The Evolution of Strategic Sophistication

By Nikolaus Robalino and Arthur Robson *

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.

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 sophisti-

* Robalino: Department of Economics, 92 Lomb Memorial Drive, Rochester Institute of Technology, Rochester, NY, (e-mail: ndrobalin@gmail.com); Robson: Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada, (e-mail: robson@sfu.ca). We thank Eddie Dekel, Dean Foster, Matt Jackson, Erik Kimbrough, Leanna Mitchell, Daniel Monte, Andrew Newman, Antonio Penta, Luis Rayo, Phil Reny, 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, NWU, Oxford, Stanford, UBC, UCR, UCSB, U of Rochester, U of Texas (Austin), 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, by the SAET in Tokyo, by the Econometric Society in Montreal, by the University of Cambridge-INET Institute, at the NSF/NBER CEME GE Conference at the University of Wisconsin, Madison, and at the workshop "The Biological Basis of Preferences and Strategic Behavior" at SFU. 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. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

cation, 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 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]

[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.

[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

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, and 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 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

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 behavior conditions simply on the circumstance that he is losing a fight. By accident perhaps, he once made the leopard warning cry 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 are more outcomes already present, there is more that needs to be learned concerning where a new outcome ranks.

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 *SR-ToP* strategy, say, that not only observes subsequent preferences but is sequentially rational, using a subgame perfect 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 *SR-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. 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.

Theorem 2 is the main result here—for an intermediate range of values for a parameter governing the rate of innovation, a unique subgame perfect equilibrium is attained, with the *SR-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 2, i) the naive players do not know the game they face while, at the same time, ii) the *SR-ToP* players do know all the relevant preferences and, furthermore, have adapted to play the subgame perfect equilibrium 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.

We first present a treatment of the simple case in which there are only two stages, two moves at each decision node, and two strategies—one naive and one sophisticated—the *SR-ToP*. The advantage of this is that the argument is simplest and most intuitively compelling in this case. This treatment is complete and self-

contained, but it is stripped-down to the bare bones, in the interests of clarity. We then turn to the general case, with any number of stages, any number of moves at each decision node and any number of naive and sophisticated strategies for each stage, one of which is the *SR-ToP*. The general argument requires more subtle assumptions, and is more complex. We also defer discussion of many important but tangential issues to the treatment of the general case.

I. The Two Stage, Two Action Case

A. The Environment

The extensive game form is a fixed tree with perfect information, two stages, and two actions at each decision node. There are then 4 terminal nodes.

There is one "player role" for each stage, $i = 1, 2$, in the game. The first player role to move is 1 and the last to move is 2. Each player role is represented by an equal-sized "large" population of agents. Independently in each iteration of the game, all players are randomly and uniformly matched with exactly one player for each of the two roles.

Each player's payoff is a scalar, lying in $[m, M]$ where $M > m$. 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 role.

Given a fixed tree structure with 4 terminal nodes, we identify each outcome with a payoff vector and each game with a particular set of such payoff vectors assigned to the terminal nodes.

ASSUMPTION 1: *The set of all two stage games is represented by $\mathcal{Q} = [m, M]^8$, for $M > m$. That is, each outcome is a payoff vector in $\mathcal{Z} = [m, M]^2$, with one component for each player role, and there are 4 such outcomes comprising each game.*

We assume the game is iterated as described in the following two assumptions. The first of these describes how the outcome set grows—

ASSUMPTION 2: *Let $n = 1, 2, \ldots,$ denote successive dates. Within each corresponding period, n, there is available a finite subset of outcomes $\mathcal{Z}_n \subset \mathcal{Z}$, as follows. There is an initial finite set of outcomes $\mathcal{Z}_0 \subset \mathcal{Z}$, of size N, say. At date $n \geq 1$, at the beginning of period n, a new outcome is added to the existing ones by drawing it independently from $\mathcal{Z}$ according to a cdf F, that has a continuous probability density that is strictly positive on $\mathcal{Z}$.*

Within each period, the set of available outcomes is then fixed, and once an outcome is introduced it is available thereafter. Also, within each period, the game is iterated an increasing number of times 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 $\alpha \geq 0$.*[5] *Each iteration involves an independent and*

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

uniform random choice of the 4 *outcomes, with replacement, from the set* $\mathcal{Z}_n$.

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.

We turn now to the specification of the strategies.

B. Strategies

In the two stage case, the strategies can be described much more simply than they can with an arbitrary number of stages. In the general case, we suppose that players in any role $i = 2, \ldots, I$ choose a dominant strategy if this is available. In the present context, this requirement binds only on players in role 2.

ASSUMPTION 4: *Consider the choice of a player in role* 2 *at a particular decision node* h. *If action* a *at* h *yields* 2 *a higher payoff than does the other action, then* 2 *takes action* a.

This requirement is in the spirit of focussing on the implications for one's behavior of knowing the preferences of *others* rather than one's own. Here it is equivalent to sequential rationality for player role 2.[6] There is essentially then no latitude left in player role 2's strategy.

Consider then strategies for player role 1. These players, when making a choice in period n and iteration t, know the history so far and the game, $\mathbf{q}_{n,t}$, 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 by players in role 2 in all previous periods and iterations.[7]

Strategies for role 1 then differ only as to how they condition on such histories.

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.

DEFINITION 1: *There is one naive strategy in role* 1. *This maps own observed payoffs to an arbitrary pure choice, whenever the game faced has never arisen previously.*

If any game faced is *not* new, there is no constraint imposed on a naive strategy.

The following example simplifies the two stage, two action case still further by describing a particular salient naive strategy. This illuminates the weaknesses of

[6]Sequential rationality would actually follow from the large number of players in each role. This is discussed and defended carefully in the next section which contains the general treatment.

[7]The general more formal treatment in the next section applies, in particular, to the present two stage case.

any naive strategy, describing the opportunity that exists for more sophisticated strategies—

Example 1.— Consider Figure 1. In view of Assumption 4, the 2s always make the equilibrium choice. The problem for the 1s 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 2s.

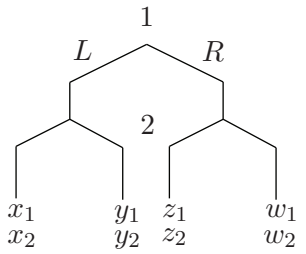

FIGURE 1. EXAMPLE 1: TWO STAGES, TWO ACTIONS.

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

A salient naive strategy for 1 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 $x_1 + y_1 > z_1 + w_1$. If either choice is dominant, this simple rule makes that dominant choice. Moreover, given risk neutrality in the payoffs, and given that F represents independent choices in the payoffs, this naive strategy is the Bayesian rational procedure initially when there is no additional information about role 2's preferences, since each of 2's choices are then equally likely given either choice for role 1.

Whenever there is not a dominant choice for 1, however, it is easy to see that *any* naive strategy must make the wrong choice with strictly positive probability, under any F with full support. This creates an opportunity for a sophisticated strategy that has the potential to outdo the naive strategy in these cases. These strategies are described next.

SOPHISTICATED PLAYERS

Recall that *ToP* (for "theory of preferences") refers to the ability to impute preferences to others. In the two stage case, *ToP* strategies for role 1 condition on knowledge of role 2's preferences. Using the sequentially rational strategy at the initial node when the preferences of the role 2s are known characterizes the *SR-ToP* (sequentially rational *ToP*) strategy that will eventually dominate the

population in role 1. For simplicity, we consider just this *SR-ToP* strategy in this section.

DEFINITION 2: *Whenever a SR-ToP player in role* 1 *knows the ordinal preferences of player role* 2 *over the set* $\mathcal{Z}_n$*, the SR-ToP strategy maps these ordinal preferences and her own to the subgame perfect equilibrium choice, if this equilibrium choice is unique.*[8]

What is meant in Definition 2 by hypothesizing that the *SR-ToP* strategies "know" the preferences of subsequent players? We use Example 1 to clarify.

Example 1 Revisited.— In this example with two stages and two choices, the *SR-ToP* role 1s learn one of role 2's binary preferences, whenever the 2s are forced to make a choice between two outcomes that has not arisen before. This follows since Assumption 4 implies that the 2s always make the sequentially rational choice. Indeed, whenever $\alpha > 1$, so that the rate of introduction of novelty is not too fast, such learning by the *SR-ToP* 1s will be shown to be essentially complete in the limit. If $\alpha < 3$, on the other hand, it is easily shown that the naive strategy sees only novel games in the limit. If $\alpha \in (1, 3)$, then, the *SR-ToP* strategy has a clear knowledge edge over the naive one.

C. Evolutionary Adaptation

The population structure and associated payoffs for the players in role 1 are as follows—

DEFINITION 3: *The total population of both strategies in role* 1 *is normalized to 1. The fraction of the population in role* 1 *that uses the SR-ToP strategy in period* $n = 1, 2, \ldots$*, and iteration* $t = 1, \ldots, \kappa(n)$ *is then denoted* $f_{n,t}$*. The average payoff obtained by the SR-ToP strategy in role* 1 *in period* n *and iteration* t *is then denoted* $\bar{z}_{n,t}(1)$*, and the average payoff obtained by the naive strategy is then* $\bar{z}_{n,t}(2)$*. We set* $\bar{\mathbf{z}}_{n,t} = (\bar{z}_{n,t}(1), \bar{z}_{n,t}(2))$*.*

The population evolves in a standard adaptive fashion between each iteration of the game. Apart from minor technicalities, the key assumption is that the fraction of individuals who play either strategy increases if it does better than the other strategy—

ASSUMPTION 5: *Consider role* 1 *in period* $n = 1, 2, \ldots$ *and at iteration* $t = 1, \ldots, \kappa(n)$*. If the fraction of SR-ToP strategies is* $f_{n,t}$ *and average payoffs are* $\bar{\mathbf{z}}_{n,t}$*, the fraction of SR-ToP strategies in the next iteration is given by* $f_{n,t+1} = \Psi(f_{n,t}, \bar{\mathbf{z}}_{n,t})$*.*[9] *This function* $\Psi : [0, 1] \times [m, M]^2 \to [0, 1]$ *has the properties i)* Ψ *is continuous, ii)* $\Psi(f_{n,t}, \bar{\mathbf{z}}_{n,t})/f_{n,t} > \eta$ *for some* $\eta > 0$*, and iii)*

[8] In the limit, the probability of ties is zero.

[9] If $t = \kappa(n)$, then $\Psi(f_{n,t}, \bar{\mathbf{z}}_{n,t}) = f_{n+1,1}$.

$$\Psi(f_{n,t}, \bar{\mathbf{z}}_{n,t}) \begin{cases} > f_{n,t} \text{ if } \bar{z}_{n,t}(1) > \bar{z}_{n,t}(2) \\ = f_{n,t} \text{ if } \bar{z}_{n,t}(1) = \bar{z}_{n,t}(2) \\ < f_{n,t} \text{ if } \bar{z}_{n,t}(1) < \bar{z}_{n,t}(2). \end{cases}$$

D. The Two Stage, Two Action Result

The main result for the two stage case is that, in the limit, the *SR-ToP* strategy in role 1 fully learns the preferences of role 2, applies this knowledge to choose the optimal action, and dominates the population.

THEOREM 1: *Suppose Assumptions 1-5 all hold. Suppose that there are two strategies for role* 1*—the SR-ToP, as in Definition 2, and a naive strategy, as in Definition 1. If* $\alpha \in (1, 3)$*, then the proportion of SR-ToP players in role* 1*,* $f_{n,t}$*, tends to 1 in probability, as* $n \to \infty$*, for all* $t = 1, \ldots, \kappa(n)$*. The observed pattern of play in each realized game converges to a subgame perfect equilibrium, in probability.*

PROOF:

It is straightforward to show that, if $\alpha < 3$, then the fraction of games that have arisen previously tends to 0, as $n \to \infty$, for all $t = 1, \ldots, \kappa(n)$. 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 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 $|\mathcal{Z}_n|^4 = (N+n)^4$ possible games. If $\alpha + 1 < 4$, then the fraction of games that are familiar must tend to zero, surely. That is, if $\alpha < 3$, then it is mechanically impossible for any naive player in role 1 to keep up with the rate of arrival of new games. The Appendix shows that this result, when combined with the observation that it is impossible for the naive strategy to play optimally in *all* new games, means that the naive strategy leaves an opportunity for a more sophisticated strategy.

A key element to complete the proof of Theorem 1 is therefore to show that, if $\alpha > 1$, then the *SR-ToP* role 1s learn the preferences of role 2 completely, in the limit. This is also relegated to the Appendix. A rough intuition is provided in the next paragraph.

Assumption 3 implies that the total number of iterations in any period n history is of order $n^{\alpha+1}$. Since one game is played at each iteration, with two decision nodes for role 2, the number of outcome pairs over which 2's preferences were exposed in period n, whether for the first time or not, is also of order $n^{\alpha+1}$. In period n, there are $|\mathcal{Z}_n|(|\mathcal{Z}_n| - 1)/2$ possible distinct outcome pairs, which is of order n^2. Given that $\alpha > 1$, if the fraction of role 2 preferences over pairs that role 1 players knew was, hypothetically, close to zero, these order of magnitude considerations imply that this fraction would grow rapidly. What complicates

the general argument is considering what happens when this fraction is strictly between 0 and 1, so that some outcome pairs are known. When $\alpha > 1$, it is nevertheless true that the stochastic process governing this fraction drifts up, whenever the fraction starts below 1. In the limit, then, this stochastic process converges to 1, in probability. This result is intuitively appealing, but a rigorous proof is technically involved, even in this simple two stage, two action case.

The key to the evolutionary success of the *SR-ToP* strategy, for an intermediate range of arrival rates of novelty, as in Theorem 1, is that the sophisticated strategy is able to keep up with such rates, whereas the naive strategy cannot. The reason for the edge that the sophisticated strategy holds is simple. Sophisticated players need to learn only *pairs* of outcomes, the number of which is of order $(N+n)^2$; whereas the naive players need to learn *games*, the number of which is of much greater order—$(N+n)^4$. Although the present model is rather specific, the simplicity of this argument implies that similar results would hold in variety of alternative models.

II. The General Case

A. The Environment

We now present suitably generalized versions of the assumptions made in the two stage, two action case. In some instances, the generalization is direct, but the sake of clarity, the set of general assumptions is presented in its entirety.

Reconsider first the underlying games. As in the two stage case the extensive game form is a fixed tree with perfect information and a finite number of stages. In the general case there are now $I \geq 2$ stages, and a fixed finite number of actions, $A \geq 2$, at each decision node.[10] There are then $A^I = T$, say, terminal nodes.

There is one "player role" for each such stage, $i = 1, \ldots, I$, in the game. Again, as in the two stage case, there is an equal-sized "large" population of agents representing each role. The agents have various strategies, which are described precisely below. These are grouped into two categories—sophisticated (*ToP*) and naive.

In each iteration of the game, players are 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 payoffs of the I player roles. Player role $i = 1, \ldots, I$ has then a function mapping all outcomes to payoffs. Again, 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

[10]The restriction that each decision 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.

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.[11]

We assume that all payoffs are scalars, lying in the compact interval $[m, M]$, for $M > m$, say. It follows that games given in terms of payoffs can be described as follows—

ASSUMPTION 6: *The set of all games is represented by $\mathcal{Q} = [m, M]^{TI}$, for $M > m$. That is, each outcome is a payoff vector in $\mathcal{Z} = [m, M]^I$, with one component for each player role, and there are $T = A^I$ such outcomes comprising each game.*

The outcome set grows as in the two stage case. The following assumption, describing the growth of the outcome set, is identical to the one made for the two stage case, except for the definitions of $\mathcal{Z}$ and $\mathcal{Z}_n$, but is reproduced here for convenience.

ASSUMPTION 7: *Let $n = 1, 2, \ldots,$ denote successive dates. Within each corresponding period, n, there is available a finite subset of outcomes $\mathcal{Z}_n \subset \mathcal{Z}$, as follows. There is an initial finite set of outcomes $\mathcal{Z}_0 \subset \mathcal{Z}$, of size N, say. At date $n \geq 1$, at the beginning of period n, a new outcome is added to the existing ones by drawing it independently from $\mathcal{Z}$ according to a cdf F, that has a continuous probability density that is strictly positive on $\mathcal{Z}$.*

Again the set of available outcomes is fixed within each period, with the number of iterations of the game within each period again given by—

ASSUMPTION 8: *The number of iterations of the game played in period n is $\kappa(n) = \lfloor (N + n)^\alpha \rfloor$, for some $\alpha \geq 0$.*[12] *Each iteration involves an independent and uniform random choice of the $T = A^I$ outcomes, with replacement, from the set $\mathcal{Z}_n$.*

Recall that the rate of arrival of novelty is inversely related to α. If the parameter α is low, there are fewer iterations within each period before the next novel outcome arrives, and thus the rate of arrival of novelty is high; if α is high, the rate of arrival of novelty is low.

This completes the basic description of the underlying game, rendered schematically in Figure 2.

[11] 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, with a continuous cdf F, as in Assumption 7 below, 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.

[12] Here again $\lfloor \cdot \rfloor$ 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.

A convenient formal description of the set of games available in each period is as follows—

DEFINITION 4: *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 $F_n(\mathbf{z})$ where $\mathbf{z} \in [m, M]^I$. The set of games in period n is the T-times product of $\mathcal{Z}_n$. This is denoted $\mathcal{Q}_n$. The empirical cdf of games in period n derives from T-fold independent sampling of outcomes according to F_n and is denoted by $G_n(\mathbf{q})$, where $\mathbf{q} \in \mathcal{Q} = [m, M]^{IT}$.*[13]

In each iteration, $t = 1, \ldots, \kappa(n)$, of the game in period n, outcomes are drawn *independently* from $\mathcal{Z}_n$ 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—

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

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

B. 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, $\mathbf{q}_{n,t}$, drawn for the current iteration. Recall that history records the outcomes available in the current period, n, the randomly drawn games and the empirical distributions of choices made in all previous 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 $\pi_{n,t}(h) \in \Delta(A)$

[13] 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 $\mathcal{Z}_n$.

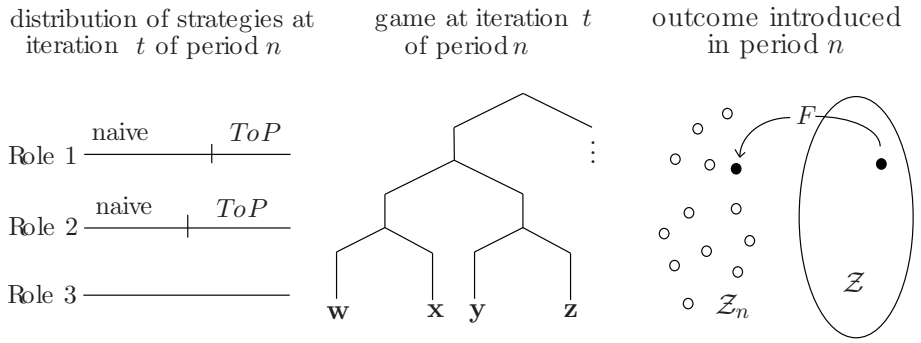

FIGURE 2. A SCHEMATIC REPRESENTATION OF THE KEY ELEMENTS OF THE MODEL.

then record the aggregate behavior of i player role at h. It follows that $H_{n,t} = \{\mathcal{Z}_n, (\mathbf{q}_{1,1}, \pi_{1,1}), \ldots, (\mathbf{q}_{n,t-1}, \pi_{n,t-1})\}$.[14] Let $\mathcal{H}_{n,t}$ be the set of period n and iteration t histories, and let $\mathcal{H} = \bigcup_{n,t} \mathcal{H}_{n,t}$.

Strategies can be formally described as follows. Let Σ_i denote the one-shot strategies available to players in role i. That is, each element of Σ_i specifies one of the A choices for each of the decision nodes of players in role i. A strategy is then a function $\sigma : \mathcal{H} \times \mathcal{Q} \longrightarrow \Sigma_i$ (recall that $\mathcal{Q}$ is the set of all possible games, $[m, M]^{IT}$).[15] An individual in period n at iteration t with strategy σ uses the one-shot strategy $\sigma(H_{n,t}, \mathbf{q}_{n,t})$ in game $\mathbf{q}_{n,t}$, $\sigma(H_{n,t+1}, \mathbf{q}_{n,t+1})$ in $\mathbf{q}_{n,t+1}$, and so on.

As part of the specification of the map σ, 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. As in the two stage case, 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.[16] 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

[14]If $n > 1$ but $t = 1$, then $H_{n,t} = \{\mathcal{Z}_n, (\mathbf{q}_{1,1}, \pi_{1,1}), \ldots, (\mathbf{q}_{n-1,\kappa(n-1)}, \pi_{n-1,\kappa(n-1)})\}$. If $n = t = 1$, then $H_{n,t} = \emptyset$.

[15]It will not be required that *ToP* players remember the entire history. All that is needed is that they make and retain certain 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.

[16]This could be proved using an argument similar to that used to establish that a subgame perfect equilibrium is ultimately obtained. See Section II.D.

only a single player in each role, for example, the player in role $i > 1$ 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 role i's choices that is observed by any role $j < i$ and thus no effect on the future behavior of the js. 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 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.

Accordingly, we have—

ASSUMPTION 9: *Consider any $i = 1, \ldots, I$ player role, and any i player role subgame $\mathbf{q}$. The action a at $\mathbf{q}$ is dominant for i if for every action $a' \neq a$, for every outcome $\mathbf{z}$ available in the continuation game after i's choice of a in $\mathbf{q}$, and every outcome $\mathbf{z}'$ available in the continuation game after i's choice of a' in $\mathbf{q}$, $z_i > z'_i$. For each $i = 2, \ldots, I$, every strategy always chooses any such dominant action.*[17]

Is this assumption reasonable, however, in the light of the putative small size of hunter-gatherer groups? If the current model were modified so there were a small number of players in each role, an optimal strategy would allow for dissembling, but would be complicated. It would require a prior distribution over the unknown payoffs of others, with Bayesian updating of this distribution in the light of observed play, which would be an onerous task. In this connection, it is of interest that Kimbrough, Robalino and Robson (2014) carried out related experiments with, usually, 12 subjects. These were randomly and anonymously matched into 6 pairs in each repetition to play a simplified two stage version of the game. A few subjects in stage 2 did describe themselves as indulging in non-myopic behavior, attempting, for example, to mislead, reward or punish the player in stage 1. It did not seem that such behavior was very successful. Further, such behavior was rare, in that 90 percent of stage 2 subjects made the optimal myopic choice. This all suggests that myopic behavior with a small number of players in each role could be viewed as a rule of thumb that does reasonably well in a variety of complex settings, even if it is not fully optimal.

Naive Players

Again, our definition of naivete binds only when the subgame is new. When the subgame is new, and there is no dominant choice, naive players condition in

[17] It is not necessary to make this assumption for role 1, but it will satisfy it in the end.

an arbitrary fashion on their own payoffs, but act in ignorance of other players' preferences.

DEFINITION 5: *All naive strategies for $i = 2, \ldots, I$ satisfy Assumption 9 in all subgames. There is a finite number of naive strategies for any role $i = 1, ..., I$ that map their own observed payoffs to an arbitrary pure choice, whenever any of the subgames faced has never arisen previously, and a dominant choice is lacking.*

As in the two stage case our relaxed definition of naivete strengthens our results. If any subgame faced is *not* new, and there is no dominant choice, the naive player can be rather smart. 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.

Sophisticated Players

There are two aspects to the *ToP* strategies. The first of these, given as part i) of Definition 6 below, concerns the utilization of the knowledge of others' preferences. The *SR-ToP* strategy, which will eventually dominate the population, makes the sequentially rational choice at each node when the preferences of subsequent players are known. The second aspect, given as ii) of Definition 6, concerns how such knowledge of the preferences of others could be acquired from observing their behavior.

DEFINITION 6: *All ToP strategies in role $i = 1, \ldots, I$ satisfy Assumption 9 in all subgames. It is convenient to describe the remaining requirements on the ToP strategies in any role $i = 1, \ldots, I$ 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 $\mathcal{Z}_n$, 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 9). A particular ToP strategy, the SR-ToP strategy, maps all of these preferences to the subgame perfect equilibrium choice at each node, if this equilibrium choice is unique. Other ToP strategies make a non-equilibrium choice in at least one subgame defined by the ordinal preferences of others and of the role in question.*[18] *ii) It is common knowledge among all ToP players in roles $i = 1, \ldots, I-1$ that there exists a positive fraction of SR-ToP players in every role.*

The appeal to common knowledge to describe the *ToP* strategies is merely for conciseness. We come back to this issue after presenting Example 2 in Section II.D.

[18]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 *SR-ToP* strategy a strict advantage over any other *ToP* strategy.

C. Evolutionary Adaptation

The population structure and associated payoffs are as follows—

DEFINITION 7: *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 $R \geq 1$, say where $r = 1$ is the SR-ToP strategy. The naive strategies are labelled as $r = R+1, \ldots, \bar{R}$, where $\bar{R} > R$.*[19] *The fraction of the total population in role $i = 1, \ldots, I$ 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 $\mathbf{f}^i_{n,t} = (f^i_{n,t}(1), \ldots, f^i_{n,t}(\bar{R}))$. The average 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 $\bar{\mathbf{z}}^i_{n,t} = (\bar{z}^i_{n,t}(1), \ldots, \bar{z}^i_{n,t}(\bar{R}))$.*

The distribution of strategies within each role evolves in an adaptive fashion, as in the two stage case. 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 10: *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 $\mathbf{f}^i_{n,t}$ with average payoffs $\bar{\mathbf{z}}^i_{n,t}$, the population structure in the next iteration is given by $\mathbf{f}^i_{n,t+1} = \mathbf{\Psi}(\mathbf{f}^i_{n,t}, \bar{\mathbf{z}}^i_{n,t})$.*[20]

This function $\mathbf{\Psi} : \Delta^{\bar{R}-1} \times [m, M]^{\bar{R}} \to \Delta^{\bar{R}-1}$, where $\Delta^{\bar{R}-1}$ is the unit simplex in $\mathbb{R}^{\bar{R}}$, has the properties i) $\mathbf{\Psi}$ is continuous, ii) $\Psi_r(\mathbf{f}^i_{n,t}, \bar{\mathbf{z}}^i_{n,t})/f^i_{n,t}(r) > \eta$ for some $\eta > 0$, and for $r = 1, \ldots, \bar{R}$[21] *, 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 $\Psi_{r^*}(\mathbf{f}^i_{n,t}, \bar{\mathbf{z}}^i_{n,t}) > f^i_{n,t}(r^*)$ and iv) if $\bar{z}^i_{n,t}(r) = \bar{z}^i_{n,t}(r')$, for all $r, r' \in \{1, \ldots, \bar{R}\}$, then $\mathbf{\Psi}(\mathbf{f}^i_{n,t}, \bar{\mathbf{z}}^i_{n,t}) = \mathbf{f}^i_{n,t}$.*[22]

Recall that Figure 2 gives a schematic representation of the model.

D. The Main Result

To gain an intuitive understanding how the generalized assumptions will generate the desired result, consider a three stage case—

Example 2.— Suppose the game has three stages, given by $i = 1, 2, 3$. The assumptions made in the two stage case have been now strengthened in two key ways. (a) Not only must all role 3 players choose a dominant strategy, but so must all role 2 players. (b) The *ToP* players in role 1 know that there are at least some *SR-ToP* players in role 2.[23] How does this enable *ToP* players in role 1 to learn the preferences of both subsequent roles?

[19] Of course, strategy r is quite different for different roles i and i'.

[20] If $t = \kappa(n)$, then $\Psi(\mathbf{f}^i_{n,t}, \bar{\mathbf{z}}^i_{n,t}) = \mathbf{f}^i_{n+1,1}$.

[21] This condition ensures that the *SR-ToP* strategy cannot become extinct in the short run when it could have lower payoffs than other strategies.

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

[23] As is true in general, common knowledge is a stronger assumption than this requirement.

Learning about $i = 3$'s preferences, whenever $\alpha > 1$, proceeds as shown for the two stage case above, only now this learning applies to the *ToP* players in both roles 1 and 2. Furthermore, 3's preferences become then common knowledge among all *ToP* players in role 1 and 2. The new issue that arises with three stages is: How do *ToP* players in role 1 learn role 2's preferences?

Suppose that a subgame is drawn in which role 2 players actually have a dominant choice, say a. (It can be shown that a *strictly positive* fraction of role 2 subgames have this property.) The *ToP* players in role 1 do not know that such a dominant choice exists for 2, even after observing that they all choose a, as they must. These *ToP*s in role 1 do know, however, that the *ToP* 2s also know 3's preferences. Hence, whether such a dominant choice exists or not, whatever other strategies might do, the *SR-ToP*s in role 2 must now be making an subgame perfect choice. Hence they have unequivocally demonstrated to all the *ToP* in role 1 that they prefer the outcome induced by a to any outcome they might have induced instead.

The assumption that $\alpha > 1$, is still enough to ensure that the *ToP* players in role 1 can keep up with the rate of innovation, and then build up a complete picture of the preferences of role 2, to add to the complete picture already obtained of those of role 3.

Once the *SR-ToP* in role 2 knows the preferences of role 3, it will clearly outdo any other strategy and come to dominate the population in role 2. Once this is so, and the *SR-ToP* in role 1 has learnt the preferences of both subsequent roles, this strategy will, in turn, dominate the population in role 1.

Example 2 also illustrates that the common knowledge assumption for the *ToP* players, as described in Definition 6 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 $I = 3$, is that it is *as if* the *ToP*s in roles 1 add to their knowledge of role 2's preferences as described above. Once a *ToP* player in role 1 has seen histories in which all of 2's binary choices have been put to the test like this, given that this is already true for role 3, the role 1 *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, σ, 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.[24]

The main result for the general case, is analogous to the main result for the two stage case. In the limit, the *SR-ToP* strategy in every role $i = 1, \ldots, I-1$ learns the preferences of others, and uses what is learnt to choose optimally, ultimately dominating the population in role i.

THEOREM 2: *Suppose Assumptions 6-10 all hold. Suppose that there are a finite number of ToP strategies, including SR-ToP in particular, as in Definition*

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

6, and a finite number of naive strategies, as in Definition 5. If $\alpha \in (1, A^2 - 1)$*, then the proportion of SR-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* $i = 1, \ldots, I - 1$*. The observed pattern of play in each realized game converges to a subgame perfect equilibrium, in probability.*

PROOF:

This is relegated to the online Appendix.

The proof derives from the result that all of the *ToP* strategies learn all the other roles' preferences if $\alpha > 1$, but all naive strategies see 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 *SR-ToP* fully exploits.

The bounds that $\alpha \in (1, A^2 - 1)$ are tight in the sense that, if $\alpha < 1$ then it is mechanically impossible for the *ToP* players to learn rapidly enough the preferences of opponents from their binary choices. Similarly, if $\alpha > A^2 - 1$, then naive players in role $I - 1$ see only familiar subgames in the limit. The condition for naive players in role i to see only familiar games becomes more stringent with lower i. That is, earlier stages generate a higher critical value of α because they face more possible subgames. The theorem then gives a sufficient condition for *all* naive players in roles $i = 1, \ldots, I - 1$ to face only unfamiliar subgames.[25]

An interesting aspect of the general result is that the advantage of the *ToP* strategies over the naive strategies is more pronounced in more complicated games.

On the one hand, $\alpha > 1$ implies that the *ToP* players in role i learn *all* of the subsequent roles' binary preferences in the limit, despite their being many of these subsequent roles, and regardless of the number of actions A. It is certainly true that the learning involved with many subsequent roles is more onerous. Nevertheless, this increased difficulty is not reflected in a higher critical value of α. As long $\alpha > 1$, *ToP* player in roles $i < I$ can first learn the preferences of role I, for exactly the same reason that this is possible with just two stages. Having established these, as common knowledge, these *ToP* players can then deduce the preferences of role $I - 1$, applying the argument sketched in Example 2. This second step relies on there being a positive fraction of games with strictly dominant choices for role $I - 1$. Intuitively, since there are order n^2 outcome pairs for role $I - 1$, this second step still only requires that $\alpha > 1$. This argument can be extended, by backwards induction, to any number of subsequent roles, all under the condition that $\alpha > 1$. The proof of this with a general number of stages is more complex than in the two stage case, since sequentially rational behavior by stages $i < I$ has to be established by backwards induction, and only holds in the limit, so there are more sources of "noise". This is the heart of the general proof.

[25] 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 $i = 1, \ldots, I - 1$.

On the other hand, the naive players have a harder task to learn all the subgames they initiate, if they are at an earlier stage, or if A is larger, simply because there then more such subgames. In the role i, that is, the cutoff value for a naive strategy is $\alpha = A^{I-i+1} - 1$, below which learning is mechanically impossible in the long run, and this decreases with i, and increases with A. Stage $I-1$ is where the naive players face the smallest number of possible subgames, but the critical value for these naive players, $A^2 - 1$, in particular, is increasing in A.[26]

What happens outside the range $\alpha \in (1, A^2 - 1)$?

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 *SR-ToP* players, the naive players will, at best, match the *SR-ToP*s.

If $\alpha > A^2 - 1$, naive players in at least role $I - 1$ have seen essentially all subgames previously, in the long run. The relative performance of the *SR-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 *SR-ToP* players. It is, in any case, not intuitively surprising that a clear advantage to the *SR-ToP* strategy relies upon there being at least a minimum rate of introduction of novelty.

Why is a subgame perfect equilibrium obtained here? Why could players in some intermediate stage not gain from misrepresenting their preferences to earlier stages?

The attainment of subgame perfection in Theorem 2 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-equilibrium 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 a subgame perfect equilibrium is attained once the preferences of others are known is analogous to Hart (2002).[27]

[26]The proof of these claims is straightforward, and extends the argument given to prove the corresponding part of Theorem 1. Observe the following. Assumption 8 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 number of outcomes. Since only one game is played at each iteration, this also gives the order of the maximum number of distinct subgames occurring along any such history. In period n, for stage i, there are $(N+n)^{A^{I-i+1}}$ possible subgames. If $\alpha+1 < A^{I-i+1}$, then the fraction of subgames that are familiar must tend to zero, surely.

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

E. Stahl Revisited

The eventual predominance of the *SR-ToP*s over all the naive strategies resolves the issue raised by Stahl (1993)—that less smart, but lucky, players can outdo smarter ones.[28] The underlying reason for the reverse result here is that in the current environment players are repeatedly confronted with novel games. 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 subgame perfect 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 *SR-ToP* that is not preprogrammed but rather adapts to the outcomes and games that are drawn, and ultimately chooses optimally essentially always.[29]

That is—

COROLLARY 1: *Under the hypotheses of Theorem 2, 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 *SR-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

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

[28] In Stahl (1993) a single game is played repeatedly by the agents—*lucky* in this context is being preprogrammed with the strategy that happens to be a best response to whatever strategy the opponents settle on in the long run—*smart* is being able to deduce this strategy directly.

[29] 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 subgame perfect 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.

If $\alpha > A^2 - 1$, however, then naive players in at least role $I - 1$ 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.

F. Further Remarks

We close this subsection with several additional 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. 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 *ToM*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.[30]

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.[31]

3) Consideration of a long run equilibrium, as in Theorem 2, is simpler analytically than direct consideration of the speed of out-of-equilibrium learning

[30]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.

[31]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.

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

G. 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) provides 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. As already mentioned above, 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.

III. 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 subgame perfect equilibrium 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.

REFERENCES

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

Bhatt, Meghana, and 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.

Billingsley, Patrick. 1986. *Probability and Measure.* 2nd ed., Chicago:John Wiley and Sons.

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

Cheney, Dorothy L., and Robert M. Seyfarth. 1990. *How Monkeys See the World: Inside the Mind of Another Species.* Chicago:University of Chicago Press.

Crawford, Vincent P., and 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.
Egghe, Leo. 1984. *Stopping Time Techniques for Analysts and Probabilists.* Cambridge, UK:Cambridge University Press.
Elker, Johann, David Pollard, and Winfried Stute. 1979. "Glivenko-Cantelli Theorems for Classes of Convex Sets." *Advances in Applied Probability*, 11(4): 820–833.
Harsanyi, John C. 1967-68. "Games with Incomplete Information Played by 'Bayesian' Players, I-III." *Management Science*, 14: 159–182, 320–334, 486–502.
Hart, Sergiu. 2002. "Evolutionary Dynamics and Backward Induction." *Games and Economic Behavior*, 41: 227–264.
Kalai, Gil. 2003. "Learnability and Rationality of Choice." *Journal of Economic Theory*, 113: 104–117.
Kimbrough, Erik, Nikolaus Robalino, and Arthur Robson. 2014. "The Evolution of Theory of Mind: Theory and Experiments." Unpublished.
Kirk, Robert. 2014. "Zombies." In *The Stanford Encyclopedia of Philosophy.* , ed. Edward N Zalta. Stanford University.
Knoepfle, Daniel T., Colin F. Camerer, and Joseph T. Wang. 2009. "Studying Learning in Games Using Eye-tracking." *Journal of the European Economic Association*, 7(2-3): 388–398.
Mohlin, Erik. 2012. "Evolution of Theories of Mind." *Games and Economic Behavior*, 75(1): 299–318.
Onishi, Kristine H., and René Baillargeon. 2005. "Do 15-Month-Old Infants Understand False Beliefs?" *Science*, 308(8): 255–258.
Robalino, Nikolaus, and Arthur J. Robson. 2012. "The Economic Approach to "Theory of Mind"." *Philosophical Transactions of the Royal Society, Biological Sciences*, 367: 2224–2233.
Robson, Arthur J. 2001. "Why Would Nature Give Individuals Utility Functions?" *Journal of Political Economy*, 109(4): 900–914.
Robson, Arthur J., and Hillard Kaplan. 2003. "The Evolution of Human Longevity and Intelligence in Hunter-Gatherer Economies." *American Economic Review*, 93(1): 150–169.
Stahl, Dale O. 1993. "Evolution of Smart$_n$ Players." *Games and Economic Behavior*, 5(4): 604–617.

Appendix A: Proof of Theorem 1

We establish here our main result for the simple version of the model, Theorem 1. Role 2 makes the sequentially rational choice in every game (Assumption 4), and thus the results here concern the player 1s. The notation is modified slightly in the proof. In particular, a single subscript is used to denote the total number of accumulated iterations, in lieu of subscripting the period n, and iteration t. For example, H_s is written in place of the history $H_{n,t}$, where s is now the number of accumulated iterations along this history. For each period $n = 1, 2, \ldots$, the

notation $s(n)$ is used to denote the iteration $s = \sum_{m=1}^{n-1} \kappa(m) + 1$. Notice, in particular, that the n-th novel outcome arrives at the beginning of iteration $s(n)$.

Assume throughout the section that Assumptions 1-5 hold.

The first step in the proof of Theorem 1 is to show that if $\alpha > 1$, the *SR-ToP*s learn their opponents' preferences completely in the limit, and therefore choose optimally against their opponents with probability tending to one.

In this simple environment with two player roles and two actions, each choice by the 2s directly reveals a pairwise preference. Specifically, the 2s make the dominant choice, as in Assumption 4, and every choice by role 2 eliminates all ambiguity about their preferred option, since there are no remaining players. One measure of how much has been revealed about 2's preferences is therefore the number of distinct 2 role subgames reached along the history. Consider in particular the following.

DEFINITION 8: *Let K_s denote the number of distinct role 2 subgames reached along H_s. There are $|\mathcal{Z}_n|^2$ role 2 subgames throughout period n. For each $s = s(n), \ldots, s(n+1) - 1$ write $L_s = K_s/|\mathcal{Z}_n|^2$ as a measure of how much can be learned about 2's preferences from H_s.*

L_s is a conservative measure of how much information is conveyed by history about 2's preferences, but it suffices for the present purpose.[32] Specifically, we have the key result that L_s converges in probability to one whenever $\alpha > 1$. The proof is immediate in the light of the next two results (Lemma 2, and Lemma 3).

LEMMA 2: *Suppose that there are two player roles, and two actions available for each role. Suppose further that L_s converges in probability to some random variable L. If $\alpha > 1$, then $L = 1$ a.e.*

PROOF:

With probability $(1 - L_s)^2$ the game at iteration s is such that neither of its role 2 subgames have occurred along the history. Such a game at s ensures the 1s observe role 2's choice in a novel subgame. Hence, for every $s = 1, 2, \ldots,$

$$E(K_{s+1} \mid H_s) - K_s \geq (1 - L_s)^2. \tag{A1}$$

That is, the smaller is the proportion of role 2 subgames seen along the history, the more likely it is that an unfamiliar one will arise.

Summing (A1) over $s = 1, 2, \ldots, \tau - 1$, and taking the unconditional expectation of the result yields

$$E(K_\tau) - E(K_1) \geq \sum_{s=1}^{\tau-1} E((1 - L_s)^2). \tag{A2}$$

[32]This ignores the transitivity of 2's preference ordering. Moreover, the denominator accounts for all of the available 2 subgames, including trivial ones in which 2s face the same two outcomes. Notice, however, that the probability of having repeated outcomes in a game tends to zero.

Let $n(\tau)$ denote the period prevailing during iteration $\tau = 1, 2, ...$ Notice, in particular, that $L_\tau = K_\tau/|\mathcal{Z}_{n(\tau)}|^2$. Next, observe that $K_1 = 0$ by definition. Dividing both sides of equation (A2) by $|\mathcal{Z}_{n(\tau)}|^2$ therefore gives

$$E(L_\tau) \geq \frac{\tau - 1}{|\mathcal{Z}_{n(\tau)}|^2} \cdot \left[\frac{1}{\tau - 1} \cdot \sum_{s=1}^{\tau-1} E((1 - L_s)^2) \right]. \tag{A3}$$

Now suppose $\alpha > 1$, and consider (A3) as τ tends to infinity. Notice first that the $(\tau - 1)/|\mathcal{Z}_{n(\tau)}|^2$ term in the expression diverges to infinity. To see this observe the following. The iteration corresponding to the arrival of the n-th novel outcome, $s(n) = \sum_{m=1}^{n-1} \kappa(m) + 1$, is non-decreasing in n, and has order of $n^{1+\alpha}$. Since each iteration τ satisfies $s(n(\tau)) \leq \tau \leq s(n(\tau) + 1) - 1$, it follows that $n(\tau)$ has order of $\tau^{\frac{1}{1+\alpha}}$, and hence that $|\mathcal{Z}_{n(\tau)}|^2 = (N + n(\tau))^2$ has order of $\tau^{\frac{2}{1+\alpha}}$. Clearly if $\alpha > 1$, then $\tau - 1$ grows at a faster rate than $|\mathcal{Z}_{n(\tau)}|^2$. Next, notice that the quantity on the right hand side of (A3) must surely be bounded above by one, uniformly in τ (this is because surely $L_s \leq 1$). The limit inferior of the bracketed term in the expression must then be zero, since otherwise the quantity on the right hand side would diverge to infinity.

Now suppose L_s converges in probability to L as hypothesized in the statement of Lemma 2. The bracketed term in (A3) will then converge to $E((1-L)^2)$. Since the limit inferior of these means is zero, it follows that $E((1-L)^2) = 0$, and hence that $L = 1$ a.e. This completes the proof of Lemma 2.

The next result is that if $\alpha > 1$, then L_s converges in probability to some random variable L. Taken together with Lemma A2, this implies that L_s converges in probability to one, whenever $\alpha > 1$. The proof of convergence is rather technical and involved even for the simple version of the model that we focus on here. A complete proof is given in this section but intuitive arguments are relied upon whenever these are thoroughly convincing. Consider first some key observations.

The crucial factor regarding the convergence of L_s is the behavior of the process along the subsequence, $s(n), n = 1, 2, \ldots$, of iterations corresponding to the arrivals of novel outcomes. In particular, if the process along this subsequence converges to some limit, then the overall sequence must converge, and moreover, it must possess the same limit. This is shown formally in the online Appendix. An intuitive treatment is as follows.

Note that L_s is non-decreasing in between the arrivals of novel outcomes. Specifically, the numerator, K_s, never decreases, and the denominator, $|\mathcal{Z}_n|^2$, is constant until the next outcome is introduced. However, the L_s process is not a sub-martingale overall. The introduction of the $n+1$-th new outcome causes the denominator to increase by a factor of n (i.e., the denominator changes from $|\mathcal{Z}_n|^2$ to $|\mathcal{Z}_{n+1}|^2$), inducing a sudden decrease in L_s.[33] It is important, however, that as the number of outcomes increases, the drop in L_s due to the arrival of yet another

[33] If L_s were a sub-martingale, the almost sure convergence of the sequence would follow immediately from the martingale convergence theorem.

outcome becomes smaller, tending to zero eventually. To see this note that

$$\begin{aligned} L_{s(n)} \geq K_{s(n)-1}/|\mathcal{Z}_n|^2 &= (K_{s(n)-1}/|\mathcal{Z}_{n-1}|^2)\cdot(|\mathcal{Z}_{n-1}|^2/|\mathcal{Z}_n|^2) \\ &= L_{s(n)-1}\cdot(|\mathcal{Z}_{n-1}|^2/|\mathcal{Z}_n|^2), \end{aligned} \tag{A4}$$

and hence that surely $\liminf\{L_{s(n)} - L_{s(n)-1}\} \geq 0$, since $|\mathcal{Z}_{n-1}|^2/|\mathcal{Z}_n|^2$ surely converges to one. The above discussion implies that, if the subsequence $L_{s(n)}, n = 1, 2, \ldots$, converges in probability to L, then so must the overall sequence, $\{L_s\}$, a result that is used in proving the next result—

LEMMA 3: *Suppose that there are two player roles, and two actions available for each role. If $\alpha > 1$, then there is a random variable L such that L_s converges in probability to L.*

PROOF:

It suffices therefore to show that if $\alpha > 1$, then the subsequence $\{L_{s(n)}\}$ converges in probability to some random variable L. Since we work exclusively with this subsequence in the proof, we simplify notation by writing $\bar{L}_n$, $\bar{K}_n$, and $\bar{H}_n$ in place of $L_{s(n)}$, $K_{s(n)}$, and $H_{s(n)}$, respectively, for each $n = 1, 2, \ldots$. To establish the convergence of $\bar{L}_n$ we use the following definition and result (Egghe (1984) [Definition VIII.1.2, and Theorem VIII.1.22]).

SUBMIL CONVERGENCE: *The $\{\bar{H}_n\}$ adapted process $\{\bar{L}_n\}$ is a sub-martingale in the limit (Submil) if for each $\eta > 0$ there is almost surely an integer M such that $n > m \geq M$ implies $E(\bar{L}_n \mid \bar{H}_m) - \bar{L}_m \geq -\eta$. If $\bar{L}_n$ is a Submil, then there exists a random variable L such that $\bar{L}_n$ converges in probability to L.*[34]

Given that Submils converge in probability, we prove Lemma 3 by showing that if $\alpha > 1$, then $\bar{L}_n$ is a Submil. Toward this end, consider two periods, m, and n, such that $n > m$. Given that $\bar{L}_n = \bar{K}_n/|\mathcal{Z}_n|^2$ it is straightforward to show that

$$E(\bar{L}_n - \bar{L}_{n-1} \mid \bar{H}_m) < 0 \implies E(\bar{K}_n - \bar{K}_{n-1} \mid \bar{H}_m) < |\mathcal{Z}_n|^2 - |\mathcal{Z}_{n-1}|^2. \tag{A5}$$

That is, $E(\bar{L}_{n-1} \mid \bar{H}_m)$ decreases only if the expected number of new subgames reached during period $n-1$ (the expected increase in the numerator of $\bar{L}_{n-1}$) is less than the number of new subgames introduced by the n-th novel outcome.

Next, revisit equation (A1), summing this time over $s = s(n), \ldots, s(n+1)-1$, to obtain

$$\begin{aligned} E(\bar{K}_n - \bar{K}_{n-1} \mid \bar{H}_m) &\geq \sum_{s=s(n-1)}^{s(n)-1} E((1-L_s)^2) \mid \bar{H}_m) \\ &\geq \kappa(n-1)\cdot E((1-L_{s(n)-1})^2 \mid \bar{H}_m). \end{aligned} \tag{A6}$$

To get the second line here we used the fact that there are $\kappa(n-1)$ terms in the

[34] For guaranteed convergence here the process in question must be uniformly integrable. L_s satisfies this requirement since $|L_s| \leq 1$ surely.

summation, and that L_s is non-decreasing as s ranges from $s(n-1)$ to $s(n)-1$. Combining (A6) with (A5) we see that

$$\begin{aligned} &E(\bar{L}_n - \bar{L}_{n-1} \,|\, \bar{H}_m) \;<0 \implies \\ &E((1-L_{s(n)-1})^2 \,|\, \bar{H}_m) \;<\; (|\mathcal{Z}_n|^2 - |\mathcal{Z}_{n-1}|^2)\,/\kappa(n-1). \end{aligned} \tag{A7}$$

Now suppose $\alpha > 1$. In this case the $(|\mathcal{Z}_n|^2 - |\mathcal{Z}_{n-1}|^2)/\kappa(n-1)$ term in equation (A7) surely converges to zero. The same equation then implies that for sufficiently large n, $E(\bar{L}_n - \bar{L}_{n-1} \,|\, \bar{H}_m)$ is negative only if $E(L_{s(n)-1} \,|\, \bar{H}_m)$ is sufficiently close to one. But as we argued before the statement of Lemma 3, $\liminf\{\bar{L}_n - L_{s(n)-1}\} = 0$ surely (i.e. equation (A4)), and thus it follows that for sufficiently large n, $E(\bar{L}_n - \bar{L}_{n-1} \,|\, \bar{H}_m)$ is negative only if $E(\bar{L}_n \,|\, \bar{H}_m)$ is close to one. More precisely, we have the following. Whenever $\alpha > 1$, for each $\eta > 0$ there is a finite integer, $M(\eta)$, such that:

$$\begin{aligned} &\text{If } n > m \geq M(\eta), \text{ then} \\ &E(\bar{L}_n - \bar{L}_{n-1} \,|\, \bar{H}_m) \;<0 \quad\implies\quad E(\bar{L}_n \,|\, \bar{H}_m) \;>\; 1-\eta. \end{aligned} \tag{A8}$$

It is this property of $\bar{L}_n$ that ensures the process has the Submil property. This will be shown next.

To see that $\bar{L}_n$ is a Submil fix η and choose $M(\eta)$ as in (A8). Consider any m, and n such that $n > m \geq M(\eta)$. Suppose $E(\bar{L}_n \,|\, \bar{H}_m) \leq 1-\eta$. Then (A8) implies $E(\bar{L}_n - \bar{L}_{n-1} \,|\, \bar{H}_m) \geq 0$, and therefore that $E(\bar{L}_{n-1} \,|\, \bar{H}_m) \leq 1-\eta$. This in turn implies (using (A8) again) that $E(\bar{L}_{n-1} - \bar{L}_{n-2} \,|\, \bar{H}_m) \geq 0$, and therefore that $E(\bar{L}_{n-2} \,|\, \bar{H}_m) \leq 1-\eta$. Proceeding recursively we see that $E(\bar{L}_n \,|\, \bar{H}_m) \leq 1-\eta$ implies $E(\bar{L}_k - \bar{L}_{k-1} \,|\, \bar{H}_m) \geq 0$, for each $k = m+1, \ldots, n$, and therefore that $E(\bar{L}_n \,|\, \bar{H}_m) - \bar{L}_m \geq 0$. Clearly $E(\bar{L}_n \,|\, \bar{H}_m) - \bar{L}_m < 0$, only if $E(\bar{L}_n \,|\, \bar{H}_m) > 1-\eta$. Since $\bar{L}_m$ is surely no greater than one it follows that $E(\bar{L}_n \,|\, \bar{H}_m) - \bar{L}_m \geq -\eta$, for all $n > m \geq M(\eta)$. Since η is an arbitrary positive number, it follows that $\bar{L}_n$ is a Submil. This completes the proof.

Lemmas 2, and 3 in combination give—

LEMMA 4: *Suppose that there are two player roles, and two actions available for each role. If $\alpha > 1$, then the history reveals role 2 preferences completely in the limit, that is, L_s converges in probability to one.*

The *SR-ToP* strategy of role 1 makes the subgame perfect choice whenever the 2s' choices in the game have been observed previously along H_s (See Definition 6 and the discussion in the example after it). Lemma 4 then sets the stage for the ultimate dominance of the *SR-ToP* strategy.

The next result is that the naive strategy makes a suboptimal choice with positive probability in the long run. Taken together with Lemma 4, the result implies that the *SR-ToP* strategy outdoes the naive strategy eventually.

LEMMA 5: *Suppose $\alpha < 3$, and that the fixed game tree has four terminal nodes. Then the payoff to the 1s from the naive strategy is dominated by the subgame*

perfect equilibrium payoff with probability that is bounded away from zero in the limit.

PROOF:

Let G denote the distribution of games implied by F (where F is defined in Assumption 2).

Note that all the games are new in the long run when $\alpha < 3$, as is proved after the statement of Theorem 1. We next show that there is a set of games with positive measure under G for which the initial reaction of the naive strategy differs from the equilibrium choice. Suppose there is a positive measure subset of games, say $\mathcal{Q}'$, that lack a dominant action for role 1, and in which role 2's payoffs are all distinct. For every game in this subset, if the initial response of the naive strategy in the game is optimal, then this choice can be rendered suboptimal by some rearrangement of 2's payoffs. Therefore, if there is a positive measure subset of $\mathcal{Q}'$ such that the initial reaction by the naive strategy is optimal, then there must also be a positive measure subset within $\mathcal{Q}'$ where the initial reaction is sub-optimal. The Glivenko-Cantelli Lemma (see Lemma 1) implies that these games, in which the initial naive reaction is suboptimal, come up with positive probability in the limit. This completes the proof of Lemma 5.

We have now shown that if $\alpha \in (1, 3)$, then 1) the *SR-ToP*s make the subgame perfect equilibrium choice with probability tending to one (Lemma 4), and 2) the payoff to the naive strategy is suboptimal with probability bounded away from zero in the limit (Lemma 5). Since role 2s always make the sequentially rational choice, the subgame perfect choice is optimal for the 1s. It follows naturally that the *SR-ToP* strategy eventually dominates. We therefore end with—

LEMMA 6: *Suppose the fraction of role* 1*s that use the SR-ToP strategy in iteration s is $f_s \in [0, 1]$. If $\alpha \in (1, 3)$, then f_s converges in probability to one.*

Although the intuition for the result is compelling, a fully rigorous proof involves rather tedious calculations. We therefore defer the formal proof to the online Appendix (see Proposition 4 there).