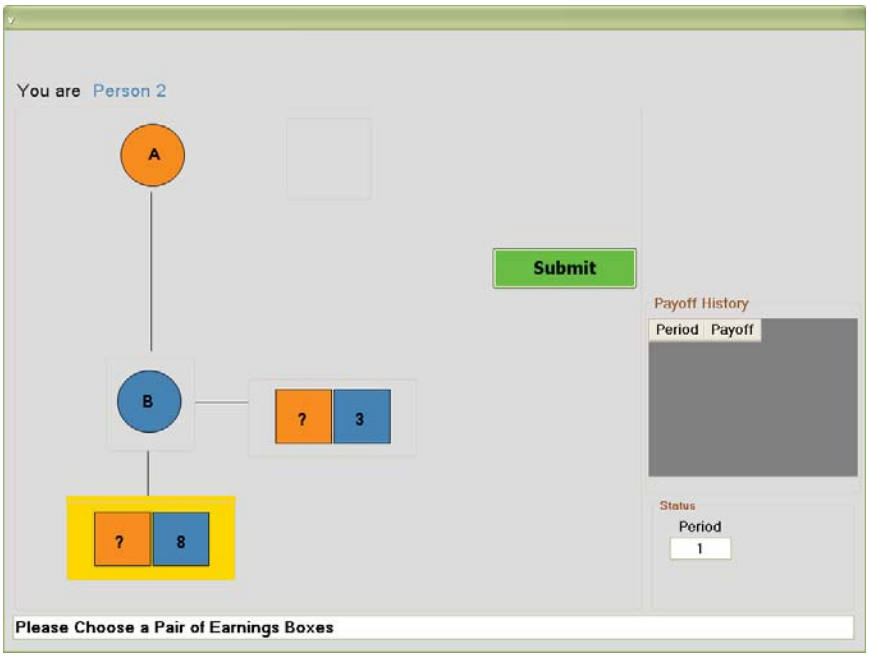

FIGURE C2: **Screenshot for Player 2.** This figure shows the screen as player 2 sees it after player 1 has chosen an action. Here, player 1 chose to move down, so the upper right portion of the game tree is no longer visible. While player 2 is making a decision, player 1 sees an identical screen except that he is unable to make a decision and the "Submit" button is invisible.

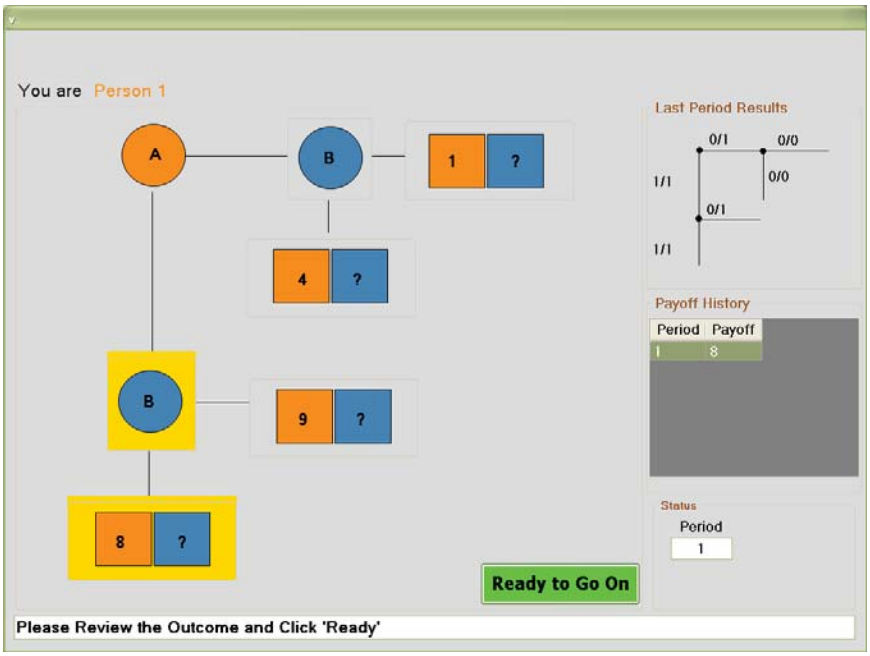

FIGURE C3: **Screenshot of Post-Decision Review.** This figure shows the final screen subjects see in each period after both player 1 and player 2 have made their decisions. The smaller game tree in the upper right portion of the figure displays information about how many pairs ended up at each outcome. For the purposes of the screenshot, the software was run with only one pair, but in a typical experiment, subjects learned about the decisions of 4 pairs (3 other than their own).

D. Autism-Spectrum Quotient Questionnaire (for online publication)

1.	I prefer to do things with others rather than on my own.	[1]	[2]	[3]	[4]
2.	I prefer to do things the same way over and over again.	[1]	[2]	[3]	[4]
3.	If I try to imagine something, I find it very easy to create a picture in my mind.	[1]	[2]	[3]	[4]
4.	I frequently get so absorbed in one thing that I lose sight of other things.	[1]	[2]	[3]	[4]
5.	I often notice small sounds when others do not.	[1]	[2]	[3]	[4]
6.	I usually notice car number plates of similar strings of information.	[1]	[2]	[3]	[4]
7.	Other people frequently tell me that what I've said is impolite, even though I think it is polite.	[1]	[2]	[3]	[4]
8.	When I'm reading a story, I can easily imagine what the characters might look like.	[1]	[2]	[3]	[4]
9.	I am fascinated by dates.	[1]	[2]	[3]	[4]
10.	In a social group, I can easily keep track of several different people's conversations.	[1]	[2]	[3]	[4]
11.	I find social situations easy.	[1]	[2]	[3]	[4]
12.	I tend to notice details that others do not.	[1]	[2]	[3]	[4]
13.	I would rather go to a library than a party.	[1]	[2]	[3]	[4]
14.	I find making up stories easy.	[1]	[2]	[3]	[4]
15.	I find myself drawn more strongly to people than to things.	[1]	[2]	[3]	[4]
16.	I tend to have very strong interests, which I get upset about if I can't pursue.	[1]	[2]	[3]	[4]
17.	I enjoy social chit-chat.	[1]	[2]	[3]	[4]
18.	When I talk, it isn't always easy for others to get a word in edgeways.	[1]	[2]	[3]	[4]
19.	I am fascinated by numbers.	[1]	[2]	[3]	[4]
20.	When I'm reading a story I find it difficult to work out the characters' intentions.	[1]	[2]	[3]	[4]
21.	I don't particularly enjoy reading fiction.	[1]	[2]	[3]	[4]
22.	I find it hard to make new friends.	[1]	[2]	[3]	[4]
23.	I notice patterns in things all the time.	[1]	[2]	[3]	[4]
24.	I would rather go to the theatre than a museum.	[1]	[2]	[3]	[4]
25.	It does not upset me if my daily routine is disturbed.	[1]	[2]	[3]	[4]
26.	I frequently find that I don't know how to keep a conversation going.	[1]	[2]	[3]	[4]
27.	I find it easy to "read between the lines" when someone is talking to me.	[1]	[2]	[3]	[4]
28.	I usually concentrate more on the whole picture, rather than the small details.	[1]	[2]	[3]	[4]
29.	I am not very good at remembering phone numbers.	[1]	[2]	[3]	[4]
30.	I don't usually notice small changes in a situation, or a person's appearance.	[1]	[2]	[3]	[4]
31.	I know how to tell if someone listening to me is getting bored.	[1]	[2]	[3]	[4]
32.	I find it easy to do more than one thing at once.	[1]	[2]	[3]	[4]
33.	When I talk on the phone, I'm not sure when it's my turn to speak.	[1]	[2]	[3]	[4]
34.	I enjoy doing things spontaneously.	[1]	[2]	[3]	[4]
35.	I am often the last to understand the point of a joke.	[1]	[2]	[3]	[4]
36.	I find it easy to work out what someone else is thinking or feeling just by looking at their face.	[1]	[2]	[3]	[4]
37.	If there is an interruption, I can switch back to what I was doing very quickly.	[1]	[2]	[3]	[4]
38.	I am good at social chit-chat.	[1]	[2]	[3]	[4]
39.	People often tell me that I keep going on and on about the same thing.	[1]	[2]	[3]	[4]
40.	When I was young, I used to enjoy playing games involving pretending with other children.	[1]	[2]	[3]	[4]
41.	I like to collect information about categories of things (e.g. types of car, types of bird, types of train, types of plant, etc.	[1]	[2]	[3]	[4]
42.	I find it difficult to imagine what it would be like to be someone else.	[1]	[2]	[3]	[4]
43.	I like to plan any activities I participate in carefully.	[1]	[2]	[3]	[4]
44.	I enjoy social occasions.	[1]	[2]	[3]	[4]
45.	I find it difficult to work out people's intentions.	[1]	[2]	[3]	[4]
46.	New situations make me anxious.	[1]	[2]	[3]	[4]
47.	I enjoy meeting new people.	[1]	[2]	[3]	[4]
48.	I am a good diplomat.	[1]	[2]	[3]	[4]
49.	I am not very good at remembering people's date of birth.	[1]	[2]	[3]	[4]
50.	I find it very easy to play games with children that involve pretending.	[1]	[2]	[3]	[4]

1 = definitely agree, 2 = slightly agree, 3 = slightly disagree, 4 = definitely disagree

E. Broad Autism Phenotype Questionnaire (for online publication)

You are about to fill out a series of statements related to personality and lifestyle. For each question, circle that answer that best describes how often that statement applies to you. Many of these questions ask about your interactions with other people. Please think about the way you are with most people, rather than special relationships you may have with spouses or significant others, children, siblings, and parents. Everyone changes over time, which can make it hard to fill out questions about personality. Think about the way you have been the majority of your adult life, rather than the way you were as a teenager, or times you may have felt different than normal. You must answer each question, and give only one answer per question. If you are confused, please give it your best guess.

1	I like being around other people.	1	2	3	4	5	6
2	I find it hard to get my words out smoothly.	1	2	3	4	5	6
3	I am comfortable with unexpected changes in plans.	1	2	3	4	5	6
4	It's hard for me to avoid getting sidetracked in conversation.	1	2	3	4	5	6
5	I would rather talk to people to get information than to socialize.	1	2	3	4	5	6
6	People have to talk me into trying something new.	1	2	3	4	5	6
7	I am "in-tune" with the other person during conversation.*	1	2	3	4	5	6
8	I have to warm myself up to the idea of visiting an unfamiliar place.	1	2	3	4	5	6
9	I enjoy being in social situations.	1	2	3	4	5	6
10	My voice has a flat or monotone sound to it.	1	2	3	4	5	6
11	I feel disconnected or "out of sync" in conversations with others.*	1	2	3	4	5	6
12	People find it easy to approach me.*	1	2	3	4	5	6
13	I feel a strong need for sameness from day to day.	1	2	3	4	5	6
14	People ask me to repeat things I've said because they don't understand.	1	2	3	4	5	6
15	I am flexible about how things should be done.	1	2	3	4	5	6
16	I look forward to situations where I can meet new people.	1	2	3	4	5	6
17	I have been told that I talk too much about certain topics.	1	2	3	4	5	6
18	When I make conversation it is just to be polite.*	1	2	3	4	5	6
19	I look forward to trying new things.	1	2	3	4	5	6
20	I speak too loudly or softly.	1	2	3	4	5	6
21	I can tell when someone is not interested in what I am saying.*	1	2	3	4	5	6
22	I have a hard time dealing with changes in my routine.	1	2	3	4	5	6
23	I am good at making small talk.*	1	2	3	4	5	6
24	I act very set in my ways.	1	2	3	4	5	6
25	I feel like I am really connecting with other people.	1	2	3	4	5	6
26	People get frustrated by my unwillingness to bend.	1	2	3	4	5	6
27	Conversation bores me.*	1	2	3	4	5	6
28	I am warm and friendly in my interactions with others.*	1	2	3	4	5	6
29	I leave long pauses in conversation.	1	2	3	4	5	6
30	I alter my daily routine by trying something different.	1	2	3	4	5	6
31	I prefer to be alone rather than with others.	1	2	3	4	5	6
32	I lose track of my original point when talking to people.	1	2	3	4	5	6
33	I like to closely follow a routine while working.	1	2	3	4	5	6
34	I can tell when it is time to change topics in conversation.*	1	2	3	4	5	6
35	I keep doing things the way I know, even if another way might be better.	1	2	3	4	5	6
36	I enjoy chatting with people.	1	2	3	4	5	6

1—very rarely
2—rarely
3—occasionally
4—somewhat often
5—often
6—very often

*Casual interaction with acquaintances, rather than special relationships such as with close friends and family members

BAPQ

F. Additional Tables and Figures (for online publication)

	8-1	8-2	10-1	10-2	8-3	6-1	8-4	10-3	8-5*	10-3*	10-4	10-5	8-6	10-6	8-7	8-8	8-9	8-10	10-7	8-11
	0.68	*0.54*	*0.62*	*0.61*	*0.63*	*0.44*	*0.51*	*0.46*	*0.68*	*0.50*	*0.32*	*0.66*	*0.53*	*0.38*	*0.54*	*0.52*	*0.57*	*0.64*	*0.25*	*0.71*
	0.72	*0.59*	*0.64*	*0.71*	*0.67*	*0.54*	*0.61*	*0.54*	*0.70*	*0.59*	*0.50*	*0.72*	*0.64*	*0.50*	*0.67*	*0.61*	*0.63*	*0.70*	*0.36*	*0.72*
	0.94	*0.91*	*0.98*	*0.87*	*0.94*	*0.83*	*0.84*	*0.86*	*0.96*	*0.85*	*0.63*	*0.92*	*0.83*	*0.77*	*0.80*	*0.85*	*0.91*	*0.92*	*0.71*	*0.98*
	0.36	*0.40*	*0.42*	*0.38*	*0.42*	*0.45*	*0.34*	*0.55*	*0.43*	*0.43*	*0.55*	*0.38*	*0.46*	*0.47*	*0.44*	*0.52*	*0.49*	*0.45*	*0.59*	*0.34*
E	*0.64*	*0.52*	*0.55*	*0.58*	*0.55*	*0.36*	*0.49*	*0.37*	*0.64*	*0.45*	*0.25*	*0.64*	*0.44*	*0.31*	*0.46*	*0.43*	*0.51*	*0.58*	*0.21*	*0.68*
E	*0.81*	*0.60*	*0.86*	*0.74*	*0.88*	*0.71*	*0.58*	*0.75*	*0.78*	*0.73*	*0.57*	*0.70*	*0.73*	*0.59*	*0.75*	*0.77*	*0.78*	*0.82*	*0.52*	*0.88*
	0.71	*0.69*	*0.65*	*0.72*	*0.70*	*0.49*	*0.56*	*0.57*	*0.74*	*0.66*	*0.63*	*0.79*	*0.60*	*0.51*	*0.73*	*0.78*	*0.73*	*0.78*	*0.52*	*0.81*
	0.94	*0.90*	*0.98*	*0.86*	*0.94*	*0.84*	*0.85*	*0.82*	*0.97*	*0.87*	*0.63*	*0.93*	*0.86*	*0.78*	*0.80*	*0.83*	*0.91*	*0.93*	*0.66*	*0.98*

oserved Probability of Outcomes by Session.

robability that we observed a particular outcome in a particular session. Pairs SPE refers to the probability that a pair P1 SPE is the probability that player 1's choice was consistent with the *SPE*. P1 Heur is the probability that Player 1 consistent with the "highest mean" rule of thumb (heuristic). P1 SPE | heur ≠ SPE is the probability that player one hen it did not correspond to the "highest mean" rule of thumb. P1 SPE | heur = SPE is the probability that player *PE* when it *did* correspond to the "highest mean" rule of thumb. P1 SPE | No Heur is the probability that player 1 hen the rule of thumb was inapplicable (i.e. equal means). P2 Dom is the probability that player 2 chose the dominant P1 SPE is the conditional probability of player 2 choosing the dominant strategy given that player 1 followed the *SPE*. in the format # of Subjects – Session ID so that 10-2 corresponds to the 2nd session with 10 subjects. * indicates e payoff set was {1,2,3,4,8,9,10}, rather than {1,2,3,4,5,6,7}.

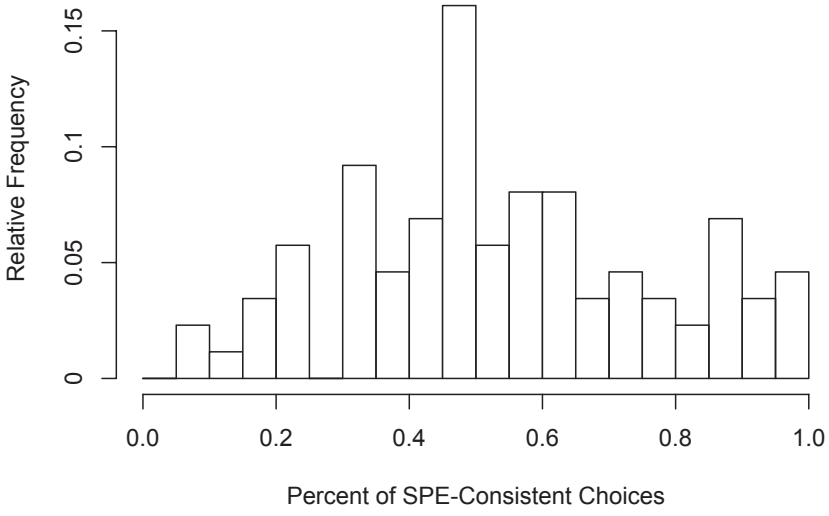

FIGURE F1: **Histogram of the Individual Rates of SPE-consistent Choices.** The figure excludes all periods in which the player had a dominant strategy and in which choice under the rule of thumb corresponded to the SPE.

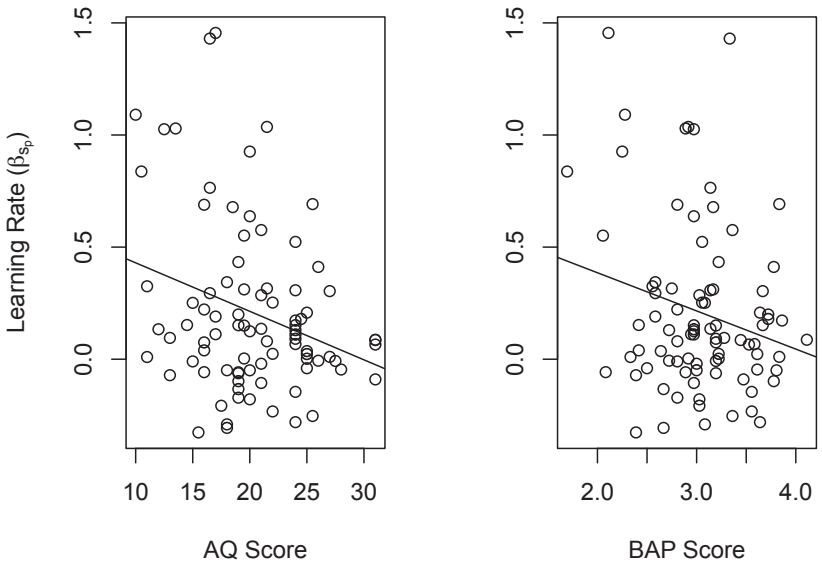

FIGURE F2: **Scatterplots Comparing Learning Rates to Theory of Mind Measures from Psychology.** The solid lines plot OLS fits of the data.

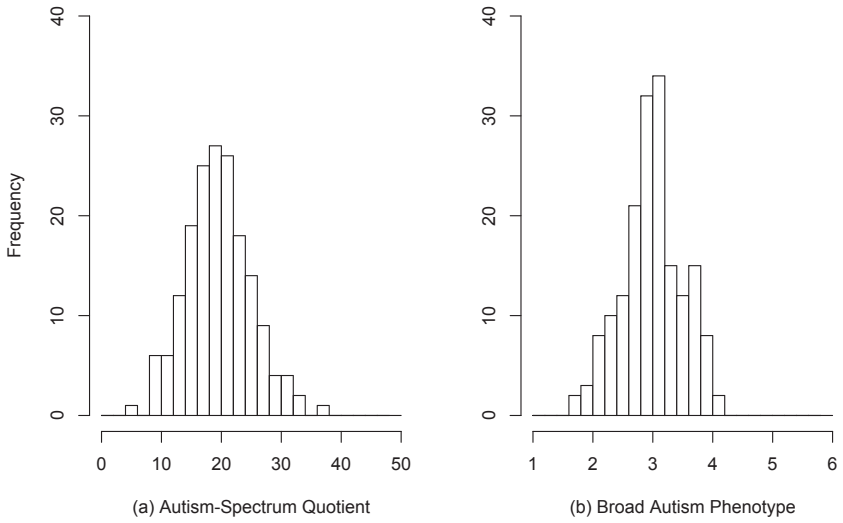

FIGURE F3: **Histograms of AQ and BAP Scores.** Each panel includes the entire range of feasible scores.