Learning and Risk Aversion*

Carlos Oyarzun
Texas A&M University

Rajiv Sarin
Texas A&M University

January 2007

Abstract

Learning describes how behavior changes in response to experience. We consider how learning may lead to risk averse behavior. A learning rule is said to be risk averse if it is expected to add more probability to an action which provides, with certainty, the expected value of a distribution rather than when it provides a randomly drawn payoff from this distribution, for every distribution. We characterize risk averse learning rules. The result reveals that the analyses of risk averse learning is isomorphic to that of risk averse expected utility maximizers. A learning rule is said to be monotonically risk averse if it is expected to increase the probability of choosing the actions whose distribution second order stochastically dominates all others in every environment. We characterize monotonically risk averse learning rules and show that such learning rules are risk averse.

1 Introduction

Expected Utility Theory (EUT) explains behavior in terms of preferences over, possibly subjective, probability distributions. Preferences are taken as

*We thank Ed Hopkins, Karl Schlag and Joel Sobel for helpful comments.

innate and exogenously given. In this paper we take an alternate approach in which the decision maker need not know that she is choosing among probability distributions. The behavior of the agent is explained by the manner in which the agent responds to experience. It is the response to experience that our approach takes as exogenously given. This paper reveals that the analysis of risk aversion can be developed, in this alternate approach, in much the same manner as in EUT.

We describe the behavior of an individual by the probabilities with which she chooses her alternative actions.[1] The manner in which the agent responds to experience is described by a *learning rule*. For given behavior today, a learning rule is a mapping from the action played in this period and the payoff obtained to behavior in the next period. We refer to the payoff distribution associated with each action as the *environment*. The environment is assumed to be unknown to the decision maker.

The focus of this paper is on how learning through experience explains behavior towards risk. In EUT, preferences are said to be risk averse if, for every probability distribution, the expected value of the distribution is preferred to the distribution itself.[2] In order to define risk aversion for a learning rule, we consider two environments that differ only in the distributions associated with one of the actions. In the first environment, the action provides a payoff drawn randomly from a given distribution, whereas in the second environment the expected payoff of that distribution is provided with certainty. We say that a learning rule is *risk averse* if, for any distribution, the learning rule is expected to add more probability to the action in the second environment than in the first. Furthermore, we require the above to hold regardless of the distributions over payoffs obtained from the other actions. Formally, the definition of when a learning rule is risk

[1] This is commonly assumed in models of learning. For a recent axiomatic foundation of probabilistic behavior in decision making see Gul and Pesendorfer (2006).

[2] See, for example, Mas-Colell, Whinston and Green (1995).

averse replaces the greater expected utility of the action that gives the expected value (rather than the distribution itself), that describes a risk averse expected utility maximizer, with being expected to add greater probability mass to the action that gives the expected value (rather than the distribution itself).

Our first result shows that a learning rule is risk averse if and only if the manner in which it updates the probability of the chosen action is a concave function of the payoff it receives. This result allows us to develop a theory of the risk attitudes of learning rules that is isomorphic to that of the risk attitudes of expected utility maximizers. Our analysis, hence, provides a bridge between decision theory and learning theory.[3]

In contrast to decision theory in which the agent selects an (optimal) action based on her beliefs about the distributions over payoffs of each action, a learning rule specifies how probability is moved among actions. Therefore, we could ask if the manner in which the learning rule shifts probability between actions results in the optimal action(s) being chosen with increased probability. This motivates us to investigate *monotonically risk averse* learning rules, which require that the learning rule is expected to add probability mass on the set of actions whose distributions second order stochastically dominate those of all other actions, in every environment.

We provide a characterization of monotonically risk averse learning rules. The result shows that how the learning rule updates the probability of unchosen actions, in response to the payoff obtained from the chosen action, plays a critical role. This response has to be a convex function of the payoff for each unchosen action. Furthermore, every monotonically risk averse learning rule satisfies a property we call *impartiality*. We say that a learning rule is impartial if there is no expected change in the probability of playing

[3]For earlier attempts to relate decision theory and learning theory see March (1996) and Simon (1956).

each action whenever all actions have the same payoff distribution. The restrictions imposed by impartiality on the functional form of the learning rule, together with the convex response of unchosen actions, imply that every monotonically risk averse learning rule is risk averse.

We also characterize *first order monotone* learning rules which require that the learning rule be expected to increase probability mass on the set of actions whose distributions first order stochastically dominate all others. We show that a learning rule is first order monotone if and only if it is impartial and the learning rule updates the probability of each unchosen action according to a decreasing function of the payoff obtained from the chosen action. This latter condition provides the analogue of the requirement that Bernoulli utility functions are increasing in EUT. This paper, therefore, provides a classification of learning rules that is much like the manner in which EUT classifies Bernoulli utility functions. In particular, our results allow us to determine how any given learning rule responds to any distribution over payoffs.

Our second and third characterizations provide a generalization of the results obtained by Börgers, Morales and Sarin (2004), who consider learning rules that are expected to increase probability on the expected payoff maximizing actions in every environment. They call such rules monotone. It is straightforward to show that monotone learning rules are monotonically risk averse and first order monotone. In Section 5 of this paper, we show by example, that the class of monotonically risk averse and first order monotone learning rules includes several well known learning rules that are not monotone.

There are some papers that investigate how learning rules respond to risk. March (1996) and Burgos (2002) investigate specific learning rules by way of simulations. Both consider an environment in which the decision maker has two actions, one of which gives the expected value of the other (risky) action

with certainty. As in our paper, the learning rules they consider update behavior using only the information on the payoff obtained from the chosen action. For the specific rules they consider, they show that they all choose the safe action more frequently over time. Denrell (2007) analytically shows that a class of learning rules choose the safe action more frequently in the long run. His result is obtained even if the decision maker follows an optimal policy of experimentation.[4]

2 Framework

Let A be the finite set of actions available to the decision maker. Action $a \in A$ gives payoffs according to the distribution function F_a. We shall refer to $F = (F_a)_{a \in A}$ as the environment the individual faces and we assume that it does not change from one period to the next. The agent knows the set of actions A but not the distributions F. The decision maker is assumed to know the finite upper and lower bounds on the set of possible payoffs $X = [x_{\min},\ x_{\max}]$. We may think of payoffs as monetary magnitudes.

The behavior of the individual is described by the probability with which she chooses each action. Let behavior today be given by the mixed action vector $\sigma \in \Delta(A)$, where $\Delta(A)$ denotes the set of all probability distributions over A. We assume that there is a strictly positive probability that each action is chosen today.[5] Taking the behavior of the agent today as given, a learning rule L specifies her behavior tomorrow given the action $a \in A$ she chooses and the monetary payoff $x \in X$ she obtains today. Hence, $L : A \times X \to \Delta(A)$. The learning rule should be interpreted as a "reduced form" of the true learning rule. The true learning rule may, for example, specify

[4] For a discussion of the manner in which evolution may affect behavior toward risk see Dekel and Schotchmer (1999) and Robson (1996a, 1996b).

[5] This assumption can be dropped with minor changes in the proofs though it would require additional notation.

how the decision maker updates her beliefs about the payoff distributions in response to her observations and how these beliefs are translated into behavior. If one combines the two steps of belief adjustment and behavior change we get a learning rule as we define.[6]

Let $L_{(a',x)}(a)$ denote the probability with which a is chosen in the next period if a' was chosen today and a payoff of x was received. For a given learning rule L and environment F, the expected movement of probability mass on action a is $f(a) := \sum_{a' \in A} \sigma_{a'} \int L_{(a',x)}(a)\, dF_{a'}(x) - \sigma_a$.

Denote the expected payoff associated with F_a by π_a. Let the distributions over payoffs of actions other than a be denoted by F_{-a}. The definition of when a learning rule is risk averse requires that if we replace the distribution F_a with another distribution $\tilde{F}_a$ which puts all probability on π_a and keep F_{-a} fixed, then the learning rule is expected to add more probability mass to a when it gives payoffs according to $\tilde{F}_a$ than when it gives payoffs according to F_a. This should be true for all a, F_a and F_{-a}. More formally, we introduce a second associated environment $\tilde{F} = \left(\tilde{F}_a, \tilde{F}_{-a}\right)$ in which $\tilde{F}_{-a} = F_{-a}$. Hence, environment $\tilde{F}$ has the same set of actions as environment F and the distribution over payoffs of all actions other than a are as in F. Let $\tilde{f}(a)$ denote the expected movement of probability mass on action a in the associated environment $\tilde{F}$.

Definition 1 *A learning rule L is* risk averse *if for all a and F, if $\tilde{F}_a$ places all probability mass on π_a then $\tilde{f}(a) \geq f(a)$.*

Risk seeking and risk neutral learning rules may be defined in the obvious manner. As the analysis of such learning rules involves a straightforward extension we do not pursue it further in the sequel. The risk aversion of

[6] Notice that we do not restrict the state space of the learning rule to be the probability simplex $\Delta(A)$. The examples in Section 5 illustrate this.

the learning rule may be considered a "local" concept as we have taken the current state as given. If the current state of learning is represented by an element $s \in S$ which is a subset of finite dimensional Euclidean space then the risk aversion of a learning rule "in the large" could be considered by defining a learning rule to be globally risk averse if it is risk averse at all states $s \in S$. This paper provides a first step in the analysis of globally risk averse learning rules.

The contrast between when a learning rule is risk averse and when an expected utility maximizer is risk averse is instructive. In EUT an individual is called risk averse if for all distributions F_a the individual prefers $\tilde{F}_a$ to F_a. Hence, the von Neumann-Morgenstern utilities v satisfy

$$v\left(\tilde{F}_a\right) = \int u(x)\, d\tilde{F}_a(x) \geq \int u(x)\, dF_a(x) = v(F_a),$$

where $u(\cdot)$ is often referred to as the Bernoulli utility function. A learning rule is called risk averse if for all actions a and distributions F_a the learning rule is expected to add more probability mass to an action that gives π_a with certainty than to an action that gives payoffs according to F_a regardless of the distributions F_{-a}. Hence, risk aversion in learning requires that

$$\tilde{f}(a) = \sum_{a' \in A} \sigma_{a'} \int L_{(a',x)}(a)\, d\tilde{F}_{a'}(x) - \sigma_a \geq \sum_{a' \in A} \sigma_{a'} \int L_{(a',x)}(a)\, dF_{a'}(x) - \sigma_a = f(a).$$

Notice that, whereas $v(.)$ in EUT depends only on the payoff distribution of a single action, $f(.)$ in learning theory depends on the distribution of the entire vector of distributions.

3 Risk Averse Learning

In this section we state our results regarding risk averse learning rules and their relationship to results concerning risk averse expected utility maximiz-

ers. The following definition provides some useful terminology.

Definition 2 *A learning rule L is* own-concave *if for all a, $L_{(a,x)}(a)$ is a concave function of x.*

A learning rule tells us how the probability of *each* action $a' \in A$ is updated upon choosing any action a and receiving a payoff x. Own-concavity of a learning rule places a restriction only on the manner in which the updated probability of action a depends upon x given that a is chosen.

Proposition 1 *A learning rule L is risk averse if and only if it is own-concave.*

Proof. We begin by proving that every own-concave learning rule is risk averse. Consider any own-concave learning rule L and environment $F = (F_a, F_{-a})$. Construct the associated environment $\tilde{F}$ in which $\tilde{F}_a$ places all probability mass on π_a (and $F_{-a} = \tilde{F}_{-a}$). By Jensen's inequality,

$$L_{(a,\pi_a)}(a) \geq \int L_{(a,x)}(a)\, dF_a(x)$$

$$\iff$$

$$\int L_{(a,x)}(a)\, d\tilde{F}_a(x) \geq \int L_{(a,x)}(a)\, dF_a(x)$$

$$\iff$$

$$\begin{aligned} & \sigma_a \int L_{(a,x)}(a)\, d\tilde{F}_a(x) + \sum_{a' \neq a} \sigma_{a'} \int L_{(a',x)}(a)\, d\tilde{F}_{a'}(x) - \sigma_a \\ \geq\ & \sigma_a \int L_{(a,x)}(a)\, dF_a(x) + \sum_{a' \neq a} \sigma_{a'} \int L_{(a',x)}(a)\, dF_{a'}(x) - \sigma_a \end{aligned}$$

$$\iff$$

$$\tilde{f}(a) \geq f(a).$$

Hence, the learning rule is risk averse.

We now turn to prove that every risk averse learning rule L is own-concave. We argue by contradiction. Suppose L is risk averse but not own-concave. Because L is not own-concave there exists an action a, payoffs $x', x'' \in [x_{\min}, x_{\max}]$ and $\lambda \in (0,1)$ such that

$$L_{(a,\lambda x'+(1-\lambda)x'')}(a) < \lambda L_{(a,x')}(a) + (1-\lambda) L_{(a,x'')}(a).$$

Now consider an environment F in which F_a gives x' with probability λ and x'' with probability $(1-\lambda)$ and the distributions of the other actions are given by F_{-a}. Consider the associated environment $\tilde{F}$ in which $\tilde{F}_a$ gives $\pi_a = \lambda x' + (1-\lambda) x''$ with probability one. Hence,

$$\begin{aligned} \int L_{(a,x)}(a)\, d\tilde{F}_a(x) &= L_{(a,\pi_a)}(a) \\ &< \lambda L_{(a,x')}(a) + (1-\lambda) L_{(a,x'')}(a) \\ &= \int L_{(a,x)}(a)\, dF_a(x). \end{aligned}$$

which implies $\tilde{f}(a) < f(a)$ by the argument above. Hence, the rule is not risk averse as we had assumed and we obtain a contradiction. ■

Proposition 1 shows that the own-concavity of a learning rule in learning theory plays an analogous role as the concavity of the Bernoulli utility function in EUT. In the latter theory the curvature properties of a Bernoulli utility function explain the individuals attitudes towards risk. In the theory of learning, the manner in which the learning rule updates the probability of the chosen action in response to the payoff it obtains explains how learning responds to risk. The proof reveals that for any action a the distributions of actions $a' \neq a$ do not play any role when we compare $\tilde{f}(a)$ and $f(a)$. This has the consequence that the theory of risk averse learning rules is isomorphic to the theory of risk averse expected utility maximizers.

For example, if $L_{(a,x)}(a)$ is a twice differentiable function of x, we can adapt the well known Arrow-Pratt measure of absolute risk aversion (Arrow (1965), Pratt (1964)) to find an easy measure of the risk aversion of a learning rule. Specifically, we define the coefficient of absolute risk aversion of a learning rule L for action a as

$$ar_{La}(x) = -\frac{\partial^2 L_{(a,x)}(a)/\partial x^2}{\partial L_{(a,x)}(a)/\partial x}.$$

In EUT a distribution F_a is said to be more risky than another $\tilde{F}_a$ if both have the same mean and every risk averse person prefers $\tilde{F}_a$ to F_a (see, e.g., Rothschild and Stiglitz (1970)). In this case it is usually said that $\tilde{F}_a$ second order stochastically dominates (sosd) F_a. The following Corollary shows that an analogous result applies in our case.

Corollary 1 *$\tilde{F}_a$ second order stochastically dominates F_a if and only if $\tilde{f}(a) \geq f(a)$ for all a for every risk averse learning rule.*

Proof. $\tilde{f}(a) \geq f(a)$ for every risk averse learning rule

$\Longleftrightarrow$

$\int L_{(a,x)}(a)\, d\tilde{F}_a(x) \geq \int L_{(a,x)}(a)\, dF_a(x)$ for every own-concave L

$\Longleftrightarrow$

$\tilde{F}_a$ second order stochastically dominates F_a. ■

For risk averse learning rules, imposing the requirement that probabilities of all actions must sum to one provides the obvious restrictions when there are only two actions. However, few restrictions are imposed when there are three or more actions. The property we study in the next section provides such restrictions.

4 Monotonic Risk Aversion

The definition of a risk averse learning rule was inspired by standard decision theory. Learning, however, differs in many respects from choice. Whereas in decision theory a single (and optimal) action is chosen, in learning theory probability is moved between actions. For learning, it then appears reasonable to ask whether probability on the optimal action is increased from one period to the next. Our next definition introduces such a property. Specifically, a learning rule is said to be monotonically risk averse if it is expected to increase probability on the best actions, in a sosd sense, in every environment.

Let A^* denote the set of actions that second order stochastically dominate all other actions. That is, $A^* = \{a : F_a \text{ sosd } F_{a'} \text{ for all } a' \in A\}$. Clearly, if $A^* = A$ we have that $F_a = F_{a'}$ for all $a, a' \in A$. For any subset $\hat{A} \subset A$, let $f\left(\hat{A}\right) := \Sigma_{a \in \hat{A}} f(a)$.

Definition 3 *A learning rule L is* monotonically risk averse *if in all environments we have that $f(A^*) \geq 0$.*

Correspondingly, we say that a learning rule is monotonically risk seeking if $f(A^*) \leq 0$ in every environment and a learning rule is monotonically risk neutral if it is monotonically risk averse and monotonically risk seeking. The analysis of such rules is analagous to the analysis of monotonically risk averse learning rules provided below. Note that, when A^* is empty, expected utility maximization by a risk averse agent places no specific restrictions on behavior. This has the analogue, in this paper, that no restrictions are placed on the movement of probability when A^* is empty.

4.1 Characterization

The following definition introduces some useful terminology.

Definition 4 *A learning rule L is* cross-convex *if for all a, $L_{(a'x)}(a)$ is convex in x for all $a' \neq a$.*

We shall see in the next result that all monotonically risk averse learning rules have the feature that if all actions have the same distribution of payoffs then there is no expected movement in probability mass on any action. We call such learning rules *impartial*.

Definition 5 *A learning rule L is* impartial *if $f(a) = 0$ for all a whenever $F_a = F_{a'}$ for all $a, a' \in A$.*

The set of impartial learning rules is related to the unbiased learning rules studied in Börgers, Morales and Sarin (2004). Unbiasedness requires that no probability mass is expected to be moved among actions when all have the same expected payoff. Clearly, the set of impartial learning rules is larger than the set of unbiased learning rules. Furthermore, it is straightforward to see that unbiased learning rules cannot respond to aspects of the distribution of payoffs other than the mean.[7]

Proposition 2 *A learning rule L is monotonically risk averse if and only if (i) $\sigma_a = \sum_{a' \in A} \sigma_{a'} L_{(a',x)}(a)$ for all a, and (ii) L is cross-convex.*

Our proof begins with two Lemmas. The first shows that all monotonically risk averse learning rules are impartial and the second characterizes impartial learning rules.

Lemma 1 *If the learning rule L is monotonically risk averse then it is impartial.*

[7]From the analysis below, it is not difficult to see that a learning rule is unbiased if and only if it is monotonically risk neutral.

Proof. The proof is by contradiction. Suppose L is monotonically risk averse but there exists an environment F with $A = A^*$ and $f(a) > 0$ for some $a \in A$. If F_a does not place strictly positive probability on (x_{min}, x_{max}), then consider the environment $\widehat{F}$ such that, for all action $a \in A$, the probabilities of $x_{\min}$ and $x_{\max}$ are $(1-\varepsilon)$ times their corresponding probabilities in the environment F, and the probability of some $x \in (x_{min}, x_{max})$ is ε. If F_a places strictly positive probability on (x_{min}, x_{max}), then let $\widehat{F} = F$. We now construct the environment $\widetilde{F}$ in which $\widetilde{F}_a$ is a mean preserving spread of $\widehat{F}_a$ and $\widetilde{F}_{a'} = \widehat{F}_{a'}$ for all $a' \neq a$. Specifically, suppose that $\widetilde{F}_a$ is obtained by assigning to every interval $I \subset [x_{min}, x_{max}]$ only $(1 - \varepsilon)$ times the probability it had under $\widehat{F}_a$ and then adding $(\widehat{\pi}_a - x_{min})\varepsilon/(x_{max} - x_{min})$ on the probability of x_{max} and $(x_{max} - \widehat{\pi}_a)\varepsilon/(x_{max} - x_{min})$ on the probability of x_{min}. By construction, $\widetilde{F}_{a'}$ sosd $\widetilde{F}_a$ for all $a' \neq a$. It follows that $\widetilde{A}^* = A\backslash\{a\}$. Since $\widetilde{f}(a)$ can be written as a continuous function in ε, there exists a small enough ε such that $\widetilde{f}(a) > 0$. This contradicts that L is monotonically risk averse. ■

Lemma 2 *A learning rule L is impartial if and only if for all $a \in A$ and $x \in X$, it satisfies $\sigma_a = \sum_{a' \in A} \sigma_{a'} L_{(a',x)}(a)$.*

Proof. *Necessity.*

Consider an environment where all the actions pay x with probability one. Then, for all $a \in A$,

$$f(a) = \sum_{a' \in A} \sigma_{a'} L_{(a',x)}(a) - \sigma_a.$$

Therefore, in order to be impartial L must satisfy

$$\sigma_a = \sum_{a' \in A} \sigma_{a'} L_{(a',x)}(a).$$

Sufficiency.

Consider the environment F such that $F_a = F_{a'}$ for all a, $a' \in A$.

$$\begin{aligned} f(a) &= \sum_{a' \in A} \sigma_{a'} \int L_{(a',x)}(a) dF_{a'}(x) - \sigma_a \\ &= \int \sum_{a' \in A} \sigma_{a'} L_{(a',x)}(a) dF_a(x) - \sigma_a \\ &= 0. \end{aligned}$$

The second statement follows from the fact that all the distributions are the same, and the third statement follows from the hypothesis. ■

Proof. We now proceed to complete the proof of Proposition 2**.**

Sufficiency.

Consider $a \in A^*$

$$\begin{aligned} f(a) &= \sigma_a \int L_{(a,x)}(a) dF_a(x) + \sum_{a' \neq a} \sigma_{a'} \int L_{(a',x)}(a) dF_{a'}(x) - \sigma_a \\ &= \int [\sigma_a - \sum_{a' \neq a} \sigma_{a'} L_{(a',x)}(a)] dF_a(x) + \sum_{a' \neq a} \sigma_{a'} \int L_{(a',x)}(a) dF_{a'}(x) - \sigma_a \\ &= \sum_{a' \neq a} \sigma_{a'} [\int L_{(a',x)}(a) dF_{a'}(x) - \int L_{(a',x)}(a) dF_a(x)] \\ &\geq 0. \end{aligned}$$

The second statement follows from Lemmas 1 and 2 and the last inequality follows from the fact that $a \in A^*$ and the convexity of the functions $L_{(a',x)}(a)$ for all $a' \in A \backslash \{a\}$.

Necessity.

We argue by contradiction. Suppose that for some $a \in A$ and some $a' \in A \backslash \{a\}$, $L_{(a',x)}(a)$ is not convex. Therefore there exists x', x'', $\lambda \in (0,1)$ and $x := \lambda x' + (1-\lambda)x''$ such that $\lambda L_{(a',x')}(a) + (1-\lambda) L_{(a',x'')}(a) < L_{(a',x)}(a)$. Consider an environment where $a' \in A \backslash \{a\}$ pays x' with probability λ, and x'' with probability $(1-\lambda)$. Action a pays x with probability one, and all

the other actions in the set, if any, pay x with probability $1-\varepsilon$, x' with probability $\varepsilon\lambda$, and x'' with probability $\varepsilon(1-\lambda)$. Clearly, $A^* = \{a\}$. From the sufficiency part we know

$$\begin{aligned} f(a) &= \sum_{a' \neq a} \sigma_{a'} [\int L_{(a',x)}(a) dF_{a'}(x) - \int L_{(a',x)}(a) dF_a(x)] \\ &= \sigma_{a'} [\lambda L_{(a',x')}(a) + (1-\lambda) L_{(a',x'')}(a) - L_{(a',x)}(a)] \\ &\quad + \varepsilon \sum_{a'' \neq a, a'} \sigma_{a''} [\lambda L_{(a'',x')}(a) + (1-\lambda) L_{(a'',x'')}(a) - L_{(a'',x)}(a)]. \end{aligned}$$

Therefore, for small enough ε, $f(a) < 0$. ■

Monotonic risk aversion places restrictions on how the learning rule updates the probability of each unchosen action as a function of the action chosen and payoff obtained. In particular, it requires this function be convex in the payoff received, for each unchosen action. Furthermore, we show that all such rules have the weak consistency property of impartiality. Because every impartial and cross-convex learning rule is own-concave we have the following Corollary.

Corollary 2 *Every monotonically risk averse learning rule L is risk averse.*

4.2 First Order Monotonicity

In EUT a distribution F_a is said to first order stochastically dominate (fosd) another $\tilde{F}_a$ if every individual with an increasing (Bernoulli) utility function prefers the former. In the context of our analysis, we would like to identify the learning rules that are expected to add probability mass on the set of actions whose distributions fosd the distributions of all the other actions, in every environment. We call such learning rules *first order monotone*. Let $A^{**} := \{a \in A : a \text{ fosd } a' \text{ for all } a' \in A\}$.

Definition 6 *A learning rule L is* first-order monotone *if $f(A^{**}) \geq 0$ in every environment.*

First order monotone learning rules can be characterized in the same manner as monotonically risk averse learning rules. In particular, these rules need to be impartial but instead of being cross-convex they require the response of the probabilities of playing the unchosen actions to be decreasing in the obtained payoff.

Definition 7 *A learning rule L is* cross-decreasing *if for all a, $L_{(a'x)}(a)$ is decreasing in x for all $a' \neq a$.*

Proposition 3 *A learning rule is first-order monotone if and only if (i) $\sigma_a = \sum_{a' \in A} \sigma_{a'} L_{(a',x)}(a)$ for all $a \in A$ and (ii) L is cross-decreasing.*

Proof. See the Appendix. ■

The notion of risk averse learning of Section 3 may be extended to first order stochastic dominance in a similar manner. We can identify a set of learning rules such that for every action a and distribution F_a, if that distribution is replaced by a distribution $\widetilde{F}_a$, such that $\widetilde{F}_a$ fosd F_a, then $\widetilde{f}(a) \geq f(a)$, in every environment F. It is easy to show that this set of learning rules is equivalent to the set of learning rules for which $L_{(a,x)}(a)$ is increasing in x for all $a \in A$.

5 Examples

The Cross learning rule (Cross (1973)) is given by

$$\begin{aligned} L_{(a,x)}(a) &= \sigma_a + (1 - \sigma_a)\, x \\ L_{(a',x)}(a) &= \sigma_a - \sigma_a x \quad \forall a' \neq a, \end{aligned}$$

where $x \in [0,1]$. It is easily seen that the Cross learning rule is impartial. Furthermore, its cross-components are affine transformations of x. Therefore this rule is monotonically risk neutral and hence it is also risk neutral. It is also clearly first order monotone.

The Roth and Erev (1995) learning rule describes the state of learning s of an agent by a vector $v \in R_{++}^{|A|}$. The vector v describes the decision makers "attraction" to choose any of her $|A|$ actions. Given v, the agents behavior is given by $\sigma_a = v_a / \Sigma_{a'} v_{a'}$ for all a. If the agent plays a and receives a payoff of x then she adds x to her attraction to play a, leaving all other attractions unchanged. Hence, the Roth and Erev learning rule is given by

$$\begin{aligned} L^v_{(a,x)}(a) &= \frac{v_a + x}{\Sigma_{a''} v_{a''} + x} \\ L^v_{(a',x)}(a) &= \frac{v_a}{\Sigma_{a''} v_{a''} + x} \qquad \forall a' \neq a, \end{aligned}$$

where $x \in [0, x_{\max}]$ and the superscript v on the learning rule defines it at that state of learning. Using Lemma 2, it is easy to check that this rule is impartial. Observing that the cross-components $L^v_{(a',x)}(a)$ are decreasing convex functions of x for all $a' \neq a$, we see that this learning rule is first order monotone and monotonically risk averse. The coefficient of absolute risk aversion of this learning rule is $ar_{L_a} = 2/(\Sigma_{a'} v_{a'} + x)$ for all a. Clearly, ar_{L_a} decreases as $\Sigma_{a'} v_{a'}$ increases and hence this learning rule exhibits declining absolute risk aversion. Note that this rule satisfies none of the properties studied by Börgers, Morales and Sarin (2004) who have shown that this rule is neither monotone nor unbiased.

Our next example considers the weighted return model studied in March (1996). This learning rule is risk averse but may not be monotonically risk averse. The state of learning is described by a vector of attractions $v \in R_{++}^{|A|}$. Given v, the agents behavior is given by $\sigma_a = v_a / \Sigma_{a'} v_{a'}$ for all a. If action a is chosen and receives a payoff of x then she adds $\beta(x - v_a)$ to her attraction

of a, where $\beta \in (0,1)$ is a parameter, leaving all other attractions unchanged. Thus, the learning rule may be written as

$$\begin{aligned} L^{\upsilon}_{(a,x)}(a) &= \frac{\upsilon_a + \beta(x - \upsilon_a)}{\sum_{a'' \in A} \upsilon_{a''} + \beta(x - \upsilon_a)} \\ L^{\upsilon}_{(a',x)}(a) &= \frac{\upsilon_a}{\sum_{a'' \in A} \upsilon_{a''} + \beta(x - \upsilon_{a'})} \ \forall a' \neq a, \end{aligned}$$

where $x \in [0, x_{\max}]$. It follows that this learning rule is risk averse (as $L^{\upsilon}_{(a,x)}(a)$ is a concave function of x). However, this learning rule is monotonically risk averse and first order monotone only if $\upsilon_a = \upsilon_{a'}$ for all $a, a' \in A$ (because otherwise it fails to be impartial). The analysis of the average return model studied by March (1996) which replaces β with $\beta_a = 1/(\kappa_a + 1)$, where κ_a is the number of times action a has been chosen in the past, is similar.

Another learning rule that has received considerable attention is the logistic fictitious play with partial feedback studied by Fudenberg and Levine (1998, section 4.8.4). The agent is described by the $|A| \times 2$ matrix (υ, κ) where κ_a denotes the number of times action a has been chosen, $\kappa = (\kappa_a)_{a \in A}$, and $\upsilon = (\upsilon_a)_{a \in A}$ gives the vector of attractions. The next period attraction of an action that was chosen today is its current attraction plus $(x - \upsilon_a)/(\kappa_a + 1)$. The attractions of unchosen actions are not updated. The learning rule is specified as

$$\begin{aligned} L^{\upsilon,\kappa}_{(a,x)}(a) &= \frac{e^{\upsilon_a + (x - \upsilon_a)/(\kappa_a + 1)}}{e^{\upsilon_a + (x - \upsilon_a)/(\kappa_a + 1)} + \sum_{a' \neq a} e^{\upsilon_{a'}}} \\ L^{\upsilon,\kappa}_{(a',x)}(a) &= \frac{e^{\upsilon_a}}{e^{\upsilon_{a'} + (x - \upsilon_{a'})/(\kappa_{a'} + 1)} + \sum_{a'' \neq a'} e^{\upsilon_{a''}}} \quad \forall a' \neq a \end{aligned}$$

This learning rule is neither risk averse nor risk seeking because the curvature of $L^{\upsilon,\kappa}_{(a,x)}(a)$ depends on the payoff obtained. Specifically, $L^{\upsilon,\kappa}_{(a,x)}(a)$ is concave in x (at σ) for the part of the domain in which $x \geq \upsilon_a +$

$(\kappa_a + 1)[\ln(1-\sigma_a) - \ln\sigma_a]$ and is convex otherwise. Nevertheless, this rule is first order monotone when $v_a = v_{a'}$ and $\kappa_a = \kappa_{a'}$ for all $a, a' \in A$. The analysis of the risk attitudes of several other learning rules (e.g. minimal information versions of Camerer and Ho (1999) and Rustichini (1999)), which use a logistic transformation of the attractions to obtain the probabilities with which each action is chosen, is closely related.

6 Discussion

In models of learning probability distributions over actions change from one period to the next in response to some experience. Monotonic risk aversion focuses only on the manner in which the probability of the best actions being chosen changes from one period to the next. This parallels decision theory's focus on the best action. However, for learning, we could look at the entire probability distribution chosen in the next period. Noting that this probability distribution on actions generates a (reduced) distribution over payoffs, which is a weighted average of the payoff distribution of each of the actions, we could ask whether the learning rule is such that expected behavior tomorrow generates a distribution over payoffs which second order stochastically dominates that of today in every environment. Such a property turns out to be too restrictive. It can be shown that, in environments with only two actions, the only learning rules which satisfy this condition are the unbiased rules studied by Börgers, Morales and Sarin (2004). Unbiased rules exhibit zero expected movement in probability mass when all actions have the same expected payoffs. Such rules satisfy the above condition in a trivial manner because the expected distribution tomorrow is the same as today.[8] Restricting the set of environments on which the improvement is

[8]It can also be shown that unbiased learning rules are the only learning rules which are continuous in x for all $a, a' \in A$ and satisfy $\sum_a (\sigma_a + f(a)) F_a$ sosd $\sum_a \sigma_a F_a$ in every environment.

required would lead us to identify a larger class of learning rules.[9] We do not pursue this approach in the current paper.

All the properties we have studied in this paper have referred to the expected movement of a learning rule. This arises naturally when describing the behavior of the population of individuals each of whom faces the same decision problem. The expected movement of a learning rule has also been studied on many previous occasions when interest has focused on the long run properties of a learning rule. As is well known under conditions of slow learning the actual movement of a learning rule closely approximate it's expected movement.[10] Combining properties of the expected movement and of the speed of learning inform us about the long term properties of learning rules.

This paper has focused on short term and local properties of learning. Often, however, the long run properties of learning rules are of interest and such an analysis requires us to look at properties that hold globally. That is, the local property would need to hold for each state of learning. This subject, which would require the analysis of risk attitudes at each state of learning, is outside the scope of the present study and we leave it for future work. For the examples we discussed in Section 5, we are often able to say if the learning rule is globally (monotonically) risk averse. This is obviously true for the Cross learning rule and the Roth and Erev learning rule which are monotonically risk neutral and monotonically risk averse at all states, respectively. The weighted return model of March is globally risk averse though not globally monotonically risk averse and the logistic fictitious play learning rule is not globally risk averse.

[9] For example, we could consider learning rules that satisfy $\sum_a (\sigma_a + f(a)) F_a$ sosd $\sum_a \sigma_a F_a$ in every environment that is completely ordered by the sosd relation. It can be shown that a learning rule is continuous in payoffs and satisfies this condition if and only if it is monotonically risk averse.

[10] See, for example, Börgers and Sarin (1997).

APPENDIX

Proof of Proposition 3

We begin with the following Lemma, the proof of which closely parallels that of Lemma 1.

Lemma 3 *Every first-order monotone learning rule L is impartial.*

Proof. We argue by contradiction. Consider an environment F with $A = A^{**}$ and suppose that $f(a) < 0$ for some $a \in A$. Now we construct an environment $\widehat{F}$ where $\widehat{A}^{**} = \{a\}$. We construct $\widehat{F}_a$ by assigning to every interval $I \subset X$ only $(1-\varepsilon)$ times the probability it had under F_a and adding ε to the probability of $x_{\max}$. We construct $\widehat{F}_{a'}$ for all $a' \in A\backslash\{a\}$ by assigning to every interval $I \subset X$ only $(1-\varepsilon)$ times the probability it had under $F_{a'}$ and then adding ε to the probability of $x_{\min}$. Clearly, $\widehat{A}^{**} = \{a\}$. Since $\widehat{f}(a)$ can be written as a continuous function in ε, for small enough ε we have $\widehat{f}(a) < 0$. Therefore L is not first-order monotone. ■

The proof of the Proposition follows.

Proof. *Necessity.*

The necessity of (i) follows from Lemma 3 and Lemma 2. To prove the necessity of (ii) we argue by contradiction. Suppose that for some $a \in A$ and $a' \in A\backslash\{a\}$, there are x and x' with $x' < x$ and $L_{(a',x)}(a) > L_{(a',x')}(a)$. Consider the environment F where action a pays x with probability one and action a' pays x' with probability one. All the other actions $a'' \in A\backslash\{a, a'\}$, if any, pay x with probability $1-\varepsilon$ and x' with probability ε. Clearly, $A^{**} = \{a\}$. From Lemma 2 and Lemma 3, we have that

$$\begin{aligned} f(a) &= \sum_{a' \neq a} \sigma_{a'} \left[\int L_{(a',x)}(a) dF_{a'}(x) - \int L_{(a',x)}(a) dF_a(x) \right] \\ &= \sigma_{a'} \left[L_{(a',x')}(a) - L_{(a',x)}(a) \right] + \varepsilon \sum_{a'' \neq a,a'} \sigma_{a''} \left[L_{(a'',x')}(a) - L_{(a'',x)}(a) \right]. \end{aligned}$$

For small enough ε, $f(a) < 0$, which contradicts first-order monotonicity.

Sufficiency.

As in the proof of Proposition 2, consider $a \in A^{**}$, then

$$\begin{aligned} f(a) &= \sum_{a' \neq a} \sigma_{a'} \left[\int L_{(a',x)}(a) dF_{a'}(x) - \int L_{(a',x)}(a) dF_a(x) \right] \\ &\geq 0. \end{aligned}$$

The last inequality follows from the fact that $a \in A^{**}$ and the fact that $L_{(a',x)}(a)$ is decreasing for all $a' \in A \backslash \{a\}$. ■

REFERENCES

1. Arrow, K.J. (1965): *Aspects of the Theory of Risk-Bearing*, Helsinki: Yrjo Hahnsson Foundation.

2. Börgers, T., A.J. Morales and R. Sarin (2004): "Expedient and monotone learning rules," *Econometrica*, 72, 383-405.

3. Börgers, T., and R. Sarin (1997): "Learning through reinforcement and replicator dynamics," *Journal of Economic Theory*, 77, 1-14.

4. Burgos, A. (2002): "Learning to deal with risk: What does reinforcement learning tell us about risk attitudes," *Economics Bulletin*, 4, 1-13.

5. Camerer, C. and T.H. Ho (1999): "Experience-weighted attraction learning in normal-form games," *Econometrica*, 67, 827-874.

6. Cross, J.G. (1973): "A stochastic learning model of economic behavior," *Quarterly Journal of Economics*, 87, 239-266.

7. Dekel, E. and S. Scotchmer (1999): "On the evolution of attitudes towards risk in winner-take-all games," *Journal of Economic Theory*, 87, 125-143.

8. Denrell, J. (2007): "Adaptive learning and risk taking," *Psychological Review*, 114, 177-187.

9. Fudenberg, D. and D. Levine (1998): *The Theory of Learning in Games*, Cambridge: MIT Press.

10. Gul, F. and W. Pesendorfer (2006): "Random expected utility," *Econometrica*, 74, 121-146.

11. March, J.G. (1996): "Learning to be risk averse," *Psychological Review*, 103, 309-319.

12. Mas-Colell, A., M.D. Whinston and J.R. Green (1995): *Microeconomic Theory*. New York: Oxford University Press.

13. Pratt, J.W. (1964): "Risk aversion in the small and in the large," *Econometrica*, 32, 122-136.

14. Robson, A.J. (1996a): "The evolution of attitudes towards risk: Lottery tickets and relative wealth," *Games and Economic Behavior*, 14, 190-207.

15. Robson, A.J. (1996b): "A biological basis for expected and non-expected utility preferences," *Journal of Economic Theory*, 68, 397-424.

16. Roth, A.E. and I. Erev (1995): "Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term," *Games and Economic Behavior*, 8, 164-212.

17. Rothschild, M. and J.E. Stiglitz (1970): "Increasing risk: I. A definition," *Journal of Economic Theory*, 2, 225-243.

18. Rustichini, A. (1999): "Optimal properties of stimulus-response learning models," *Games and Economic Behavior*, 29, 244-273.

19. Simon, H.A. (1956): "A comparison of game theory and learning theory," *Psychometrika*, 21, 267-272.