RULES OF THUMB AND DYNAMIC PROGRAMMING*

By Martin Lettau and Harald Uhlig
CentER for Economic Research
Tilburg University
P.O. Box 90153
5000 LE Tilburg
The Netherlands

March 28, 1995

*This paper grew out of an earlier investigation on "How well do Artificially Intelligent Agents Eat Cake?" (1992). We have received many helpful comments in particular from Tom Sargent, Seppo Honkapohja, Aldo Rustichini, Johan Stennek, Nabil Al Najjar, Tim Van Zandt, Laurence Baker, Kenneth Corts, Michihiro Kandori and seminar participants at CORE and Princeton University.

Abstract

This paper studies the relationships between learning about rules of thumb (represented by classifier systems) and dynamic programming. Building on a result about Markovian stochastic approximation algorithms, we characterize all decision functions that can be asymptotically obtained through classifier system learning, provided the asymptotic ordering of the classifiers is strict. We demonstrate in a robust example that the learnable decision function is in general not unique, not characterized by a strict ordering of the classifiers, and may not coincide with the decision function delivered by the solution to the dynamic programming problem even if that function is attainable. As an illustration we consider the puzzle of excess sensitivity of consumption to transitory income: classifier systems can generate such behavior even if one of the available rules of thumb is the decision function solving the dynamic programming problem, since bad decisions in good times can "feel better" than good decisions in bad times.

JEL Classification: E00, C63, C61, E21

1 Introduction

Agents faced with an intertemporal optimization problem are usually assumed to use dynamic programming methods to derive their optimal decisions. However, observed data is often hard to reconcile with intertemporal optimization. As an alternative approach, Campbell and Mankiw (1989, 1991), DeLong and Summers (1986) and Ingram (1990) have demonstrated that observed behavior is often more consistent with agents following some simple ad hoc rules of thumb. Typically these models postulate a single rule of thumb and demonstrate its implications. In this paper we study systematically how agents learn to choose between many different rules of thumb in general dynamic choice problems. We represent a collection of rules of thumb as a classifier system following Holland (1986). Learning about these competing rules of thumb takes place via a simple accounting schmes which nonetheless enables the learning agent to deal with the dynamic structure of the model. Modelling boundedly rational agents in this way is appealing because it relies only on simple calculations rather than complex and forward-looking reasoning. Using stochastic approximation methods we characterize analytically all possible asymptotic learning outcomes and compare them to the dynamic programming solution. We show that certain aspects of the classifier system are closely related to the value function in dynamic programming. However, in general the learnable decision function is not unique and may not coincide with the optimal decision function even if that function is attainable. As an illustration we consider the puzzle of excess sensitivity of consumption to transitory income as documented by Flavin (1981), Hall and Mishkin (1982), Zeldes (1989) and Carroll and Summers (1991) among others. DeLong and Summers (1986) and Campbell and Mankiw (1989, 1991) explain this puzzle by allowing for irrational consumers who always consume their current income. We show analytically that this specific rule of thumb can be the asymptotic outcome of classifier system learning despite the fact that the optimal decision function is part of the system.

Intuitively, classifier system learning works as follows. A classifier system is motivated by a model of the brain as a collection of competing "if .. then .." statements. These condition - action pairs are called classifiers or, in our words, rules of thumb. [1] They compete via a single number attached to each of the classifiers, which is called its strength. At any given date t, the precondition of several classifiers (the if-part) might be satisfied. The brain will then have to choose among these applicable classifiers. It will do so by selecting the classifier with the highest strength. Learning takes place by adjusting the strength over time in the following manner. First, the strength is reduced by a certain percentage, thus penalizing the

[1] See Edelman (1992) for a useful reference on how a biologist/psychologist relates the functioning of the human brain to classifier system like structures.

classifier for being chosen. Second, the classifier is rewarded by adding the instantaneous utility generated at that date as well as a certain percentage of the strength of classifier chosen at the next date $t+1$ to its strength. This updating scheme accomplishes several things. It is simple and thus compatible with modelling boundedly rational behaviour. If a classifier does not generate any immediate or future benefits, it will quickly drop out of competition due to the penalty of loosing some of its strength when chosen. If a classifier generates very little instantaneous utility, but helps to improve conditions in the future, it can be rewarded via the percentage of the strength of the classifier chosen at date $t+1$. Clearly, this is a crucial feature in any scheme which attempts to solve dynamic decision problems reasonably well. The strength updating scheme has been called the "bucket brigade algorithm" since one can think of the classifier chosen at date $t+1$ as handing part of its strength down to the classifier chosen at date t like a bucket of water. One can also think of this as an auction, where classifiers offer a payment proportional to their strength for the privilege of determining the choice at the present date, which is paid to the classifier chosen at the previous date.

Classifier systems belong to the class of artificial intelligence methods. Some recent applications of artificially intelligent learning in economic models include Marimon, McGrattan, and Sargent (1990) who use classifier system learning in Kiyotaki and Wright's (1989) model of money as a method to compute equilibria. They demonstrate that classifier system learning offers an addition to the toolkit of numerically solving dynamic optimization problems surveyed in Taylor and Uhlig (1990). Arthur et.al. (1994) simulate a complete stock market with many agents, each endowed with a classifier system. Lettau (1994) has shown that artificially intelligent learning can explain observed inflows and outflows of mutual funds. Arthur (1993) derives some theoretical results, which are also based on stochastic approximation results, for a simplified version of classifier system learning in a stationary model where there is no dynamic link between periods. He also compares classifier system learning to behavior observed by human subjects in economic experiments and concludes that this learning method matches many features of human behavior in experiments very well.

The rest of the paper is organized as follows. In the second section, we define a general discrete dynamic choice problem and solve it using standard dynamic programming. In the third section, we introduce classifier systems and and explain how they learn. In section four, we analyze what classifier systems learn asymptotically. To this end we recast the bucket brigade algorithm for updating the strengths of the classifiers as a stochastic approximation algorithm. Building on a result about Markovian stochastic approximation procedures by Metivier and Priouret (1984) (see Appendix B) we completely characterize all possible strict orderings of the asymptotic strengths, i.e. all decision functions which a given classifier system can learn, provided the classifiers are strictly ordered by their strength asymptotically. Using these results we are able to compare the asymptotic outcome of classifier system learning to the solution obtained by dynamic programming. We show that the strength measure is closely related to the value function in dynamic programming, and that, in fact, they coincide in some special situations (see Proposition 5). The fifth section demonstrates in a simple but revealing example that the learnable decision function is in general not unique, not characterized by a strict ordering of the classifier strengths, and may not coincide with the optimal decision function even if that function is attainable. As an illustration we consider in section six the puzzle of excess sensitivity of consumption to transitory income as documented by Flavin (1981), Hall and Mishkin (1982), Zeldes (1989) and Carroll and Summers (1991) among others. DeLong and Summers (1986) and Campbell and Mankiw (1989, 1991) explain this puzzle by allowing for irrational consumers who always consume their current income. We show analytically that this specific rule of thumb can be the asymptotic outcome of classifier system learning despite the fact that the optimal decision function is part of the system. The basic intuition is the following. Spending more in good times may simply "feel better" on average than following the optimal decision function at all times. The last section concludes.

2 Dynamic Programming

Many recursive stochastic dynamic optimization problems can be discretized at least approximately, and written as a dynamic program in the following form:

$$v(s) = \max_{a \in \mathcal{A}} \left\{ u(s,a) + \beta E_{\pi_{s,a}} v(s') \right\}, \tag{1}$$

where $\mathcal{A} = \{a_1, \dots, a_m\}$ is the set of actions an agent can take, $\mathcal{S} = \{s_1, \dots, s_n\}$ is the set of possible states, $u(s,a) \in \mathbf{R}$ is the instantaneous utility derived from choosing action a in state s, $0 < \beta < 1$ is a discount factor and $\pi_{s,a}$ is a probability distribution on $\mathcal{S}$, which is allowed to depend on s and a. We assume throughout that $\pi_{s,a}(s') > 0$ for all $s' \in \mathcal{S}$. Define a decision function to be a function $h : \mathcal{S} \to \mathcal{A}$ and let $\mathcal{H}$ be the set of all decision functions. A standard contraction mapping argument as in Stokey and Lucas, with Prescott (1989), section 3.2, shows that there is a unique v^* solving the dynamic programming problem in (1). The solution is characterized by some (not necessarily unique) decision function $h^* : \mathcal{S} \to \mathcal{A}$ which prescribes some action $h^*(s)$ in state s.

For any decision function h define the associated value function v_h as the solution to the equation

$$v_h(s) = u(s, h(s)) + \beta E_{\pi_{s,h(s)}} v_h(s') \tag{2}$$

or as

$$v_h = (I - \beta \Pi_h)^{-1} u_h, \tag{3}$$

where v_h is understood as the vector $[v_h(s_1), \ldots, v_h(s_n)]'$ in $\mathbf{R}^n$, Π_h is the $n \times n$-matrix defined by

$$\Pi_{h,i,j} = \pi_{s_i, h(s_i)}(s_j)$$

and u_h is the vector $[u(s_1, h(s_1)), \ldots, u(s_n, h(s_n))]'$ in $\mathbf{R}^n$. Clearly, $v^* = v_{h^*}$. The next proposition tells us that no randomization is needed to achieve the optimum (This will contrast with some classifier systems in the example in section 5 below, which require randomizing among classifiers even in the limit).

PROPOSITION 1 *For all $s \in \mathcal{S}$,*

$$v^*(s) = \max_{h \in \mathcal{H}} v_h(s).$$

PROOF: *Define $\bar{v}$ via $\bar{v}(s) = \max_{h \in \mathcal{H}} \; v_h(s)$. We have to show that $v^*(s) \geq \bar{v}(s)$ for all states s (the reverse inequality is trivial). Define an operator $T : \mathbf{R}^n \to \mathbf{R}^n$ as follows: for any $v \in \mathbf{R}^n$, let Tv be the right-hand side of (1). Since $\bar{v} \geq v_h$ for any decision function h, we have $(T\bar{v})(s) \geq (Tv_h)(s) \geq v_h(s)$ for any decision function h. In particular, $(T\bar{v})(s) \geq \bar{v}(s)$ for all states s. Iterating this argument, we find that $T^j \bar{v}(s) \geq \bar{v}(s)$ for all states s. By the usual contraction mapping argument, $T^j \bar{v} \to v^*$. It therefore follows that $v^*(s) \geq \bar{v}(s)$ for all states s.* •

For future reference, let $\underline{u} = \min_{s,a} u(s,a)$ and $\bar{u} = \max_{s,a} u(s,a)$ be the minimum and the maximum one-period utility attainable. Furthermore, define μ_h to be the unique invariant probability distribution on $\mathcal{S}$ for the transition law Π_h, i.e. μ_h is the solution to $\mu_h = \Pi_h^T \mu_h$ with $\sum_s \mu_h(s) = 1$. The uniqueness of μ_h follows with standard results about Markov chains from the strict positivity of all $\pi_{s,a}(s')$.

3 Classifier System Learning

Let $\mathcal{A}^0 = a_0 \cup \mathcal{A}$, where a_0 is meant to stand for "no action specified". A rule of thumb is a function $r : \mathcal{S} \to \mathcal{A}^0$ with $r(S) \neq \{a_0\}$. A classifier c is a pair (r, z) consisting of a rule of thumb r and a strength $z \in \mathbf{R}$. A classifier $c = (r, z)$ is called applicable in state s, if $r(s) \neq a_0$. [2] A classifier system is a list $\mathcal{C} = (c_1, \ldots, c_K)$ of classifiers, so that for every

[2] Note that Holland (1986) proposed a binary decoding of the state space. The two formualtions are equivalent since one can always appropriately redefine the state space.

state s, there is at least one applicable classifier. Given a classifier system $\mathcal{C}$ and a state s, let $k(s;\mathcal{C})$ denote the index of the classifier with the highest strength of all applicable classifiers in $\mathcal{C}$, i.e. in state s classifier $k(s;\mathcal{C})$ is activated and action $r_{k(s;\mathcal{C})}(s)$ is carried out. Use randomization with some arbitrarily chosen probabilities to break ties. A classifier system $\mathcal{C}$ thus gives rise to a decision function $h(s;\mathcal{C}\ ) \equiv r_{k(s;\mathcal{C})}(s)$ by selecting the strongest among all possible classifiers at each state.[3]

Classifier system learning is a stochastic sequence of states $(s_t)_{t=1}^{\infty}$, indices $(k_t)_{t=1}^{\infty}$ of activated classifiers and classifier systems $(\mathcal{C}_t)_{t=1}^{\infty}$. Choose a decreasing cooling sequence $(\gamma_t)_{t\geq 0}$ of positive numbers satisfying

$$\sum_{t=1}^{\infty} \gamma_t^p \quad < \quad \infty \quad \text{for some } p \geq 2, \tag{4}$$

$$\sum_{t=1}^{\infty} \gamma_t \quad = \quad \infty \tag{5}$$

an initial classifier system $\mathcal{C}\ _1$ and an initial state s_1. We assume throughout that all initial strengths z_c for the classifiers in $\mathcal{C}_1$ are bounded below by $\underline{u}/(1-\beta)$.

Before updating the strengths at date t, the current state s_t and the current classifier system $\mathcal{C}_t$ are known. Choosing an action takes place at the end of date t, whereas the updating step for the strength of the active classifer in period t takes place at the beginning of date $t+1$. In detail:

1. (in date t) The classifier in $\mathcal{C}_t$ with highest strength subject to being applicable in state s_t is selected. Denote the index of the winning classifer by $k_t = k(s_t;\mathcal{C}_t)$.

2. (in date t) The action $r_{k_t}(s_t)$ is carried out.

3. (in date t) The instantaneous utility $u(s_t, r_{k_t}(s_t))$ is generated.

4. (in date $t+1$) The state transits from s_t to s_{t+1} according to the probability distribution $\pi_{s_t,r(s_t)}$ on $\mathcal{S}$.

5. (in date $t+1$) Determine the index $k(s_{t+1};\mathcal{C}_{t+1})$ of the strongest classifier $\mathcal{C}_t$ Denote its strength by z'. Update the strength of classifier with index k_t to

$$\tilde{z} = z - \gamma_{t+1}\left(z - u_t - \beta z'\right).^4 \tag{6}$$

[3] Another method to determine the decision function is to randomize among applicable classifiers according to their relative strengths, see e.g. Arthur (1993).

[4] Marimon, McGrattan and Sargent (1990) introduce an adjustment factor in the bidding to account for differences in the "generality" of classifiers. We will consider this extension in section (5.1).

The classifier system $\mathcal{C}_{t+1}$ is then defined to be the classifier system $\mathcal{C}_t$ with c replaced by $\tilde{c} = (r, \tilde{z})$.

The updating of the strength of the classifier activated in period t occurs at stage 5 when u_t and s_{t+1} are known. Note, that the updating equation (6) uses the period t strenghts to determine z' which is added to the strength of classifier which is active in period t. After finishing with stage 5, we go on to stage 1 in time $t+1$. The classifier chosen at stage 1 in period $t+1$ might differ from the "hypothetical" one which was used to complete the updating in stage 5. The updating algorithm is formulated in such a way that the updating does not require calculating the strengths at $t+1$ first, which would otherwise give rise to complications in cases where the activated classifiers at date t and $t+1$ have the same index. This is an attractive way to model boundedly rational learning since it relies only on simple calculations and avoids complex forward-looking reasoning like forming expectations.

The strength updating equation (6) is often referred to as a *bucket brigade*, since each activated classifier has to pay or give away part of its strength z in order to get activated, but in turn receives not only the instantaneous reward u_t for his action, but also the "payment" by the next activated classifier $\beta z'$, discounted with β. Note the formal similarity to equation (1) which hints at the intuition why the bucket brigade is able to deal with the dynamic structure of the maximization problem. Numerical procedures to calculate the value function in (1) are often based on iterations of approximation of the value function. A guess is plugged into the right hand side of (1) which yields a new guess. The new guess is in turn plugged into the right hand side, and so on. The contration mapping theorem ensures that this process converges to the true value function. The bucket brigade algorithm accomplishes a similar approximation for the strengths of the classifiers instead of the state dependent value function.

We should stress that an agent who is equipped with a classifier system only performs very simple computations. She does not have to deal with a complex dynamic programming problem but learns via the simple bucket brigade algorithm for updating the strenghts. Moreover, she only has to memorize K numbers, the strength for each classifier, which for large state spaces might be much smaller than the number of states n. The dimension of the value function in the dynamic program is of course equal to n. Note, however, that the number of classifers can in general be larger than the number of states.

4 Asymptotic Behavior

Classifier system learning leads to a stochastic sequence of decision functions $(h_t)_{t=1}^{\infty}$ given by $h_t = h(s_t; \mathcal{C}_t)$. We are interested in determining the asymptotic behavior of this sequence, i.e. which decision functions are eventually learned, and whether they coincide with an optimal decision function for (1).

It is convenient for the further analysis to rewrite the accounting scheme (6) as a stochastic approximation algorithm in the following way (for a general overview and introduction to stochastic approximation algorithms, see Sargent (1992) and Ljung, Pflug and Walk (1992)). [5] Let θ_t be the K-dimensional vector of all strengths at date t. Let $Y_t = [s_{t-1}, k_{t-1}, s_t, k_t]$ be the vector of the past and present state and the indices k_{t-1} and k_t of the activated classifiers. Given Y_t and θ_t, the updating procedure as laid out above generates Y_{t+1}. The vector θ_{t+1} is computed via

$$\theta_{t+1} = \theta_t - \gamma_{t+1} f(\theta_t, Y_{t+1}) \tag{7}$$

with

$$f(\theta_t, Y_{t+1}) = \mathbf{e}_{k_t}\, g(\theta_t, Y_{t+1}), \tag{8}$$

where $\mathbf{e}_{k_t}$ is the K-dimensional unit vector with a one in entry k_t and zeros elsewhere, and where

$$g(\theta_t, Y_{t+1}) = \theta_{t,k_t} - u(s_t, r_{k_t}(s_t)) - \beta\theta_{t,k_{t+1}}. \tag{9}$$

Note that f is linear in θ_t. Furthermore, if $\underline{u}/(1-\beta) \leq \theta_{t,k} \leq \bar{u}/(1-\beta) + \bar{\rho}$ for all k and some $\bar{\rho} \geq 0$, then $\underline{u}/(1-\beta) \leq \theta_{t+1,k} \leq \bar{u}/(1-\beta) + \beta\bar{\rho}$ for all k. Thus, the elements of θ_t will be bounded below and above by $\underline{u}/(1-\beta)$ and $\bar{u}/(1-\beta)$ in the limit as $t \to \infty$. The sequences $(\theta_{t,k})_t$ are in general not monotonically decreasing, even if the starting strengths z_k of the first classifier system $\mathcal{C}_1$ are bounded below by $\bar{u}/(1-\beta)$, but they are typically close to being monotonically decreasing in numerical applications.

Consider a vector of strengths θ_∞ so that, conditional on some strength θ_{t_0} and some value for Y_{t_0} at some date t_0, we have $\theta_t \to \theta_\infty$ with positive probability. If all elements of θ_∞ are distinct, we call θ_∞ a limit strength vector. We aim at characterizing all limit strength vectors. To do that, we first consider a special situation and formulate necessary conditions for a vector to be a limit strength vector in that situation, see Proposition 2 and the consistency condition below. We then show in Theorem 1, that these conditions are a complete characterization in general.

[5] Marimon et. al. (1990, sec. 5) already suggest to use stochastic approximation results to study the limit behavior of classifier system.

Consider the special situation where convergence to θ_∞ is almost sure for any value of Y_{t_0}, and where the ordering of the elements in θ_t coincides with the ordering of the entries in θ_∞ almost surely for all $t \geq t_0$. Using the list of rules from the classifier system $\mathcal{C}_1$ and attaching strengths according to a strength vector $\theta \in \mathbf{R}^K$ allows one to identify a strength vector θ with a classifier system $\mathcal{C}_\theta$. Given a limit strength vector θ_∞, find the associated classifier system $\mathcal{C}_\infty$. Find the index $k(s) = k(s; \mathcal{C}_t)$ of the strongest classifier for each state s, the resulting decision function $h(s) = h(s; \mathcal{C}_t)$, the associated transition matrix Π_h and thus the invariant distribution μ_h on $\mathcal{S}$. Note that $k(s)$ coincides with $k(s_t; \mathcal{C}_t)$ for all $t \geq t_0$ almost surely by our special assumption. Thus, the transition law for the state vector Y_t to Y_{t+1} can be restated as drawing s_{t+1} according to the transition law Π_h and setting the index k_{t+1} to be the index $k(s_{t+1})$: denote this transition law with $\hat{\Pi}$. Since μ_h is the unique invariant distribution for Π_h, it follows that there is a unique invariant distribution Γ for $\hat{\Pi}$. The marginal distribution of Γ with respect to the index k_t in Y_{t+1} yields an invariant distribution ν on the set $\{1, \ldots, K\}$ of classifier indices. This distribution can alternatively be computed directly via $\nu(k) = \text{Prob}(k = k(s)) = \sum_{\{s|k=k(s)\}} \mu_h(s)$. Call classifiers asymptotically active if they are a winning classifier for at least one state, i.e. if $\nu(k) > 0$. Call all other classifiers asymptotically inactive. Define

$$\phi(\theta) \equiv E_\Gamma \left[f(\theta, Y) \right],$$

where the expectation is calculated with respect to the invariant distribution Γ.

PROPOSITION 2 *In the special situation where convergence to θ_∞ is almost sure for any value of Y_{t_0}, and where the ordering of the entries in θ_t coincides with the ordering of the entries in θ_∞ almost surely for all $t \geq t_0$, a necessary condition for a limit strength vector is*

$$\phi(\theta_\infty) = 0. \tag{10}$$

PROOF: *Take expectations with respect to the invariant distribution Γ over the initial state Y_{t_0} in equation (7) and sum from $t = t_0$ to some T. Since Γ is the invariant distribution, this amounts to taking expectations with respect to the invariant distribution over each future Y_t as well. Exploiting almost sure convergence yields*

$$\theta_\infty = \theta_{t_0} - \sum_{t=t_0}^{\infty} \gamma_{t+1} \phi(\theta_t). \tag{11}$$

Assume now that (10) does not hold and that instead, say, $\phi(\theta_\infty) > \epsilon > 0$. Since $\phi(\theta_t) \to \phi(\theta_\infty)$ and hence $\phi(\theta_t) > \epsilon/2$ for $t \geq T$, some T, a contradiction follows from (11), (5) and the finiteness of θ_∞. •

The necessary condition (10) can be studied a bit further. It is easy to see that

$$\phi(\theta) = \nu \circ (\theta - \vec{u} - \beta B \theta), \tag{12}$$

where $\circ$ denotes element-by-element multiplication of two vectors of equal dimension, where

$$\vec{u} = \begin{bmatrix} \vdots \\ E_{\mu_h}\left[u(s, r_k(s)) \mid k(s) = k\right] \\ \vdots \end{bmatrix} \tag{13}$$

(and arbitrary entries, whenever the conditional expectation is not well defined, i.e. whenever $\nu(k) = 0$) and where B is a matrix with

$$\begin{aligned} B_{k,l} &= \mathrm{Prob}(\{s' \mid k(s') = l\} \mid k(s) = k) \\ &= \frac{\sum_{\{i \mid k = k(s_i)\}} \sum_{\{j \mid l = k(s_j)\}} \mu_h(s_i) \Pi_{h,i,j}}{\nu(c_k)} \end{aligned} \tag{14}$$

for all indices k indexing classifiers with $\nu(k) \neq 0$ (choose some arbitrary number between 0 and 1 otherwise).

Suppose one were given only the ordering of the classifiers according to θ_∞ rather than the strength vector itself. Given the ordering, one can recover the index of the winning classifier $k(s)$ as well as the decision function h, the utilities $u(s, r_k(s))$ and the probabilities $\Pi_{h,i,j}$, $\mu_h(s)$ and $\nu(k)$. Equation (10) can thus be used to solve for θ_∞ by solving a system of K linear equations in K unknowns. For the asymptotically active classifiers, a simple contraction mapping argument shows that their strengths is uniquely given by

$$\theta_\infty = (I - \beta B)^{-1} \vec{u}, \tag{15}$$

where (in slight abuse of notation) the rows and columns corresponding to asymptotically inactive classifiers are meant to be eliminated in that equation. Equation (10) does not impose any restrictions on the strength of asymptotically inactive classifiers. However their value can be bound by inspecting the construction of $k(s)$ from the ordering implied by θ_∞: the strength of any classifier, including asymptotically inactive classifiers, is bound above by the strengths of the asymptotically active classifiers which are applicable in states where the given classifier is applicable as well. In other words, the following consistency condition applies.

CONSISTENCY CONDITION:
for all $s \in \mathcal{S}$ and all classifiers c_k in $\mathcal{C}_\theta$ with $k \neq k(s)$ and $c(s) \neq a_0$, we have
$\theta_{\infty,k} < \theta_{\infty,k(s)}$.

We call a vector $\theta_\infty \in \mathbf{R}^K$ a candidate limit strength vector, if θ_∞ satisfies equation (10) as well as the consistency condition. Note that the calculations leading up to and following (10) essentially only require knowledge of the ordering of the classifiers resulting from θ_∞. As a result of the discussion above we thus have

Proposition 3 1. *Under the conditions of proposition 2, every limit strength vector is a candidate limit strength vector.*

2. *For each of the $K!$ possible strict orderings of the K classifiers, there is a vector $\theta_\infty \in \mathbf{R}^K$ satisfying (10). θ_∞ is unique up to the assignment of strengths to asymptotically inactive classifiers. θ_∞ is a candidate limit strength vector, if it satisfies the consistency condition.*

It is important when interpreting the second part of this proposition, that the given ordering of the classifiers is used to calculate $k(s)$, not the solved-for strength vector θ_∞, and the resulting decision function and probabilities. The vector θ_∞ may give rise to a different index $\tilde{k}(s)$ of the winning classifier and it is the task of the consistency condition to check for the equality of $k(s)$ with $\tilde{k}(s)$.

For two special cases, the candidate limit strength vectors are easy to construct and are directly related to the dynamic programming calculations in section 2.

Proposition 4 *Suppose there is only one rule $r \in \mathcal{R}$. Then there is a unique candidate limit strength vector $\theta_\infty \in \mathbf{R}$ and it satisfies $\theta = E_{\mu_r} v_r$. In particular, if $r = h^*$, then $\theta_\infty = E_{\mu_{h*}} v^*$.*

Proof: *In this case (15) reduces to*

$$\theta_\infty = (I - \beta)^{-1} E_{\mu_h}[u_h]. \tag{16}$$

We therefore need to show that

$$E_{\mu_h}[v_h] = (1 - \beta)^{-1} E_{\mu_h}[u_h] \tag{17}$$

or

$$(1 - \beta)\mu^T v_h = \mu_h u_h. \tag{18}$$

But this follows immeadiately from (3) and from the fact that

$$\mu^T = \mu^T \Pi_h, \tag{19}$$

which completes the proof. •

However, in general

$$z_k \neq E_{\mu_h}\left[v_h(s)|k(s)=k\right]. \tag{20}$$

PROPOSITION 5 *Let h^* be a decision function with $v^* = v_{h^*}$ and suppose that h^* is unique. Suppose, furthermore, that all K rules are applicable in at most one state and that for each $s \in \mathcal{S}$, there is exactly one rule with $r(s) = h^*(s)$; denote its index with $k^*(s)$. Define θ by assigning for each state s strength $v^*(s)$ to the classifier with index $k^*(s)$. For all other classifiers c applicable in some state s, assign some strength strictly strictly below $v^*(s)$. Then θ_∞ is a candidate limit strength vector which implements the dynamic programming solution.*

PROOF: *Compare equation (15) to equation (3) and note that $B = \Pi_h$ and $\vec{u} = u_h$.* •

Theorem 1 shows that our characterization is general.

THEOREM 1 *Every candidate limit strength vector is a limit strength vector and vice versa.*

The proof of this theorem can be found in Appendix A. It draws on a result by Metivier and Priouret (1984) about Markov stochastic approximation procedures, restated in Appendix B for convenience. The theorem indicates how classifier system learning happens over time. For some initial periods, the orderings of the classifiers may change due to chance events. Eventually, however, the system has cooled down enough and a particular ordering of the strengths is fixed for all following periods. As a result, the asymptotically inactive classifiers will no longer be activated, and the system converges to the limit strength vector as if the transition from states to states was exogenously given: the classifier system has learned the final decision rule. Alternatively, one can train a classifier system to learn a particular decision rule corresponding to some candidate limit strength vector by forcing the probabilistic transitions from one state to the next to coincide with those generated by the desired decision rule: after some initial training periods, the strengths will remain in the desired ordering and will not change the imprinted pattern. The number of initial training periods and the number of the cooling periods is path-dependent; however, given a particular history, the theorem and its proof do not rule out that the strengths break free once more to steer towards a different limit. In fact, this will typically happen with some probability due to (5). If there is a sufficiently long string of "unusual events", these events can have a large effect on the updating of the strengths in (6) and thus change an existing ordering.

It should be noted that the characterization only applies to limit strength vectors with a strict ordering of the strengths. As we will see in the next section, this is not just ruling out knife-edge cases. A robust example will be constructed, where equality of the strengths of two classifiers is necessary asymptotically.

5 Examples

We provide an example which demonstrates the similarities and the differences between classifier systems and the dynamic programming approach. It also demonstrates why the case of several asymptotically active classifiers for one state cannot be ruled out. The example is abstract and meant for illustration only; it has therefore been kept as simple as possible.

Suppose $\mathcal{S} = \{1;2;3\}$, $\mathcal{A} = \{1;2\}$ and that the transition to the next state is determined by the choice of the action only, regardless of the current state s:

$$\begin{array}{cccc} & s' = 1 & s' = 2 & s' = 3 \\ a = 1: & \pi_{s,1}(1) = 1/3 & \pi_{s,1}(2) = 1/3 & \pi_{s,1}(3) = 1/3 \\ a = 2: & \pi_{s,2}(1) = 0 & \pi_{s,2}(2) = 1 & \pi_{s,2}(3) = 0 \end{array}$$

Note, that some probabilities are zero, in contrast to our general assumption. This is done to simplify the algebra for this example. We further have a discount factor $0 < \beta < 1$ and utilities $u(s,a)$, $s = 1,2,3$, $a = 1,2$. We assume without loss of generality that $u(2,1) = 0$. We impose the restriction that $u(3,a) = u(1,a)$ for $a = 1,2$, so that state $s = 3$ is essentially just a "copy" of state $s = 1$. Thus, there are three free parameters, $u(1,1)$, $u(1,2)$ and $u(2,2)$.

The difference between state $s = 1$ and state $s = 3$ is in how they are treated by the available rules. Assume that there are two rules, r_1 and r_2, described by

$$\begin{array}{ccc} & r_1 & r_2 \\ s = 1: & r_1(1) = 1 & r_2(1) = 2 \\ s = 2: & r_1(2) = 0 & r_2(2) = 1 \\ s = 3: & r_1(1) = 1 & r_2(1) = 0 \end{array}$$

with "0" denoting the action a_0, i.e. non-applicability. Note that rule 2 is applicable in state $s = 1$ but not in state $s = 3$.

We aim at calculating all candidate limit strength vectors. Since there are only two rules, there can be only two strict rankings of the corresponding classifier strengths, namely

$z_1 > z_2$ (Case I) and $z_2 > z_1$ (Case II). We will also have reason to consider the case $z_1 = z_2$ with nontrivial randomization between the classifiers (Case III), a situation not covered by our theoretical analysis above. Each of these cases are analyzed below. Note that the two given rules never lead to action $a = 2$ in state $s = 2$: the value of $u(2, 2)$ is thus irrelevant for the comparison of the classifiers. Each of the cases below will thus be valid only under some restrictions on the values for the remaining two free parameters $u(1, 1)$ and $u(1, 2)$. The results are summarized in Table 1 and Figure 1.

Case I: $z_1 > z_2$: In this case, classifier 1 is activated in states $s = 1$ and $s = 3$ and classifier 2 is activated in state $s = 2$. Thus, action $a = 1$ is taken in all three states: $h(s) \equiv 1$. It follows that $\mu_h(1) = \mu_h(2) = \mu_h(3) = 1/3$. For the strengths, one needs to solve the equations

$$\begin{aligned} z_1 &= u(1,1) + \frac{\beta}{3}(2z_1 + z_2) \\ z_2 &= \frac{\beta}{3}(2z_1 + z_2). \end{aligned}$$

This case can thus be obtained if and only if

$$u(1,1) > 0. \tag{21}$$

Case II: $z_2 > z_1$: In this case, rule 2 is applied in states $s = 1$ and $s = 2$ whereas rule 1 is applied in state $s = 3$. Hence, $h(1) = 2$, $h(2) = 1$, $h(3) = 1$ and consequently $\mu_h(1) = \mu_h(3) = 1/4$ and $\mu_h(2) = 1/2$. The strengths are calculated from

$$\begin{aligned} z_1 &= u(1,1) + \frac{\beta}{3}(z_1 + 2z_2) \\ z_2 &= \frac{1}{3}(u(1,2) + \beta z_2) + \frac{2\beta}{9}(z_1 + 2z_2). \end{aligned}$$

It is easily checked that this case can be obtained if and only if

$$u(1,2) > 3u(1,1). \tag{22}$$

Case III: $z_1 = z_2 = z$: . We provide a "solution" for this case, even though our theory above does not cover cases without strict ranking of the classifiers. The reasoning employed here should be rather intuitive, however. Given state $s = 1$, we guess that classifier c_1 is activated with some probability p, whereas classifier c_2 is activated with probability 1 - p, i.e. there is randomization between the classifiers. The resulting decision function is random. Given $s = 1$, states $s' = 1$ and $s' = 3$ will be reached with probability

$p/3$ each. The invariant distribution μ_h is therefore $\mu_h(1) = \mu_h(3) = 1/(4-p)$, $\mu_h(2) = (2-p)/(4-p)$. Let $\mu_h(s,k)$ be the joint probability that state s occurs and classifier k is activated. We have

$$\begin{array}{lll} & c_1 & c_2 \\ s=1: & \mu_h(1,1) = p/(4-p) & \mu_h(1,2) = (1-p)/(4-p) \\ s=2: & \mu_h(2,1) = 0 & \mu_h(2,2) = (2-p)/(4-p) \\ s=3: & \mu_h(3,1) = 1/(4-p) & \mu_h(3,2) = 0. \end{array}$$

The common strength z should satisfy both equations arising from (10), one for classifier c_1 and one for classifier c_2 yielding

$$\frac{1}{1-\beta} u(1,1) = z = \frac{1}{1-\beta} \frac{1-p}{3-2p} u(1,2), \tag{23}$$

which can be solved for p. Note that p is a viable probability if and only if $0 \leq p \leq 1$. Thus, case III is valid, if and only if one of the following two inequality restrictions is satisfied:

- $u(1,2) \leq 3u(1,1) \leq 0$ or
- $u(1,2) \geq 3u(1,1) \geq 0$.

The calculated strengths and probabilities calculated in this case are unique except if $u(1,1) = u(1,2) = 0$. The inequalities have to be strict in order for p to be nontrivial: otherwise, the decision rule obtained coincides with the one derived from case I or case II.

Table 1 shows that for any given values of $u(s,a)$ there is at least one applicable case. However, the only case available may be Case III and thus the solution prescribed by the classifier system involves randomizing between the classifiers.

It is interesting to compare these possibilities with the solution to the dynamic programming problem. If $u(2,2)$ is large enough, the optimal decision function will always prescribe action $a = 2$ in state $s = 2$, which cannot be done with the classifiers above. Any classifier system with the rules given above will result in a suboptimal solution simply because the correct solution is not within reach.

Assume instead that $u(2,2)$ is small enough, so that the decision function h^* solving the dynamic programming problem takes action $h^*(2) = 1$ in state $s = 2$. By symmetry, $h^*(1) = h^*(3)$ and $v^*(1) = v^*(3)$. Furthermore, $h^*(1) = h^*(3) = 1$ if and only if

$$u(1,1) + \beta/2\,(v^*(1) + v^*(2)) \geq u(1,2) + \beta v^*(2) \tag{24}$$

and $h^*(1) = 2$ otherwise.[6] Directly calculating $v^* = v_{h^*}$ with equation (3) for the two choices yields

- $h^*(1) = 1$, if $u(1,2) \leq \left(1 + \frac{2}{3}\beta\right) u(1,1)$,
- $h^*(1) = 2$, if $u(1,2) \geq \left(1 + \frac{2}{3}\beta\right) u(1,1)$.

A summary of all possible situation is found in Table 1 and Figure 1. The learnable decision function may not be unique (area C2). The learnable decision function may involve asymptotic randomization between the available rules (area C1). The learnable decision function can also be different from the solution to the dynamic programming problem, even if that solution is attainable by ranking the classifiers appropriately: this is the case in area B1. The intuitive reason for this last observation is quickly found: since $u(2,1)$ has been normalized to zero, $u(1,1)$ measures how much classifier c_1 gains against classifier c_2 by being applicable in state 2 rather than state $s = 1$. If $u(1,1)$ is positive, state $s = 3$ corresponds to "good times" and state $s = 2$ corresponds to "bad times". Since the accounting system for calculating the strength of classifiers does not distinguish between rewards generated from the right decision and rewards generated from being in good times, a classifier that is applicable only in good times "feels better" to the artificially intelligent agent than it should. Thus, if $u(1,1) > 0$, classifier c_1 may be used "too often" and if $u(1,1) < 0$, classifier c_1 may be used "too little". This is what happens in regions B and C.

5.1 Adjusting for Generality

In the bidding and accounting scheme as laid out in section (3) each classifier is treated equally independent of their generality. The example above shows that rules which are applicable only in a small number of states, ie. specific rules, can dominate "better" general rules even if their are inferior. This raises the immediate question whether it is possible to adjust the scheme so that this drawback can be eliminated. Marimon, McGrattan and Sargent (1990) adjust the payment of the classifiers with a proportional factor that depends on the number of states in which the classifier is active. We will allow for the more general case where there is a general correction factor for each classifier.

Consider the above example with only two classifiers. Let κ_1, κ_2 be the adjustment factor for the classifier 1 and 2, respectively and let $x_i = \kappa_i z_i$ be the adjusted strength for classifier i. The first obvious change in the scheme is as follows. The strongest classifier is found by comparing the values of κz instead of just z and the strength is updated according to

[6] If equation (24) holds with equality, both choices for $h^*(1)$ are optimal.

$\tilde{z} = z - \gamma_{t+1}\left(\kappa z - u_t - \beta\kappa' z'\right)$. However, the difference to the accounting system presented above is small and immaterial asymptotically. To see that rewrite the equation as $\tilde{x} = x - \kappa\gamma_{t+1}\left(x - u_t - \beta x'\right)$ and therefore differs from (6) only by a classifier-individual scalar adjustment for the updating step size. Asymptotically, only the expression in brackets matters, and there is no difference. This shows that for the adjustment to work in the limit, one has to distinguish between the bidding and the payment between the classifiers. Thus we propose the following adjustment.

The bid of each applicable classifier is still equal to its strength. Thus the strongest classifier still wins the bidding. However, the winning classifier only pays back $x = \kappa z$ instead of z. Hence, the adjustment factors κ determine how much of its strength the winning classifier has to give away. A classifier with a low κ has thus an advantage over a classifier with a high κ. Next we check whether it is possible to find a set of κ's which guarantee that the classifier learning solution is identical to the dynamic programming solution. We normalize κ_2 to unity thus leaving κ_1 as free parameter. Assume for the moment that the dynamic programming solution prescribes $h^*(1) = 1$ (the condition for this case is given in the preceeding section). Thus, we have to find κ_1 so that classifier 1 is stronger than classifier 2 in the limit, ie. $z_1 > z_2$. Note that since the payments are now x_1 and x_2, equation (10) x_2. Using $x_1 = \kappa_1 z_1$ and $x_2 = z_2$ we can get the consistency condition in terms of the strengths. We get

$$z_1 - z_2 = u(1,1)\left[\frac{\frac{3-\beta}{\kappa} - 2\beta}{3(1-\beta)}\right]. \tag{25}$$

Hence, if $u(1,1) > 0$ any κ satisfying

$$\kappa < \frac{3-\beta}{2\beta} \tag{26}$$

will give the desired result that classifier 1 is stronger than classifier 2. If $u(1,1) < 0$ the inequality is reverses. The analysis for the other case $z_2 > z_1$ is analogous. This shows that we always can find appropriate payment adjustments that can lead the classifier system solution to coincide with the dynamic programming solution if it is attainable. Note however, that this is only possible after having solved the dynamic program. In other words, it is not possible to select the correct payment adjustment factors without knowing the dynamic programming solution.[7]

Furthermore, it is not clear whether letting the adjustment parameters depend on the generality of the classifiers is optimal in every case. It might solve the problem of selecting

[7] Marimon, McGrattan and Sargent's "bids" correspond to our payment κz. The winning classifier in their paper is, as in ours, determined by the strength and not by the as bid, as one might be led to believe from their teerminology. From this it follows that more general rules benefit from their adjustment scheme and not specific ones as they say on p. 338.

inferior specific rules in some cases. But one also could imagine cases where a specific but superior rule which is only applicable in bad states s dominated by an general but inferior rule which is applicable in good and bad states. Instead one could define adjustment schemes which depend on whether a classifier is applicable in "good" or in "bad" states. In general, the above analysis shows that rules that are applicable in "good" states should receive a high κ thus penalizing them for being applicable in these good times. On the contrary, classifiers which are mostly applicable in bad states should receive a low κ. Of course, in some problems it is ex ante not clear to determine which state is "good" and which one is "bad". In these cases it is hard to find suitible adjustment parameters.

6 Generating Excess Sensitivity of Consumption to Transitory Income

Consider the puzzle of excess sensitivity of consumption to transitory income as documented by Hall and Mishkin (1982), Zeldes (1989) and Carroll and Summers (1991) among others. DeLong and Summers (1986) and Campbell and Mankiw (1989, 1991) propose ad hoc rule of thumb consumers to explain this feature of aggregate consumption. Their rule of thumb consumers are not as sophisticated as our learning agents. They estimate models which allow for a fixed proportion of consumers who just consume their current income and find that the specification including rule of thumb consumers is capable of producing excess sensitivity. In an alternative approach Laibson (1993) explains the puzzle as stemming from the inability to precommit future selves not to spend too much out of transitory income.

We will construct an example to show that classifier system learning can generate such behavior even if the optimal decision function is part of the system since bad decisions in good times can "feel better" than good decisions in bad times. Specifically we show that the ad-hoc rule of thumb "consume current income" can be the asymptotic outcome of classifier system learning. In order to get analytical results we make some simplifying assumptions, but the general flavor of the results should be intuitive even in more complex models.

There is an infinitely lived agent who derives utility $u(c_t)$ from consuming in period t and who discounts future utility at the rate $0 < \beta < 1$. The agent receives random income y_t each period. Suppose there are two income levels, $\overline{y} > \underline{y} > 0$ and that income follows a Markov process with transition probabilities $p_{\underline{y}\,\underline{y}} = \mathrm{Prob}(y_t = \underline{y} \mid y_{t-1} = \underline{y})$, etc.. The agent enters period t with some wealth w_t. Next periods wealth is given by $w_{t+1} = w_t + y_t - c_t$. A borrowing constraint is imposed so that $c_t \leq w_t + y_t$. Furthermore, we assume that the agent is born with zero wealth: $w_0 = 0$.

To cast this model into our framework, we discretize the model and assume that all variables take only integers values: $w \in \{0, 1, 2, \ldots, \bar{w}\}$, etc.. The state of the system is given by $s_t = (w_t, y)$, where y is the present income level, $y \in \{\underline{y}, \overline{y}\}$. Note that the implied transition probabilities π_{ij} from state i to state j are not strictly positive, in contrast to our assumption in section 2. This assumption is for simplification of the algebra only. [8]

The dynamic programming problem is given by

$$v(w,y) = \max_{c \in \{0,1,\ldots,w+y\}} \left(u(c) + \beta \sum_{y' \in \{\underline{y},\overline{y}\}} p_{yy'} \left[v(w+y-c, y') \right] \right). \tag{27}$$

Given particular choices for the (increasing) utility function and the parameters of this model, this dynamic programming problem can be solved with the techniques in section 2. Let $h^*(s) = c^*(w,x)$ be the decision function solving this dynamic program. Note that $c^*(0, y_l) = \underline{y}$. Hence when the agent has zero wealth and current income is low, she spends all her income.

Now consider two rules, r_1 and r_2 with strengths z_1 and z_2 respectively. Rule r_1 is applicable in all states and coincides with the optimal decision function h^*. Rule r_2 is applicable only in states when the income is high, i.e. in "good" states. We assume that

$$r_2(w, \overline{y}) = w + \overline{y}, \tag{28}$$

so that rule r_2 prescribes consumption of the maximal amount when income is high.

Will the suboptimal rule r_2 be asymptotically active when it is applicable despite the fact that the optimal rule r_1 is always applicable? This coincides with the ranking $z_2 > z_1$. Note that in this regime the agent always spends her total current income and never saves This is the ad-hoc rule of thumb consumer considered by DeLong and Summers (1986) and Campbell and Mankiw (1989, 1991). Thus the invariant distribution over states μ_h has zero weight on all states $s_t = (w, y)$ with $w > 0, y \in \{\underline{y}, \overline{y}\}$. This makes the equation for calculating the strengths very simple:

$$\begin{aligned} z_1 &= u(y_l) + \beta(p_{\underline{y}\,\underline{y}} z_1 + (1 - p_{\underline{y}\,\underline{y}}) z_2) && (29) \\ z_2 &= u(y_h) + \beta(p_{\overline{y}\,\overline{y}} z_1 + (1 - p_{\overline{y}\,\overline{y}}) z_2). && (30) \end{aligned}$$

Solving these two equations for z_1 and z_2 gives the limit strengths. To see whether this

[8] We could modify the model in the following way so that $\pi_{ij} > 0 \, \forall i, j$. Let $p > 0$ be the probability that next periods wealth is given by $w_{t+1} = w_t + y_t - c_t$ whereas with probability $1 - p$, an arbitrary wealth level w_{t+1} is drawn next period uniformly from $0, 1, ..., \bar{w}$. While all the results are valid for p close to unity, the algebra is much more cumbersome. For simplicity we choose $p = 1$.

ranking is feasible we have to check if $z_2 > z_1$ since

$$z_2 - z_1 = \frac{u(\overline{y}) - u(\underline{y})}{1 - \beta(p_{\underline{y}\,\underline{y}} - p_{\overline{y}\,\overline{y}})} > 0, \;\; 0 < p_{\underline{y}\,\underline{y}}, p_{\overline{y}\,\overline{y}}, \beta < 1, \tag{31}$$

rule r_1 (r_2) will be active when income is low (high). The resulting consumption decision is $h(s) = c(w, y) = y$. The intuition behind this suboptimal behavior is as in the preceding example: rule r_2 may be asymptotically stronger than rule r_1 since it only applies in "good times" and thus "feels better" on average than rule r_2. Thus we have demonstrated that rule of thumb consumption behavior can be the outcome of learning behavior and thus should not be dismissed as completely ad-hoc.

This example is only intended as an illustration. We should stress that the choice of classifiers is ad hoc. However, the resulting behavior should be fairly robust in more complex problems.

7 Conclusion

In this paper we have discussed the asymptotic behavior of rules of thumb learning represented by Holland's (1986) classifier system. We have shown how a bucket brigade algorithm enables classifier systems to deal with general discrete recursive stochastic dynamic optimization problems. We reformulated the evolution of the strengths as a stochastic approximation algorithm. Using a theorem by Metevier and Priouret (1984) we are able to obtain a general characterization of all possible limit outcomes provided that the classifiers are strictly ordered in the limit. A simple example shows that the attainable decision function is neither necessarily unique nor characterized by a strict ordering of classifiers. With these results we are also able to compare classifier learning to the dynamic programming solution of the dynamic optimization problem. Due to the bucket brigade payment scheme, the classifier strengths are in close relationship to the value function of the dynamic problem, in fact, in certain situations they coincide.

The example also shows in what circumstances classifier system learning may lead to suboptimal behavior even when the optimal decision rule is an element of the classifier system. Since the optimal classifier is applicable in all possible states of nature, a suboptimal classifier might dominate the optimal one if is applicable only in "good" states of the world. Bad decisions in good times can "feel better" than good decision in bad times. We also show that this effect might lead a consumer in an intertemporal consumption problem to consume too much in periods of high income generating excess sensitivity to transitory income.

Appendix

A Proof of Theorem 1

Proof: *We first show that a given candidate limit strength vector is a limit strength vector. To that end, we analyze first an alteration of the stochastic approximation scheme above and characterize its limits in the claim below. We then show that the limit to this altered scheme corresponds to a limit strength vector in the original scheme.*

Claim: Consider a candidate limit strength vector $\theta_\infty a$ and its associated decision function h. Fix the transition probabilities Π_h. Consider the following altered updating system: let the classifier system consist only of the asymptotically active classifiers according to θ_∞. Fix some starting date t_0, initial strength vector θ_{t_0} with $\theta_{t_0,l} \geq \underline{u}/(1-\beta)$ for all l and an initial state Y_{t_0}. Let $\tilde{\theta}_t$ be the vector of strengths of this reduced classifier system at date $t \geq t_0$ and let $\tilde{\theta}_\infty$ be the corresponding subvector of θ_∞ of strengths of only the asymptotically active classifiers. Furthermore, let the transition from state s_t to s_{t+1} always be determined by the transition probabilities Π_h. Then $\tilde{\theta}_t \rightarrow \tilde{\theta}_\infty$ almost surely. Furthermore, for almost every sample path, the transition probabilities $\pi_{s_t,r_{k_t}}$ coincide with the transition probabilities given by π_h for all but finitely many t.

Proof of the Claim: *The updating scheme is still given by (7), (8) and (9). The transition law for Y_t is given by $\hat{\Pi}$, where $\hat{\Pi}$ was defined in section 4. In particular, $\hat{\Pi}$ does not depend on $\tilde{\theta}_t$ due to our alteration of the updating process. The random variables Y lie in a finite, discrete set. Note that $\tilde{\theta}_t$ always remains in the compact set $\Theta = [\underline{u}/(1-\beta), \bar{z}]^d$, where d is the number of asymptotically active classifiers and $\bar{z}$ is the maximum of all initial starting strengths in $\tilde{\theta_{t_0}}$ and $\bar{u}/(1-\beta)$. With our remarks after theorem 2 in Appendix B, this theorem thus applies if we can verify assumptions (F), (M1), (M5c) and the additional assumptions listed in the theorem itself.*

Assumption (F) is trivial, since f is continuous. Assumption (M1), the uniqueness of Γ, follows from the uniqueness of μ_h. For (M5c), note that $I - \hat{\Pi}$ is continuously invertible on its range and that $f(\tilde{\theta}, Y)$ is linear and thus Lipschitz continuous in $\tilde{\theta}$. For the additional assumptions of the theorem, note first that $p = \infty$ in (M2) is allowed according to our remarks following the theorem in Appendix B, so that the restriction $\sum_n \gamma_n^{1+(p/2)} < \infty$ is simply the restriction that the sequence (γ_n) is bounded. For the conditions on the differential equation, consider ϕ as given in equation (12) and $\vec{u}$ and B given in equations (13) and (14) restricted to entries of asymptotically active classifiers. Note that the differential equation

$$\begin{aligned}\frac{d\tilde{\theta}(t)}{dt} &= -\phi(\tilde{\theta}(t)) \\ &= \nu \circ \left(\tilde{\theta}_\infty - \vec{u} - \beta B \tilde{\theta}_\infty\right)\end{aligned}$$

is linear with the unique stable point $\tilde{\theta}$ given by (15). The differential equation is globally stable since the matrix $-\nu \circ (I - \beta B)$ has only negative eigenvalues (note that $0 < \beta < 1$ and that $B_{k,l}$ is a stochastic matrix). Theorem 2 thus applies with $A = \Theta$ and we have

$$\lim_{t \to \infty} \tilde{\theta}_t = \tilde{\theta}_\infty \text{ a.s.},$$

as claimed.

The claim that the transition probabilities $\pi_{s_t, r_{k_t}}$ coincide with the transition probabilities given by π_h follows from the almost sure convergence to the limit. Given almost any sample path, all deviations $\tilde{\theta}_{t,k} - \tilde{\theta}_{\infty,k}$ will be smaller than some given $\epsilon > 0$ for all $t \geq T$ for some sufficiently large T, where T depends in general on the given sample path and on ϵ. Make ϵ less than half the minimal difference between the limit strengths of any two different classifiers $\mid \tilde{\theta}_{\infty,k} - \tilde{\theta}_{\infty,l} \mid$. We then have that the ranking of the classifiers by strength will not change from date T onwards. But that means that the transition probabilities $\pi_{s_t, r_{k_t}}$ coincide with the transition probabilities given by π_h, concluding the proof of the claim. $\circ$

Given a candidate limit strength vector θ_∞, find $\epsilon > 0$ such that 4ϵ is strictly smaller than the smallest distance between any two entries of θ_∞. Denote the underlying probability space by $(\Omega, \Sigma, \mathcal{P})$ and states of nature by $\omega \in \Omega$. Consider the altered updating scheme as described in the claim above with $t_0 = 1$ and the given initial state. Find the subvector $\tilde{\theta}_\infty$ of θ_∞, corresponding to the asymptotically active classifiers according to θ_∞. We can thus find a date t_1, a state Y and a strength vector $\bar{\theta}$ for only the asymptotically active classifiers so that given some event $\Omega' \subset \Omega$ of positive probability, sample paths satisfy $Y_{t_1} = Y$, $\mid \tilde{\theta}_{t_1,l} - \bar{\theta}_l \mid < \epsilon$ for all l, $\mid \tilde{\theta}_{t,l} - \tilde{\theta}_{\infty,l} \mid < \epsilon$ for all l and all t and $\tilde{\theta}_t \to \tilde{\theta}_\infty$. For any sample path $(\tilde{\theta}_t)_{t \geq t_1}$ (i.e. not just those obtained for states of nature in Ω') , find the "shifted" sample path $(\hat{\theta}_t)_{t \geq t_1}$ obtained by starting from $\hat{\theta}_{t_1} = \bar{\theta}$ instead of $\tilde{\theta}_{t_1}$, but otherwise using the same realizations u_t and states Y_t for updating. This resets the initial conditions and shifts the starting date to t_1, but leaves the probabilistic structure otherwise intact: the claim thus applies and we have again $\hat{\theta}_t \to \tilde{\theta}$a.s.. Furthermore, given Ω', an induction argument applied to (6) yields $\mid \hat{\theta}_{t,l} - \tilde{\theta}_{t,l} \mid \leq \mid \hat{\theta}_{t_1,l} - \tilde{\theta}_{t_1,l} \mid < \epsilon$ for all $t \geq t_1$ and all l. As a result, $\mid \hat{\theta}_{t,l} - \tilde{\theta}_{\infty,l} \mid < 2\epsilon$ for all $t \geq t_1$ and all l, given Ω'. Extend $\hat{\theta}_t$ to a strength vector $\check{\theta}_t$ for all classifiers by assigning the strengths given by θ_∞ to inactive classifiers. By our assumption about ϵ, the *ordering of the strengths given by any $\check{\theta}_t$, $t \geq t_1$ coincides with the ordering of the strengths given by θ_∞. Thus starting the classifier system learning at t_1, strength vector $\theta_{t_1} = \check{\theta}_{t_1}$ and state $Y_{t_1} = Y$ at t_1, the evolution of the strengths θ_t is described by $\theta_t = \check{\theta}_t$ for all $\omega \in \Omega'$ and we therefore have that $\theta_t \to \theta_\infty$ with positive probability. This shows that θ_∞ is a limit strength vector, completing the first part of the proof.*

Consider in reverse any limit strength vector θ_∞: we have to show that θ_∞ satisfies (10), since the consistencyconsistencyconsistencyconsistency condition is trivially satisfied by definition of $k(s)$. Find $\epsilon > 0$ so that 4ϵ is strictly smaller than the smallest distance between any two entries of θ_∞. Find a date $t_1 \geq t_0$ so that on a set Ω' of positive probability, we have $\mid \theta_{t,l} - \theta_{\infty,l} \mid < \epsilon$ for all $t \geq t_1$ and all l, and $\theta_t \to \theta_\infty$. Given the strict ordering of the strengths in θ_∞, there is a candidate limit strength vector θ'_∞ which is unique up to the assignment of strength to asymptotically inactive classifiers, see proposition 3. Given any particular state of nature $\bar{\omega} \in \Omega'$ and thus values for θ_{t_1} and Y_{t_1} at date t_1, consider the altered updating scheme as outlined in the claim with that starting value (and $t_0 \equiv t_1$ for the notation in the claim). Via the claim, $\tilde{\theta}_t \to \tilde{\theta}_\infty$ a.s., where $\tilde{\theta}'_\infty$ is the subvector of the candidate limit strength vector θ'_∞ corresponding to the asymptotically active classifiers. Thus, the strengths in $\tilde{\theta}(\omega)$ coincide with the strengths of the asymptotically active classifiers in $\theta_t(\bar{\omega})$ for almost all ω and it is now easy to see that therefore the strengths of the asymptotically active classifiers in θ_∞ has to coincide with the strength of the asymptotically active classifiers in θ'_∞, finishing the proof of the second part. To make the last argument precise, observe that $\tilde{\theta}_t(\omega) \to \tilde{\theta}_\infty$ except on a measurable nullset $\omega \in \Xi_{\bar{\omega}} \in \Sigma$. Note that the exceptional set is the same whenever the initial conditions θ_{t_1} and Y_{t_1} are the same. Since there are only finitely many such initial conditions that can be reached, given the discrete nature of our problem and the fixed initial conditions at date t_0, the exceptional set $\Xi = \{(\bar{\omega}, \omega) \mid \bar{\omega} \in \Omega', \omega \in \Xi_{\bar{\omega}}\}$ is a measurable subset of zero probability of $\Omega' \times \Omega$ in the product probability space on $\Omega \times \Omega$. It follows that the strengths of the asymptotically active classifiers in θ_∞ and θ'_∞ coincide for all $(\bar{\omega}, \omega) \in \Omega' \times \Omega / \Xi$, which is a set of positive probability. Since these strengths are not random, we must have equality with certainty. •

B A Theorem about Markov Stochastic Approximation Algorithms

In this section, we use the notation of Metivier and Priouret (1984). For a general overview and introduction to stochastic approximation algorithms, see Sargent (1992) and Ljung, Pflug and Walk (1992).

For each $\theta \in \mathbf{R}^d$ consider a transition probability $\hat{\Pi}_\theta(y;dx)$ on $\mathbf{R}^k$. This transition probability defines a controlled Markov chain on $\mathbf{R}^d$.

Define a stochastic algorithm by the following equations:[9]

$$\theta_{n+1} = \theta_n - \gamma_{n+1} f(\theta_n, Y_{n+1}) \tag{32}$$

where $f(\theta, y)$ is a function $f : \mathbf{R}^d \times \mathbf{R}^k \to \mathbf{R}^d$,

$$P\left[Y_{n+1} \in B \mid \mathcal{F}_n\right] = \hat{\Pi}_{\theta_n}(Y_n; B) \tag{33}$$

where $P\left[Y_{n+1} \in B \mid \mathcal{F}_n\right]$ is the conditional probability of the event $\{Y_{n+1} \in B\}$ given $\theta_0, \ldots, \theta_n$, $Y_0, \ldots, Y_n$.

We call $\psi \to \hat{\Pi}_\theta \psi$ the operator $\hat{\Pi}_\theta \psi(x) \equiv \int \psi(y) \hat{\Pi}_\theta(x; dy)$. Assume the following:

(F) For every $R > 0$ there exists a constant M_R such that

$$\sup_{|\theta| \leq R} \sup_x \mid f(\theta, x) \mid \leq M_R.$$

(M1) For every θ, the Markov chain $\hat{\Pi}_\theta$ has a unique invariant probability Γ_θ.

(M2) There exist $p \geq 2$ and positive constants $\alpha_R < 1$, K_R for which $\sup_{|\theta| \leq R} \int \mid y \mid^p \hat{\Pi}_\theta(x; dy) \leq \alpha_R \mid x \mid^p + K_R$.

(M3) For every function[10] v with the property $\mid v(x) \mid \leq K(1+ \mid x \mid)$ and every θ, θ', $\mid \theta \mid \leq R$, $\mid \theta' \mid \leq R$,

$$\sup_x \mid \hat{\Pi}_\theta v(x) - \hat{\Pi}_{\theta'} v(x) \mid \leq \tilde{K}_R \mid \theta - \theta' \mid \sup_{x \neq x'} \frac{\mid v(x) - v(x') \mid}{\mid x - x' \mid}.$$

(M4) For every θ the Poisson equation

$$(1 - \hat{\Pi}_\theta) v_\theta = f(\theta, \cdot) - \int f(\theta, y) \Gamma_\theta(dy) \tag{34}$$

has a solution v_θ with the following properties of (M5).

[9] The algorithm here is subscripted with n rather than t.

[10] The functions v here and in the next two assumptions have no (or at least no apparent) connection with the value functions in the main body of the paper.

(M5) For all R there exist constants M_R and C_R so that

- **a)** $\sup_{|\theta|\leq R} \mid v_\theta(x) - v_\theta(x') \mid\leq M_R \mid x - x' \mid$,
- **b)** $\sup_{|\theta|\leq R} \mid v_\theta(x) \mid\leq C_R(1+ \mid x \mid)$,
- **c)** $\mid v_\theta(x) - v_{\theta'}(x) \mid\leq C_R \mid \theta - \theta' \mid (1+ \mid x \mid)$ for $\mid \theta \mid\leq R$, $\mid \theta' \mid\leq R$.

Let

$$\phi(\theta) \equiv \int f(\theta, y)\Gamma_\theta(dy) = E_{\Gamma_\theta}[f(\theta, y)].$$

Metivier and Priouret (1984) have shown the following theorem.

THEOREM 2 *Consider the algorithm defined by (32) and (33) and assume that (F) and (M1) through (M5) are satisfied. Suppose that* (γ_n) *is decreasing with* $\sum_n \gamma_n = +\infty$ *and* $\sum_n \gamma_n^{1+(p/2)} < \infty$, *where* $p \geq 2$ *is the constant entering (M2). Let* $\Omega_1 \equiv \{\sup_n \mid \theta_n \mid< \infty\}$. *Then there is a set* $\tilde{\Omega}_1 \subset \Omega_1$ *such that* $P(\Omega_1 \backslash \tilde{\Omega}_1) = 0$ *and with the following property: for every* θ^* *that is a locally asymptotically stable point of the equation*

$$\frac{d\theta(t)}{dt} = -\phi(\theta(t))$$

with domain of attraction $D(\theta^*)$ *and for every* $\omega \in \tilde{\Omega}_1$ *such that for some compact* $A \subset D(\theta^*)$, $\theta_n(\omega) \in A$ *for infinitely many n, the following holds:*

$$\lim_n \theta_n(\omega) = \theta^*$$

Remarks:

1. Suppose Y is always a member of some finite set $\{y_1, \ldots, y_q\}$ and assume that (M1) is satisfied. The operator $\hat{\Pi}_\theta$ can then simply be understood as a matrix operating on $\mathbf{R}^q$ via $\hat{\Pi}_\theta v_i = \sum_j (\hat{\Pi}_\theta)_{ij} v_j$, where $v_i \equiv v(y_i)$ for any given function $v : \mathbf{R}^k \to \mathbf{R}$. In particular, the q-dimensional vector corresponding to the function v_θ in (M4) can always be found by inverting the matrix $(I - \hat{\Pi}_\theta)$ on its range and applying it to the q-dimensional vector corresponding to the right hand side of equation (34), noting that the right hand side of that equation is indeed in the range of $I - \hat{\Pi}_\theta$, since it is orthogonal to the q-dimensional vector representing the unique invariant probability Γ_θ.

2. Suppose θ is always in some compact subset of $\mathbf{R}^d$ and Y is always a member of some discrete, finite subset of $\mathbf{R}^k$. Then assumptions (M2), (M4) and (M5a) and (M5b) are trivially satisfied and p in (M2) can be chosen to be $p = \infty$.

3. Suppose, $\hat{\Pi}_\theta$ is independent of θ. Then assumption (M3) is trivially satisfied.

References

Arthur, W. B. (1993): "On designing economic agents that behave like human agents," *Journal of Evolutionary Economics,* 3, 1-22.

—, J. Holland, B. LeBaron, R. Palmer, and P. Taylor (1994): "An artificial stock market," work in progress, Santa Fe Institute.

Banks, J. S. and R. K. Sundaram (1992): "Denumerable-Armed Bandits," *Econometrica,* 5, 1071-1096

Blume, Lawrence and David Easley (1993): "Rational Expectations and Rational Learning," mimeo, Cornell University.

Campbell, J. Y., and N. G. Mankiw (1989):"Consumption, Income, and Interest Rates: Reinterpreting the Time Series Evidence," *NBER Macroeconomics Annual,* 185-215

—, and — (1991): "The response of consumption to income: a cross-country investigation," *European Economic Review,* 35, 715-21.

Carroll, C. D., and L. H. Summers (1991): "Consumption growth parallels income growth: some new evidence," in *National Saving and Economic Performance,* ed. by B. D. Bernheim and J. Shoven, Chicago: Chicago University Press.

DeLong, B. J., and L. H. Summers (1986): "The changing cyclical variability of economic activity in the US," in *The American business cycle: continuity and change,* ed. by R. J. Gordon, Chicago: Chicago University Press.

Edelman, G. (1992): *Bright Air, Briliant Fire - On the matter of mind,* London: Penguin Books.

Flavin, M. (1981): "The Adjustment of Consumption to changing Expectations about future Income," *Journal of Political Economy*, 89, 974-1009.

Hall, R. E., and S. Mishkin (1982): "The sensitivity of consumption to transitory income: estimates from panel data on households," *Econometrica,* 50, 461-481.

Holland, J.H. (1986): *Adaptation in Natural and Artificial Systems*, 2. ed., Cambridge, MA: MIT Press.

Ingram, B. (1990): "Equilibrium Modeling of Asset Prices: Rationality Versus Rules of Thumb," *Journal of Business and Economic Statistics,* 1, 115-126.

KIYOTAKI, N., AND R. WRIGHT (1989): "On Money as a Medium of Exchange," *Journal of Political Economy*, 97, 927-954.

LAIBSON, D. (1993): "Golden eggs and hyperbolic discounting," draft, MIT.

LETTAU, M. (1994): "Risk-Taking Bias in a Financial Market with Adaptive Agents," draft, Princeton University.

—, AND H. UHLIG (1992): "How well do Artificially Intelligent Agents Eat Cake?" draft, Princeton University.

LJUNG, L., G. PFLUG, AND H. WALK (1992): *Stochastic Approximation and Optimization of Random Systems,* Basel: Birkhäuser Verlag.

MARCET, A.T AND T. J. SARGENT (1989): "Least Squares Learning and the Dynamics of Hyperinflation," in *Chaos, Sunspots, Bubbles and Nonlinearities*, ed. by W. A. Barnett, J. Geweke and K. Shell, Cambridge: Cambridge University Press.

MARIMON, R., E. MCGRATTAN, AND T.J. SARGENT (1990): "Money as a Medium of Exchange in an Economy with Artificially Intelligent Agents," *Journal of Economic Dynamics and Control,* 14, 329-373.

METIVIER, M., AND P. PRIOURET (1984): "Applications of a Kushner and Clark Lemma to General Classes of Stochastic Algorithms," *IEEE Transactions on Information Theory,* IT-30, 2, 140-151.

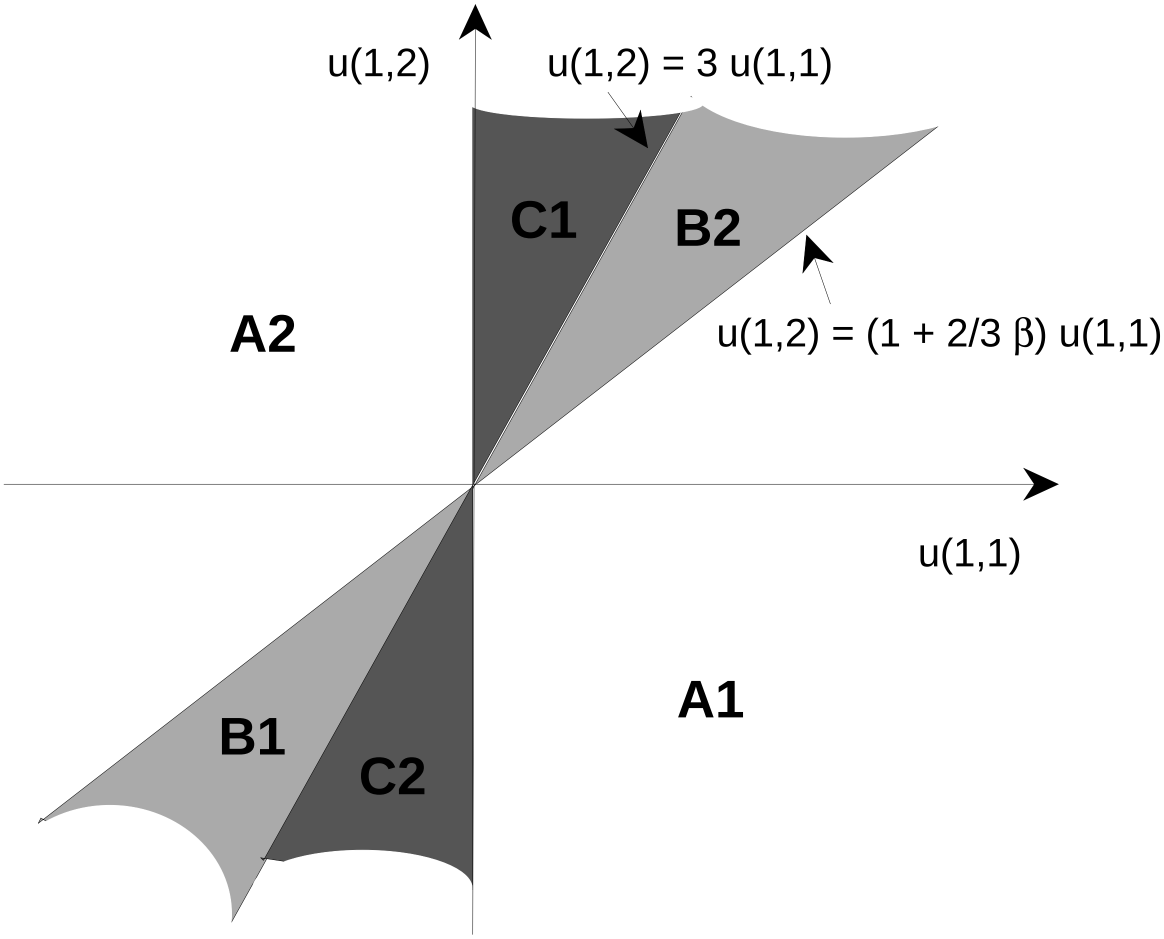
SARGENT, T. J. (1992): *Bounded Rationality in Macroeconomics,* draft, University of Chicago.

STOKEY, N. L., AND R. E. LUCAS, JR., WITH E. C. PRESCOTT (1989): *Recursive Methods in Economic Dynamics,* Cambridge, MA: Harvard University Press.

TAYLOR, J. B., AND H. UHLIG (1990): "Solving Nonlinear Stochastic Growth Models: A Comparison of Alterantive Solution Methods," *Journal of Business and Economic Statistics*, 8, 1 - 18.

ZELDES, S. P. (1989): "Consumption and liquidity constraints: an empirical investigation," *Journal of Political Economy*, 97, 305-346.

Figure 1



Area	Restriction	Case I $z_1 > z_2$	Case II $z_1 < z_2$	Case III $z_1 = z_2$	$h^*(1)$	Dyn. Prog. = CS ?
A1	$u(1,1) > 0$ $u(1,2) \leq 3u(1,1)$ $u(1,2) \leq (1 + 2/3\,\beta)u(1,1)$	Yes	No	No	1	yes
B2	$u(1,1) > 0$ $u(1,2) \leq 3u(1,1)$ $u(1,2) \geq (1 + 2/3\,\beta)u(1,1)$	Yes	No	No	2	no
A2	$u(1,1) \leq 0$ $u(1,2) > 3u(1,1)$ $u(1,2) \geq (1 + 2/3\,\beta)u(1,1)$	No	Yes	No	2	cannot
B1	$u(1,1) \leq 0$ $u(1,2) > 3u(1,1)$ $u(1,2) \leq (1 + 2/3\,\beta)u(1,1)$	No	Yes	No	1	no, but could
C1	$u(1,1) \leq 0$ $u(1,2) \leq 3u(1,1)$ $u(1,2) \geq (1 + 2/3\,\beta)u(1,1)$	No	No	Yes	random	cannot
C2	$u(1,1) > 0$ $u(1,2) > 3u(1,1)$ $u(1,2) \leq (1 + 2/3\,\beta)u(1,1)$	Yes	Yes	Yes	not unique	maybe

Table 1: Summary of Cases I, II, III