Yale University

EliScholar – A Digital Platform for Scholarly Publishing at Yale

Cowles Foundation Discussion Papers

Cowles Foundation

9-1-2013

The Evolution of 'Theory of Mind:' Theory and Experiments

Erik O. Kimbrough

Nikolaus Robalino

Arthur Robson

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

By

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

September 2013

COWLES FOUNDATION DISCUSSION PAPER NO. 1907

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

Simon Fraser University

Nikolaus Robalino

Simon Fraser University

Arthur J. Robson

Simon Fraser University

PRELIMINARY AND INCOMPETE: PLEASE DO NOT CIRCULATE OR CITE

August 31, 2013

Abstract. This paper provides an 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. Such environments are represented here by multistage games of perfect information with randomly chosen outcomes. "Theory of Mind" then yields a sharp, unambiguous advantage over less sophisticated, behavioral approaches to strategic interaction. In related experiments, we show the subscale for social skills in standard tests for autism is a highly significant determinant of the speed of learning in such games.

We thank participants at a large number of departmental seminars for useful comments. 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.

1. Introduction

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

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

The argument made here in favor of theory of mind is a substantial generalization and reformulation of the argument in Robson (2001) concerning the advantage of having 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 focus here to a dynamic model in which players repeatedly interact with one another but in which the set of games that they might face becomes larger and larger with time. We presume individuals have an appropriate own utility function. The focus is then on the advantage to an agent of conceiving of her opponents as rational actors—as having preferences, in particular, and understanding that they act optimally in the light of these. Having a template into which the preferences of an opponent can be fitted enables a player to deal with a higher rate of innovation than can a behavioral type of individual that adapts to each game as a distinct set of circumstances. In other words, the edge to *ToM* derives from a capacity to extrapolate to novel circumstances information that was learned about preferences in a specific case.

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

Consider the argument in greater detail. We begin by fixing a game tree with perfect information, with stages $i = 1, ..., I$. 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 $i = 1, .., I$. The outcomes needed to complete the game are drawn randomly and uniformly in each period from a finite outcome set. Each player has a preference ordering over the entire infinite set of possible outcomes. 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 outcome set, where each new outcome is drawn independently from a given distribution. The number of outcomes available grows to infinity at a parametric rate. We use this device of a growing outcome set as a simple way of deriving the comparative rates at which the various behavioral types learn.

All players are given the full history of the games played—the outcomes that were chosen, and the choices that were made by all player roles. 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 *ToM* types are disposed to learn others' preferences. 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. The naive types are characterized by reinforcement learning, and treat each new game as a unfamiliar set of circumstances. All types are assumed to avail themselves of a dominant choice in the subgame they start, whenever such a dominant choice 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 aspect of the *ToM* behavior is that, in the long run, once the history of the game has revealed the preferences of all subsequent players, they map these preferences to an action. There is a particular *ToM* type, the *SPE-ToM* type, who maps these preferences to the SPE choice for the subgame. 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.

The crucial feature of naive types is simply that they play the same strategy in response to any new game. Even if the naive types are ultra-fast learners, and use the SPE strategy the second time a game is played, as is usually literally impossible, 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. This characterization of naive types is in line with "evolutionary game theory," which was inspired, in turn, by the psychological theory of reinforcement learning.

Another simplification is that we assume that the *ToM* types do not avail themselves of the transitivity of opponents' preferences. The *ToM* types build up a description of others' preferences only by learning all the pairwise choices. Again, relaxing this assumption would only strengthen our results.

Theorem 1 is the basic 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.

In principle, 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 as the main result in Theorem 2—that the unique SPE is attained, with the *SPE-ToM* type ultimately dominant, over all other *ToM* types, as well as over all the naive types.

We also report here closely related experiments on Theory of Mind that buttress the current approach. In these experiments, we fix the outcome set, rather than allow this to grow over time. Also for simplicity, we consider only two-stage games where each player role has two moves at each decision node. As is the crucial feature of the theoretical model, all players in a given role had the same (induced) preferences, and players knew only their own payoff at each outcome and not that of their opponent. We randomly and anonymously paired subjects in each of 90 repetitions to observe the ability of players 1 to learn (and exploit knowledge of) the preferences of players 2.

Our design alleviates two potential confounds which arise in this setting. First, such a game might have a dominant strategy for player 1; any such game provides no opportunity for the experimenter to infer whether player 1 has learned player 2's preferences. Second, there are some such two stage games where a simple "highest mean" rule of thumb adequately guides player 1—choose between your two actions based on a 50-50 expectation of the resulting payoffs for you. Before player 1 has observed any of player 2's choices, this rule of thumb will be indeed be optimal, given that players' preferences are uncorrelated. This is because player 2 will randomize 50-50 from player 1's perspective. To get greater mileage from the experiments, we thus divided each experimental session into three treatments designed to gradually eliminate these confounds by varying the outcome set from which games were drawn. The four outcomes in the first treatment were chosen in an unconstrained way so dominant strategies did sometimes arise and the rule of thumb sometimes gave the right choice. The outcomes in the second treatment were chosen so that dominant strategies could not arise but the rule of thumb still sometimes worked. The outcomes in the third treatment, which comprised 2/3 of the session, were chosen so that the rule of thumb never gave the right answer (and thus there were also no dominant strategies). Thus, the final 60 periods provide a particularly difficult setting in which the only way to perform well in the role of player 1 is to learn the preferences of player 2.

At the end of each experimental session, we asked the students to complete two short multiple-choice surveys measuring autism. One was the Autism-Spectrum Quotient (AQ) survey; the other was the Broad Autism Phenotype Questionnaire (BAP). Of particular relevance here is the subscale on each of these two tests that rates social skills (Social Skills - in the BAP and Aloofness in the AQ).

The results of the experiments were striking. First, we observed statistically and quantitatively significant learning of player 2 preferences by players 1. This learning is evident during the third treatment when the rule of thumb cannot work. Second, a particularly salient finding was that player 1's with lower scores on the social subscales (i.e. players who exhibit greater social skills) were highly likely (with p=0.01) to learn player 2's choices faster.

These experiments corroborate the present approach on two grounds. In the first place, the theoretical model apparently captures important features of real world decision-making. To ascribe preferences to an opponent and to endeavor to learn these, to one's own advantage, is a real world ability, and one that is present to a varying extent in individuals. Secondly, correlations between behavior in the experiment and survey measures of autism spectrum suggest that preference learning ability in strategic settings is closely tied to a central aspect of theory of mind, as this term is used in psychology.

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.[1] There is one "player role" for each such stage, $i = 1, ..., I$, in the game. Each player role is represented by an equally large population of agents. These agents may have different "strategic types", in a way that is described below, but all such types have identical payoff functions. To describe the basic elements of the game, all that matters, then, is how each player role maps outcomes to expected offspring. 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.[2] There is a fixed overall set of outcomes, each with consequences for the reproductive success of the I types of agents. 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.

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.[3]

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

---

[1] 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.

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

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

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.[4] 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 $Z_1 \cup \{z_1, \ldots, z_k\}$, whenever $t_k \leqslant t < t_{k+1}$, where $z_k \in Z$ denotes the introduced outcome at arrival date t_k.

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.

A3: *Fix* $\alpha \geqslant 0$. *The arrival date sequence* $\{t_k\}$ *satisfies, for each* $k = 1, 2, \ldots$, $t_k = \lfloor (|Z_1| + k)^{\alpha} \rfloor$.[5]

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* $z \in [m, M]^I$. *Similarly, the empirical cdf of games at date* t *is denoted by* $G_t(q)$, *where* $q \in Q = [m, M]^{IT}$.[6] *The set of games available at date* t *is denoted by* Q_t.

We suppose that, at each date t, an extensive form game denoted q_t is drawn at random from Q_t—uniformly and independently from the sequence of previously realized games. The players in each match then independently play q_t. Players of each

[4] 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, given that 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 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.

[5] 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.

[6] 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.

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.[7]

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, p. 275, and Eike, Pollard and Stute, 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.

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

2.2. **Strategic Types.**

We allow a finite number of different "strategic types" within each role. When making a choice at date t every player 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.[8] 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 $\boldsymbol{H} = \cup_{t \geqslant 1} \boldsymbol{H}_t$.

[7] 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.

[8] Although each player observes the previous games in the sense of seeing 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.

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 groups differing with respect to 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. 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 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.[9]

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 courses of action a player mixes evenly between these actions.*

We assume there are two main categories of types of players—

Naive players. We adopt a relaxed concept of naivete, which serves to make the ultimate results stronger. The only additional requirement on the map c is that a naive player must choose a fixed arbitrary strategy whenever the game is novel. For specificity, suppose naive players in this situation mix uniformly over all available actions.

---

[9] 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 also choose any strictly dominant action in the remaining subgame, as in A4.) Although it is highly implausible, we could indeed allow naive players to play the SPE strategy that is appropriate for the underlying unknown preferences of subsequent players, in the *second* repetition of each game. This is highly implausible since it is not generally possible to deduce the payoffs that rapidly and neither could convergence to an SPE once preferences are known but other agents make non-SPE choices be that rapid. Nevertheless, even if the naive players are ultra-fast learners, the sophisticated *ToM* players will out-compete them, given only the naive players' inability to adapt in any way to a new game. To the extent that naive players fail to be such ultra-fast learners, our results would simply be strengthened.

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 need not agree at the outset about what these preference orderings are, but they all know there are some preferences influencing j players' choices in every period, and they learn what these preference are.

Theory of Mind Players.

What this means precisely is as follows. The important long run aspect of the behavior of these *ToM* types is that, if the history of the game has revealed the preferences of all subsequent players, perhaps in the way that is described in detail below, 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. 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, in order to learn others' preferences, it is important that 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 are relatively weak. For example, 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. Further, 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.

We can now describe the strategic environment as follows

$$\mathscr{E} = (I, A, Z, F, \alpha, \{M_{i1}\}_{i \in I}),$$

where the new variable is M_{i1} giving the initial distribution of strategic types in player role i. Recall that I and A describe the fixed game tree, F is the distribution on $Z = [m, M]^I$ over introduced outcomes, and α governs the rate of introduction of novel outcomes.

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 2: *The history H_t reveals players in role i strictly prefer z to z' if for all $\mathscr{E}$ satisfying A4 whenever H_t occurs it becomes common knowledge among ToMs that $z_i > z'_i$.*

The notion is well defined. In particular—

Lemma 2: *Suppose A4 holds and that there are ToMs in every role $i > 1$. Then for every finite subset $X \subset Z$ there exists a finite history $H \in \boldsymbol{H}$ such that for each $j \in I$, and $z, z' \in X$, if $z_j > z'_j$, then H reveals players in role j strictly prefer z to z'.*

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 $\{z, z'\}$.[10] 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.[11] It follows that T is the cut-off point for the naive types. We then have the following key result that sets the stage for establishing the evolutionary dominance of the *SPE-ToM* 's over all other players—

Theorem 1: *Suppose $\mathscr{E}$ satisfies assumptions A1-A4. If $\alpha < 2$, then L_{it} surely converges to zero, $i = 1, \ldots, I$. If $\alpha < T$, then γ_t surely converges to zero. On the other hand, if $\alpha > T$, then γ_t converges in probability to one. If $\alpha > 2$, and additionally A4 holds, then L_{it} converges in probability to one for each $i = 1, \ldots, I$.*

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

[10] Recall that for simplicity, we assume that players mix whenever indifferent and that this too is common knowledge among the *ToM* types. This means that such indifference can be ruled out. However, such indifference has probability 0.

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

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.[12]

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 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... 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 $i - 1$.[13]

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 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'

[12] Naive players do not always make inappropriate choices, because they adopt dominant strategies, whenever these are available. They are often not available, of course.

[13] The above learning process relies on relatively weak assumptions concerning the sophistication of the *ToM*s, as is desirable in this evolutionary context. As a result, it also relies on improbable events and so seems bound to be undesirably slow. However, this process suffices to establish that complete learning by the *ToM*s occurs whenever $\alpha > 2$. Further, since it is mechanically impossible to learn others' preferences when $\alpha < 2$, a more sophisticated process cannot significantly improve the result.

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 generically unique SPE.

We now have the main result—

Theorem 2: *Consider an environment, $\mathscr{E}$, satisfying assumptions A1-A4. Suppose that there are a finite number of types—naive and ToM, one of which is the SPE-ToM. Suppose further that every alternative to the SPE-ToM, that is a also a ToM type, differs from the SPE-ToM at every reached decision node in a set of games that arises with positive probability under the distribution F. If $\alpha \in (2, T)$, then the proportion of SPE-ToM in role i, R_{it}, say, tends to 1 in probability, $i = 2, ..., I$.*

Note that we focus here on the case that $\alpha \in (2, T)$.[14]

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. 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 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).

[14] If $\alpha < 2$, the *ToM* players are also overwhelmed with novelty. The *ToM*s may still outperform the naive players, but this depends on the detailed behavior of the *ToM*s when facing games in which the payoffs of other players are unknown. In the case that $\alpha > T$, the relative performance of the two types depends on the detailed behavior of the naive players. If the naive players are ultra-fast learners, and play an SPE strategy the second time they encounter a given game, they might keep up with the *ToM*s. More realistically, they would not learn so quickly, and so would lag the *ToM*s.

When making their decisions, participants only observe their own payoff at each outcome and are originally uninformed of the payoff for the other participant. 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 "?". 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. 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, by observing how role 2 players respond to various choices presented to them in the repetitions of the game. 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 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 seven of our sessions, payoff possibilities for each participant consisted of integers between 1 and 7, and in our final two sessions the set was {1,2,3,4,8,9,10}.[15] The one exception was our first session which used only 6 payoff pairs, where payoffs were integers between 1 and 6.[16]

Each session lasted for 90 periods in which, in the first 15 periods, the game set included 15 randomly chosen games from the set Q. 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 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.

One potential issue with our design is that learning 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 eliminate all but the mechanics of the inference problem faced by player 1—in essence the instructions would be providing the theory of mind.

A second potential concern is that foregone payoffs (due to role 1 player's choice) may lead to non-rational 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 related our results directly to theory of mind, as this is measured by two short survey instruments. At the conclusion of the experiment, participants completed the

[15] This variation was employed in our final two sessions and was intended to reduce noise by more strongly disincentivizing 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.

[16] We include this session in our data analysis, but our results are qualitatively unchanged if we exclude it.

Autism-Spectrum Quotient (AQ) survey designed by Baron-Cohen [1], 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. Participants also completed the Broad Autism Phenotype Questionnaire (BAP) due to Hurley [5], which provides a similar measure of autism spectrum behavior. 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. Copies of the questionnaires are available in Appendices D and E.

We report data from 11 experimental sessions with a total of 86 participants (43 in each role). Each experimental session consisted of 6-10 participants, recruited from the student body of Simon Fraser University. Participants entered the lab and were seated at visually isolated computer terminals where they privately read self-paced instructions. A monitor 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 (2) 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 $27.22, with a maximum of $45 and a minimum of $13. 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.**

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 chose an action consistent with the SPE of the game. For a given game tree, and with repeated play with fixed matching and private information about individual payoffs, it is known that pairs frequently converge to non-cooperative equilibrium outcomes over time [4, 6]. In such a setting, an individual merely need learn her counterpart's preferences over two pairwise comparisons. Our setting is more strategically complex, and hence we are able to observe heterogeneity in *ToM* capabilities and exploit this in our data analysis. First, we describe overall learning trends and we show that learning is correlated with both Autism Quotient and Broad Autism Phenotype scores. Finally, we analyze individual rates of learning and show that learning speed is highly correlated with AQ and BAP subscales associated with social skills.

3.2.1. **Overall Averages.**

Figure 1 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 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.[17]

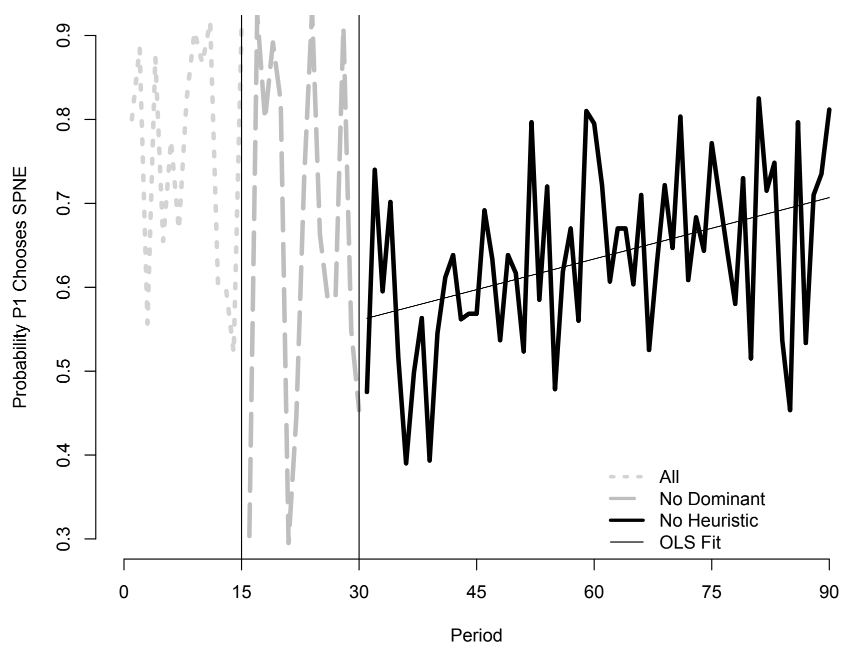

Figure 1: **Time Series of Learning Opponent's Preferences.**

Table 1 reports linear probability panel regressions where the dependent variable takes a value of 1 if player 1 chose 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. We also include an individual's AQ score in column

[17] Table F1 in Appendix F also reports summary statistics for each experimental session.

(2) and BAP score in column (3) to investigate the effect of autism spectrum scores on learning rates. Higher scores on both instruments indicate increasing presence of autism spectrum behaviors. We include random effects for each subject to control for repeated measures, and we cluster standard errors at the session level.

	(1)	(2)	(3)
	P1 Chose SPNE	P1 Chose SPNE	P1 Chose SPNE
Autism Quotient		-0.009**	
		(0.004)	
Broader Autism Phenotype			-0.069*
			(0.041)
Period	0.003***	0.003***	0.003***
	(0.001)	(0.001)	(0.001)
No Dominant Options	0.132***	0.132***	0.132***
	(0.027)	(0.027)	(0.027)
All Treatments	0.276***	0.276***	0.276***
	(0.043)	(0.043)	(0.043)
Constant	0.466***	0.573***	0.669***
	(0.043)	(0.071)	(0.138)
Observations	3960	3960	3960
Wald Chi-Sq.	41.57	50.27	68.99

Clustered standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 1: **Regression Analysis of Learning.**

The positive and significant estimated coefficient on Period indicates that participants are increasingly likely to choose optimally over time. Thus, even in this complex setting, individuals are able to learn the preferences of others. Positive and significant coefficients on the treatment dummy variables support our decision to screen out the games with dominant and highest mean rule of thumb strategies. Moreover, negative and significant estimated coefficients on the AQ and BAP scores provide the first evidence that our games reflect *ToM* as implicated in autism. A one standard deviation increase in AQ score is associated with a 4.1 percentage point reduction in the probability of choosing optimally, on average. Similarly, a one standard deviation increase in BAP score is associate with a 3.2 percentage point reduction. We summarize these findings below:

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

Finding 2: This increase in learning is correlated with the AQ and BAP questionnaires.

3.2.2. **Individual Heterogeneity in Learning Rates.**

Our finding that the AQ and BAP scores are correlated with average performance is suggestive, but the most important aspect of ToM in our setting is *learning* the preferences of others. The faster player 1 learns the preferences of role 2 players, the higher his payoff. Hence, we now turn our attention to heterogeneity in individual learning rates. In particular, we explore the relationship between individuals' learning rates in our experiment and their AQ and BAP scores.

To estimate individual learning rates for each player 1 we estimate a linear regression where the dependent variable takes a value of 1 when the player chose a node consistent with SPE and 0 otherwise and the independent variable is a period trend and a constant term. We exclude all games with dominant strategies and all games in which the "highest mean" rule of thumb yields the SPE choice. The coefficient on the period trend, β, provides an estimate of each individual's rate of learning. We them compute simple correlation coefficients of individual βs and measures of *ToM* from the AQ and BAP questionnaires. Recall that on both instruments, a higher score indicates increased presence of more autism spectrum behaviors. Thus, negative correlations will indicate that our rate of learning measure provides similar information to the AQ and BAP surveys, while the absence of correlation or positive correlations will indicate that *ToM* in strategic environments differs from *ToM* in other social contexts.

Table 2 reports these simple correlations between estimated β's and various measures of autism spectrum intensity, and we include 90% confidence intervals for the correlation coefficients. To indicate what portion of the variance in learning rates is attributable to differences in autism spectrum scores, we also report R^2 estimates from separate OLS regressions where the dependent variable is an individual's estimated β and the independent variable is an individual's score on a particular scale or sub-scale of the AQ and BAP questionnaires.

From the table, we can see that learning rates are correlated with average AQ and BAP scores, and this provides additional evidence that our games measure *ToM* . However, most striking is the strong correlation between learning rates and the two subscales that emphasize *social skills*: AQ_Social and BAP_Aloof. In particular, these

	Correlation Coef.	10%	90%	R-Squared
BAP	-0.25*	-1	-0.05	0.06
BAP_Rigid	0.05	-1	0.25	0.00
BAP_Aloof	-0.43***	-1	-0.25	0.18
BAP_Prag	-0.14	-1	0.06	0.02
AQ	-0.21*	-1	-0.01	0.04
AQ_Social	-0.47***	-1	-0.30	0.22
AQ_Switch	-0.09	-1	0.11	0.01
AQ_Detail	0.19	-1	0.38	0.04
AQ_Commun	-0.15	-1	0.05	0.02
AQ_Imagin	-0.21*	-1	-0.01	0.05
N	43			

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, one-sided tests.

TABLE 2: **Correlations between Autism Spectrum Measures and 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.

scales are concerned with the extent to which individuals understand and enjoy social interaction. 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 precisely the idea of *ToM* in a strategic setting.

Learning is also correlated with the AQ_Imagin subscale which 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. Most of the other subscales exhibit negative but insignificant correlation coefficients. Interestingly, the two subscales that exhibit non-negative coefficients (BAP_Rigid and AQ_Detail) emphasize 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 ToM deficits, perhaps accounting for the lack of correlation.

Importantly, our survey data reveal measured *ToM* in the normal range. Thus, differences in the strategic aspects of *ToM* vary significantly across individuals in the normal range of social intelligence. Figure 2 displays histograms of AQ and BAP scores over the range of feasible scores.

Finding 3: Individual learning is highly correlated with the social skills subscales of the AQ and BAP questionnaires.

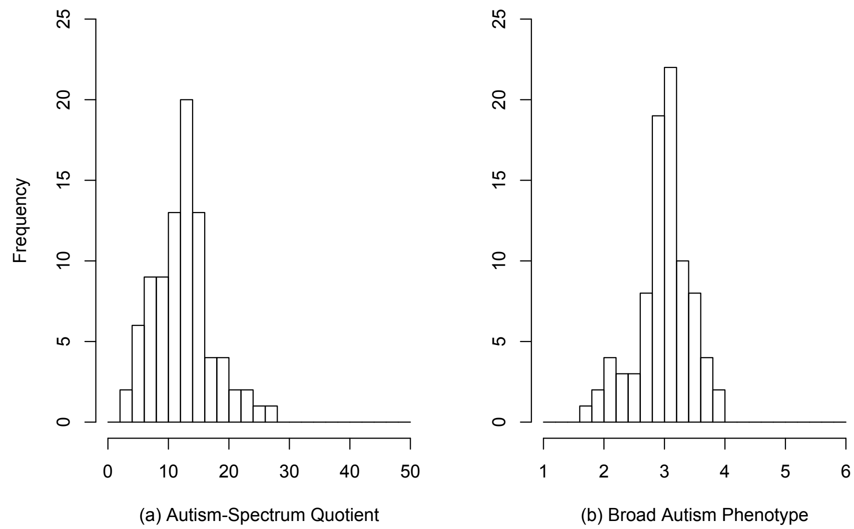

Figure 2: **Histograms of AQ and BAP Scores.** Each panel includes the entire range of feasible scores.

4. Conclusions

This paper presents a theoretical model of the evolution of theory of mind. The model demonstrates the advantages to predicting opponents' behavior in simple games of perfect information. A departure from standard game theory is to 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 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 significant learning over time, and 2) strong correlations between behavior in these experiments and responses to two well-known survey instruments measuring *ToM* from psychology, thus justifying the use of the term "theory of mind" in the present context. We show, in particular, that there is a highly significant correlation between the social skill subscales on standard short tests for autism and the rate at which player 1's learn player 2's preferences.

Appendices

A. Proofs of the Theorems

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

A1- A4 will be assumed throughout without further mention.

We first establish (in Lemma 3 below) the negative claims of Theorem 1, that is, we consider first arrival rates of novelty for which either L_{it} or γ_t converge to zero. After this, the positive claim of Theorem 1 will be proved.

Lemma 3: *Each of the following is true.*

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

ii) Recall there are T terminal nodes. If $\alpha \in [0, T)$, then $\gamma_t \longrightarrow 0$ surely.

Proof. Consider any environment in which the underlying game tree has T end-nodes. Clearly $L_{it} \leqslant t \cdot T/|Z_t|^2$ everywhere, since the maximal number of binary preference orderings that can be revealed for any player at any date is bounded above by T. Similarly, since only one game is played in a 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,

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

Surely $|Z_t| \longrightarrow \infty$. The lemma follows immediately since obviously whenever $\alpha < 2$, for instance, (1) implies $L_{it} \longrightarrow 0$ surely. ■

In to prove the positive claims of Theorem 1 (that L_{it} converges to one when $\alpha > 2$) we proceed by induction using the following three results. First a required notation—

Definition 3: *Let the random variable K_{it} denote the number of pairs (z, z') such that i's preferred outcome in $\{z_i, z_i'\}$ has been revealed.*

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

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

for all $t \geqslant 1$, where each $\{\theta_{1t}(\xi)\}$ is non-increasing between arrival dates, and converges in probability to a function $\theta_1(\xi)$, which converges to zero as ξ approaches zero.

Lemma 5: *Consider a fitness type $i \in I$. 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 $\{\theta_{it}(\xi)\}$ is non-increasing between arrival dates, and converges in probability to a function $\theta_i(\xi)$, which converges to zero as ξ approaches zero. Then $\{L_{it}\}$ converges to one in probability.

Lemmas 4 and 5 together imply that, if $\alpha > 2$, then, for player role 1, L_{1t} converges to one in probability. Theorem 1 then follows by induction from the next claim.

Lemma 6: *Consider a fitness type $i > 1$. Suppose $\alpha > 2$. Suppose further that L_{jt} converges to one in probability for each $j \leqslant i - 1$. Then, for each $\xi \in (0, 1]$ there is a sequence of random variables $\{\theta_{it}(\xi)\}$, non-increasing between arrival dates, such that*

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

for all $t \geqslant 1$, where each $\{\theta_{it}(\xi)\}$ converges in probability to a function $\theta_i(\xi)$, which converges to zero as ξ approaches zero.

That is, if L_{jt} converges to one in probability for each $j < i \leqslant I$. Then, in the limit, the probability of revealing new information about i preferences is small *only if* the fraction of extant knowledge about i preferences, L_{it}, is close to one.[18]

A.1.1. **Proof of Lemma 6.**

The proof of Lemma 4, which is a special case of Lemma 6 except for its focus on the *1* player role in particular, goes through with slight modifications to the notation developed here for the $i > 1$ players. Thus, Lemma 6 will be proved directly but not Lemma 4. For the remainder of this section fix a player role $i > 1$.

We focus on dominance-solvable i player subgames. Rather than keep track of all of these, fix two end-nodes of the i player extensive form 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

[18] 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 6, 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 two reasons—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, and because of outcomes that are avoided by the remaining opponents, making it difficult to reveal information about i's preferences over those outcomes.

"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 fixed way.

Definition 4: *Let $Q_i(z, z')$ be the set of i player subgames with outcome z appended to "one", and z' appended to "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 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.* [19]

Definition 5: *Let Q_{it} denote all the i player subgames possible in period t,*[20] *and denote by $Q^*_{it} \subseteq Q_{it}$ the subgames for which all $1, \ldots, i-1$ player preferences have been revealed along H_t.*

Definition 6: *Let the random variable N_{it} denote pairs of outcomes (z, z') available in period t for which i's favored among z and z' has not been revealed along H_t.*

Suppose a subgame $q \in Q_i(z, z') \cap Q^*_{it}$ is reached, with $z_i > z'_i$. By A4 every i player there will choose a_1, and moreover, every *ToM* observer will know that every i player chose a_1. By the definition of $Q_i(z, z') \cap Q^*_{it}$, the backward induction outcomes of the subgames following a_1 and a_2 are common knowledge among *ToM* players. Every *ToM* can 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.

It follows then that if $(z, z') \in N_{it}$ and the i players in period t reach a subgame in $Q_i(z, z') \cap Q^*_{it}$, their choice there reveals new information about their preferences. The fraction of i subgames, among those in Q_{it}, where i choice reveals new information about i preferences is then 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}. There are A^{I-i} i player decision nodes. Therefore,

$$E(K_{it+1} - K_{it} \mid H_t) \geqslant \left(\sum_{(z,z') \in N_{it}} \frac{|Q_i(z, z') \cap Q^*_{it}|}{|Q_{it}|} \right)^{A^{I-i}}. \tag{2}$$

Consider now some additional required notation.

[19] In proving Lemma 4, since there is no subsequent player after the 1 fitness type, define $Q^*_1(z, z')$ as the *1* player subgames with z available at a_1 and z' at a_2, where player 1's are not indifferent between z and z'.

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

Definition 7: *Let $\boldsymbol{Z}$, and $\boldsymbol{Z}_t$ denote the $[A^i - 2]$-times product of Z, and Z_t, respectively. Bold letters $\boldsymbol{x}$, $\boldsymbol{z}$, etc., denote elements of $\boldsymbol{Z}$. Define $\boldsymbol{Z}(z,z') \subset \boldsymbol{Z}$ as follows. The tuple of outcomes $\boldsymbol{x} = (x^1, \ldots, x^r) \in \boldsymbol{Z}(z,z')$ if and only if the subgame with z at node "one", z' at node "two", x^1 at "three", ..., x^r at node A^i, is in $Q_i(z,z')$. Define then the functions $E_t(z,z') = |\boldsymbol{Z}_i(z,z') \cap \boldsymbol{Z}_t|/|\boldsymbol{Z}_t|$, and $E(z,z') = \int_{\boldsymbol{x}\in\boldsymbol{Z}_i(z,z')} f(x^1)\cdots f(x^r)\,d\boldsymbol{x}$.*

Since each i player subgame has A^i endnodes, $|Q_{it}| = |Z_t|^2 \cdot |\boldsymbol{Z}_t|$. Thus, for any finite subset $X \subseteq Z \times Z$,

$$\begin{aligned}
\sum_{(z,z')\in X} \frac{|Q_i(z,z') \cap Q_{it}^*|}{|Q_{it}|} &\geqslant \sum_{(z,z')\in X} \frac{|Q_i(z,z') \cap Q_{it}|}{|Q_{it}|} - \frac{|Q_{it} \setminus Q_{it}^*|}{|Q_{it}|} \\
&= \sum_{(z,z')\in X} \frac{|Q_i(z,z') \cap Q_{it}|}{|Z_t|^2 |\boldsymbol{Z}_t|} - \frac{|Q_{it} \setminus Q_{it}^*|}{|Q_{it}|} \qquad (3) \\
&\geqslant \frac{1}{|Z_t|^2} \sum_{(z,z')\in X} E(z,z') - \sup_{(z,z')\in Z\times Z} \left\{E(z,z') - E_t(z,z')\right\} - \frac{|Q_{it} \setminus Q_{it}^*|}{|Q_{it}|}.
\end{aligned}$$

Write $\boldsymbol{S}(\xi) = \{(z,z') \in Z \times Z : E(z,z') < \xi\}$, and let

$$\phi_t = \sup_{(z,z')\in Z\times Z} \left\{E(z,z') - E_t(z,z')\right\} + \frac{|Q_{it} \setminus Q_{it}^*|}{|Q_{it}|}.$$

Equations (2) and (3) yield

$$\begin{aligned}
E(K_{it+1} - K_{it} \mid H_t) &\geqslant \left[\frac{1}{|Z_t|^2} \sum_{(z,z')\in N_{it} \setminus \boldsymbol{S}(\xi)} E(z,z') - \phi_t\right]^{A^{I-i}} \\
&\geqslant \left[\frac{\xi}{|Z_t|^2} \left(N_{it} - |\boldsymbol{S}(\xi) \cap \{Z_t \times Z_t\}\,|\right) - \phi_t\right]^{A^{I-i}} \qquad (4) \\
&= \left[\xi \cdot \left(1 - L_{it} - \frac{|\boldsymbol{S}(\xi) \cap \{Z_t \times Z_t\}\,|}{|Z_t|^2}\right) - \phi_t\right]^{A^{I-i}}.
\end{aligned}$$

The terms $E(z,z')$, $E_t(z,z')$ and $|Q_{it}|$ are constant in between arrival dates; $|Q_{it}^*|$ is non-decreasing. We thus define the random variable $\theta_{it}(\xi)$ from the claim,

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

Lemma 1 implies $|\boldsymbol{S}(\xi) \cap \{Z_t \times Z_t\}\,|/|Z_t|^2$ converges to $\int_{\boldsymbol{S}(\xi)} f(z)f(z')dzdz'$, which tends to zero as ξ approaches zero (recall A2). Furthermore, if $L_{jt} \longrightarrow 1$ in probability, then $|Q_{it} \setminus Q_{it}^*|/|Q_{it}| \longrightarrow 0$ in probability. With that in mind set $\theta_i(\xi)$ from the claim to $\int_{\boldsymbol{S}(\xi)} f(z)f(z')dzdz'$. The claim is then established by the following result.

Claim 1: *$E_t(z, z')$ almost surely converges uniformly to $E(z, z')$. That is*

$$\sup_{(z,z')\in Z\times Z} |E_t(z, z') - E(z, z')| \longrightarrow 0$$

almost surely.

Proof. We rely on the result from Potscher and Prucha [7]—

> *Suppose 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 required asymptotically uniform equicontinuity condition is:

$$\limsup_{t\longrightarrow\infty} \sup_{(z,z')\in Z\times Z} \sup_{(x,x')\in B(z,z',\eta)} |E_t(z, z') - E_t(x, x')| \longrightarrow 0 \ \text{ a.s.} \quad \text{as} \quad \eta \longrightarrow 0, \tag{5}$$

where we use $B(z, z', \eta)$ to denote the open ball $\{(z, z') \in Z \times Z : \rho((z, z'), (x, x')) < \eta\}$.

By Lemma 1 $E_t(z, z')$ almost surely converges pointwise to $E(z, z')$. A2 gives $E(z, z')$ continuous. We complete the proof by verifying (5).

Define the set $\boldsymbol{Z}(z, z', \eta) \subset \boldsymbol{Z}_t$ such that $\boldsymbol{x} \in \boldsymbol{Z}(z, z', \eta)$ if and only if for every coordinate x of $\boldsymbol{x}$, both $\min_{i\in I} |x_i - z_i| \geqslant \eta$, and $\min_{i\in I} |x_i - z'_i| \geqslant \eta$. Then, for each $(z, z') \in Z \times Z$, and each $(x, x') \in B(z, z', \eta)$,

$$|E_t(z, z') - E_t(x, x')| \leqslant \frac{|\boldsymbol{Z}_t \setminus \boldsymbol{Z}(z, z', \eta)|}{|\boldsymbol{Z}_t|} = U_t(z, z', \eta).$$

Lemma 1 implies $U_t(z, z', \eta)$ almost surely converges uniformly to

$$U(z, z', \eta) = \int_{\boldsymbol{x}\in\boldsymbol{Z} \setminus \boldsymbol{Z}(z,z',\eta)} f(x^1) \cdots f(x^r)\, d\boldsymbol{x}.$$

Hence,

$$\begin{aligned} &\limsup_{t\longrightarrow\infty} \sup_{(z,z')\in Z\times Z} \sup_{(x,x')\in B(z,z',\eta)} |E_t(z, z') - E_t(x, x')| \\ &\leqslant \lim_{t\longrightarrow\infty} \sup_{(z,z)\in Z\times Z} U_t(z, z', \eta) \longrightarrow \sup_{(z,z')\in Z\times Z} U(z, z', \eta) \quad \text{a.s.} \end{aligned} \tag{6}$$

A2 ensures the limit converges to zero as $\eta \longrightarrow 0$. ■

A.1.2. **Proof of Lemma 5.**

Fix a preference type $i \in I$ and assume the hypotheses of the lemma. 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 establish 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 this connection [(Egghe, 1984), Definition VIII.1.3 and Theorem VIII.1.22].

W-SUBMIL CONVERGENCE: *The adapted process (L_t, H_t) is a* ***weak sub-martingale in the limit (w-submil)*** *if almost surely, for each $\eta > 0$, 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 we show that $\{L_t\}$ is a w-submil under the hypotheses of Lemma 5. As first step we prove that the arrival date subsequence $\{L_{t_k}\}$ is a w-submil. Toward that end, consider consecutive arrival dates t^*, τ^*, with $\tau^* > t^*$. By the definition of L_t,

$$L_{\tau^*} - L_{t^*} = \frac{1}{|Z_{\tau^*}|^2} \sum_{t=t^*}^{\tau^*-1} [K_{t+1} - K_t] - \frac{|Z_{\tau^*}|^2 - |Z_{t^*}|^2}{|Z_{\tau^*}|^2} \cdot L_{t^*}. \tag{7}$$

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

$$\begin{aligned} &\sum_{t=t^*}^{\tau^*-1} E(K_{t+1} - K_t \mid H_{t^*}) \\ &\geqslant \sum_{t=t^*}^{\tau^*-1} \left[\xi \cdot E\left(1 - L_t - \theta_{it}(\xi) \,\middle|\, H_{t^*}\right)\right]^{A^{I-i}} \\ &> [\tau^* - t^*] \cdot \left[\xi \cdot E\left(1 - L_{\tau^*-1} - \theta_{it^*}(\xi) \,\middle|\, H_{t^*}\right)\right]^{A^{I-i}} \end{aligned} \tag{8}$$

The third line uses the fact that L_t is non-decreasing between arrival dates, that $\theta_{it}(\xi)$ is non-increasing between arrival dates (see the definition of $\theta_{it}(\xi)$ in the statement of Claim 5).

Combining equations (7) and (8), and using the fact that $L_{\tau^*} \geqslant L_{\tau^*-1} \cdot |Z_{t^*}|^2/|Z_{\tau^*}|^2$ surely, yields—after some algebra—

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

$$E(L_{\tau^*} \mid H_{t^*}) > \frac{|Z_t|^2}{|Z_{\tau^*}|^2}\left[1 - \theta_{it^*}(\xi) - \frac{1}{\xi}\cdot\left[\frac{|Z_{\tau^*}|^2 - |Z_{t^*}|^2}{\tau^* - t^*}\right]^{\frac{1}{A^{I-i}}}\right] \equiv A_{t^*}(\xi). \tag{9}$$

The left-hand-side of (9) surely converges to $1 - \theta_{it}(\xi)$ for each $\xi \in (0, 1]$. To see this note that $|Z_{t^*}|^2/|Z_{\tau^*}|^2 \longrightarrow$ surely converges to one, and observe that if t^* is the arrival date of the k-th new outcome then

$$\frac{|Z_{\tau^*}|^2 - |Z_{t^*}|^2}{\tau^* - t^*} = \frac{(|Z_1| + k + 1)^2 - (|Z_1| + k)^2}{\lfloor(|Z_1| + k + 1)^\alpha\rfloor - \lfloor(|Z_1| + k)^\alpha\rfloor},$$

which surely converges to zero when $\alpha > 2$.

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

We use (9) to prove the following. For each $\eta > 0$ there exists an M such that for all arrival dates t_m, t_n, such that $n > m \geqslant M$,

$$P\{E(\hat{L}_n \mid \hat{H}_m) - \hat{L}_m \geqslant -\eta\} > 1 - \eta. \tag{10}$$

To that end suppose for some m and n, that $n > m$, $E(\hat{L}_n \mid \hat{H}_m) < \hat{L}_m$. 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 $\{m, , \dots, n-1\}$ 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 $k = r+1, \dots, n-1$. According to (9)

$$E(\hat{L}_{r+1} \mid \hat{H}_m) \;>\; E(A_{t_r}(\xi) \mid \hat{H}_m).$$

Hence,

$$E(\hat{L}_n \mid \hat{H}_m) - \hat{L}_m \;>\; E(A_{t_r}(\xi) \mid \hat{H}_m) - \hat{L}_m \;\geqslant\; E(A_{t_r}(\xi) \mid \hat{H}_m) - 1.$$

Now, recall that $A_t(\xi)$ converges surely to $1 - \theta_{it}(\xi)$ which, by the hypothesis of the claim, converges in probability to $1 - \theta_i(\xi)$. We can then choose an arrival M large enough so that

$$P\{E(A_{t_k}(\xi) \mid \hat{H}_m) - 1 > -2\cdot\theta_i(\xi)\} > 1 - 2\cdot\theta_i(\xi)$$

for all k and m with $k > m \geqslant M$, and thus for $k > m \geqslant M$,

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

By assumption $\theta_i(\xi)$ converges to zero as ξ approaches zero. Hence (10) can be obtained by choosing ξ so that $\theta_i(\xi) < \eta/2$ and thus $\{\hat{L}_k\}$ is a w-submil.

Having established that $\{\hat{L}_k\}$ is a w-submil, it remains to show that $\{L_t\}$ is also a w-submil. Consider any dates t and τ where $t < \tau$. Then

$$L_\tau - L_t \geqslant L_{\tau^*} - L_{t^*} \frac{|Z_{t^*}|^2}{|Z_{\tau^*}|^2}$$

everywhere, when t^* is the first arrival date after t and τ^* is the greatest arrival date less than or equal to τ. Since $\{\hat{L}_k\}$ is a w-submil and each $|L_t|$ bounded above by 1, the w-submil convergence result implies $L_{\tau^*} - L_{t^*} \longrightarrow 0$ in probability. Furthermore, $|Z_t|^2/|Z_{t+1}|^2 \longrightarrow 1$. 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 5,

$$\begin{aligned} E(L_t) &= \sum_{t=0}^{\tau-1} E\left(K_{t+1} - K_t\right)/|Z_\tau|^2 \\ &\geqslant \xi^{A^{I-i}} \cdot \frac{\tau}{|Z_\tau|^2} \cdot \left[\frac{1}{\tau}\sum_{t=1}^{\tau-1} \cdot E\left(1 - L_{it} - \theta_{it}(\xi)\right)^{A^{I-i}}\right]. \end{aligned} \tag{11}$$

It is straightforward to show that

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

Then, since $\tau/|Z_\tau|^2 \longrightarrow \infty$ whenever $\alpha > 2$,[21] (11) implies, since L_t is everywhere bounded by one,

$$E\left(1 - L - \theta_i(\xi)\right) \leqslant 0,$$

for all $\xi \in (0, 1]$. By assumption $\theta_i(\xi)$ tends to zero as ξ approaches zero, implying $E(1 - L) = 0$, and thus $L = 1$ a.e.

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 8: *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, $\mu_t(S) = |S \cap Q_t|/|Q_t|$.*

---

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

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.[22] 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.[23] 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—

A5: *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.*

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 9: *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 10: *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 $\delta \geqslant 0$.*

---

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

[23] 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.

Definition 11: *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 12: *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

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

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. ∎

Consider the realized one period growth rate of the alternative *ToM* type relative to that of the *SPE-ToM* at date t. (We omit the detailed argument for a naive alternative since it is nearly identical.) In view of Claim 2, this rate is bounded above by

$$\begin{aligned} & I_t(\delta)J_t(\varepsilon)D_t \cdot \ln\left(1 + \frac{\varepsilon}{1-\varepsilon}\frac{M}{m} - \frac{\delta}{M}\right) \\ & + I_t(0)(1 - I_t(\delta))J_t(\varepsilon) \cdot \ln\left(1 + \frac{\varepsilon}{1-\varepsilon}\frac{M}{m}\right) \\ & + (1 - I_t(0) + 1 - J_t(\varepsilon)) \cdot \ln M/m. \end{aligned} \tag{12}$$

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 $I_t(0) = 1$, 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, (12) 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 (12) is bounded above by

$$\begin{aligned} & \Delta_t(\delta,\varepsilon) \equiv (D_t + J_t(\varepsilon) - 1) \cdot \ln\left(1 + \frac{\varepsilon}{1-\varepsilon}\frac{M}{m} - \frac{\delta}{M}\right) \\ & + (1 - I_t(\delta)) \cdot \ln\left(\left[1 + \frac{\varepsilon}{1-\varepsilon}\frac{M}{m}\right] \Big/ \left[1 + \frac{\varepsilon}{1-\varepsilon}\frac{M}{m} - \frac{\delta}{M}\right]\right) \\ & + (1 - I_t(0) + 1 - J_t(\varepsilon)) \cdot \ln M/m. \end{aligned} \tag{13}$$

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) Suppose* $\alpha < T$ *and that the alternative is a naive type. Let* μ^* *be the measure of games where* i *has no dominant action at any node. Then* $\frac{1}{\tau}\sum_{t=1}^{\tau} D_t$ *converges in probability to* μ^*. *iii) Assume A5. If 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 A5.*

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

Consider now ii). Assume the alternative is a naive type. 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

$$\frac{1}{\tau}\sum_{t=1}^{\tau}(A_t - 1) + \frac{1}{\tau}\sum_{t=1}^{\tau} B_t \leqslant \frac{1}{\tau}\sum_{t=1}^{\tau} D_t \leqslant \frac{1}{\tau}\sum_{t=1}^{\tau} B_t. \tag{14}$$

$E(A_t)$ 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 (14) 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 $E(B_t|H_\infty)$ converges to μ^* for almost every complete history H_∞. Then note that the random variables $(B_t|H_\infty)$ are independent, for each H_∞, and that the sequence satisfies Kolmogorov's criterion. Thus, $\frac{1}{\tau}\sum_{t=1}^{\tau}[(B_t|H_\infty) - E(B_t|H_\infty)] \longrightarrow 0$, so that $\frac{1}{\tau}\sum_{t=1}^{\tau}(B_t|H_\infty) \longrightarrow \mu^*$, so $\frac{1}{\tau}\sum_{t=1}^{\tau} B_t \longrightarrow \mu^*$, almost surely. ■

Claims 3 and 4 (using also (13) and the fact that $\mu(Q_\delta)$ converges to one as δ tends to zero) then give that

$$\frac{1}{\tau}\sum_{t=1}^{\tau}\Delta_t(\delta,\varepsilon) \longrightarrow \mu(S)\cdot\ln\left(1+\frac{\varepsilon}{1-\varepsilon}\frac{M}{m}-\frac{\delta}{M}\right) \\ + (1-\mu(Q_\delta))\cdot\ln\left(\left[1+\frac{\varepsilon}{1-\varepsilon}\frac{M}{m}\right]\Big/\left[1+\frac{\varepsilon}{1-\varepsilon}\frac{M}{m}-\frac{\delta}{M}\right]\right), \tag{15}$$

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 (15) 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. Instructions

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 persons 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

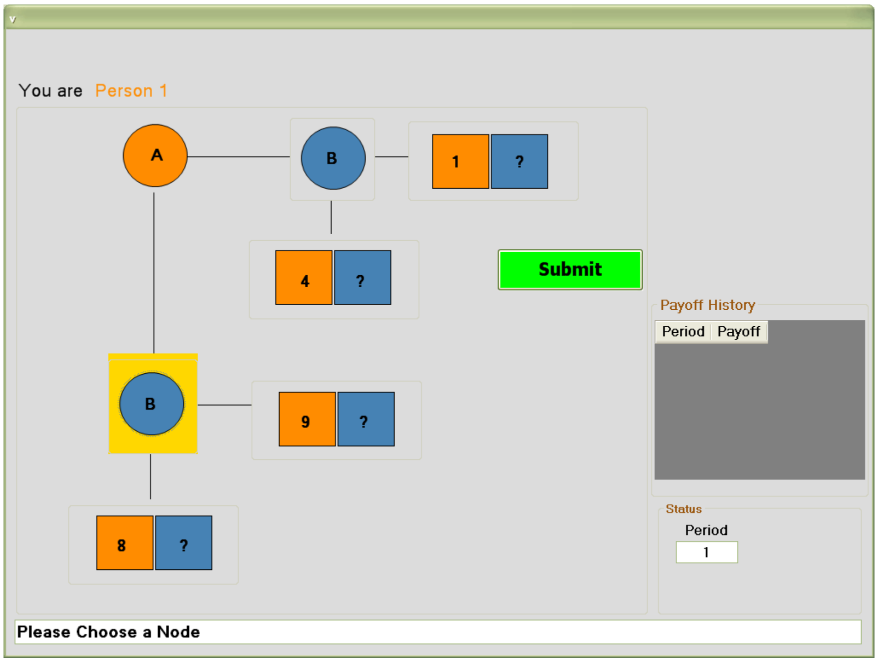

Figure C1: **Screenshot for Player 1.** This figure shows the screen as player 1 sees it prior to submitting his choice of action. The yellow highlighted node indicates that player 1 has provisionally chosen the corresponding action, but the decision is not final until the submit button is clicked. While 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.

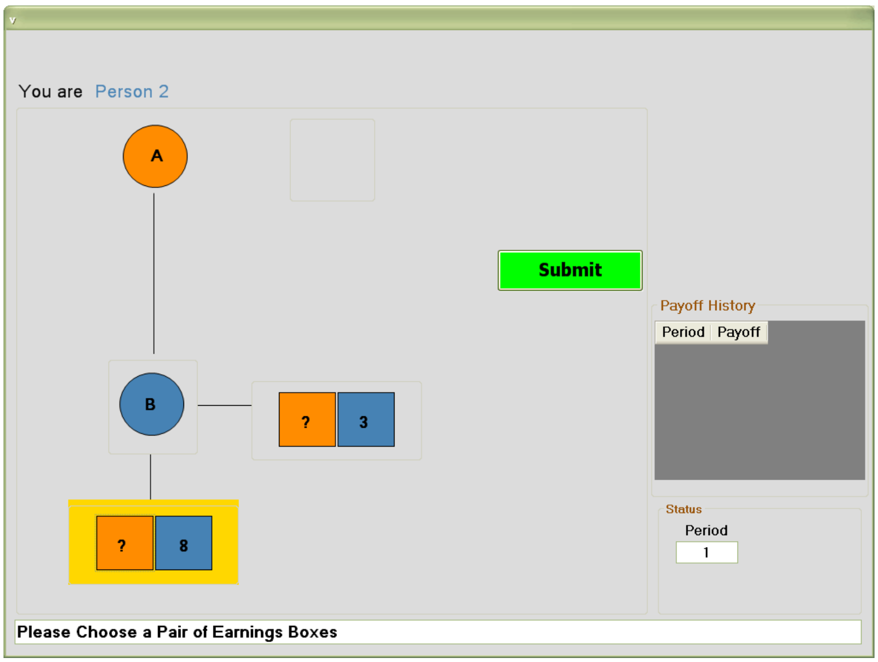

FIGURE C2: **Screenshot for Player 2.** This figure shows the screen as player 2 sees it after player 1 has chosen an action. Here, player 1 chose to move down, so the upper right portion of the game tree is no longer visible. While player 2 is making a decision, player 1 sees an identical screen except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

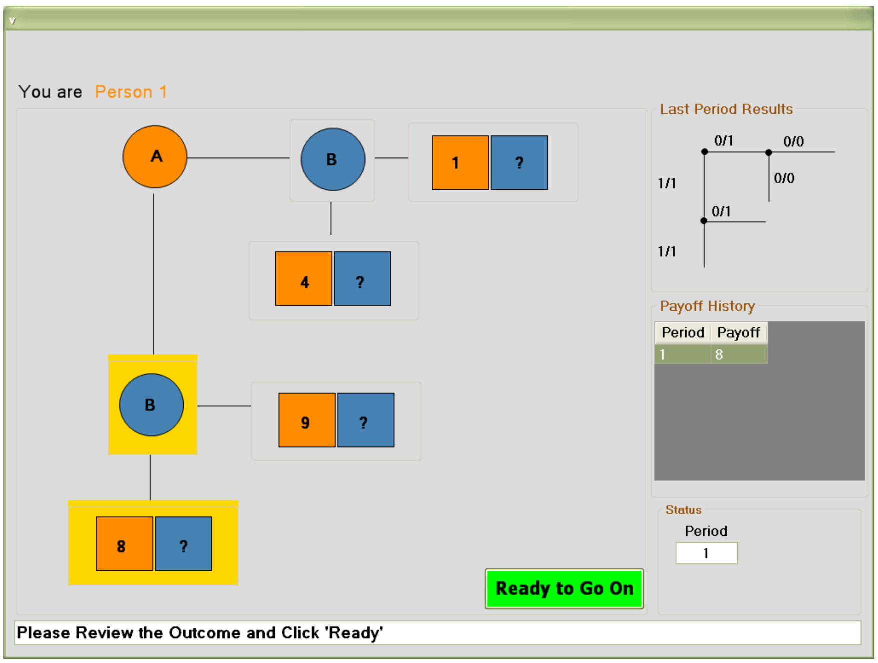

Figure C3: **Screenshot of Post-Decision Review.** This figure shows the final screen subjects see in each period after both player 1 and player 2 have made their decisions. The smaller game tree in the upper right portion of the figure displays information about how many pairs ended up at each outcome. 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

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

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

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

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

	BAPQ

F. Additional Tables and Figures

	6-10-1	7-8-1	7-8-2	7-10-1	7-10-2	7-8-3	7-6-1	7-8-4	7-10-3	7-8-5
Pairs SPE	0.84	0.68	0.54	0.62	0.61	0.63	0.44	0.51	0.46	0.68
P1 SPE	0.87	0.72	0.59	0.64	0.71	0.67	0.54	0.61	0.54	0.70
P1 Heur	0.52	0.36	0.40	0.42	0.38	0.42	0.45	0.34	0.55	0.43
P1 SPE $\mid$ heur $\neq$ SPE	0.79	0.64	0.52	0.55	0.58	0.55	0.36	0.49	0.37	0.64
P1 SPE $\mid$ heur = SPE	0.89	0.81	0.60	0.86	0.74	0.88	0.71	0.58	0.75	0.78
P1 SPE $\mid$ No Heur	0.88	0.71	0.69	0.65	0.72	0.70	0.49	0.56	0.57	0.74
P2 Dom	0.95	0.94	0.90	0.98	0.86	0.94	0.84	0.85	0.82	0.97
P2 SPE $\mid$ P1 SPE	0.96	0.94	0.91	0.98	0.87	0.94	0.83	0.84	0.86	0.96

Table F1: **Observed Probability of Outcomes by Session.**

Each entry is a probability that we observed a particular outcome in a particular session. Pairs SPE refers to the probability that a pair ended at the SPE. P1 SPE is the probability that player 1's choice was consistent with the SPE. P1 Heur is the probability that Player 1 chose in a manner consistent with the "highest mean" rule of thumb (heuristic). P1 SPE $\mid$ heur $\neq$ SPE is the probability that player one followed the SPE when it did not correspond to the "highest mean" rule of thumb. P1 SPE $\mid$ heur = SPE is the probability that player one followed the SPE when it *did* correspond to the "highest mean" rule of thumb. P1 SPE $\mid$ No Heur is the probability that player 1 followed the SPE when the rule of thumb was inapplicable (i.e. equal means). P2 Dom is the probability that player 2 chose the dominant strategy. P2 SPE $\mid$ P1 SPE is the conditional probability of player 2 choosing the dominant strategy given that player 1 followed the SPE. Sessions are labeled in the format # of Payoffs – # of Subjects – Session ID so that 7-10-2 corresponds to the 2nd session with 7 payoff pairs and 10 subjects.

References

[1] Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J. and Clubley, E. (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), pp. 5–17.

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

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

[4] Fouraker, L.E. and Siegel, S. (1963): *Bargaining Behavior*. McGraw Hill.

[5] Hurley, R.S., Losh, M., Parlier, M., Reznick, J.S. and Piven, J. (2007): "The broad autism phenotype questionnaire." *Journal of autism and developmental disorders*, 37(9), pp. 1679–1690.

[6] McCabe, K.A., Rassenti, S.J. and Smith, V.L. (1998): "Reciprocity, Trust, and Payoff Privacy in Extensive Form Bargaining." *Games and Economic Behavior*, 24(1), pp. 10–24.

[7] Pötscher, B.M. and Prucha, I.R. (1994): "Generic Uniform Convergence and Equicontinuity Concepts for Random Functions." *Journal of Econometrics*, 60(1), pp. 23–63.

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

E.O. Kimbrough: ekimbrough@gmail.com
N. Robalino: nkasimat@sfu.ca
A.J. Robson: robson@sfu.ca