The Empirical Content of Adaptive Models

Jonathan Bendor Daniel Diermeier Michael Ting[1]

December 20, 2002

[1] Jonathan Bendor is from the Graduate School of Business, Stanford University; Daniel Diermeier is from the Kellogg School of Management, Northwestern University, and Michael Ting is from the Department of Political Science and SIPA, Columbia University. We would like to thank Adam Meirowitz for his very helpful comments on an earlier draft, and Sunil Kumar, Dilip Mookherjee, and Debraj Ray for many illuminating conversations over the years.

Abstract

Models with adaptive agents have become increasingly popular in computational sociology (e.g. Macy 1991, Macy and Flache 2002). In this paper we show that at least two important kinds of such models lack empirical content. In the first type players adjust via reinforcement learning: they adjust their propensities to undertake actions based on the kind of feedback they receive. In the second type players satisfice—i.e., retain the same action if the payoff is satisfactory—and search when payoffs are unsatisfactory. In both types of models feed-back is coded as satisfactory if it exceeds some aspiration level, where aspirations may themselves adjust to reflect prior payoffs. We show that outcomes in either type of model are highly sensitive to initial parameters; that is, *any* outcome of the stage game can be supported as a stable outcome. Intuitively, this occurs because players may be endowed with initial aspirations that make any outcome satisfactory, and thus the actions producing that outcome can be reinforced by all players. These results hold even when players' aspirations are endogenous. We also present two solutions to this problem. First, we show that stochastic versions of the model ensure ergodicity: i.e., the players' action-propensities and aspirations converge to a unique limiting distribution that is independent of their initial values. Second, we show that if players engage in social comparisons—specifically, an agent's aspiration depends on the payoffs of his peers, in addition to his own—then far fewer outcomes can be sustained in equilibrium.

1 Introduction

Computational modeling and theories of adaptive behavior have been closely linked for many years, as a recent review (Macy and Wilner 2002) has made clear. In organization theory, for example, the connection goes back almost 40 years (Cyert and March 1963). More recently, agent-based models—now probably the most common approach in computational sociology—typically assume boundedly rational actors who adjust their behavior via simple rules of thumb.

Because computational models in sociology rely so heavily on theories of adaptation, understanding the latter's methodological properties is a vital part of any serious study of the former. And one key methodological property of a class of theories is its empirical content. Some theories say "almost anything" can happen; some predict a much narrower range of outcomes. The latter naturally make stronger predictions than the former, and one needn't be a strict Popperian to value the difference.

It is now fairly well-known that the problem of nearly vacuous predictions plagues standard game theoretic models of repeated interactions. This problem is identified by the so-called "folk theorems," which state roughly that any outcome in a repeated game in which each person gets her maximin payoff is a (subgame-perfect) Nash equilibrium, if the future is sufficiently important (Fudenberg and Maskin 1986).[1] For example, in the Prisoner's Dilemma this includes any outcome that gives each player at least his or her mutual defection payoff. So, the canonical game-theoretic analysis of the repeated PD has little empirical content.

This problem of a lack of empirical content—known by game theorists for over 40 years but increasingly recognized by nonspecialists as well—has been rightfully considered such a serious issue for game theory that some of the best minds in the field have devoted a lot of time and effort to solving it (e.g., Harsanyi and Selten 1988). Although no one proposed solution commands universal acceptance, game theorists do continue to agree that it is a problem of the first order.[2]

[1] Because the maximin payoff is the one that a player can guarantee herself, regardless of other player's actions, it is sometimes aptly called a player's "security level."

[2] This view is so widespread that game theorists today typically regard a model that asserts only that a particular outcome in a repeated game is an equilibrium as uninteresting or even trivial. This evaluation has relevance

However, it is less widely understood that similar problems plague many theories of adaptively rational behavior. Indeed, such models are frequently motivated by the lack of predictive power in classical game theory (e.g. Macy and Flache 2002). In this paper we establish the validity of this claim for two types of theories that have long been important in the overall research program of bounded rationality: theories of trial-and-error learning (e.g., Bush and Mosteller 1955) and theories of satisficing-and-search (e.g., Simon 1955).[3] It turns out that the two kinds of theories have a strong and significant overlap, and very similar results can be established for them.

2 The Predictive Content of Models of Adaptation: Several "Folk Theorems"

We first establish the result for the case of exogenously fixed aspirations, in the context of theories of learning (Theorem 1) and of theories of satisficing (Theorem 2). These results cover a very large set of games: in particular, (unlike the classical game theoretic folk theorems) they allow payoffs to be stochastic and nonstationary. Then we will show that the problem reappears, for games with deterministic and stationary payoffs, even in the context of endogenous aspirations. (Theorem 3 covers learning models; Theorem 4, satisficing models.)[4]

Let t denote discrete time periods and i denote actors ($i = 1, \ldots, n$). A player i's generic action is denoted α_i. Let Ω denote an outcome function such that $o = \Omega(\alpha_1, \ldots, \alpha_n)$ denotes some generic outcome. We write $\alpha(o)$ to denote any action profile that generates o and $\alpha_i(o)$ to denote i's action within that profile. That is, $\Omega(\alpha_1, \ldots, \alpha_i, \ldots, \alpha_n) = o$ for some profile $\alpha(o)$.

Let $\Pi_{i,t}$ denote the set of i's feasible payoffs in t, and let $\pi_{i,t(o)}$ denote i's realized payoff in period t, given outcome o. The generality of the notation reflects two substantively important properties. First, in a given period

for the current discussion, as we shall see shortly.

[3] For examples of reinforcement models see Macy (1989, 1991a, 1993, 1995), Kanazawa (2000), and Bendor, Diermeier and Ting (2003); for examples of satisficing models see Cyert and March (1963), Winter (1971), and Nelson and Winter (1982).

[4] We have not seen these results in the literature. However, we cheerfully acknowledge that they may already be understood by the small coterie of formal modellers of reinforcement learning. If so, they should be regarded as "folk theorems."

payoffs may be stochastic rather than deterministic: i.e., Theorems 1 and 2 allow for the possibility that a given vector of actions does not determine a unique set of payoffs to the players but instead generates *distributions* of payoffs. Second, over time payoffs may change. That is, Theorems 1 and 2 do not require stationary (time-homogeneous) payoffs. Indeed, the only structure we impose on payoffs for Theorems 1 and 2 is that every player has a minimum feasible payoff, i.e., $\Pi_{i,t}$ has a minimum for all i and t, and let $\underline{\pi}_i$ denote the smallest such minimum. Naturally, the twin assumptions of deterministic and stationary payoffs—standard in repeated game theory and common in computational modeling—are also permitted by Theorems 1 and 2. If payoffs are stationary we drop the time subscript and simply write Π_i and $\pi_i(o)$.

A player's propensity to play action α_i at t is denoted $p_{i,t}(\alpha_i) = 1$ where $p_{i,t}(\cdot)$ is a probability measure over i's set of actions. We write $p_{i,t}$ to denote i's measure over all actions at t. Aspiration levels are denoted by $a_{i,t}$. We assume that there is an aspiration level for each possible payoff level, i.e. for each i, t and o there exists some $a_{i,t}$ such that $a_{i,t} = \pi_{i,t}(o)$. In the case of exogenous ("fixed") aspiration levels we also write a_i. That is, we have $a_{i,t} = a_i$ for all t. We denote agent i's minimum and maximum feasible aspirations $\underline{a}_i$ and $\overline{a}_i$, respectively.

For the solution concept of all four Theorems we use Macy and Flache's concept (2002) of a self-reinforcing equilibrium (SRE). (The following discussion describes the solution concept in terms of learning theories; the extension to satisficing models is straightforward.) In an SRE with exogenous aspirations players' propensities to try certain actions generate outcomes and hence feedback which are consistent with those original propensities. Thus the combination of propensities, aspirations, actions, and payoffs form an equilibrium in which all these different elements reinforce each other. (Hence, in terms of an underlying stochastic process, an SRE is an absorbing state.) More precisely, suppose that in period t every player has a propensity of 1.0 to try some action. Then this set of propensities is, in combination with a vector of fixed aspirations, a (pure) SRE if each player gets feedback in t such that all of these propensities continue to hold in $t + 1$ (and, hence, thereafter). Thus this set of propensities is indeed self-reinforcing: the actions produce payoffs that in turn reinforce the underlying propensities that generated those actions in the first place. (In an SRE with endogenous aspirations, both propensities and aspirations must be self-replicating.)

Definition 1 *A tuple* $(p_{i,t}; a_{i,t})_i^t$ *is a* ***Self-Reinforcing Equilibrium*** *iff for all* i, t *and* α_i*:*
(i) $p_{i,t+1}(\alpha_i) = p_{i,t}(\alpha_i)$*, and*
(ii) $a_{i,t+1} = a_{i,t}$.

Note that (ii) is trivially satisfied if aspirations are exogenous.

Reinforcement learning. Models of reinforcement learning (including the well-known Bush-Mosteller model) are designed to capture the "Law of Effect" (Thorndike 1911): positive reinforcement increases the tendency to play an action, negative reinforcement decreases it. To capture this general idea both of our Theorems on learning models use a very general axiom of positive reinforcement, Axiom 1, which says that if an action produces a satisfactory payoff in the current period then the agent will not decrease his propensity on that action.[5] Note that this axiom is weaker than the Law of Effect since it does not make any assumptions about the impact of negative feedback.

Axiom 1 *(positive feedback): For all* i, t*, and action* α_i *chosen by player* i *in period* t*, if* $\pi_{i,t} \geq a_{i,t}$ *then* $p_{i,t+1}(\alpha_i) \geq p_{i,t}(\alpha_i)$.

Theorem 1 *Consider any repeated game, with either stationary or nonstationary payoffs, in which players adjust their action-propensities by any arbitrary mix of adaptive rules that satisfy Axiom 1 and where aspirations are exogenously fixed. Then* any *outcome* o *of the stage game can be sustained as a stable outcome by some pure SRE.*

(For the proof of Theorem 1 and of all other results, see the appendix.)

Proving Theorem 1 is easy because one is free to exogenously fix aspirations.[6] Thus a person could be content with even the lowest payoff in a game,

[5] The assumption is stated for the case of endogenous aspirations. To reformulate it for exogenous aspirations is easy: one just replaces the endogenous aspiration, $a_{i,t}$, by the exogenous one, a_i, in the assumption's key inequality, below.

[6] As mentioned above, the proof of the theorem uses the assumption that for each payoff there is a corresponding aspiration level. (Theorem 2's proof uses the same assumption.) Not only is this assumption very mild, relaxing it does not significantly change the results. For example, suppose that each player i has a lowest aspiration level satisfying $\underline{a}_i > \underline{\pi}_i$. Then the conclusion

if his aspiration level is low enough. Exactly the same mechanism works in the context of satisficing-and-search models (with exogenous aspirations). We turn to these models now.

Satisficing Theories. Behavioral theories of search are based on the premise that search is problem-driven: decision makers search for new alternatives if and only if today's action is unsatisfactory, i.e., yields a payoff that falls below the decision maker's aspiration level. When this informal theory is formalized as a mathematical or computational model, the basic premise naturally leads one to construct a stochastic process in which the state space is the vector of actions chosen by the agents. Thus, if the process is Markovian, then the actor stays where s/he is (satisfices) if the current payoff is satisfactory, and with some positive probability transits to a different alternative tomorrow if today's payoff is unsatisfactory (Simon 1955). Hence a key difference—possibly the major one—between learning theories and satisficing-and-search theories is how they specify their state spaces: whereas the latter describes its states directly in terms of decision makers' actions, the former's state space is more complex, being actors' propensities over actions. (Each type of theory includes aspirations in the state space if aspirations are endogenous; see Theorems 3 and 4, below.)

We continue to use exactly the same assumptions about the nature of the stage game as we used for learning theories. (In particular, this means that Theorem 2, like Theorem 1, allows for payoffs that are stochastic or nonstationary or both.) But instead of Axiom 1 we use Axiom 2.[7]

Axiom 2 *(satisficing): For all i, t, and action α_i chosen by player i in period t, if $\pi_{i,t} \geq a_{i,t}$ then player i will choose action α_i in period $t+1$.*

of both theorems must be modified in the following way: "...any outcome o in which $\pi_i(o) \geq \underline{a}_i$ can be sustained as a stable outcome." Intuitively, the condition that each player's payoff is at least as high as her aspiration level is more-or-less the adaptive analogue to the requirement in the full-rationality folk theorems of game theory that every person get at least his maxmin payoff. An optimizing player will never settle for getting less than her maxmin; analogously, an adaptively rational player will not be satisfied with something less than her aspiration level. As long as these criteria are satisfied, anything is stable.

[7] Note that just as Theorem 1 does not need any assumption about how agents respond to negative feedback, so Theorem 2 only needs an assumption about what happens when payoffs are satisfactory. It is not tied to a specific model of search.

It is important to note that Axiom 2 is strictly stronger than Axiom 1: if the decision maker satisfices then with probability one he retains the action he used in t. Because his propensity to choose that action in period $t+1$ is 1, it cannot be less than it was in t. This fact makes it easy to show that Theorem 2—which is identical to Theorem 1 except that Axiom 2 replaces Axiom 1—holds. (The proof is therefore omitted.)

Theorem 2 *Consider any repeated game, with either stationary or nonstationary payoffs, in which players satisfice-and-search by any arbitrary mix of adaptive rules that satisfy Axiom 2 and where aspirations are exogenously fixed. Then any outcome o can be sustained as a stable outcome by some pure SRE.*

Endogenous Aspirations

A natural criticism of positing fixed aspirations is that that premise precludes an important kind of learning: aspirations should reflect one's payoff-experience. Indeed, to assume otherwise—to keep aspirations constant in the face of discrepant evidence—seems inconsistent with the spirit of the underlying research program: people are boundedly rational but they do adapt (learn from experience). Thus endogenizing aspirations should be a major part of the modeling effort for substantive reasons. Surprisingly, however, in an important class of circumstances (games), endogenizing aspirations *does not* by itself solve the methodological problem of low predictive power. That is, if we confine ourselves to the context that is standard in game theory and common in agent-based models—deterministic and stationary payoffs—it remains true that "anything can happen" even when aspirations adjust to experience.[8] This is the content of our next results: Theorems 3 and 4.

Since aspirations are endogenous in these results, we must specify some properties of adjustment. It turns out that we only need one assumption, Axiom 3, below.

Axiom 3 *For all i and t, if $\pi_{i,t} = a_{i,t}$ then $a_{i,t+1} = a_{i,t}$.*

Note that Axiom 3 is consistent with any model of aspiration-adjustment in which an agent's aspiration level tomorrow is a weighted average of today's

[8] It turns out that the *combination* of endogenous aspirations and stochastic payoffs is one solution to the problem of low predictive power (see Theorem 5). We postpone a discussion of this issue to the next section.

level and today's payoff: i.e., if $a_{i,t+1} = \lambda a_{i,t} + (1 - \lambda)\pi_{i,t}$, where $\lambda \in [0, 1]$, then $a_{i,t+1} = a_{i,t}$ whenever $\pi_{i,t} = a_{i,t}$, for any value of λ.

We first consider models of reinforcement learning (with endogenous aspirations).

Theorem 3 *Consider any repeated game with deterministic and stationary payoffs in which players adjust their action-propensities by any arbitrary mix of adaptive rules that satisfy Axiom 1 and adjust their aspirations by any arbitrary mix of rules that satisfy Axiom 3. Then any outcome of the stage game can be sustained as a stable outcome by some pure SRE.*

The proof of Theorem 3 (see the appendix) is similar to that of Theorem 1. The only twist arises because here aspirations adjust to experience, and so one must be more careful in specifying a vector of aspirations that will self-replicate.

As with learning theories, we can now make aspirations in behavioral theories of search endogenous. Here, we use Axiom 2 just as before. (Theorem 4 also inherits Theorem 3's assumptions on payoffs: they must be deterministic and stationary.) We then get the following result. (The proof closely follows that of Theorem 3 and so is omitted.)

Theorem 4 *Consider any repeated game with deterministic and stationary payoffs in which players satisfice-and-search by any arbitrary mix of adaptive rules that satisfy Axiom 2 and adjust their aspirations by any arbitrary mix of rules that satisfy Axiom 3. Then any outcome of the stage game can be sustained as a stable outcome by some pure SRE.*

Although the domains of Theorems 3 and 4 are smaller than those of 1 and 2 in being confined to games with deterministic and stationary payoffs, it is worth noting that their domains remain large in several important respects. First, they hold for any number of players, including one-person decision problems. Second, each player can have any number of actions, finite or infinite. Third, the game can be symmetric or asymmetric. Hence the players may, for example, have completely different sets of actions. Fourth, the results do not even require that players must keep using the same adaptive rule over time: a person could switch to different methods of adjusting his/her action-propensities or aspirations, provided only that new rules continued to satisfy the relevant axioms (i.e., Axioms 1 or 2 and Axiom 3).

So *both* game theory and behavioral theories of learning and of search suffer from similar defects of weak empirical content. Indeed, in some ways the adaptive theories are in worse shape, qualitatively speaking, than the full-rationality theory: *any* outcome of the stage game can be sustained as an SRE, even those below the maximin level. Hence in terms of predicting the range of possible outcomes, the empirical content of the adaptive theories is even less than game theory's.[9] For example, in the iterated PD neither adaptive theory can exclude even the extreme outcome in which one person is always suckered by his partner, whereas noncooperative game theory excludes this case because in such circumstances the exploited player could always unilaterally guarantee his security level by defecting.

The key to overcoming this methodological problem is to ensure that the adaptive process does not "lock-in" too easily. That is, the theorist must "tie his/her own hands." The reward for this self-control is a model with much more empirical content. We turn to this idea next.

3 The Individual Solution – Randomness

In this part of the paper we examine a diverse set of assumptions that ensure that a model of aspiration-based adaptation has empirical content, i.e., it does not predict that "anything can happen." In fact, these changes ensure that the model generates *unique* predictions. The key modification is to introduce randomness at the level of individual decision making and adaptation. Of course, the model already *permits* probabilistic decision-making (the propensity to play an action need not equal one or zero). However, now we *require* individual decisions to have a random component.

[9] If, of course, we could independently measure (observe) aspiration levels, then even deterministic models of adaptation could sometimes make sharp, falsifiable predictions. Consider, for example, Macy and Flache's (2002) analysis of the prisoner's dilemma. If in an experiment we could *induce* (fixed) aspirations in the (P, R) interval (where P is the payoff to mutual defection and R is the payoff for mutual cooperation), then their prediction that mutual cooperation is the only stable outcome would be testable. What renders the theory hard to falsify is the testing of a *joint* hypothesis that includes both the adaptive processes and the agent's aspiration level. Theorems 1 and 2 imply that for *any* observation of any kind of stable behavior, there always exists a joint hypothesis that cannot be rejected. This creates a temptation for the analyst to use *ad hoc* maneuvers ("Ah ha! So the agent must have had an aspiration level of such-and-such.") that are methodologically suspect.

This modification ensures that reinforcement learning can be modeled as a Markov chain. Adding randomness thus requires the use of a different solution concept of our model. Rather than solving for SREs we now need to characterize the process's long-run behavior. That is, the appropriate solution concept is now a *stable distribution*. As we show below, our stochastic process has a *unique* stable distribution. Moreover, the process must eventually converge to that unique distribution from any starting point, i.e., from any initial configuration of agents' propensities and aspirations. Thus, the process yields a unique, albeit stochastic, prediction.[10]

It is important to understand the nature of this claim. We do *not* mean that the players' behavior necessarily settles down in the long run (e.g. to cooperation). This would happen only if their propensities settled down to the corresponding pure values (e.g., all pairs of players reached a cooperation-propensity of one). This need not happen, even for arbitrarily rare trembles. Instead, consider a large number of identical games. The state space can then be thought of as a finite "grid" of propensity and aspiration values. Then in the long run, the frequency distribution over this grid will settle down to a probability distribution. In the simple case where each player eventually plays only the same pure strategy (e.g., mutual cooperation in the Prisoner's Dilemma), this distribution is degenerate. As we shall see shortly, however,

[10] It is possible to relate the idea of a distribution as a solution concept to the perhaps more familiar idea of an equilibrium as a solution of the model, whether the equilibrium is Nash or Self-Reinforcing. This can be done by considering the limit of sequences of probabilistic distributions. Suppose we start out with any (benchmark) tremble probability (i.e., with some exogenously fixed probability agents randomly do something other than they had intended) and then gradually reduce it toward zero, holding all other parameters constant. This yields a *sequence* of (unique) limiting distributions and their associated statistics (e.g., the population's average propensity to cooperate). As the tremble probabilities get sufficiently small, by continuity, further diminutions in these probabilities can have only negligible effects on the associated limiting distributions. In short, as the trembles go to zero the limiting distributions themselves converge. In the limit, we are left with a distribution that assigns non-zero probability only to finitely many states (usually a unique state). Note, however, that when the tremble probability is *exactly* zero (not arbitrarily small, in the limit), the corresponding learning rule would be subject to Theorems 3 and 4. That is, it would lack empirical content. So, we require some randomness to ensure a unique prediction, but the amount of randomness can be arbitrarily small. See Foster and Young (1990) or Bendor, Diermeier and Ting (2003) for details.

there may be cases in which the frequency distribution of players over states is not degenerate. Note that the notion of a limiting distribution does *not* imply that in any particular game the players' actions settle down on one point of the state space grid. Rather, while the specific action profiles taken may change indefinitely, the *population of players' action profiles* will settle down.[11]

We now need to define a stochastic version of reinforcement learning. Since we are interested in proving general existence and uniqueness results the model needs to be defined in a fairly general fashion. Specifically, we construct a stationary Markov chain that describes the behavior of N ($N \geq 1$ and finite) players over time. Each agent, i, has finitely many feasible actions in the stage game, $\alpha_1^i, \ldots, \alpha_{m(i)}^i$, where $m(i) > 1$ for all i. (The set and number of actions could be different for each agent; hence the dependence on i is necessary. But for the sake of economy of expression we shall virtually always suppress the relevant subscripts) We assume throughout that each agent has finitely many possible *propensity values* for each action, where each propensity value is non-negative. Thus, in every period each agent i has a finitely many vectors of propensity values over her possible actions of the form: $(p_{i,t}(\alpha_1^i), \ldots, p_{i,t}(\alpha_{m(i)}^i))$, where $\sum_{j=1}^{m(i)} p_{i,t}(\alpha_j^i) = 1$. So, each vector constitutes a probability measure.

Agents also have aspiration levels that partition current payoffs into *satisfactory* and *unsatisfactory*. Each agent has finitely many feasible aspiration levels. Agent i's aspiration levels are denoted by $a_1^i, \ldots, a_{n(i)}^i$, where $n(i) > 1$.

An agent's sets of feasible propensity vectors and feasible aspiration levels are the same in every period. Let $\underline{p}_i$ denote i's minimal propensity value for any action, and let $\overline{p}_i$ denote i's maximal propensity value. We assume that for every agent i feasible propensity values satisfy the following condition: $\overline{p}_i + (m(i) - 1)\underline{p}_i = 1$, where $m(i)$ is the number of actions that agent i has in the stage game. This condition must hold if $\underline{p}_i = 0$ and $\overline{p}_i = 1$. Note, however that we do not require (but in general permit) that $\underline{p}_i = 0$ and $\overline{p}_i = 1$. That is, we allow for adaptive models where propensities are never "pure." Indeed, as part (2) of Theorem 5 shows, this is one of the ways in which we can ensure a unique limiting distribution.

Thus the state of the Markov process is a vector of N elements, where each

[11] This misunderstanding—believing that the behavior of a single sample path will settle down as time goes to infinity—is so common that in his seminal work on stochastic processes Feller (1950) devoted a good part of one chapter to dispelling it.

element (for, say, agent i) is composed of i's vector of current propensities over her feasible actions and her current aspiration level. Because there are finitely many actions, finitely many possible propensities over these actions, finitely many aspirations, and finitely many agents, this is a finite state Markov chain.

To help visualize the state space, one might think about a game in which each person has only two actions. In this case the state space for every player is a two dimensional finite grid, where the horizontal axis is the current propensity for playing action 1 and the vertical axis is the player's current aspiration level; see figure 1. (The current propensity for action 2 is just the complement of action 1's propensity, so it can be omitted.) Hence, for $N-$player games the state space consists of N such grids.

[figure 1 about here]

Note that the model allows for two forms of randomness. First, as in the base-line model discussed in the previous section, randomness may be present because agents use propensities strictly between 0 and 1. This randomness is captured by the propensity levels in each grid. We will call a vector of propensities *totally mixed* if every propensity-value in the vector exceeds zero. Second, learning and adapting corresponds to the transitions from one part of the grid to another. The system's transition rules describe how agents adjust their propensities and aspirations, based on what they have done in t and their current payoffs. These transition rules themselves may be probabilistic. Indeed, to exhibit a unique limiting distribution these transitions must be structured such that we do not get "stuck" easily on a particular point on the grid. Intuitively, this corresponds to a randomly perturbed learning process. We now discuss some of the features of such processes.

Inertia. In our model we wish to capture agents that learn by trial-and-error, i.e., propensities and aspirations may adjust to payoff experience. However, since an actor's attention may be on other matters, these codings do not invariably lead to adjustments in propensities or aspirations. Consistent with the spirit of bounded rationality, we allow for the possibility that humans are sometimes inertial: they do not invariably adapt or learn. That is, we assume that each agent may be inertial with respect to either adjustment mechanism with a probability that is in $(0, 1)$. If the agent is not inertial with respect to an adjustment mechanism, then we say that s/he is *alert* with respect to that mechanism. The probabilities of being inertial with respect to propensity-adjustment and aspiration-adjustment are assumed to be i.i.d. across players and across periods, but they need not be independent. It is only required that all four possibilities—(active, active), (active, inertial), (inertial, active), and (inertial, inertial)—occur with positive probability.

Assumptions about propensity-adjustment. The following assumptions about the two adjustment mechanisms hold whenever the agent in question is alert.

A4 (positive feedback): If i used action α^i in t and if $\pi_{i,t} \geq a_{i,t}$ then $\Pr(p_{i,t+1}(\alpha^i) \geq p_{i,t}(\alpha^i)) = 1$; if $p_{i,t}(\alpha^i)) < \overline{p}_i$ and $\pi_{i,t} > a_{i,t}$ then $\Pr(p_{i,t+1}(\alpha^i) > p_{i,t}(\alpha^i)) = 1$.

A5 (negative feedback – direct effect): If i used action α^i in t and if $\pi_{i,t} < a_{i,t}$ then $\Pr(p_{i,t+1}(\alpha^i) \leq p_{i,t}(\alpha^i)) = 1$; if $p_{i,t}(\alpha^i) > \underline{p}_i$ then $\Pr(p_{i,t+1}(\alpha^i) < p_{i,t}(\alpha^i)) = 1$.

A6 (negative feedback – indirect effect): If i used action α^i_r in t and if $\pi_{i,t} < a_{i,t}$, then for every other action α^i_s (where $s \neq r$), with positive probability i moves to some new propensity vector in $t+1$ in which α^i_s has positive weight.

Note that A4 is stronger than A1; i.e., A4 implies A1 but the converse does not hold. Intuitively, it requires reinforcement to be strictly positive, if possible. A5 is the corresponding axiom on negative feedback. In conjunction they are simply a probabilistic version of the Law of Effect. A6 additionally requires that no action is *a priori* excluded. Rather, each other action must be reachable with some (possibly arbitrarily small) probability. Note that A6 does *not* require that there be any new propensity vector that i moves to in $t+1$ in which *all* actions (other than the one used in t) receive positive weight. Instead, there could be a set of propensity vectors, one in which α_1 gets positive weight, another in which α_2 does, and so on. Note also that in the case where an agent has only two actions A6 is already implied by A5.

Assumptions about aspiration-adjustment. We continue to use A3, as defined above, to stipulate what happens when current payoffs equal current aspirations. For the reader's convenience it is restated here. The new aspiration-adjustment assumptions, A7 and A8, stipulate what happens when current payoffs do *not* equal current aspirations. Again, both axioms are simply probabilistic versions of endogenous aspirations.

A3: If $\pi_{i,t} = a_{i,t}$ then $\Pr(a_{i,t+1} = a_{i,t}) = 1$.

A7: If $\pi_{i,t} > a_{i,t}$ then $\Pr(a_{i,t} < a_{i,t+1} \leq \pi_{i,t}) = 1$.

A8: If $\pi_{i,t} < a_{i,t}$ then $\Pr(\pi_{i,t} \leq a_{i,t+1} < a_{i,t}) = 1$.

Because these adjustment properties are stationary (time homogeneous), the resulting transition probabilities of the Markov chain are also stationary. We refer to any process that satisfies these axioms as an "aspiration-based adaptive process."

Stochastic Payoffs A third potential source of randomness may originate from stochastic payoffs, i.e. the assumption that the payoff to a player is not completely determined by the choices of all players, but also has a random component.[12] That is, payoffs are modeled as a non-degenerate (conditional) probability distribution with finite support for each action profile. For each action profile outcome we denote realized payoffs by $\pi_{i,t}(o)$ with corresponding random variables $\Pi_{i,t}$. Let $\underline{\pi}_i(o)$ denote agent i's minimal possible payoff given outcome o, and $\overline{\pi}_i(o)$ her maximal payoff. For example, in the two-person prisoner's dilemma $\pi_{i,t}(C, C)$ denotes agent i's payoff at time t given that both agents have cooperated.[13] We assume that payoff realizations are mutually independent across agents and time.

Different payoff assumptions then correspond to different restrictions on the respective distribution, such as assumptions on the ordering of expectations or the supports of the random variables. These restrictions can be applied to different aspects of the distribution. For example, one may require that each agent's expected payoff from mutual cooperation in the 2-person PD is strictly higher than the expected payoff from mutual defection. Formally, for all t,

$$E[\Pi_{i,t}(C, C)] > E[\Pi_{i,t}(D, D)]$$

Alternatively, one may assume that distributions are ordered in terms of their best or worst possible realizations. For example, one may assume that

[12] This approach is sometimes called a "random utility model". It is common in empirical studies of decision making, e.g. consumer decisions, and can provide a decision-theoretic foundation for logit regressions. See e.g. McFadden (1973). The approach is less well-known in game-theoretic approaches.

[13] In this example we simply identify strategy combinations with outcomes. That is, Ω is the identity function. In general, it simplifies the analysis to use general outcome functions.

each agent's *maximal* payoff from mutual cooperation in the 2-person PD is strictly higher than the maximal payoff from mutual defection:

$$\overline{\pi}_i^{\max}(C, C) > \overline{\pi}_i(D, D).$$

Of course, which one of these assumptions makes sense depends on the phenomenon being modeled.

Theorem 5 *An aspiration-based adaptive process has a unique limiting distribution if any of the following conditions hold:*

1. ***Action trembles****: with a positive probability (which is i.i.d. across periods and independent across players), player i, instead of doing what he intended to do, "experiments" by randomly playing some action given by a totally mixed vector of probabilities over feasible actions. (This vector is i.i.d across periods and independent across players.) Further, in the stage game there is an outcome, o, in which nobody gets their minimal payoff (i.e. $\pi_i(o) > \underline{\pi}_i$ for all i).*

2. ***Extreme propensities excluded****: neither 0 nor 1 are feasible propensity values for any action for any player. Further, in the stage game there is an outcome in which nobody gets their minimal payoff (i.e., ($\pi_i(o) > \underline{\pi}_i$ for all i).*

3. ***Stochastic payoffs****: every vector of actions produces a (nondegenerate) distribution of payoffs for every agent, where each distribution is finitely valued. Payoffs are i.i.d. across periods and independently distributed across players.*

4. ***State trembles****: with positive probability (again i.i.d. over periods and independently across players) i's state can randomly tremble to any neighboring state on his grid.*

Thus various forms of randomness in the transition from one state to another ensure the existence of a unique limiting distribution. What all these approaches have in common is that they act on the level of individual actions or learning. They do not involve any reference to other actors. The next section examines such a "social" solution to the problem of low predictive content.

4 The Social Solution – Reference Groups

In this section we introduce a quite different theoretical change in models of adaptation, one which has direct sociological content: how reference groups affect the adjustment of aspirations (e.g. Merton and Rossi 1950). It turns out that assuming that people are tied into social networks which influence their aspirations has benign methodological effects, quite similar to the effects of introducing randomness at the individual level. Hence, in order to demonstrate cleanly that it is the introduction of these social comparison processes that restores empirical content to the model, in this section we return to the base-line deterministic model from section 2 with stationary stage-game payoffs. Accordingly we know, via Theorems 3 and 4, that absent social influence on aspirations "anything could happen."

In this model, agent i's aspiration in period $t+1$ will be a weighted average of the payoffs of some of the other players in the game. We call such people i's *reference group*, and we say that i *emulates* his reference group. (We continue to assume that i's aspiration in $t+1$ is influenced by $a_{i,t}$ and $\pi_{i,t}$, i.e., by his current aspiration and current payoff.) Thus we write

$$a_{i,t+1} = \lambda_{i,0} a_{i,t} + (\lambda_{i,1}\pi_{1,t} + \cdots + \lambda_{i,N}\pi_{N,t}) \tag{1}$$

where for all i, $\sum_{j=1}^{N} \lambda_{i,j} = 1$ and $\lambda_{i,j} \geq 0$. For the sake of continuity with the model underlying Theorem 2, we assume that $\lambda_{i,0}$ and $\lambda_{i,i}$ are strictly positive for all i. We further assume that all $\lambda_{i,j}$'s are exogenously fixed for each player, over the entire game. (We plan to investigate systems with endogenous reference groups in subsequent work.) Note that if for all i, $\lambda_{i,j} = 0$ (for all $0 \neq j \neq i$), then we recover an individualistic process of aspiration adjustment that is consistent with property (A3).

There are many different kinds of Theorems that one can generate via models in which social comparisons influence aspirations. Here we examine a very broad class of such issues, by looking at all N-person games that satisfy two properties: (1) N is finite and (2) each player's action set is finite. (Again we do *not* require that the game be symmetric.) Because computers are finite state machines, all computational models that are actually run on computers presume finitely many players and finite action sets, and so satisfy assumptions (1) and (2). Thus, these conditions are not very restrictive.

The crucial property used in our next result is that the social networks of emulation be dense enough so that "no man (or group) is an island," i.e., no subset of players completely ignores everyone else's payoffs. We now define this property precisely.

Definition 2 *N is* ***nondecomposable*** *if, for every partition of N into two disjoint, nonempty subsets, at least one person in each subset emulates someone in the other subset.*

Formally, N is ***nondecomposable*** *if for every J with $J \subsetneqq N$ and $J \neq \emptyset$:*
there exists some $j \in J$ and $i \in N \backslash J$ such that: $\lambda_{j,i} \neq 0$ and
there exists some $i' \in N \backslash J$ and $j' \in J$ such that: $\lambda_{i',j'} \neq 0$.

Theorem 6 *Consider any N-person repeated game with deterministic payoffs in which N is finite and in which each person's action set is finite. Further, axiom A4 governs propensity-adjustment and (1) governs aspiration-adjustment. If N is nondecomposable, then in any stable stage-game outcome all players get the same payoff.*

Thus when the set of players is nondecomposable, it is impossible to sustain, as a stable pattern, outcomes in which some people do worse than others. Hence, rather than saying that "anything can happen," a model of adaptation with social comparisons predicts that many things *cannot* happen.[14]

With just one more assumption about the nature of the stage-game, this model also makes a sharp prediction about what combinations of actions are stable.

(A9): If two players take different actions then they get different payoffs.[15]

This property holds in many of the games studied in agent-based models, e.g. in the study of collective action. Obviously A9 holds in virtually all variants of the prisoner's dilemma. If, e.g., a player's payoff is monotonically

[14] Theorem 6 describes a property—equal payoffs—that is *necessary* for stability. It is easy to show that this property is also *sufficient* for stability (i.e., for an outcome to be supported by a pure SRE). Putting necessity and sufficiency together yields a characterization result: under the hypotheses of Theorem 6, a stage-game outcome is stable *if and only if* all payoffs in that outcome are equal.

[15] More formally, if $\alpha_{i,t}$ denotes player i's action in period t, then A8 says that for all $i \neq j$, if $\alpha_{i,t} \neq \alpha_{j,t}$ then $\pi_{i,t} \neq \pi_{j,t}$.

falling in his/her degree of cooperation, then whoever cooperates more in a given period must get a lower payoff than those who cooperate less. A bit more subtly, it holds in most "threshold" games (e.g., Palfrey and Rosenthal 1984, Macy 1990, 1991b) as well. Consider, for example, a standard N-person threshold game where each person has a binary choice of either contributing or shirking. Suppose in period t player i contributed while j shirked. If i is pivotal—without his/her contribution the group would fall short of the required number of contributions—then i prefers contributing over shirking. Thus unlike the PD defecting is *not* a dominant strategy in this threshold game. Rather there are multiple Nash equilibria, one where nobody cooperates and others where some or all players contribute. Suppose player j took a different action in t than i, e.g. j did not contribute, then, because the collective good was provided yet j "free-rode" on other people's contributions, j must have gotten a higher payoff than i did. Therefore (obviously) i and j got different payoffs, so A9 holds in this kind of game.

Corollary 1 *Suppose the hypotheses of Theorem 6 are satisfied, and in addition A9 holds. Then in any stable stage-game outcome everyone takes the same action (and receives the same payoff).*

To see some of what this result implies, reconsider the binary choice, N-person threshold game described above. Suppose that all the players belong to a village that is trying to sustain a collective good (e.g., maintaining a commons; see Ostrom 1991). The process is governed by adaptation with endogenous aspirations with reference groups. Further, the community networks are sufficiently dense so that the community is nondecomposable. Then there are *only two* stable outcomes: either everyone contributes to the collective good or no one does.[16] All the other stage game outcomes that are stable under the purely individualistic adjustment processes of Theorem 2 (one person contributes, two contribute, . . . , $N - 1$ contribute) are destabilized by social comparison processes.

Theorem 6 and its corollary also generate some interesting implications about social hierarchies. Suppose in the preceding threshold game that a

[16] One is reminded of Putnam's description (1993) of the good equilibrium in northern Italy, with its high level of collective goods, and the bad equilibrium in southern Italy where people contribute to public goods to much lesser extent.

community *is* decomposable: it can be partitioned into two distinct subgroups, A and B, such that nobody in A emulates anyone in B or vice versa. Then exploitation of one group by the other is a stable outcome: all A's contribute (suppose that this meets the threshold criterion) while all B's shirk. Thus, although everyone enjoys the collective good, only the downtrodden A's bear the burden of providing the good. This outcome is supported by an SRE in which the endogenous (but very different) aspirations of A's and B's are consistent with these unequal payoffs. The key is disjoint reference groups: people emulate only members of their own group.

It is a plausible hypothesis that elites throughout history have tried to keep the aspirations of the lower classes in check and consistent with their station in life. Theorem 6 and its corollary provide a micro-foundation (of adaptive behavior) for this hypothesis.

These two solutions to the empirical content problem of adaptive models seem quite different but they actually work via common mechanisms. To see why this is so, let us compare the noise created by stochastic payoffs with socially based aspirations. Consider the standard two-person, binary-choice prisoner's dilemma. Without noise or social comparison we can stabilize the extreme outcome in which Row is always exploited by Column; hence Row gets S, the lowest ("sucker's") payoff, and Column gets T, the highest ("temptation") payoff in every period. Now let us consider stochastic payoffs. Specifically, there are now two payoff realizations for every pair of actions ("high" and "low"), so that Row can get either S^- or S^+ and Column can get either T^- or T^+. Focus on Row. Suppose she is exploited and in a series of periods gets S^+. By assumption (A6) eventually her aspiration level must exceed S^-. Now suppose that she is exploited and gets the lower payoff, S^-. Because this is below her aspiration level she is dissatisfied, and by A5 her propensity to cooperate must fall. Thus, *negative feedback* induces her to experiment with defecting in the next period.[17]

Now consider a deterministic PD in which both players emulate each other. Again, suppose that we try to stabilize the asymmetric outcome in which Column always exploits Row. But if that outcome were stable then Row's steady state aspiration would be a weighted average of S *and* T. Hence it must exceed S. Consequently, getting exploited would be dissatisfying for Row. Thus she would once again experience negative feedback when

[17] As Theorem 5 demonstrates this insight generalizes to other forms of randomness.

cooperating—*just as she did in the case of stochastic payoffs.* And as before, this negative feedback would induce her to experiment with another action.

We see, then, that social comparisons provide *vicarious experience* which can lead players who are getting the short end of the stick to become dissatisfied. A similar process occurs, individualistically, when payoffs are random. Thus although the two approaches appear at first glance to be quite different, fundamentally they are closely related.

5 Conclusions

This paper has shown that an important class of behavioral models of adaptation has little empirical content: models in this class imply that virtually "anything can happen." However, we have also demonstrated that two approaches can restore predictive content: one can either introduce enough noise so that agents will be shaken out of arbitrary patterns of behavior, or one can keep the models deterministic but make aspirations depend on social comparisons. The key common feature of both solutions lies in the fact that actors are exposed to experiences other than those present in the current equilibrium. This leads to feedback that destabilizes unsustainable equilibria and restores empirical content to the model. Indeed, in the case of stochastic learning, the model makes a unique probabilistic prediction.

Given the high predictive power of (appropriately specified) adaptive models one may then want to investigate the relation between adaptive and rational-choice based game-theoretic approaches. As shown in related work (Bendor, Diermeier and Ting n.d., 2003) adaptive models should not be understood as complements of classical game theory. That is, the limiting distributions of e.g. reinforcement learning models *do not* in general put high probability on Nash equilibria. For example, they may predict a high probability of cooperation in the prisoners' dilemma (Bendor, Diermeier and Ting 2003) or high turnout in the turnout paradox in the study of voting (Bendor, Diermeier and Ting n.d.). That is, adaptive models offer a conceptually and empirically distinct alternative to the classical, rational-choice based approach. This suggests that one may use reinforcement learning models to solve persistent anomalies confronting game theoretic models.

Appendix

Proof of Theorem 1

Pick an arbitrary outcome, o, of the stage game and some period t. To show that o is stable we construct the following SRE:

$$\text{For all } i\text{: let } p_{i,t}(\alpha_i(o)) = 1 \text{ and } a_i = \underline{\pi}_i.$$

Then since $\pi_{i,t}(o) \geq \underline{\pi}_i = a_i$, Axiom 1 implies that $p_{i,t+1}(\alpha_i(o)) \geq p_{i,t}(\alpha_i(o))$. But since $p_{i,t}(\alpha_i(o)) = 1$, we must also have $p_{i,t+1}(\alpha_i(o)) = 1$. Hence $p_{i,t+1}(\alpha_i(o)) = p_{i,t}(\alpha_i(o))$. QED.

Proof of Theorem 2

Suppose payoffs are fixed: i.e. $\pi_{i,t}(o) = \pi_i(o)$ for all t. Pick an arbitrary outcome, o, of the stage game and some period t. To show that o is stable we construct the following SRE:

$$\text{For all } i\text{: let } p_{i,t}(\alpha_i(o)) = 1 \text{ and } a_{i,t} = \pi_i(o).$$

Hence outcome o is attained in period t for sure and (given fixed payoffs) we have $\pi_i(o)$ for each player i. Since $a_{i,t} = \pi_i(o)$, by Axiom (2) we have $a_{i,t+1} = a_{i,t} = \pi_i(o)$. This satisfies condition (ii) for an SRE. Moreover, $\pi_i(o) \geq a_{i,t}$, we must have (as in the proof of Theorem 1) that $p_{i,t+1}(\alpha_i(o)) = p_{i,t}(\alpha_i(o))$. QED.

Proof of Theorem 3

Below we follow the convention of using "with positive probability" (abbreviated by 'wpp') as a shorter way of saying "with strictly positive probability."

A standard result in the theory of finite Markov chains is that if a stationary (finite) chain is irreducible and aperiodic then it has a unique limiting distribution (Feller 1950, p.393-394). Given the assumptions described above, our Markov chain must be aperiodic, since agents are inertial with positive probability at every state. Hence it only remains to show that it is irreducible for every case (i.e., for parts (1)-(4) of the Theorem). To show this, the following lemma is very useful.

Lemma 1 *Any finite state Markov chain is irreducible if it has a state which is accessible from all states.*

Proof: Call such a state s^*. Since s^* is accessible from all states it must belong to every closed set of states. (A set of states C is "closed" if once the process enters C then it must stay there forever.) By Theorem 3 of Feller (vol. 1, chapter 15, p. 392), any Markov chain can be partitioned, in a unique way, into nonoverlapping sets $T, C_1, C_2, \ldots$, where T is composed of all transient states and the C_k are closed sets. Since these closed sets are disjoint and s^* belongs to each of them, our chain must have only one closed set, and a finite chain with a unique closed set is irreducible. QED.

We now return to the proof of the main result. (For parts (1), (2) and (4) these proofs apply to payoffs with degenerate distributions, for a given vector of actions. Extending them to cover nondegenerate distributions is straightforward.) Parts (1)-(3) exploit the lemma by identifying a state s^*, which we will also call a *distinguished state.* This, together with the aperiodicity provided by inertia, ensures existence of a limiting distribution.

Proof of (1) and (2)

By assumption there is at least one outcome, say o^*, in which everyone's payoff strictly exceeds their minimal payoff, i.e. $\pi_i(o^*) > \underline{\pi}_i$ for all i. Now consider the following state of the Markov process:

Associated with o^* is a state in the Markov process, s^*, in which all players put maximal propensity on the action corresponding to o^*, and have an aspiration equal to the payoff they get from o^*. That is, $s^* := (p^*_{i,t}; a^*_{i,t})_i$ such that for all i and some t: $p^*_{i,t}(\alpha_i(o^*)) = \overline{p}_i$ and $a^*_{i,t} = \pi_i(o^*)$. We nominate s^* as a distinguished state, in the sense of lemma 1. Hence we must show that given any arbitrary starting state, the process can reach s^* wpp.

The assumption in both parts (1) and (2) ensure that each agent in every period can play any action wpp, and hence also every outcome can occur wpp in every period. Hence, every finite string of outcomes can occur wpp.

Consider an arbitrary starting state s. Now construct an arbitrarily long, but finite string of outcomes in which player 1 gets his minimal payoff $\underline{\pi}_1$. Such a string must occur wpp. For some finite t we then must have $a_{1,t} = \underline{\pi}_1$. To see why such a state must exits, note that because of A6 and A7 $|\underline{\pi}_1 - a_{1,t}|$ is decreasing in t. Equality is ensured by the fact that there are only finitely many aspiration levels and that there exists an aspiration level for each individual payoff level. Call such a state s_1. Now consider a second string of outcomes, commencing at s_1, where player 2 receives $\underline{\pi}_2$ and where player 1 is inert with respect to his aspiration level. Since the event where player 1 is inert occurs wpp and is independent of the updating process of other players there must exist some t where $a_{2,t} = \underline{\pi}_2$ and (by inertia) $a_{1,t} = \underline{\pi}_1$. Repeat this procedure for all n players until we reach state s_N at some t where for all $i : a_{i,t} = \underline{\pi}_i$.

Next, consider an outcome o^* and state transition where all i are inertial with respect to their aspiration level, i.e. where we continue to have $a_{i,t} = \underline{\pi}_i$ for all i. Such a state must be reached wpp. Then, we have $\pi_i(o^*) > \underline{\pi}_i = a_{i,t}$ for all i. Hence, by A4, $p_{i,t+1}(\alpha_i(o^*)) > p_{i,t}(\alpha_i(o^*))$. Repeat this outcome until for all i $p_{i,t}(\alpha_i(o^*)) = \overline{p}_i$ while maintaining $a_{i,t} = \underline{\pi}_i$ for all i. Then apply an analogous process for each i's aspiration level. That is, consider a sequence of states with outcome o^* where agents are inert with respect to their propensity; i.e. each agent's propensity is frozen at $p_{i,t}(\alpha_i(o^*)) = \overline{p}_i$. Since for all i, $\pi_i(o^*) > a_{i,t}$, (A6) implies that $\underline{\pi}_i < a_{i,t+1} \leq \pi_i(o^*)$. By an analogous argument, consider a finite sequence of states until for some t we have $a_{i,t} = \pi_i(o^*)$ for all i. But this is exactly the distinguished state, s^*. QED.

Proof of (3)

Here we will show that the state in which everyone puts maximal propensity on some action α_i^* and in which everyone's aspiration is their minimal possible payoff from the outcome produced by $(\alpha_1^*, \alpha_2^*, \ldots, \alpha_n^*)$ is a distinguished state. That is, there exists a profile of actions $(\alpha_1^*, \alpha_2^*, \ldots, \alpha_n^*)$ with $o^* := \Omega(\alpha_1^*, \alpha_2^*, \ldots, \alpha_n^*)$ such that $s^* := (p_{i,t}^*; a_{i,t}^*)_i^t$ where for all i and t we have $p_{i,t}^*(\alpha_i^*(o^*)) = \overline{p}_i$ and $a_{i,t}^* = \underline{\pi}_i(o^*)$.

From any state s in period t_0, consider some agent i and action α_i.

Case (a): Suppose $p_{i,t_0}(\alpha_i) > 0$. Then, construct an arbitrarily long but finite string of states where in each state (i) every agent's propensity is frozen, in particular i's propensity is constant at $p_{i,t_0}(\alpha_i)$ (this occurs wpp by inertia), (ii) player i's realized action is α_i (this must occur wpp by (i)), (iii) the realized outcome is some fixed o (this must occur wpp by (i)), (iv) agent i's realized payoff is minimal, i.e. $\underline{\pi}_i(o)$ (this occurs wpp given random payoffs). Such a string must occur wpp. For some finite t_1 we then must have $a_{i,t_1} = \underline{\pi}_i(o)$. Then construct a second string of states with (i) every agent j's (including i's) aspiration level frozen at $a_{j,t_1}(\alpha_i)$ (again this occurs wpp by inertia), (ii) agent i's realized payoff is maximal, i.e. $\overline{\pi}_i(o)$; since $\overline{\pi}_i(o) > a_{i,t} = \underline{\pi}_i(o)$ for all $t > t_1$, $p_{i,t}(\alpha_i)$ is strictly increasing until $\overline{p}_{i,t}(\alpha_i)$

(by A3), (iii) the realized outcome is some fixed o (this must occur wpp by (ii)), and (iv) player i's realized action is α_i (this must occur wpp by (ii)). Thus at some t_2 we must have $p_{i,t_2}(\alpha_i(o)) = \overline{p}_i$ and $a_{i,t_2} = \underline{\pi}_i(o)$. These are the desired α_i and o^*.

Case (b): Suppose instead that $p_{i,t_0}(\alpha_i) = 0$. Then for some other action α_i' we must have $p_{i,t_0}(\alpha_i') > 0$. Now construct an arbitrarily long but finite string of states where in each state (i) every agent j's (including i's) propensity is frozen; in particular i's propensity is frozen at $p_{i,t_0}(\alpha_i')$ (this occurs wpp by inertia), (ii) player i's realized action is α_i' (this must occur wpp by (i)), (iii) the realized outcome is some fixed o' (this must occur wpp by (i)), (iv) agent i's realized payoff is maximal, i.e. $\overline{\pi}_i(o')$ (this occurs wpp given random payoffs). Such a string must occur wpp. For some finite t_1 we then must have $a_{i,t_1} = \overline{\pi}_i(o')$. Now consider a state at $t_1 + 1$ where (i)-(iii) hold, but where agent i's realized payoff is minimal, i.e. $\underline{\pi}_i(o')$. But since $\underline{\pi}_i(o') < a_{i,t_1} = \overline{\pi}_i(o')$, agent i is dissatisfied. Hence, A6 implies that wpp i reaches a propensity vector in which $p_{i,t_2+2}(\alpha_i) > 0$. But then we are back in case (a).

Hence in both cases there must be some t where $p_{i,t}(\alpha_i(o)) = \overline{p}_i$ and $a_{i,t_2} = \underline{\pi}_i(o)$. But because this holds for all i and because each i's aspiration and propensity can be frozen by inertia we must eventually reach s^*. QED.

Proof of (4)

By assumption every player can tremble wpp to any neighboring state. Hence all states communicate. Since we also have aperiodicity (via inertia), ergodicity follows immediately. QED.

Proof of Theorem 6

The proof is by contradiction. Suppose that an outcome o^* is stable (i.e., it is supported by a pure SRE), but that in the associated steady state there are at least two people who get different payoffs. That is, there exists an SRE such that

$$\text{for all } i \text{ and } t : \text{let } p_{i,t}(\alpha_i(o^*)) = 1 \text{ and } a_{i,t} = a_i^*,$$

(where a_i^* is some constant), and there exist some distinct i and j such that $\pi_i(o^*) \neq \pi_j(o^*)$. Let $\pi_{\min}(o^*) = \min\{\pi_i(o^*), \pi_j(o^*)\}$. Suppose first that for some $t : a_{i,t} > \pi_{\min}(o^*)$. Then $p_{i,t+1}(\alpha_i(o^*)) < 1$. But then we cannot have a pure SRE. Since for all t, $a_{i,t} = a_i^*$ we have $a_i^* \leq \pi_{\min}(o^*)$.

Now we can partition N into two nonempty subsets, A and B, such that everyone in A gets the minimal payoff $\pi_{\min}(o^*)$ and everyone in B gets a higher one, i.e. for every $k \in B : \pi_k(o^*) > \pi_{\min}(o^*)$. Now consider some $i \in A$. Since N is nondecomposable, we can rewrite equation (1) as:

$$\begin{aligned} a_{i,t+1} &= \lambda_{i,0}a_{i,t} + (\lambda_{i,1}\pi_{1,t}(o^*) + \cdots + \lambda_{i,N}\pi_{N,t}(o^*)) = \\ a_i^* &= \lambda_{i,0}a_i^* + \sum_{j \in A} \lambda_{i,j}\pi_{\min}(o^*) + \sum_{k \in B} \lambda_{i,k}\pi_k(o^*). \end{aligned}$$

But since $a_i^* \leq \pi_{\min}(o^*) < \pi_j(o^*)$ for all $k \in B$, we also have

$$\lambda_{i,0}a_i^* + \sum_{j \in A} \lambda_{i,j}\pi_{\min}(o^*) + \sum_{k \in B} \lambda_{i,k}\pi_k(o^*) > a_i^*$$

which is a contradiction. QED.

Bibliography

[1] Bendor, Jonathan, Daniel Diermeier and Michael Ting. 2001. “A Behavioral Model of Turnout.” Stanford University: Unpublished manuscript.

[2] Bendor, Jonathan, Daniel Diermeier and Michael Ting. 2003 “Recovering Behavioralism: Adaptively Rational Strategic Behavior with Endogenous Aspirations.” In K. Kollman and S. Page (eds.), *Computational Political Economy.* Cambridge: MIT Press:213-269.

[3] Bush, Robert, and Frederick Mosteller. 1955. *Stochastic Models of Learning.* New York: John Wiley and Sons.

[4] Cyert, Richard, and James G. March. 1963. *A Behavioral Theory of the Firm.* Englewood Cliffs: Prentice-Hall.

[5] Feller, William. 1950. *An Introduction to Probability Theory and its Applications.* New York: Wiley.

[6] Foster, Dean and H. Peyton Young. 1990. “Stochastic Evolutionary Game Dynamics.” *Theoretical Population Biology* 38: 219-232.

[7] Fudenberg, Drew, and Eric Maskin. 1986. “The Folk Theorem in Repeated Games with Discounting and with Incomplete Information.” *Econometrica* 61:547-73.

[8] Harsanyi, John, and Reinhard Selten. 1988. *A General Theory of Equilibrium Selection in Games.* Cambridge: MIT Press

[9] Kanazawa, Satoshi. 2000. “A New Solution to the Collective Action Problem: The Paradox of Voter Turnout.” *American Sociological Review* 65:433-42.

[10] Macy, Michael. 1989. "Walking Out of Social Traps: A Stochastic Learning Model for the Prisoner's Dilemma." *Rationality and Society* 1:197-219.

[11] Macy, Michael. 1990. "Learning Theory and the Logic of the Critical Mass." *American Sociological Review* 55:809-26.

[12] Macy, Michael. 1991a. "Learning to Cooperate: Stochastic and Tacit Collusion in Social Exchange." *American Journal of Sociology* 97:808-43.

[13] Macy, Michael. 1991b. "Chains of Cooperation: Threshold Effects in Collective Action." *American Sociological Review* 56:730-47.

[14] Macy, Michael. 1993. "Backward-Looking Social Control." *American Sociological Review* 58:819-36.

[15] Macy, Michael. 1995. "PAVLOV and the Evolution of Cooperation: An Experimental Test." *Social Psychology Quarterly* 58:74-87.

[16] Macy, Michael, and Andreas Flache. 2002. "Learning Dynamics in Social Dilemmas." *Proceedings of the National Academy of Sciences,* forthcoming.

[17] Macy, Michael, and Robb Willer. 2002. "From Factors to Actors: Computational Sociology and Agent-Based Modeling." *Annual Review of Sociology,* Vol.28.

[18] McFadden, David. 1973. "Conditional Logit Analysis of Qualitative Choice Behavior." in P. Zarembka, ed., *Frontiers in Econometrics*. New York: Academic Press.

[19] Merton, Robert, and A. Rossi. 1950. "Contributions to the Theory of Reference Group Behavior." In R. Merton and P. Lazarsfeld (eds.), *Continuities in Social Research.* Glencoe IL: Free Press.

[20] Ostrom, Elinor. 1991. *Governing the Commons : The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.

[21] Nelson, Richard, and Sidney Winter. 1982. *An Evolutionary Theory of Economic Change*. Cambridge: Harvard University Press.

[22] Palfrey, Thomas R., and Howard Rosenthal. 1984. ”Participation and the Provision of Discrete Public Goods: A Strategic Analysis.” *Journal of Public Economics* 24:171-193.

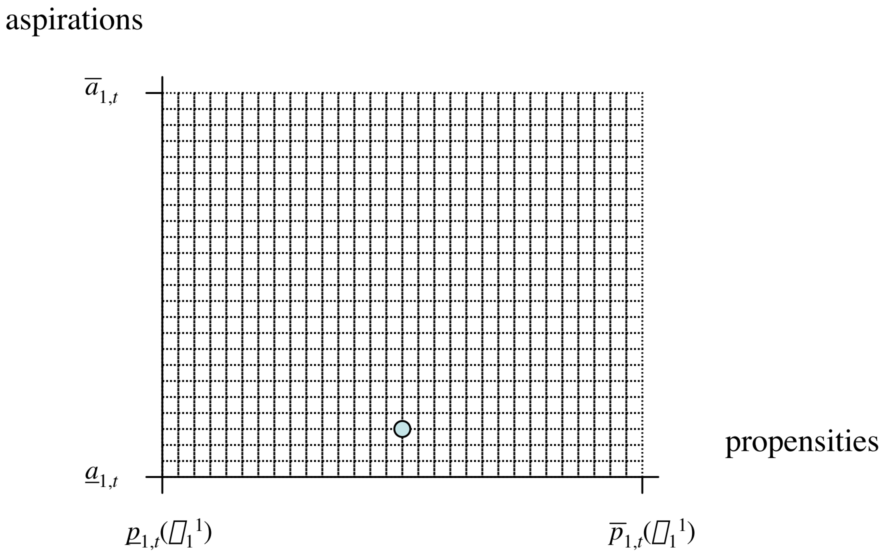
[23] Putnam, Robert D. 1993. *Making Democracy Work: Civic Traditions in Modern Italy.* Princeton, NJ: Princeton University Press.

[24] Simon, Herbert. 1955. “A Behavioral Model of Rational Choice.” *Quarterly Journal of Economics* 69:99-118.

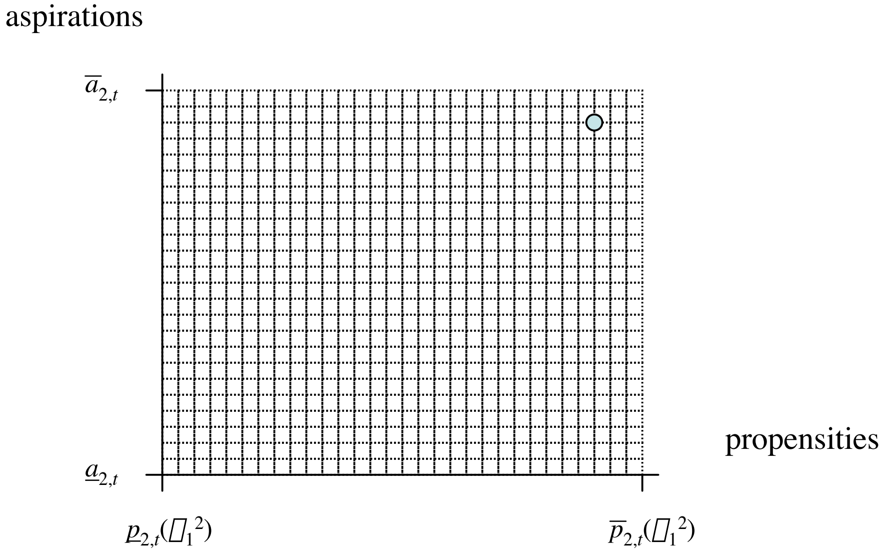
[25] Thorndike, Edward L. 1911. *Animal Intelligence: Experimental Studies.* New York: MacMillan.

[26] Winter, Sidney. 1971. “Satisficing, Selection and the Innovating Remnant.” *Quarterly Journal of Economics* 85:237-61.

Figure 1: State Space in a 2x2 Game



Player 1: middling propensities, low aspirations.



Player 2: high propensity for action 1, high aspirations.