The Evolution of Strategic Sophistication*

Nikolaus Robalino

Arthur J. Robson

December 2, 2014

Abstract. This paper investigates the evolutionary foundation for our ability to attribute preferences to others, an ability that is central to conventional game theory. We argue here that learning others' preferences allows individuals to efficiently modify their behavior in strategic environments with a persistent element of novelty. Agents with the ability to learn have a sharp, unambiguous advantage over those who are less sophisticated because the former agents extrapolate to novel circumstances information about opponents' preferences that was learned previously. This advantage holds even with a suitably small cost to reflect the additional cognitive complexity involved.

*Affiliation and mailing address for Robalino—Department of Economics, Rochester Institute of Technology, Rochester, NY; for Robson—Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada. Email addresses—ndrobalin@gmail.com and robson@sfu.ca, respectively. We thank Dean Foster, Matt Jackson, Erik Kimbrough, Leanna Mitchell, Daniel Monte, Andrew Newman, Antonio Penta, Luis Rayo, Phil Reny, Larry Samuelson, Bill Sandholm, Ricky Vohra, and three referees for helpful discussions. The paper also benefited from comments by participants in seminars at BU, Caltech, Cal Poly, Stanford, UBC, UCR, and UCSB, at conferences sponsored by the Becker-Friedman Institute at the University of Chicago, by the Toulouse School of Economics, by the Max Planck Institute at Ringberg Castle, and at the NSF/NBER CEME GE Conference at the University of Wisconsin, Madison. Robalino and Robson thank the Human Evolutionary Studies Program at SFU; Robson also thanks the Guggenheim Foundation, the Canada Research Chair Program and the SSHRC Insight Grants Program.

1. Introduction

Conventional game theory relies on agents correctly ascribing preferences to the other agents. Unless an agent has a dominant strategy, that is, her optimal choice depends on the choices of others and therefore indirectly on their preferences. We consider here the genesis of the strategic sophistication necessary to acquire others' preferences.

We address the questions: *Why* and *how* might this ability to impute preferences to others have evolved? In what types of environments would this ability yield a distinct advantage over alternative, less sophisticated, approaches to strategic interaction? In general terms, the answer we propose is that this ability is an evolutionary adaptation for dealing with strategic environments that have a persistent element of novelty.

Our interpretation of strategic sophistication is dynamic in that it entails *learning* other agents' preferences from their observed behavior. It also extends the theory of revealed preference in that knowing *others'* preferences has consequences for one's own actions. Throughout the paper, we refer to such strategic sophistication, for simplicity, as *ToP*, for "theory of preferences".[1]

The argument made here in favor of such strategic sophistication 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 the particular gamble in question and so 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

[1] Our "theory of preferences" is an aspect of "theory of mind", as in psychology. An individual with theory of mind has the ability to conceive of herself, and of others, as having agency, and so to attribute to herself and others mental states such as belief, desire, knowledge, and intent. It is generally accepted in psychology that human beings beyond infancy possess theory of mind. The classic experiment that suggests children have theory of mind is the "Sally-Ann" test described in Baron-Cohen, Leslie, and Frith (1985). According to this test, young children begin to realize that others may have beliefs they know to be false shortly after age four. This test relies on children's verbal facility. Onishi and Baillargeon (2005) push the age back to 15 months using a non-verbal technique. Infants are taken to express that their expectations have been violated by lengthening the duration of their gaze. The presence of this capacity in such young individuals increases the likelihood that it is, to some degree at least, innate.

the correct gamble to take; conversely, if she acts correctly over a sufficiently rich set of gambles, she must possess, at least implicitly, the appropriate utility function.

We consider here a dynamic model in which players repeatedly interact. Although the perfect information game tree is fixed, with fixed terminal nodes, there are various physical outcomes that are assigned to these terminal nodes in a flexible fashion. More particularly, the outcomes are randomly drawn in each iteration of the game from a finite outcome set, where this outcome set grows over time, thus introducing suitable novelty.

Individuals know how their own utility functions are defined on all these physical outcomes, but do not know the preferences of their opponents. There will be an advantage to an agent of sophistication—of effectively understanding that her opponents act optimally in the light of their preferences. Such a sophisticated agent can then learn opponents' preferences in order to exploit this information.

The sophisticated players are contrasted with naive players who are reinforcement learners, viewing each subgame they initiate as a distinct indivisible circumstance. Naive players condition in an arbitrary fashion on their own payoffs in each novel subgame. That is, their reinforcement learning is *initialized* in a general way.

Sophistication enables players to better deal with the innovation that arises from new outcomes than can such "naive" players that adapt to each subgame as a distinct circumstance.[2] The edge to sophistication derives from a capacity to extrapolate to novel circumstances information that was learned about others' preferences in a previous situation.[3]

---

[2] The novelty here is circumscribed, but it is clear that evolution would be unable to deal with completely unrestricted novelty.

[3] The distinction between the *ToP* and naive players might be illustrated with reference to the following observations of vervet monkeys (Cheney and Seyfarth 1990, p. 213). If two groups are involved in a skirmish, sometimes a member of the losing side is observed to make a warning cry used by vervets to signal the approach of a leopard. All the vervets will then urgently disperse, saving the day for the losing combatants. The issue is: What is the genesis of this deceptive behavior? One possibility, corresponding to our *ToP* strategy, is that the deceptive vervet effectively appreciates what the effect of such a cry would be on the others, acts as if, that is, he understands that they are averse to a leopard attack and exploits this aversion deliberately. The other polar extreme corresponds to our naive reinforcement learners. Such a type has no model whatever of the other monkeys' preferences and beliefs. His alarm cry

Consider now our strategic environment in greater detail. We view the particular environment here as a convenient test-bed on which we can derive the speeds with which the various players can learn. The basic results do not seem likely to be specific to this particular environment, so these differences in relative learning speeds would be manifested in many alternative models.

We begin by fixing a game tree with perfect information, with I stages, say. There are I equally large populations, one for each of the stages or the associated "player roles." In each iteration of the game, a large number of random matches are made, with each match having one player in each role. The physical outcomes assigned to the terminal nodes are drawn randomly and uniformly in each iteration from the finite outcome set that is available then.

Players have preference orderings over the set of outcomes that are ever possible, and so preferences over the finite subset of these that is actually available in each period. Each player is fully aware of her own utility function but does not directly know the preference ordering of his opponents.

At each date, at the start of each period, a new outcome is added to the set of potential outcomes, where each new outcome is drawn independently from a given distribution. The number of times the game is played within each period grows at a parametric rate, potentially allowing the preferences of other players to be learned.[4]

All players see the history of the games played—the outcomes that were chosen to attach to the terminal nodes in each iteration of the game, the choices that were made by all player roles (but not, directly, the preferences of others). Players here differ with respect to the extent and the manner of utilization of this information.

All strategies use a dominant action in any subgame they face, if such an action is available. This is for simplicity, in the spirit of focussing on the implications of others's preferences, while presuming full utilization of one's own

---

behavior conditions simply on the circumstance that he is losing a fight. By accident perhaps, he once made the leopard warning in such a circumstance, and it had a favorable outcome. Subsequent reapplication of this strategem continued to be met with success, reinforcing the behavior.

[4] When there more outcomes already present, there is more to be learned about where a new outcome ranks.

preferences. However, the current set up would permit such sequentially rational behavior to be obtained as a result rather than as an assumption.

Although the naive strategies can condition in an arbitrary way on their own observed payoffs in a novel subgame, it is crucial that they condition only on these payoffs. The other details of these naive strategies are not relevant to the main result. Indeed, even if the naive players apply a fully Bayesian rational strategy the *second* time a subgame is played, they will still lose the evolutionary race here to the *ToP* players. A slower and therefore more reasonable rate of learning for the naive players would only strengthen our results.

Once history has revealed the ordinal preferences of all subsequent players in any subgame to the *ToP* players, they choose a strategy that is a function of these ordinal preferences and their own. Furthermore, there is a particular *ToP* strategy, the *SPE-ToP* strategy, say, that not only observes subsequent preferences but uses the *SPE* strategy associated with these preferences and their own.

The *ToP* players know enough about the game that they can learn the preferences of other player roles, in the first place. In particular, it is common knowledge among all the *ToP* players that there is a positive fraction of *SPE-ToP* players in every role.

It is not crucial otherwise how the *ToP* players behave—they could even *minimize* their payoffs according to a fully accurate posterior distribution over all the relevant aspects of the game, when the preferences of all subsequent players are not known.

We do not assume that the *ToP* players use the transitivity of opponents' preferences.[5] The *ToP* players build up a description of others' preferences only by observing all the pairwise choices. Generalizing this assumption could only strengthen our results by increasing *ToP* players' learning speed.

Between each iteration of the game, the fraction of each role that plays each strategy is updated to reflect the payoffs that this strategy obtains. This updating rule is subject to standard weak assumptions. In particular, the strategy that performs the best must increase at the expense of other strategies.

[5] Indeed, the results here would apply even if preferences were not transitive.

Theorem 1 is the main result here—for an intermediate range of values for a parameter governing the rate of innovation, a unique *SPE* is attained, with the *SPE-ToP* strategy ultimately taking over the population in each role, at the expense of all other strategies—naive or *ToP*.

Moreover, our results hold if the *ToP* incur a fixed per game cost. This is a key finding of the present paper since the previous literature has tended to find an advantage to (lucky and) less smart players over smarter players—see, for example, Stahl (1993). The underlying reasons for the reverse (and more plausible) result here are that, in the limit considered in Theorem 1, i) the naive players do not know the game they face while, at the same time, ii) the *SPE-ToP* players do know all the relevant preferences and, furthermore, have adapted to play the *SPE* strategy.

It is unambiguously better then to be "smart"—in the sense of *ToP* —than it is to be naive, no matter how lucky—even for the relatively mild form of naivete here.

2. A Model

2.1. **The Environment.**

Consider first the underlying games. The extensive game form is a fixed tree with perfect information and a finite number of stages, $I \geqslant 2$, and a fixed finite number of actions, $A \geqslant 2$, at each decision node.[6] Every complete history of the game then has I decision nodes and there are A^I terminal nodes.

There is one "player role" for each such stage, $i = 1, \ldots, I$, in the game. (In a reversal of the usual convention, the first player role to move is I and the last to move is 1. This simplifies the notation used in the proof. Role i therefore has a subgame of rank i in that there are i successor nodes in each path to a terminal node.) Each player role is represented by an equal-sized "large" population of agents, where these agents differ in their choice of strategy. The strategies are described precisely below, but they will be grouped into two "categories"—sophisticated (*ToP*) and naive.

[6] The restriction that each node induce the same number of actions, A, can be relaxed. 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.

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

There is a fixed overall set of physically observable outcomes, each with consequences for the material payoffs of the I player roles. Player role $i = 1, \ldots, I$ has then a function mapping all outcomes to material payoffs. A fundamental novelty is that, although each player role knows her own payoff at each outcome, she does not know the payoffs for the other player roles.

For notational simplicity, however, we avoid the explicit construction of outcomes, with payoff functions defined on these. Given a fixed tree structure with T terminal nodes, we instead simply identify each outcome with a payoff vector and each game with a particular set of such payoff vectors assigned to the terminal nodes. We assume that all material payoffs are scalars, lying in the compact interval $[m, M]$, for $M > m > 0$, for simplicity.[7]

Assumption 1: *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 $n = 1, 2, \ldots,$ denote successive dates. Within each corresponding period, n, there is available a finite subset of outcomes $Z_n \subset Z$, determined in the following way. There is an initial finite set of outcomes $Z_0 \subset Z$, of size N, say, where each of these outcomes is drawn independently from Z according to a cumulative distribution function F as follows.

Assumption 2: *The cdf over outcomes F has a continuous probability density that is strictly positive on Z.*

At date $n \geqslant 1$, at the beginning of period n, a new outcome is added to the existing ones by drawing it independently from Z according to the same cdf F. Within each period, the set of outcomes is then fixed, and once an outcome is

[7] 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. However, with the current set-up, when the cdf F is continuous, the probability of any role's payoff arising more than once, but with different payoffs for the other roles, is zero. Each player i can then safely assume that a given own payoff is associated to a unique (but initially unknown) vector of other roles' payoffs. We then adopt this simpler set-up.

introduced it is available thereafter. Figure 1 is a schematic representation of the game.

We specify the number of games that are played within each period as follows.

Assumption 3: *The number of iterations of the game played in period n is $\kappa(n) = \lfloor (N+n)^{\alpha} \rfloor$, for some $\alpha \geqslant 0$.*[8]

If the parameter α is low, the rate of arrival of novelty is high in that there are not many games within each period before the next novel outcome arrives; if α is high, on the other hand, the rate of arrival of novelty is low.

Consider now a convenient formal description of the set of games available in each period.

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

In each iteration, $t = 1, ..., \kappa(n)$, of the game in period n, outcomes are drawn *independently* from Z_n according to the cdf F_n, so the game is chosen *independently* in each iteration according to G_n.

The cdf's F_n and G_n are well-behaved in the limit. This result is elegant and informative and so is included 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 that is strictly positive on $[m, M]^{IT}$. These two later cdf's are then the limits of the cdf's F_n and G_n—

---

[8] Here $\lfloor \cdot \rfloor$ denotes the floor function. It seems more plausible, perhaps, that the number of games per period 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.

[9] Note that F_n and G_n are random variables measurable with respect to the information available in period n, in particular the set of available outcomes Z_n.

Lemma 1: *It follows that* $F_n(z) \to F(z)$ *and* $G_n(q) \to G(q)$ *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 1986, p. 275, and Elker, Pollard and Stute 1979, p. 825, for its extension to many dimensions.) ■

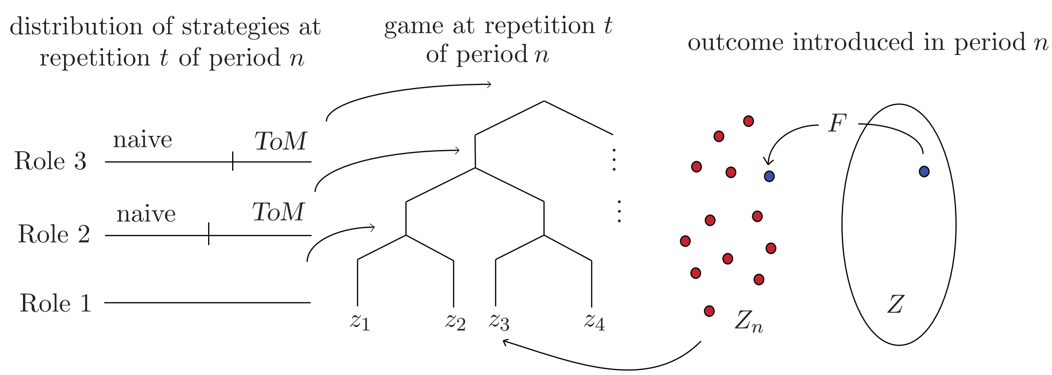

Figure 1: **A Schematic Representation of the Key Elements of the Model.**

We turn now to the specification of the strategies for each player role.

2.2. Strategies.

When making a choice in period n and iteration t, every player, whether naive or *ToP*, knows the history so far, $H_{n,t}$, say, and the game, $q_{n,t}$, drawn for the current iteration. The history records the outcomes available in the current period, n, the randomly drawn games and the empirical distributions of choices made in all previous periods and iterations. Although each player observes the outcome 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.

More precisely, for each player role i, given that decision-node h is reached by a positive fraction of players in period n and iteration t, let $\pi_{n,t}(h) \in \Delta(A)$ then record the aggregate behavior of i player roles at h. It follows that $H_{n,t} =$

$\{Z_n, (q_{1,1}, \pi_{1,1}), \ldots, (q_{n,t-1}, \pi_{n,t-1})\}$.[10] Let $\boldsymbol{H}_{n,t}$ be the set of period n and iteration t histories, and let $\boldsymbol{H} = \bigcup_{n,t} \boldsymbol{H}_{n,t}$.

Strategies can be formally described as follows. Let Σ_i denote the set of choices available to the player role i's. A strategy is then a function $c : \boldsymbol{H} \times Q \longrightarrow \Sigma_i$.[11] An individual in period n at iteration t with strategy c uses $c(H_{n,t}, q_{n,t})$ in game $q_{n,t}$, $c(H_{n,t+1}, q_{n,t+1})$ in $q_{n,t+1}$, and so on.

As part of the specification of the map c, we assume that all strategies choose a strictly dominant action in any 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 assumption is in the spirit of focussing upon the implications of other players' payoffs rather than the implications of one's own payoffs. Indeed, if players are to learn other players' preferences from observing their choices, other players cannot be completely free to act contrary to their own preferences.

More importantly, in the present model, using any such dominant choice could be made a result rather than an assumption. The key part of this assumption is sequential rationality, since such a dominant choice is optimal conditional upon having reached the node in question.

It is the large population in each role that is crucial in this connection. With only a single player in each role, the player in role $i < I$ might well prefer to not choose such a dominant action in order to misrepresent her preferences to some player $j > i$, so inducing j to choose in a way that is beneficial to i. However, when there is a large number of players in every role, who are randomly matched in each iteration of the game, each role i player has no effect on the distribution of i's choices that is observed by any role $j > i$ and thus no effect on j's future behavior. In these circumstances, not only is the best choice by each i myopic, in the sense of neglecting the future, but it is also sequentially rational.

Strategies that failed to use such dominant choices would eventually be pushed to an arbitrarily low level. Once this was so, we would approximate the current

---

[10] If $n > 1$ but $t = 1$, then $H_{n,t} = \left\{Z_n, (q_{1,1}, \pi_{1,1}), \ldots, (q_{n-1,\kappa(n-1)}, \pi_{n-1,\kappa(n-1)})\right\}$. If $n = t = 1$, then $H_{n,t} = \varnothing$.

[11] It will not be required that *ToP* players remember the entire history. All that is needed is that they make and retain the exact inferences about other roles' binary preferences that are possible from observing the aggregate choices made in each period. It is not important whether naive players remember the entire history or not, in familiar subgames.

model. There is no reason then to be suspicious of the current assumption, but the approximation would make the proofs more complicated, so we do not pursue this option.

All strategies then satisfy—

ASSUMPTION 4: *Consider any i player role, and any 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 strategy always chooses any such dominant action.*

2.2.1. **Naive Players.**

We adopt a definition of naivete that binds only if the subgame is new. This serves to make the ultimate results stronger, since the naive players can be otherwise rather smart. When the subgame is new, and there is no dominant choice, naive players condition in an arbitrary fashion on their own payoffs, but act in ignorance of other players' preferences.

DEFINITION 2: *All naive strategies satisfy Assumption 4 in all subgames. There is a finite number of naive strategies that map their own observed payoffs to an arbitrary pure choice, whenever any of the subgames faced has never arisen previously, and a dominant strategy is lacking.*

If any subgame faced is *not* new, and there is no dominant strategy, there is no constraint imposed on any naive strategy. Although it makes an implausible combination, the naive players could then be fully Bayesian rational with respect to all of the relevant characteristics of the game—updating the distribution of opponents' payoffs, for example.

The following example illuminates the strengths and weaknesses of naive strategies, describing the opportunity that exists for more sophisticated strategies—

EXAMPLE 1: *Consider Figure 2. In view of Assumption 4, the P1's always make the SPE choice. The problem for the P2's is to make the appropriate choice for each of the games they face, but where the outcome for each choice depends on the unknown preferences of the P1's.*

The key consideration in the long-run concerns how the various strategies perform when payoffs are chosen independently according to the cdf F. Suppose,

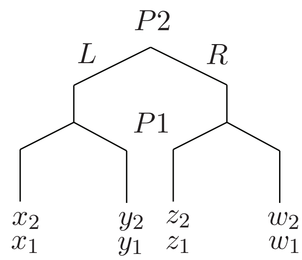

FIGURE 2: $I = A = 2$.

for simplicity, that the cdf F represents independent choice of the two payoffs from the uniform distribution on [1,2].

A salient naive strategy for $P2$ is to choose L, for example, if and only if the 50-50 average of the own payoffs after choosing L exceeds the 50-50 average of the own payoffs after choosing R, in any novel game. That is, choose L if and only if $x_2 + y_2 > z_2 + w_2$. If either choice is dominant, this simple rule makes that dominant choice. Moreover, given risk neutrality in the payoffs, this naive strategy is the Bayesian rational procedure initially when there is no additional information about $P1$'s preferences, since each of $P1$'s choices are then equally likely given either choice for $P2$.

Whenever there is not a dominant choice for $P2$, however, any *naive strategy must make the wrong choice with strictly positive probability, under any F with full support. To show this for this F, it is enough to note the following. There is a clearly a positive probability that neither L nor R is dominant. Further, when there is no dominance, one of the following typical patterns of $P2$'s payoffs must arise i) $x_2 > z_2 > w_2 > y_2$* or *ii) $x_2 > z_2 > y_2 > w_2$.*[12] *In case i), L is optimal if and only if $x_1 > y_1$, which has probability $1/2$. In case ii), L is optimal if* and *only if $x_1 > y_1$* or *both $y_1 > x_1$* and *$w_1 > z_1$. This has overall probability $3/4$. That is,* any *naive rule makes the wrong choice in a nontrivial subset of novel games, given the actual pattern of $P1$'s payoffs.*

Furthermore, we subsequently show that, whenever $\alpha < 3$, the naive $P2$s see only *novel games, in the limit.*[13] *Hence any naive strategy makes a suboptimal*

[12] That is, it is without loss of generality to assume that the highest payoff is x_2. If there is no dominance, the next highest cannot be y_2 and so can be taken as z_2 without loss of generality. The only issue is how the two outcomes left rank relative to one another.

[13] To see this, observe the following. Assumption 3 implies that the total number of iterations in any period n history is bounded above by $n \cdot (N+n)^{\alpha} < (N+n)^{\alpha+1}$ where N is the initial

choice in a positive fraction of games, in the long run. This creates an opportunity for sophisticated strategies for P2 that observe the choices made by the P1's and thereby build up a picture of P1's preferences.

2.2.2. **Sophisticated Players.**

There are two aspects to the *ToP* strategies. The first of these, given as part i) of Definition 3 below, concerns the utilization of the knowledge of others' preferences. Picking the *SPE* choice at each node when the preferences of subsequent players are known characterizes the *SPE-ToP* strategy that will eventually dominate the population in every role. The second aspect, given as ii) of Definition 3, concerns how such knowledge of the preferences of others could be acquired from observing their behavior.

DEFINITION 3: *All ToP strategies always satisfy Assumption 4 in all subgames. It is convenient to describe the remaining requirements on the ToP strategies in the reverse order to the temporal order in which they apply. i) If a ToP player in role i knows the ordinal preferences of all subsequent players over the set Z_n, each such ToP player maps the array of own preferences plus those of subsequent players to a pure action at each decision node (still subject to Assumption 4). A particular ToP strategy, the SPE-ToP strategy, maps all of these preferences to the SPE choice at each node, if this SPE choice is unique. Other ToP strategies make a non-SPE choice in at least one subgame defined by the ordinal preferences of others and of the player in question.*[14] *ii) It is common knowledge among all ToP players that there exists a positive fraction of SPE-ToP players in every role.*

What is meant in Definition 3 i) by hypothesizing that the *ToP* strategies "know" the preferences of subsequent players? That is, what patterns of play reveal these preferences under Definition 3 ii)? We use Example 1 to illustrate

---

number of outcomes. Since only one game is played at each iteration, this provides also an upper bound on the number of distinct games occurring along any such history. Further, in period n, there are $|Z_n|^4 = (N+n)^4$ possible games. It follows that if $\alpha + 1 < 4$, then the fraction of games that are familiar tends to zero, surely. This result is stated as Lemma 7 in the Appendix.

[14] This requirement is merely to avoid triviality. It has the following implication. Since the preferences involved are ordinal, the probability of such a subgame is positive under F. Indeed, the probability of a game that repeats this subgame for every decision node of the role in question is also positive. Such games will then give the *SPE-ToP* strategy a strict advantage over any other *ToP* strategy.

these issues and then indicate how the argument can be generalized by considering the case where $I = 3$.

Example 1 Revisited: *In this example with $I = A = 2$, all of the ToP P2's learn one of P1's binary preferences, whenever the P1 are forced to make a choice between two outcomes that has not arisen before. This follows since Assumption 4 implies that the P1's always make the SPE choice. Indeed, whenever $\alpha > 1$, so that the rate of introduction of novelty is not too fast, such learning by the ToP P2's will be essentially complete in the limit.*[15] *If $\alpha \in (1, 3)$, the ToP strategies then have a clear knowledge edge over the naive strategies.*

Each ToP strategy maps the preferences of P1, once these are known, as well as own preferences, to an action. The SPE-ToP strategy fully exploits the knowledge edge of the ToP strategies over the naive strategies, by mapping these two preference profiles to the SPE choice. It is obvious in Example 1 that this SPE-ToP strategy will then eventually outdo all other strategies, naive or sophisticated, since only the SPE-ToP comes to correctly anticipate all of the choices of the P1's.

It will not matter that in the interim—when the sophisticated strategies do not know P1 preferences—that they make inappropriate choices, as these instances occur with vanishing probability. Similarly, neither will it matter how sophisticated the naive strategies are on familiar games, since these also arise with vanishing probability.

We now further illustrate this mechanism, by extending the argument in Example 1 to the case $I = 3$. Learning about $i = 1$'s preferences remains straightforward, whenever $\alpha > 1$, and proceeds as before. Indeed, role 1's preferences then become common knowledge among all *ToP* players in role 2 and 3. The

[15] This is a key theoretical result of the paper, given in the Appendix as Lemma 2. Although the proof there is complicated by the need to allow more than two stages, it is nontrivial even for $I = 2$. It is not hard to see, however, that $\alpha > 1$ means that complete learning is not ruled out, as follows. The introduction of the n-th novel outcome results in $N + n - 1$ new pairwise choices. There are $\kappa(n) > (N + n - 1)^{\alpha}$ iterations of the game before the introduction of further novelty. Therefore, whenever $\alpha > 1$ the number of iterations between outcomes outstrips the number of new pairwise choices introduced. As n grows to infinity, this shows it is at least *possible* that the P2 *ToP*s will see nearly all of the P1 choices before the next outcome arrives. The more difficult task is to show that this possibility is realized, for the *ToP* strategies given in Definition 3.

new interesting case then concerns how the *ToP* players in role 3 can learn role 2's preferences. Suppose then a game is drawn in which some subgame has a dominant choice a, say, for role 2, as Assumption 4, and this subgame is reached.[16] It follows from Assumption 4 that all players in role 2 take this dominant action. The *ToP* players in role 3 do not know that such a dominant choice exists for 2. They do know, however, that the 2's also know 1's preferences. Hence, whether such a dominant choice exists or not, the *SPE-ToP*'s in role 2 have unequivocally demonstrated to the *ToP* in role 3 that they prefer the outcome induced by a to any outcome they might have induced instead. Still under the assumption that $\alpha > 1$, the *ToP* players in role 3 can then build up a complete picture of the preferences of role 2.[17]

The common knowledge assumption for the *ToP* players, as described in Definition 3 ii), can be stripped to its bare revealed preference essentials. It is unimportant, that is, what or whether the *ToP* players think, in any literal sense. All that matters, in the case that $I = 3$, for example, is that it is *as if* the *ToP*s in roles 3 add to their knowledge of role 2's preferences as described above. Once a *ToP* player in role 3 has seen histories in which all of role 2's binary choices have been put to the test like this, given that this is already true for role 1, the role 3 *ToP* players effectively know all that is relevant about the ordinal preferences of subsequent players and can act on this basis. This is essentially purely a mechanical property of the map, c, used by the *ToP* players. That is, not merely can the naive players be "zombies", in the philosophical sense, but so too can the *ToP* players.[18]

2.3. **Evolutionary Adaptation.**

The population structure and associated payoffs are as follows—

Definition 4: *The total population of all strategies is normalized to 1 for every role i. The sophisticated (ToP) strategies are labelled as $r = 1, \ldots, R$, for $R \geqslant 1$, say where $r = 1$ is the SPE-ToP strategy. The naive strategies are labelled as $r = R + 1, \ldots, \bar{R}$, where $\bar{R} > R$. The fraction of the total population in role*

[16] That this subgame is reached could be forced by assuming that this subgame arises for all of 3's choices.

[17] Lemma 2 in the Appendix proves the key result that $\alpha > 1$ ensures complete learning.

[18] That is, the revealed preference approach adopted here is agnostic about internal mental processes. For a philosophical treatment of "zombies", see Kirk (2014).

$i = 1, \ldots, I$ *that uses strategy* $r = 1, \ldots, \bar{R}$ *in period* $n = 1, 2, \ldots$ *and iteration* $t = 1, \ldots, \kappa(n)$ *is then denoted* $f^i_{n,t}(r)$, *where* $f^i_{n,t} = (f^i_{n,t}(1), \ldots, f^i_{n,t}(\bar{R}))$. *The average material payoff obtained by such a strategy* r *in role* i *in period* n *and iteration* t *is then denoted* $\bar{z}^i_{n,t}(r)$, *where* $\bar{z}^i_{n,t} = (\bar{z}^i_{n,t}(1), \ldots, \bar{z}^i_{n,t}(\bar{R}))$.

The population evolves in a standard adaptive fashion between each iteration of the game. This has the property, in particular, that the fraction of individuals who use a strategy that is best increases, given only that there is some suboptimal strategy—

Assumption 5: *Consider role* $i = 1, \ldots, I$ *in period* $n = 1, 2, \ldots$ *and at iteration* $t = 1, \ldots, \kappa(n)$. *If the population structure is* $f^i_{n,t}$ *with average payoffs* $\bar{z}^i_{n,t}$, *the population structure in the next iteration is given by* $f^i_{n,t+1} = \Psi(f^i_{n,t}, \bar{z}^i_{n,t})$.[19] *This function* $\Psi : \Delta^{\bar{R}-1} \times [m, M]^{\bar{R}} \to \Delta^{\bar{R}-1}$, *where* $\Delta^{\bar{R}-1}$ *is the unit simplex in* $\mathbb{R}^{\bar{R}}$, *has the properties i)* Ψ *is continuous, ii)* $\Psi_r(f^i_{n,t}, \bar{z}^i_{n,t})/f^i_{n,t}(r) > \eta$ *for some* $\eta > 0$, *and for* $r = 1, \ldots, \bar{R}$, *iii) if* $\bar{z}^i_{n,t}(r^*) = \max_{r=1,\ldots,\bar{R}} \bar{z}^i_{n,t}(r) > \bar{z}^i_{n,t}(r')$, *for some* $r' \in \{1, \ldots, \bar{R}\}$, *then* $\Psi_{r^*}(f^i_{n,t}, \bar{z}^i_{n,t}) > f^i_{n,t}(r^*)$ *and iv) if* $\bar{z}^i_{n,t} = \bar{z}^i_{n,t}(r')$, *for all* $r, r' \in \{1, \ldots, \bar{R}\}$, *then* $\Psi(f^i_{n,t}, \bar{z}^i_{n,t}) = f^i_{n,t}$.[20]

Recall that Figure 1 gives a schematic representation of the model.

2.4. **The Main Result.**

The main result is that, in the limit, the *SPE-ToP* strategy fully learns the preferences of others, applies this knowledge to choose the optimal action, and dominates the population.

Theorem 1: *Suppose Assumptions 1-5 all hold. Suppose that there are a finite number of ToP strategies, including SPE-ToP in particular, as in Definition 3, and a finite number of naive strategies, as in Definition 2. If* $\alpha \in (1, A^2 - 1)$, *then the proportion of SPE-ToP players in role* i, $f^i_{n,t}(1)$, *tends to 1 in probability, as* $n \to \infty$, *for all* $t = 1, \ldots, \kappa(n)$, *and for all* $i = 2, \ldots, I$. *The observed pattern of play in each realized game converges to an SPE, in probability.*

The proof of this is relegated to the Appendix.

The following specific remarks apply—

i) The result for $i = 1$ holds trivially by Assumption 4.

---

[20] Ψ_r denotes the rth component of the vector Ψ, $r = 1, \ldots, \bar{R}$.

ii) The proof shows that all of the *ToP* strategies learn others' preferences essentially always if $\alpha > 1$, but all naive strategies see essentially only new subgames if $\alpha < A^2 - 1$, in the long run. If both inequalities hold, as above, there is an opportunity for the *ToP* strategies to outdo the naive strategies, one that the *SPE-ToP* fully exploits.

iii) We focus here on the case that $\alpha \in (1, A^2 - 1)$. These bounds are tight in the sense that, if $\alpha < 1$ then it is mechanically impossible for the *ToP* players to learn the preferences of opponents from their binary choices.[21] On the other hand, if $\alpha > A^2 - 1$, then naive players in role 2 see only familiar subgames in the limit.[22]

iv) If the role i is earlier in the game, so i is larger, it is harder for naive strategies to learn all the subgames they initiate. In the role i, that is, the cutoff value for a naive strategy is $\alpha = A^i - 1$, below which learning is impossible in the long run, and this increases with i. However, the task faced by the *ToP* strategies does not become more complex in the same way, in that the cutoff value of $\alpha = 1$ is unaffected by the role i involved.

v) If it were assumed that naive players need to have experienced the *entire* game, and not just a subgame they initiate, before they can learn it, the upper bound for α would be $A^I - 1$, uniformly in $i = 2, \ldots, I$.

vi) If $\alpha < 1$, so that all the *ToP* players are overwhelmed with novelty, as are the naive players, the outcome of the evolutionary contest hinges on the default behavior of the naive and *ToP* strategies when these face their respective novel circumstances. As long as the naive players are not given a more sophisticated default strategy than the *SPE-ToP*s players, the naive players will, at best, match the *SPE-ToP*s players.

vii) If $\alpha > A^2 - 1$, naive players in at least role 2 have seen most subgames previously, in the long run. The relative performance of the *SPE-ToP* and the naive players then depends on the detailed long run behavior of the naive players. If the naive players play a Bayesian rational strategy the second time they encounter a given subgame, they might tie the *SPE-ToP* players. It is, in

[21] The proof is analogous to that of Lemma 7 in the Appendix which establishes the corresponding property for naive players.

[22] The proof is analogous to that of Lemma 2 in the Appendix which establishes the corresponding property for the *ToP* players.

any case, not intuitively surprising that a clear advantage to *SPE-ToP*s relies upon there being at least a minimum rate of introduction of novelty.

The eventual predominance of the *SPE-ToP*s over all the naive strategies resolves the issue raised by Stahl (1993) in this context. Consider any particular naive strategy that maps own payoffs to an action, where this choice cannot, of course, condition on the future realization of the sequence of games. If there is a dominant strategy in any subgame, this naive strategy chooses that by assumption. Otherwise, although there may be a set of subgames, with positive probability under F conditional on the observed own payoffs, in which the naive strategy makes the *SPE* choice, there must also be a set of subgames, also with positive conditional probability under F, for which this is not true. Since any particular naive strategy must therefore, with probability one, choose suboptimally in a positive fraction of games, in the limit, it is outdone, with probability one, by the *SPE-ToP* that is not preprogrammmed but rather adapts to the outcomes and games that are drawn, and ultimately chooses optimally essentially always.[23]

That is—

Corollary 1: *Under the hypotheses of Theorem 1, any particular naive strategy will, with probability one, choose suboptimally in a positive fraction of new subgames in the limit.*

Further, *ToP* strategies could be extended to deal with occasional shifts in preferences over outcomes. Such a generalized model would be noisier than the current model, and therefore harder to analyze, but this potential flexibility of the *ToP* strategies would constitute a telling additional argument in their favor.

It follows, significantly, that the evolutionary dominance of the *SPE-ToP* is robust to the introduction of sufficiently small cost, completing the resolution of the issue raised by Stahl (1993). Suppose that all *ToP* strategies entail a per game cost of $\omega > 0$, to reflect the cognitive cost associated with deriving the preferences of others from observation. Then we have

[23] This argument has the following subtlety. Consider a particular *realized* sequence of games. With probability one, each observed own payoff is associated with a unique vector of payoffs for the other roles. It follows that, with probability one, there exists a naive strategy that maps own payoffs to an action that is the *SPE* choice in every such realized subgame. To choose this naive strategy in advance is to condition on the future, however, given that there are uncountably many possible naive strategies.

COROLLARY 2: *Theorem 1 remains valid when all ToP strategies entail a per game cost ω (where the naive players have zero cost), if ω is small enough.*

If $\alpha > A^2 - 1$, however, then naive players in at least role 2 are usually familiar with the subgame they initiate, in the long run. The presence of a fixed cost might then tip the balance in favor of the naive players. If $\alpha < 1$, so all players, naive or sophisticated, are overwhelmed with novelty, this might also be true, when the default play of the naive and sophisticated players is comparable.

The presence of such a per game cost, that is independent of the number of outcomes, is not unreasonable since the *ToP* strategies would require the maintenance of a brain capable of sophisticated analysis. However, the *memory* demands of the naive players here are likely to be greater than the memory demands of *ToP*. The naive players need to remember each game; the *ToP*s need only remember preferences over each pairwise choice for opponents, and if memory is costly then these costs would be lower for the *ToP*s whenever there are a large number of outcomes. In this sense, consideration of all costs might well reinforce the advantage of the *ToP* players over the naive players.

The attainment of an *SPE* in Theorem 1 relies on the assumption that there is a large population in each role, with random matching for each iteration of the game. Even though a non-*SPE* choice by all role i players might benefit all role i players since it could advantageously influence the choice of a role $j > i$, this benefit is analogous to a public good. The choice by just one role i player has no effect on j's information bearing on i's preferences. Thus, the optimal choice by any particular role i player is sequentially rational. (The large population in each role, together with random matching, also ensures choices are myopic, ignoring, that is, future iterations of the game.) This argument that an *SPE* is attained once the preferences of others are known is analogous to Hart (2002).[24]

We close this subsection with several additional general remarks.

1) The key issue here is how *ToP*s deal with *novelty*—the arrival of new outcomes—rather than with *complexity*—the unbounded growth of the outcome set. Indeed, the model could be recast to display the role of novelty as follows.

[24] Hart considers a finite population in each role, with mutation ensuring all subgames are reached. His result is that the *SPE* is attained for a large enough common population size and small enough mutation rate.

Suppose that a randomly chosen outcome is dropped whenever a new outcome is added, at each date n, so the size of the outcome set is fixed, despite such updating events. There will then be a critical value such that, if the number of games played between successive dates is less than this critical value, the naive players will be mechanically unable to keep up with the flow of new games. There will also be an analogous but lower critical value for the *ToP*s. If the fixed interval between updating events is chosen to lie between these two critical values, the naive players will usually be faced with novel subgames; the *ToP*s will face a stochastic but usually positive fraction of subgames in which the preferences of subsequent player roles are known. This provides a version of the current results, although one that is noisier and therefore more awkward than the current approach.[25]

2) The sophisticated players here do not use the transitivity of others' preferences. If they were to do so, this could only extend the range of α over which complete learning of opponents' preferences would arise, and therefore the range over which the sophisticated strategies would outcompete the naive strategies.[26]

3) Consideration of a long run equilibrium, as in Theorem 1, is simpler analytically than direct consideration of the speed of out-of-equilibrium learning of the various strategies. More importantly, it also permits the use of minimal restrictions on the naive and *ToP* strategies, as is desirable in this evolutionary context.

4) Our results show how an increase in the rate of introduction of novelty might precipitate a transition from a regime in which there is no advantage to strategic sophistication to one in which a clear advantage is evident. This is consistent with theory and evidence from other disciplines concerning the evolution of intelligence. For example, it is argued that the increase in human intelligence was in part due to the increasing novelty of the savannah environment into which we were thrust after we exited our previous arboreal niche. (For a discussion of

---

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

[26] Although they do not apply directly, the results of Kalai (2003) concerning PAC-learning and P-dimension, Theorem 2.1 and Theorem 3.1, in particular, suggest that the use of transitivity might lower the critical value of α as far as *0*.

the intense demands of a terrestrial hunter-gatherer lifestyle, see, for example, Robson and Kaplan, 2003.)

2.5. **Related Literature.**

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

A strand of literature is concerned to model individuals' beliefs in a more realistic fashion than does the general abstract approach. An early paper in this strand is Stahl (1993) who considers a hierarchy of more and more sophisticated strategies analogous to iterated rationalizability. A smart$_n$ player understands that no smart$_{n-1}$ player would use a strategy that is not $(n-1)$-level rationalizable. A key aim of Stahl is to examine the evolution of intelligence in this framework. He obtains negative results—the smart$_0$ players who are right in their choice of strategy cannot be driven out by smarter players in a wide variety of plausible circumstances. Our positive results, in Corollary 2, in particular, stand in sharp contrast to these previous results.

Mohlin (2012) provides a recent substantial generalization of the closely related level-k approach that allows for multiple games, learning, and partial observability of type. Nevertheless, it remains true that lower types coexist with higher types in the long-run. This is not to deny that the level-k approach might work well in fitting observations. For example, Crawford and Iriberri (2007) provide an explanation for anomalies in private-value auctions based on this approach.

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

3. Conclusions

This paper presents a model of the evolution of strategic sophistication. The model investigates the advantages to learning opponents' preferences in simple games of perfect information. An unusual feature is that the outcomes used in the game are 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 their own observed payoffs in novel situations, remain in the dark. We impose plausible conditions under which some sophisticated individuals, who choose the *SPE* action, dominate all other strategies—naive or sophisticated—in the long run. That is, we establish a clear sense in which it is best to be smart, in contrast to previous results.

Kimbrough, Robalino and Robson (2014) presents experiments that measure the ability of real-world individuals to learn the preferences of others in a strategic setting. The experiments implement a simplified version of the theoretical model, using a two-stage game where each decision node involves two choices. We find 1) evidence of highly significant learning of opponents' preferences over time, but not of complete games, and 2) significant correlations between behavior in these experiments and responses to two well-known survey instruments from psychology intended to tentatively diagnose autism, as an aspect of theory of mind.

Appendix

A. Proof of Theorem 1

The proof of the theorem is given in two parts. Recall that α determines the number of games played in each period as described in Assumption 3. The first part shows that if $\alpha > 1$, the *ToP* players learn their opponents' preferences in the limit.[27] The ultimate dominance of the *SPE-ToP*s is established in the second part.

Part 1. *ToP*s Learn Opponent Preferences. The result proved here is that the *ToP*s learn their opponents' preferences whenever $\alpha > 1$. This sets the stage for the ultimate dominance of the *SPE-ToP*s.[28] We first introduce some notation.

Definition 5: $\Delta_n^i(z, z')$ *is the set of i role subgames, available in period n, that satisfy the following. The subgame q is in $\Delta_n^i(z, z')$ if and only if, for two actions, say a, $a' \in A$, z is the unique SPE outcome of the subgame following i's choice of a, and z' is the unique SPE outcome of the subgame following i's choice of a', and moreover one of the actions a, a' is stricly dominant for the i players themselves.*

The subgames of Definition 5 play a special role in how *ToP*s learn preferences. Recall that player roles are enumerated in reverse order of play. Consider a situation in which player 1's reach a subgame $q \in \Delta_n^1(z, z')$. Suppose that z is strictly preferred by player 1's to z'. Assumption 4 implies all the player 1's reaching q will there choose the action resulting in z. Any *ToP* observing this choice will then know that player 1's prefer z to z'. To establish as common knowedge then player 2's preferences over, say, z, z' it suffices for player 2's to be observed making a choice in a subgame $q \in \Delta_n^2(z, z')$ when all the player 1 pairwise choices in q had already been established as common knowledge among the *ToP*s. With this in mind consider the following. Suppose $z, z' \in Z$ is such

---

[27] If $\alpha < 1$, it is easily shown that it is mechanically impossible for the *ToP* players to learn opponents' preferences.

[28] In particular, the *SPE-ToP*s will then eventually choose an *SPE* in each game. Although in general this *SPE* choice is sub-optimal initially, it is the appropriate strategy in the long run (as will be established later).

that the 1 players prefer z to z'. Say that the history $H_{n,t}$ *reveals* players in role 1 prefer z to z' if along $H_{n,t}$ a subgame $q \in \Delta_n^1(z, z')$ was reached and all the 1 role players there chose the action delivering z. Proceeding inductively, suppose $z, z' \in Z$ is such that player i's prefer z to z'. Say that $H_{n,t}$ *reveals* players in role $i > 1$ prefer z to z' if along $H_{n,t}$ a subgame $q \in \Delta_n^i(z, z')$ was reached, after all the pairwise preferences of the $i-1, \ldots, 1$ role players in q had been revealed, and there all of the i role players chose into the $i-1$ subgame delivering z in the *SPE*.

To keep an account of how much information has been revealed along a given history we define—

Definition 6: *For each date $n = 1, 2, \ldots,$ and each iteration $t = 1, \ldots, \kappa(n)$, the random variable $K_{n,t}^i$ is number of outcome pairs $(z, z') \in Z_n \times Z_n$ such that $H_{n,t}$ reveals i role preferences on $\{z, z'\}$. Write $L_{n,t}^i = K_{n,t}^i / |Z_n|^2$—the fraction of i role pairwise preferences that are revealed along $H_{n,t}$.*[29]

A key step in establishing the eventual dominance of the *SPE-ToPs* is to show that if $\alpha > 1$, then $L_{n,t}^i$ tends to one in probability, for each $i \geqslant 1$. Here, convergence in probability of $L_{n,t}^i$ to L, for instance, is taken to mean that for each η there is a $\bar{n}$ such that $P\left\{|L_{n,t}^i - L| < \eta\right\} > 1 - \eta$ for each $n \geqslant \bar{n}$, and $t = 1, \ldots, \kappa(n)$. The aim of the remainder of this section is to prove the following result.

Lemma 2: *Suppose $\alpha > 1$, then $L_{n,t}^i$ tends in probability to one, $i = 1, \ldots, I$.*[30]

The proof relies on two preliminary results (Lemmas 3 and 4 below). First two definitions—

Definition 7: *Consider $i > 1$. For each period $n = 1, 2, \ldots,$ and each iteration $t = 1, \ldots, \kappa(n)$, $I_{n,t}^i \in \{0, 1\}$ is such that $I_{n,t}^i = 0$ if and only if the game drawn at iteration t of date n is such that all the available pairwise choices of players $j = 1, \ldots, i-1$ have been revealed along $H_{n,t}$.*

[29] Assumption 2 implies that with probability one, throughout period n, the number of pairs $(z, z') \in Z_n \times Z_n$ such that $z_i \neq z_i'$ is $|Z_n| \cdot [|Z_n| - 1]/2$. Thus, with probability one this is the maximal number of preference revelations possible up to period n. We opt in favor of the simpler expression $|Z_n|^2$ in the denominator of $L_{n,t}^i$, and to allow this set $K_{n,1}^i = |Z_n|$ for each n and require that $K_{n,t}^i$ increase by 2 with each revelation of an i role preference ordering.

[30] If $\alpha < 1$, it follows that $L_{n,t}^i \to 0$ surely. This follows from an adaptation of the proof of Lemma 7.

Definition 8: *For each $\varepsilon > 0$,*

$$\boldsymbol{S}^i_n(\varepsilon) = \left\{(z,z') \in Z_n \times Z_n : |\Delta^i_n(z,z')|/|Z_n|^{A^i - 2} < \varepsilon\right\}.$$

When ε is small, the number of subgames in $\bigcup_{\boldsymbol{S}^i_n(\varepsilon)} \Delta^i_n(z,z')$ is a small fraction of the number of possible i role subgames.[31] The roles of $I^i_{n,t}$, and $\boldsymbol{S}^i_n(\varepsilon)$ are clarified in the following lemma.

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

1. *Consider $i > 1$. For each $\varepsilon > 0$, $n = 1, 2, \ldots,$ and $t = 1, \ldots, \kappa(n)$,*

$$\begin{aligned} &E(K^i_{n,t+1} \,|\, H_{n,t}) - K^i_{n,t} \geqslant \\ &- E(I^i_{n,t} \,|\, H_{n,t}) + \left[\varepsilon \cdot \max\left\{0, 1 - L^i_{n,t} - \frac{|\boldsymbol{S}^i_n(\varepsilon)|}{|Z_n|^2}\right\}\right]^{A^{I-i}}. \end{aligned}$$

   If $i = 1$, the above expression holds for i with $I^i_{n,t}$ identically equal to zero.

2. *There exists an $S(\varepsilon)$ such that $S(\varepsilon)$ tends to zero as ε tends to zero, and such that for each $\varepsilon > 0$, $|\boldsymbol{S}^i_n(\varepsilon)|/|Z_n|^2$ almost surely converges to $S(\varepsilon)$.*

Proof. Consider the first enumerated claim. Fix an $i \geqslant 1$, and let $U_{n,t}$ denote the outcome pairs $(z,z') \in Z_n \times Z_n$ such that i's preferences on $\{z,z'\}$ have not been revealed along $H_{n,t}$. Notice that $L^i_{n,t} = 1 - |U_{n,t}|/|Z_n|^2$. Define $J_{n,t} \in \{0,1\}$ such that $J_{n,t} = 1$ if and only if, at iteration t of date n, for each i player subgame q there is some $(z,z') \in U_{n,t}$ such that q is in $\Delta^i_n(z,z')$. For $i = 1$, since there are no players after i, set $I^i_{n,t} \equiv 0$ in all of the following expressions. Note that if $[1 - I^i_{n,t}] \cdot J_{n,t} = 1$, then i's choice at any i subgame reveals i preferences over some pair of outcomes $(z,z') \in U_{n,t}$. Therefore, $K^i_{n,t+1} - K^i_{n,t} \geqslant [1 - I^i_{n,t}] \cdot J_{n,t}$. Since $[1 - I^i_{n,t}] \cdot J_{n,t} \geqslant J_{n,t} - I^i_{n,t}$, it follows that for each $t = 1, \ldots, \kappa(n) - 1$,

$$E(K^i_{n,t+1} \,|\, H_{n,t}) - K^i_{n,t} \geqslant E(J_{n,t} \,|\, H_{n,t}) - E(I^i_{n,t} \,|\, H_{n,t}). \tag{1}$$

---

[31] Recall that the set of i role subgames in period n is $Q^i_n \equiv |Z_n|^{A^i}$, and thus

$$\frac{\left|\bigcup_{(z,z') \in \boldsymbol{S}^i_n(\varepsilon)} \Delta^i_n(z,z')\right|}{|Q^i_n|} < \varepsilon \cdot \frac{|\boldsymbol{S}^i_n(\varepsilon)|}{|Z_n|^2} \leqslant \varepsilon.$$

Next observe that $E(J_{n,t} \,|\, H_{n,t}) = P\{J_{n,t} = 1 \,|\, H_{n,t}\}$, and

$$P\{J_{n,t} = 1 \,|\, H_{n,t}\} \geqslant \left[\sum_{(z,z') \in U_{n,t}} \frac{|\Delta_n^i(z,z')|}{|Q_n^i|} \right]^{A^{I-i}}. \tag{2}$$

This is because the distribution over games at date n can induced by the A^{I-i}-times independent sampling of i player subgames, uniformly from Q_n^i, while the fraction of i role subgames in $\Delta_n^i(z,z')$ at date n is $|\Delta_n^i(z,z')|/|Q_n^i|$. Using $\boldsymbol{S}_n^i(\varepsilon)$ in equation (2) gives,

$$\begin{aligned} P\{J_{n,t} = 1 \,|\, H_{n,t}\} &\geqslant \left(\sum_{(z,z') \in U_{n,t} \setminus \boldsymbol{S}_n^i(\varepsilon)} \frac{|\Delta_n^i(z,z')}{|Q_n^i|} \right)^{A^{I-i}} \\ &\geqslant \left(\varepsilon \cdot \frac{|U_{n,t} \setminus \boldsymbol{S}_n^i(\varepsilon)|}{|Z_n|^2} \right)^{A^{I-i}} \\ &\geqslant \left[\varepsilon \cdot \max \left\{ 0, \frac{|U_{n,t}|}{|Z_n|^2} - \frac{|\boldsymbol{S}_n^i(\varepsilon)|}{|Z_n|^2} \right\} \right]^{A^{I-i}} \\ &= \left[\varepsilon \cdot \max \left\{ 0, 1 - L_{n,t}^i - \frac{|\boldsymbol{S}_n^i(\varepsilon)|}{|Z_n|^2} \right\} \right]^{A^{I-i}}. \end{aligned} \tag{3}$$

Equations (1) and (3) together deliver the desired result.

The second enumerated claim follows by direct application of Lemma 1 in the light of Assumption 2. ■

LEMMA 4: *Let x_s, $s = 1, 2, \ldots$, be a sequence taking values in $[0,1]$. Given $\bar{\varepsilon} > 0$, consider a family of sequences, $\{\theta_s(\varepsilon)\}$, $\varepsilon \in (0, \bar{\varepsilon}]$, that satisfy the following conditions. For each $\varepsilon \in (0, \bar{\varepsilon}]$, $\lim_{s \to \infty} \theta_s(\varepsilon) = \theta(\varepsilon) \in \mathbb{R}_+$, where $\theta(\varepsilon)$ tends to one as ε tends to zero. Suppose $\liminf[x_{s+1} - x_s] \geqslant 0$, and that $x_{s+1} - x_s < 0$ only if $x_s > \theta_s(\varepsilon)$. Then x_s converges to some limit $\hat{x} \in [0,1]$.*

Proof. Fix an arbitrary $\eta > 0$. Choose a $\varepsilon \in (0, \bar{\varepsilon}]$ such that $\theta(\varepsilon) > 1 - \eta/3$, and then choose T_1 so that if $s \geqslant T_1$, then $\theta_s(\varepsilon) > 1 - 2 \cdot \eta/3$. Choose T_2 such that $x_{s+1} - x_s > -\eta/3$ for every $s \geqslant T_2$. Define $T = \max\{T_1, T_2\}$, and $\sigma = \inf\{s \geqslant T : x_{s+1} - x_s < 0\}$. If $\sigma = \infty$, then clearly x_s converges. Suppose then that σ is finite. Observe the following. Whenever $s \geqslant T$, if $x_{s+1} - x_s < 0$, then $x_s > \theta_s(\varepsilon) > 1 - 2 \cdot \eta/3$. Thus, since $x_{s+1} - x_s > -\eta/3$, for all $s \geqslant T$, it follows that $x_{s+1} - x_s < 0$ implies $x_{s+1} > 1 - \eta$, whenever $s \geqslant T$. It follows then that $x_s > 1 - \eta$ for each $s > \sigma$. Clearly x_s converges since η can be chosen arbitrarily small. ■

Lemma 4 plays a key role in showing that the sequences $\{E(L^i_{n,t})\}$ are convergent (Lemma 6 below). Convergence of $E(L^i_{n,t})$ to $\bar{L}$, say, means that for each η there is a $\bar{n}$ such that $|E(L^i_{n,t}) - \bar{L}| < \eta$ for each $n \geqslant \bar{n}$, and $t = 1, \ldots, \kappa(n)$. First another preliminary result—

Lemma 5: *If the subsequence $\{E(L^i_{n,1})\}$ converges, then $\{E(L^i_{n,t})\}$ convererges and possesses the same limit. If the subsequence $\{L^i_{n,1}\}$ converges in probability to L^i, say, then so does $\{L^i_{n,t}\}$.*

Proof. The result pertaining to convergence in probability implies the result about the expectations converging. It thus suffices to prove the claim concerning convergence in probability. With that in mind assume the subsequence $\{L^i_{n1}\}$ converges in probability to L^i. Recall that $L^i_{n,t} = K^i_{n,t}/|Z_n|^2$. Then, notice that since $K^i_{n,t}$ is non-decreasing in n, and t, $L^i_{n,t} \geqslant L^i_{n1}$ for each $t = 1, \ldots, \kappa(n)$. The number of outcomes increases by one at the beginning of each period, thus $L^i_{n+1,1} \geqslant K^i_{n,t}/[|Z_n| + 1]^2 \geqslant L^i_{n,t} - 2/|Z_n|$, for each n, and each $t = 1, \ldots, \kappa(n)$. Hence, for all n, and $t = 1, \ldots, \kappa(n)$,

$$L^i_{n,1} - L^i \leqslant L^i_{n,t} - L^i \leqslant L^i_{n+1,1} - L^i + \frac{2}{|Z_n|}.$$

Given $\varepsilon > 0$ we may choose a $\bar{n}$ such that for all $n \geqslant \bar{n}$, $2/|Z_n| < \varepsilon/2$, and $P\left\{|L^i_{n,1} - L^i| < \varepsilon/2\right\} > 1 - \varepsilon$ (since $\{L^i_{n,1}\}$ converges in probability to L^i). Clearly now for all $n \geqslant \bar{n}$, and $t = 1, \ldots, \kappa(n)$, $P\{|L^i_{n,t} - L^i| < \varepsilon\} > 1 - \varepsilon$. $\varepsilon > 0$ is arbitrary and thus $L^i_{n,t}$ converges in probability to L^i. ■

Lemma 6: *Suppose $\alpha > 1$. Then $E(L^1_{n,t})$ converges to some $\bar{L}^1 \in [0, 1]$. If $L^j_{n,t}$ converges in probability to one for each $j = 1, \ldots, i - 1$, then $E(L^i_{n,t})$ converges to some $\bar{L}^i \in [0, 1]$.*

Proof. Fix $i \geqslant 1$. The notation $\hat{L}_n$, and $\hat{K}_n$, will from now on be used as shorthand for $L^i_{n,1}$, and $K^i_{n,1}$, respectively. In view of Lemma 5 it suffices to prove that the subsequence $\{E(\hat{L}_n)\}$ converges. With that aim in mind notice first that

$$\hat{L}_{n+1} - \hat{L}_n = \frac{1}{|Z_n|^2} \cdot \sum_{t=1}^{\kappa(n)-1} [K^i_{n,t+1} - K^i_{n,t}] + \frac{\hat{K}_{n+1}}{[|Z_n| + 1]^2} - \frac{K^i_{n\kappa(n)}}{|Z_n|^2} \tag{4}$$

$$\geqslant \frac{1}{|Z_n|^2} \cdot \sum_{t=1}^{\kappa(n)-1} [K^i_{n,t+1} - K^i_{n,t}] - \frac{2}{|Z_n|}.$$

Write $Y_{n,t}(\varepsilon) = 1 - L^i_{n,t} - |\boldsymbol{S}^i_n(\varepsilon)|/|Z_n|^2$, take the expectation in (4), and apply Lemma 3 to obtain

$$\begin{aligned} E(\hat{L}_{n+1}) - E(\hat{L}_n) \geqslant & - \frac{2}{|Z_n|} - \frac{1}{|Z_n|^2} \sum_{t=1}^{\kappa(n)-1} E\left(I^i_{n,t}\right) \\ & + \frac{1}{|Z_n|^2} \cdot \sum_{t=1}^{\kappa(n)-1} E\left(\left[\varepsilon \cdot \max\left\{0, Y_{n,t}(\varepsilon)\right\}\right]^{A^{I-i}}\right). \end{aligned} \tag{5}$$

Twice applying Jensen's inequality (i.e., first $E(X^N) \geqslant E(X)^N$, given a random variable X, and then $E(\max\{X_1, Y_2\}) \geqslant \max\{E(X_1), E(X_2)\}$ given X_1, X_2) yields

$$\begin{aligned} E(\hat{L}_{n+1}) - E(\hat{L}_n) \geqslant & - \frac{2}{|Z_n|} - \frac{1}{|Z_n|^2} \sum_{t=1}^{\kappa(n)-1} E\left(I^i_{n,t}\right) \\ & + \frac{1}{|Z_n|^2} \cdot \sum_{t=1}^{\kappa(n)-1} \left[\varepsilon \cdot \max\left\{0, E\left(Y_{n,t}(\varepsilon)\right)\right\}\right]^{A^{I-i}}. \end{aligned} \tag{6}$$

$L^i_{n,t}$ is everywhere non-decreasing in $t = 1, \ldots, \kappa(n)$, and thus $Y_{n\kappa(n)}(\varepsilon) \leqslant Y_{n,t}(\varepsilon)$, for each $t = 1, \ldots, \kappa(n)$. Using this in (6) delivers

$$\begin{aligned} E(\hat{L}_{n+1}) - E(\hat{L}_n) \geqslant & - \frac{2}{|Z_n|} - \frac{1}{|Z_n|^2} \sum_{t=1}^{\kappa(n)-1} E\left(I^i_{n,t}\right) \\ & + \frac{\kappa(n) - 1}{|Z_n|^2} \cdot \left[\varepsilon \cdot \max\left\{0, E\left(Y_{n\kappa(n)}(\varepsilon)\right)\right\}\right]^{A^{I-i}}. \end{aligned} \tag{7}$$

Thus $E(\hat{L}_{n+1}) - E(\hat{L}_n) < 0$ implies the expression on the right-hand-side of (7) is negative. After some rearranging it becomes clear that $E(\hat{L}_{n+1}) - E(\hat{L}_n) < 0$ only if

$$E\left(Y_{n\kappa(n)}(\varepsilon)\right) < \frac{1}{\varepsilon}\left[\frac{2\cdot|Z_n|}{\kappa(n)-1} + \frac{1}{\kappa(n)-1}\sum_{t=1}^{\kappa(n)-1} E\left(I_{n,t}^i\right)\right]^{1/A^{I-i}}. \tag{8}$$

By definition, $Y_{n,t}(\varepsilon) = 1 - L_{n,t}^i - |\boldsymbol{S}_n^i(\varepsilon)|/|Z_n|^2$. Since $\hat{L}_{n+1} \geqslant L_{n,t}^i - 2/|Z_n|$ for each $t = 1,\ldots,\kappa(n)$, it follows that

$$Y_{n,t}(\varepsilon) \geqslant 1 - \hat{L}_{n+1} + 2/|Z_n| - |\boldsymbol{S}_n^i(\varepsilon)|/|Z_n|^2,$$

for each $t = 1,\ldots,\kappa(n)$. Using this in (8) and then solving for $E(\hat{L}_{n+1})$ yields: $E(\hat{L}_{n+1}) - E(\hat{L}_n) < 0$ only if $E(\hat{L}_{n+1}) > \theta_n(\varepsilon)$, where

$$\begin{aligned}\theta_n(\varepsilon) \equiv & 1 - \frac{2}{|Z_n|} - E\left(\frac{|\boldsymbol{S}_n^i(\varepsilon)|}{|Z_n|^2}\right) \\ & - \frac{1}{\varepsilon}\left[\frac{2\cdot|Z_n|}{\kappa(n)-1} + \frac{1}{\kappa(n)-1}\sum_{t=1}^{\kappa(n)-1} E\left(I_{n,t}^i\right)\right]^{1/A^{I-i}}.\end{aligned} \tag{9}$$

$\theta_n(\varepsilon)$, is defined here with the sequence $\{\theta_s(\varepsilon)\}$ from Lemma 4 in mind. Consider then the following. If $\alpha > 1$, then $|Z_n|/[\kappa(n)-1]$ tends to zero as n tends to infinity. For $i > 1$, if $L_{n,t}^j$ converges to one in probability, $j = 1,\ldots,i-1$, then $I_{n,t}^i$ tends to zero in probability, and thus so does $\sum_{t=1}^{\kappa(n)-1} E(I_{n,t}^i)/[\kappa(n)-1]$ (recall that $I_{n,t}^i$ is identically zero for $i = 1$). Therefore, $\theta_n(\varepsilon)$ tends to $1 - S(\varepsilon)$, where $S(\varepsilon)$ is the almost sure limiting value of $\boldsymbol{S}_n^i(\varepsilon)/|Z_n|^2$ (from Lemma 3). $S(\varepsilon)$ tends to zero as ε tends to zero (Lemma 3 again). We have already argued that $E(\hat{L}_{n+1}) - E(\hat{L}_n) \geqslant -2/|Z_n|$, and thus $\liminf[E(\hat{L}_{n+1}) - E(\hat{L}_n)] = 0$. Lemma 4 now gives the desired result. ∎

We are now in a position to prove Lemma 2.

Proof of Lemma 2. Fix $\alpha > 1$. The proof is by induction. Consider first $i > 1$. The induction hypothesis is: If $L_{n,t}^j$ tends to one in probability, $j = 1,\ldots,i-1$, then $L_{n,t}^i$ converges to one in probability. As in the proof of Lemma 6 write $\hat{L}_n = L_{n1}^i$. It suffices, to prove the induction claim, that the subsequence $\hat{L}_n$ converges to one in probability (Lemma 5). Toward that end, first write $Y_{n,t}(\varepsilon) = 1 - L_{n,t}^i - |\boldsymbol{S}_n^i(\varepsilon)|/|Z_n|^2$ (for role i as in the proof of Lemma 6). Write (with equation (7) in mind) $\bar{Y}_n(\varepsilon) = \left[\varepsilon \cdot \max\left\{0, E(Y_{n\kappa(n)}(\varepsilon))\right\}\right]^{A^{I-i}}$, and

$$\bar{X}_n = \frac{2 \cdot |Z_n|}{\kappa(n) - 1} + \frac{1}{\kappa(n) - 1} \sum_{t=1}^{\kappa(n)-1} E(I^i_{n,t}).$$

Consider dates s, m, such that $s > m$. Summing the terms of equation (7) from m to s gives,

$$\begin{aligned} E(\hat{L}_s) - E(\hat{L}_m) &= \sum_{n=m}^{s-1} [E(\hat{L}_{n+1}) - E(\hat{L}_n)] \\ &\geqslant \sum_{n=m}^{s-1} \frac{\kappa(n) - 1}{|Z_n|^2} \left[\bar{Y}_n(\varepsilon) - \bar{X}_n\right]. \end{aligned} \tag{10}$$

Lemma 6 gives that the sequence $\{E(\hat{L}_n)\}$ converges, and therefore implies that $\lim_{m\to\infty} \sup_{s\geqslant m}[E(\hat{L}_s) - E(\hat{L}_m)] = 0$. When $\alpha > 1$ the series $\sum_{n=m}^{s}[\kappa(n) - 1]/|Z_n|^2$ diverges to infinity as s tends to infinity.[32] It follows then from (10) that $\liminf[\bar{Y}_n(\varepsilon) - \bar{X}_n] \leqslant 0$ for each $\varepsilon > 0$. Observe now that if $L^j_{n,t} \longrightarrow 1$ in probability, then $I^i_{n,t}$ tends to zero in probability, and thus $E(I^i_{n,t})$ tends to zero in probability. The Cesaro means, $\sum_{t=1}^{\kappa(n)-1} E(I^i_{n,t})/[\kappa(n) - 1]$ thus tend to zero also, and hence $\bar{X}_n$ tends to zero. It follows that $\liminf \bar{Y}_n(\varepsilon) = 0$, for all $\varepsilon > 0$. In view of the definition of $\bar{Y}_n(\varepsilon)$, this implies $\liminf Y_{n,t}(\varepsilon) = 0$ for all $\varepsilon > 0$, and thus $\liminf[1 - E(\hat{L}_n) - |\boldsymbol{S}^i_n(\varepsilon)|/|Z_n|^2] = [1 - \bar{L} - S(\varepsilon)] = 0$, where $\bar{L}$ is the limiting value of $E(\hat{L}_n)$. Since $S(\varepsilon)$ can be made arbitrarily small by choice of ε, it follows that $\bar{L} = 1$. Since $\hat{L}_n$ is surely bounded above by one, we have that $\hat{L}_n$ converges to one in probability. This completes the proof of the induction claim.

What is needed now to complete the proof is to show that $L^1_{n,t}$ tends to one in probability. This follows by applying the previous arguments in establishing the convergence of $L^i_{n,t}$ for $i > 1$. In particular, in the definition of $\bar{X}_n$ set $I^i_{n,t} = 0$, and then proceed as above. ■

Part 2. Eventual Dominance of *SPE-ToPs*. Fix an $i > 1$ throughout. From now on a single subscript will denote the total number of iterations. For example, rather than writing $H_{n,t}$ for the history at iteration t of date n, H_s will be used where $s = \sum_{m=1}^{n-1} \kappa(m) + t$. In this section it will be proved that if

---

[32] That is, that $\lim_{s\to\infty} \sum_{n=1}^{s} \frac{1}{n} = \infty$, and $\sum_{n=m+N}^{s+N} \frac{1}{n} = \sum_{n=m}^{s} \frac{1}{|Z_n|} < \sum_{n=m}^{s}[\kappa(n) - 1]/|Z_n|^2$.

the arrival rate of novelty α lies in the range $(1, A^2 - 1)$, and if the fraction of players in role $j = 2, \dots, i-1$ that are the *SPE-ToP* tends to one in probability, then the fraction of players in role i that is the *SPE-ToP* will tend to one in probability also. Consider first some required definitions and results.

DEFINITION 9: *The game q is new to i at H_s if no i subgame of q has occured along H_s. $N_s \in \{0, 1\}$ is such that $N_s = 1$ if and only if the game in iteration s is new to i.*

LEMMA 7: *Suppose $\alpha < A^i - 1$, for $i = 2, \dots, I$, then every subgame is new to i in the limit. In particular, $P\{N_s = 1 \mid H_s\}$ converges to one almost surely.*[33]

Proof. First observe the following. If s is the total number of iterations along $H_s = H_{n,t}$, then $s \leqslant \sum_{m \leqslant n} \kappa(m)$. Assumption 3, where N is the number of outcomes initially, then gives $s \leqslant n \cdot (N+n)^{\alpha}$, and hence $s \leqslant (N+n)^{\alpha+1} = |Z_n|^{\alpha+1}$. Since there are A^{I-i} i role subgames in the fixed game tree, the number of i role subgames that have been encountered along H_s is surely bounded above by $A^{I-i} \cdot |Z_n|^{\alpha+1}$, and therefore the fraction of i subgames encountered previously along H_s is no greater than $A^{I-i} \cdot |Z_n|^{\alpha+1} / |Z_n|^{A^i}$, which clearly converges to zero whenever $\alpha + 1 < A^i$. This establishes the result as the distribution over games at iteration t of date n can be induced by drawing the appropriate number of i subgames uniformly from the A^i-times product of Z_n. ■

DEFINITION 10: *The measure induced by F on the full set of games Q is μ, and the measure induced by F on the full set of i role subgames, Q^i, is μ^i.*

LEMMA 8: *For each strategy r of role i that is not the SPE-ToP strategy there exists a set of games $Q(r)$ with positive measure under μ such that if $q \in Q(r)$ and q is new to i at H_s, then for every subgame q' of q, the choice made by r in q' at H_s is not part of an SPE of q'.*

Proof. It suffices to show that for any alternative r to the *SPE-ToP* there exists a set of i subgames, $Q^i(r)$, with positive measure under μ^i, such that for all $q \in Q^i(r)$, if q is new, then r's choice in q is not part of an *SPE*.[34] If $r > 1$ is a *ToP* alternative to the *SPE-ToP* this follows by definition. Thus,

---

[33] If $\alpha > A^i - 1$, then the fraction of games that are new to i converges to 0 in probability. The proof is analogous to that of Lemma 2.

[34] Any game that has each i subgame in $Q^i(r)$ will belong to $Q(r)$, and thus $Q(r)$ has measure at least $[\mu^i(Q^i(r))]^{A^{I-i}} > 0$, since there are A^{I-i} subgames of player i in each game and since μ is derived from the A^I-fold independent sampling from F, while μ^i is derived from the A^i-fold independent sampling from F.

assume r is a naive strategy. Recall that each naive strategy maps own payoffs to a fixed choice whenever making a choice in a new subgame. Although this mapping might correspond to an *SPE* in some i subgames, the richness of the set of possible games ensures it does not correspond to an *SPE* choice on a set of i subgames with positive measure under μ^i. To see this fix an action $a \in A$. Suppose i's choice of a is part of an *SPE* in every subgame in a subset $\tilde{Q}$ of i role subgames. Then, for almost every $q \in \tilde{Q}$, lacking a dominant choice for i, the action a can be rendered suboptimal in some q' obtained from q through a re-assignment of the remaining player $j = i - 1, \ldots, 1$ payoffs. The result then follows since the set of subgames contained in $\tilde{Q}$ such that i has no dominant choice has positive measure under μ^i. ■

Definition 11: *For each i role strategy $r = 2, \ldots, \bar{R}$, $B_s(r) \in \{0, 1\}$ is such that $B_s(r) = 1$ if and only if the game drawn at iteration s is new to i and belongs to $Q(r)$.*

Definition 12: *Let $C_s \in \{0, 1\}$ be such that $C_s = 1$ if and only if at iteration s some alternative to the SPE-ToM in role i outdoes the SPE-ToM in any i role subgame reached by the i players.*

Definition 13: *Q_δ is the set of games such that the absolute difference between any payoffs of the game is at least δ. For each $\xi > 0$, and $\delta > 0$, $D_s(\xi, \delta) \in \{0, 1\}$ is such that $D_s(\xi, \delta) = 1$ if and only if at iteration s each of the following hold: 1) the game is in Q_δ, 2) the fraction of remaining players after i, $j = 1, \ldots, i-1$, that chooses an SPE in every subgame is at least $1 - \xi$, and 3) the SPE-ToMs in role i themseves make an SPE choice at each node they reach.*

Now the last of the preliminary results—

Lemma 9: *1) If $\alpha < A^2 - 1$, then $P\{B_s(r) = 1\}$ converges to $\mu(Q(r))$. In addition suppose $\alpha > 1$, and that the fraction of players in roles $j = 1, \ldots, i-1$ that is the SPE-ToM tends to one in probability. Then, 2) C_s tends to zero in probability, and 3) for each $\xi > 0$, $P\{D_s(\xi, \delta) = 1\}$ tends to $\mu(Q_\delta)$.*

Proof. 1) Let $I\{\cdot\}$ denote the indicator function. Recall that $B_s(r)$ is equal to one if and only if the game at iteration s is new to i and in $Q(r)$ (Definition 11), and that $N_s = 1$ if and only if the game at iteration s is new to i. Thus, where q_s denotes the game at iteration s, $B_s(r) = N_s \cdot I\{q_s \in Q(r)\}$, and therefore $N_s + I\{q_s \in Q(r)\} - 1 \leqslant B_s(r) \leqslant I\{q_s \in Q(r)\}$. Taking expectations gives $E(B_s(r)) \longrightarrow E(I\{q_s \in Q(r)\})$ (i.e., Lemma 7 implies $E(N_s) \longrightarrow 1$). Notice that $P\{S\} = E(I\{S\})$ given an event S. Thus, $P\{B_s(r) = 1\} \longrightarrow P\{q_s \in Q(r)\}$. Lemma 1 delivers the desired result since it implies $P\{q_s \in Q(r)\}$ converges to $\mu(Q(r))$.

2) If $\alpha > 1$, then Lemma 2 implies the *SPE-ToP*s in role i choose an *SPE* of the underlying game with probability tending to one. By hypothesis, the remaining players also choose an *SPE* with probability tending one. Thus, in the long run, the *SPE-ToP*s at role i choose optimally.

3) Let $\hat{D}_s \in \{0, 1\}$ equal one if and only if the *SPE-ToP*s in role i choose an *SPE* of the game drawn at iteration s, and at least $1 - \xi$ of the remaining players behave as in an *SPE* of the game. Then, $D_s(\xi, \delta) = I\{q_s \in Q_\delta\} \cdot \hat{D}_s$, and thus $I\{q_s \in Q_\delta\} + \hat{D}_s - 1 \leqslant D_s(\xi, \delta) \leqslant I\{q_s \in Q_\delta\}$. Taking expectations gives $E(D_s(\xi, \delta)) \longrightarrow E(I\{q_s \in Q_\delta\})$ (since by hypothesis $E(\hat{D}_s) \longrightarrow 1$). Thus, $P\{D_s(\xi, \delta) = 1\} \longrightarrow P\{q_s \in Q_\delta\}$. Lemma 1 gives $P\{q_s \in Q_\delta\} \longrightarrow \mu(Q_\delta)$, and hence $P\{D_s(\xi, \delta) = 1\} \longrightarrow \mu(Q_\delta)$ as claimed. ■

We are now in position to prove the key result of this section.

LEMMA 10: *Suppose $\alpha \in (1, A^2 - 1)$. If the fraction of players in role $j = 2, \ldots, i-1$ that is the SPE-ToM tends to one in probability, then the fraction of players in role i that is the SPE-ToM tends to one in probability.*

Proof. It will first be proved that $E(f_s^i(1))$ converges by showing that the sequence $\{E(f_s^i(1))\}$ satisifies the hypotheses imposed on $\{x_s\}$ from Lemma 4. With that in mind notice the following, which is implied by Assumption 5 (Parts i, and iii). For each $\varepsilon > 0$ and $\delta > 0$ there are positive numbers Δ and ξ such that the following is true for any strategies of i, r and r'. Suppose the fraction of i's using strategy r' exceeds ε at iteration s. Suppose at the same time that 1) strategy r' choses an *SPE* in every subgame reached by the i's, 2) the game is such that the minimal absolute payoff difference between any i payoffs is greater than δ, 3) the proportion of remaining players $j = i-1, \ldots, 1$ that choose an *SPE* in each subgame is at least $1 - \xi$, and 4) the strategy r makes a non-*SPE* choice in every i subgame. Then, 5) the fraction of i players that use strategy r'

increases by at least Δ. [35] Notice that the previous facts imply the following. For each $\varepsilon > 0$ there is a triple of positive numbers $(\delta(\varepsilon), \Delta(\varepsilon), \xi(\varepsilon))$ that give the above implications with $\delta = \delta(\varepsilon), \Delta = \Delta(\varepsilon)$, and $\xi = \xi(\varepsilon)$, where $\delta(\varepsilon)$ can be chosen so that $\lim_{\varepsilon \to 0} \delta(\varepsilon) = 0$. [36] In the remainder let $(\delta(\varepsilon), \Delta(\varepsilon), \xi(\varepsilon))$ be as just described.

Next, recall the definitions of $B_s(r)$, and $D_s(\xi, \delta)$ (Definitions 11, and 13, respectively). For each $\varepsilon > 0$, and strategy $r > 1$, if the fraction of role i players that use strategy r at iteration s is no less than ε, then the fraction of i players that use the *SPE-ToP* strategy increases by at least $\Delta(\varepsilon)$, whenever $D_s(\xi, \delta) \cdot B_s(r) = 1$, for any $\xi \leqslant \xi(\varepsilon)$, and $\delta \geqslant \delta(\varepsilon)$ (i.e., $B_s(r) \cdot D_s(\xi, \delta) = 1$ if and only if the 1)-4) above with $r' = 1$, the *SPE-ToP* strategy). Thus, where $I\{\}$ denotes the indicator function, if $B_s(r) \cdot D_s(\xi(\varepsilon), \delta(\varepsilon)) \cdot I\left\{f_s^i(r) \geqslant \varepsilon\right\} = 1$, then $f_{s+1}^i(1) - f_s^i(1) > \Delta(\varepsilon)$. Next, recall that $C_s = 1$ if and only if some alternative to the *SPE-ToP* at role i outdoes the *SPE-ToP* in some subgame at iteration s. Since $f_{s+1}^i(1) - f_s^i(1) \geqslant -1$, if follows that $f_{s+1}^i(1) - f_s^i(1) \geqslant -C_s$ (Assumption 5 (iii-iv) implies this since $C_s = 1$ if and only if some strategy obtains a higher payoff than does the *SPE-ToP* in some subgame at iteration s). Hence, for each $r = 2, \ldots, \bar{R}$ and $s = 1, 2, \ldots,$

$$
\begin{aligned}
& f_{s+1}^i(1) - f_s^i(1) \\
\geqslant & \Delta(\varepsilon) \cdot B_s(r) \cdot D_s(\xi(\varepsilon), \delta(\varepsilon)) \cdot I\left\{f_s^i(r) \geqslant \varepsilon\right\} - C_s \\
\geqslant & \Delta(\varepsilon) \cdot \left[B_s(r) \cdot I\left\{f_s^i(r) \geqslant \varepsilon\right\} + D_s(\xi(\varepsilon), \delta(\varepsilon)) - 1\right] - C_s \\
= & \Delta(\varepsilon) \cdot \left[I\left\{f_s^i(r) \geqslant \varepsilon\right\} \cdot [B_s(r) + \mu(Q(r)) - \mu(Q(r))] + D_s(\xi(\varepsilon), \delta(\varepsilon)) - 1\right] - C_s \\
\geqslant & \Delta(\varepsilon) \cdot \left[I\left\{f_s^i(r) \geqslant \varepsilon\right\} \cdot \mu(Q(r)) - \mid B_s(r) - \mu(Q(r)) \mid + D_s(\xi(\varepsilon), \delta(\varepsilon)) - 1\right] - C_s.
\end{aligned}
\tag{11}
$$

---

[35] That is, given a minimal payoff difference of δ, if ξ is sufficiently small, a large enough fraction of the remaining players choose the unique *SPE* so that the *SPE* choice is optimal for i. Since, by assumption, r deviates from the *SPE* in every reached subgame, the payoff to r is dominated stricty by the payoff to r'. Assumption 5 then implies an increase in the fraction of the i population that uses r'. Note, however, that even when the average payoff to r' exceeds the payoff to r, the rate at which role i's abandon r in favor of r' is limited by the fraction of i's that use r at iteration s, and hence the requirement that $f_s^i(r) > \varepsilon$.

[36] As asserted initially, Assumption 5 gives: For each $\varepsilon > 0$ and $\delta > 0$ there are positive numbers Δ and ξ such that, if 1)-4) above hold, then 5) holds. To obtain the desired $(\delta(\varepsilon), \Delta(\varepsilon), \xi(\varepsilon))$, choose the function $\delta(\varepsilon)$ first, so that $\lim_{\varepsilon \to 0} \epsilon = 0$. Then, choose $\Delta(\varepsilon)$ and $\xi(\varepsilon)$ as required to to make 1)-4) imply 5).

In the third line of (11) we use the fact that $B \cdot D \geqslant B + D - 1$, for any binary variables $B, D \in \{0, 1\}$. Taking expectations in equation (11) gives

$$\begin{aligned} E\left(f^i_{s+1}(1)\right) - E\left(f^i_s(1)\right) \geqslant \Delta(\varepsilon) \cdot \Big[& P\left\{f^i_s(r) \geqslant \varepsilon\right\} \cdot \mu(Q(r)) \\ & - |E\left(B_s(r) - \mu(Q(r))\right)| + E(D_s(\xi(\varepsilon), \delta(\varepsilon))) - 1 \Big] - E(C_s). \end{aligned} \tag{12}$$

Therefore, $E\left(f^i_{s+1}(1)\right) - E\left(f^i_s(1)\right) < 0$ implies, after rearranging in (12), that for each $r' = 2, \ldots, \bar{R}$,

$$\begin{aligned} P\left\{f^i_s(r') \geqslant \varepsilon\right\} < \max_{r \geqslant 2} \Bigg\{ \frac{1}{\mu(Q(r))} \cdot \Bigg[& 1 - E(D_s(\xi(\varepsilon), \delta(\varepsilon))) + \frac{E(C_s)}{\Delta(\varepsilon)} \\ & + |E\left[\mu(Q(r)) - B_s(r)\right]| \Bigg] \Bigg\}. \end{aligned} \tag{13}$$

Let $\phi_s(\varepsilon)$ denote the value of the maximum in (13). Since $E\left(f^i_{s+1}(1)\right) - E\left(f^i_s(1)\right) < 0$ implies

$$E(f^i_s(r)) < \varepsilon \cdot (1 - \phi_s(\varepsilon)) + \phi_s(\varepsilon), \;\; r = 2, \ldots, \bar{R}, \tag{14}$$

we have

$$\begin{aligned} & E\left(f^i_{s+1}(1)\right) - E\left(f^i_s(1)\right) < 0, \;\; \text{only if} \\ & E(f^i_s(1)) \geqslant 1 - [\bar{R} - 1] \cdot [\varepsilon \cdot (1 - \phi_s(\varepsilon)) + \phi_s(\varepsilon)]. \end{aligned} \tag{15}$$

With (15) in mind set $\theta_s(\varepsilon)$ from the statement of Lemma 4 to $1 - [\bar{R} - 1] \cdot [\varepsilon \cdot (1 - \phi_s(\varepsilon)) + \phi_s(\varepsilon)]$. Lemma 9 implies $\phi_s(\varepsilon) \longrightarrow [1 - \mu(Q_{\delta(\varepsilon)})]/\min_{r \geqslant 2}\{\mu(Q(r))\}$. Then, set $\theta(\varepsilon)$ from Lemma 4 to

$$1 - [\bar{R} - 1] \cdot \left[\varepsilon \cdot \left(1 - \frac{1 - \mu(Q_{\delta(\varepsilon)})}{\min_r\{\mu(Q(r))\}} \right) + \frac{1 - \mu(Q_{\delta(\varepsilon)})}{\min_{r \geqslant 2}\{\mu(Q(r))\}} \right],$$

so that $\theta_s(\varepsilon) \longrightarrow \theta(\varepsilon)$. In view of our choice of $\delta(\varepsilon)$, $\theta(\varepsilon)$ thus defined tends to zero as ε approaches zero (since $\lim_{\delta \to 0} \mu(Q_\delta) = 1$, by Assumption 2). Next, since $f^i_{s+1}(1) - f^i_s(1) \geqslant -C_s$, $\liminf[E(f^i_{s+1}(1)) - E(f^i_s(1))] = 0$, and thus Lemma 4 gives that $E(f^i_s(1))$ is a convergent sequence.

To see that $f^i_s(1)$ must converge in probability to one, note that Lemma 9 implies that the right-hand-side of (12) converges to $\Delta(\varepsilon) \cdot [P\left\{f^i_s(r) \geqslant \varepsilon\right\} \cdot \mu(Q(r)) + \mu(Q_{\delta(\varepsilon)}) - 1]$. Since $E(f^i_{s+1}(1)) - E(f^i_s(1)) \longrightarrow 0$, in view of equation (12), it follows that $\limsup[P\left\{f^i_s(r) \geqslant \varepsilon\right\} + \mu(Q_{\delta(\varepsilon)}) - 1] \leqslant 0$, for all $\varepsilon > 0$, and $r = 2, \ldots, \bar{R}$. This gives $\limsup P\left\{f^i_s(r) \geqslant \varepsilon\right\} \leqslant 0$, for all $\varepsilon > 0$, and $r = 2, \ldots, \bar{R}$, which establishes the result. ■

Theorem 1 now follows by induction.

References

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

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

[3] **Patrick Billingsley** (**1986**): *Probability and Measure*. 2nd edition, Chicago: John Wiley and Sons.

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

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

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

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

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

[9] **Sergiu Hart** (**2002**): Evolutionary Dynamics and Backward Induction. *Games and Economic Behavior*, 41, 227–264.

[10] **Gil Kalai** (**2003**): Learnability and Rationality of Choice. *Journal of Economic Theory*, 113, 104–117.

[11] **Erik Kimbrough, Nikolaus Robalino & Arthur Robson** (**2014**), The Evolution of Theory of Mind: Theory and Experiments.

[12] **Robert Kirk** (**2014**): Zombies. In *The Stanford Encyclopedia of Philosophy*, edited by E.N. Zalta, Stanford University, URL `<http://plato.stanford.edu/archives/fall2014/entries/zombies/>`.

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

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

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

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

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

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

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