# Learning and Risk Aversion*

Carlos Oyarzun
Texas A&M University

Rajiv Sarin
Texas A&M University

September 2005

**Abstract**

A learning rule is risk averse if, for all distributions, it is expected to add more probability mass to an action that gives the expected value of a distribution with certainty than to an action that gives the distribution itself. We provide several characterizations of risk averse learning rules. Our analysis reveals that the theory of risk averse learning rules is isomorphic to the theory of risk averse expected utility maximizers. We consider two additional properties of the risk attitudes of a learning rule. We show that both are sufficient for a learning rule to be risk averse and characterize the set of all learning rules that satisfy the more stringent of these sufficient conditions.

## 1 Introduction

The literature on choice under uncertainty defines an individual as risk averse if, for all distributions over monetary outcomes, the decision maker prefers the action which gives the expected value of the distribution with certainty to an action that gives the distribution itself.[1] Consequently, the expected utility maximizer assigns a higher (expected) utility to the action with the former distribution. When can learning be said to be risk averse?

*We are happy to acknowledge the comments of Joel Sobel on a previous version of this paper.

[1]See, e.g., Mas-Colell, Whinston and Green (1995).

We adopt the view, common in psychology, that learning consists of the change in an individual's behavior that occurs in response to some experience.[2] The behavior of an individual is described by the probability with which the individual chooses her alternative actions. We consider learning rules in which the experience of the individual comprises of the action chosen and the payoff it obtains. The agent is not assumed to know the probability according to which any action gives different payoffs.

A natural notion of when a learning rule is risk averse is if, for all distributions over monetary outcomes, it adds more probability to an action which gives the expected value of the distribution with certainty than to an action that gives the distribution itself. Given that both choices and payoffs are random it does not seem possible for any learning rule to satisfy this notion with probability 1. To define the risk attitude of a learning rule we shall consider the expected change in behavior it implies. By considering the *expected* change in probability, the uncertainty in the direction of learning resulting from both payoffs and behavior being stochastic is integrated out.

Specifically, we call a learning rule *risk averse* if, for all distributions of payoffs, the learning rule is expected to add more probability mass to an action that gives the expected value of a distribution with certainty than to an action that gives the distribution itself. 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 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).

To illustrate that the definition of risk averse learning rules allows us to develop a theory that is analogous to the expected utility theory of risk we introduce definitions of the certainty equivalent of a learning rule and the risk premium of a learning rule. These definitions also replace the preference functional of expected utility theory with the functional describing the expected movement of probability mass. Roughly speaking, the certainty equivalent of a learning rule for a distribution asks how much an action would have to pay with certainty for it to result in the same expected movement of

[2] A similar approach is adopted in Börgers, Morales and Sarin (2004).

probability mass as the action which gives the distribution, when all other things between the two environments are held constant. The risk premium of a learning rule for a distribution is the difference between the expected value of the distribution and the amount of money that an action would have to offer with certainty for the expected movements to coincide.

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. The proof of the result reveals that when comparing the expected movement of probability mass on an action with two alternative possible distributions we do not need to be concerned about the distributions of the other actions. This feature 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. In particular, we show that a learning rule is risk averse if and only if the certainty equivalent of the learning rule is no greater than the expected value of the distribution, for all distributions. Furthermore, we show that risk aversion of a learning rule is equivalent to the risk premium of the learning rule being non-negative for all distributions.

Next, we provide a result of when one learning rule is more risk averse than another that is parallel to the well known results of Arrow (1971) and Pratt (1964). Here we introduce the coefficient of absolute risk aversion of a learning rule which is defined analogously to the Arrow-Pratt measure of absolute risk aversion. We also show that one distribution second order stochastically dominates another if and only if all risk averse learning rules are expected to add more probability mass to the former. This result parallels the well known result of Rothschild and Stiglitz (1970).

The notion of risk aversion of a learning rule we formalize in this paper allows us to develop a theory of the risk attitudes of learning rules that parallels the development of the expected utility theory of risk. However, in contrast to decision theory in which the agent selects an action based on the knowledge of the distributions of payoffs each action generates, a learning rule "selects" a probability distribution on the set of actions given the behavior today, the action chosen and payoff received. This probability distribution on actions generates a (compound) distribution over payoffs which is a weighted average of the payoff distribution of each of the actions. We could hence ask whether the learning rule is such that expected behavior tomorrow generates a distribution of payoffs tomorrow which second order stochastically dominates that of today. We call a learning rule that satisfies this property in a specific class of environments *super risk averse*. We also investigate a related property that requires 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. We call the set of learning rules which satisfy this property *monotonically risk averse*.

We provide some necessary and some sufficient conditions for a learning rule to be super risk averse. We show that every super risk averse learning rule is risk averse. Our proof reveals that a necessary condition for super risk averse rules, that may not be shared by risk averse learning rules, is that they are not expected to move probability mass on any action when all of them give the same distribution of payoffs. A sufficient condition for a learning rule to be super risk averse is that it is monotonically risk averse. A characterization of monotonically risk averse learning rules is provided.

Our analysis of super risk averse and monotonically risk averse learning rules extends the analysis provided by Börgers et al (2004) who study absolutely expedient and monotone learning rules. A learning rule is absolutely expedient if, whatever the environment, behavior tomorrow is expected to have a higher expected payoff than today. A learning rule is monotone if in every environment it is expected to add probability mass to the set of expected payoff maximizing actions. Absolutely expedient and monotone learning rules are risk-neutral as they cannot respond to differences in riskiness between two distributions. The necessary and sufficient conditions derived by Börgers et al for absolutely expediency have analogues in our results on super risk averse learning rules. The relation is particularly clear between monotone learning rules and monotonely risk averse learning rules. Whereas monotone learning rules allow for affine transformations of payoffs before applying the Cross (1973) learning rule, monotonely risk averse learning rules allow for concave transformations of payoffs before applying the Cross learning rule.[3]

We conclude by discussing the risk attitudes of some well-known learning rules. We show that the Cross (1973) learning rule is risk neutral (risk averse

[3] In each case the transformation of payoffs is allowed to depend on the action that was chosen and the action whose probability is being updated.

and risk seeking), super risk neutral and monotonely risk neutral. The Roth-Erev (1995) learning rule is risk averse, super risk averse and monotonely risk averse. Lastly, we also study the risk attitudes of the stochastic payoff assessment rule of Fudenberg and Levine (1998).

There are some papers that investigate how learning rules respond to risk. March (1996) and Burgos (2002) investigate these properties for 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. Della Vigna and Li Calzi (2001) analytically study the long term properties of a learning rule that supposes the agent maximizes the probability of getting a payoff above an aspiration level. The aspiration level is adjusted in the direction of the experienced payoff. They show that this learning rule converges to make risk neutral choices in the long run provided that the distribution of rewards from an action are symmetric.[4]

This paper is structured as follows. In the next section we introduce the framework for the analysis. Section 3 provides several characterizations of risk averse learning rules. It also gives results that develop the relation between the risk attitudes of learning rules with those of expected utility maximizers. Section 4 studies super risk averse learning rules and monotonely risk averse learning rules. Section 5 discusses the risk attitudes of some well known rules. Section 6 concludes.

# 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 $F_a$.[5] 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

[4]From an evolutionary viewpoint, Robson (1996) shows that risk neutral behavior is selected, if the environment is fixed.

[5]Hence, each action $a$ can be viewed as a lottery.

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 shall 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.

Taking the behavior of the agent today as given, a learning rule $L$ specifies the behavior of the agent 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 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. We assume that $L_{(a,x)}(a)$ is a strictly increasing function of $x$ for all $a$. 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 $\pi_a$. 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 $\pi_a$ and keep $F_{-a}$ fixed, then the learning rule would expect to add more probability mass to $a$ when it gives $\tilde{F}_a$ than when it gives $F_a$. 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

[6] Notice that we do not restrict the state space of the learning rule to be the probability simplex $\Delta(A)$.

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 $F_a$, if $\tilde{F}_a$ places all probability mass on $\pi_a$ 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 of the current paper we do not pursue it further in the sequel. Note that risk aversion of the learning rule may be considered a "local" concept as we have taken behavior today as given. Hence, it pertains only to the current state of learning or behavior $\sigma$. We could also consider risk aversion of a learning rule "in the large" by defining a learning rule to be globally risk averse if it is risk averse at all states $\sigma \in \Delta(A)$. 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 expected utility theory 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 $\pi_a$ with probability one than to an action that gives $F_a$ regardless of the distributions $F_{-a}$. Hence, risk aversion in the theory of learning requires that

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

Whereas $v(.)$ in expected utility theory 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 Rules

In this section we state our results regarding risk averse learning rules and their relationship with results concerning risk averse expected utility maximizers. 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. The following result characterizes risk averse learning rules and shows that they are own-concave.

**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 $\pi_a$ (and $F_{-a} = \tilde{F}_{-a}$). By Jensen's inequality

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

$\iff$

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

$\iff$

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

$\iff$

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

$\iff$

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

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 that it is 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 $\lambda$ 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 in 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 expected utility theory. In the latter theory the curvature properties of a Bernoulli utility function explains 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. Notice that we did not use the assumption that $L_{(a,x)}(a)$ is an increasing function of $x$ for all $a$ in the proof.

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. To illustrate how such a theory can be developed we now introduce definitions of the certainty equivalent of a learning rule $L$ and of the risk premium of a learning rule.

Recall that in expected utility theory, the certainty equivalent of a distribution for an individual is the amount of money that another distribution would have to offer with certainty for the individual to be indifferent between the two distributions. The definition of the certainty equivalent of a learning rule, intuitively, asks how much $\tilde{F}_a$ would have to pay with certainty for it to result in the same expected movement of probability mass as $F_a$, when all other things between the two environments are held constant.

In expected utility theory, the risk premium of a distribution $F_a$ is the difference between $\pi_a$ and the amount of money that $\tilde{F}_a$ would have to offer with certainty for the individual to be indifferent between the two. We define the risk premium of a learning rule for a distribution $F_a$ as the difference between $\pi_a$ and the amount of money that $\tilde{F}_a$ would have to offer with certainty for $\tilde{f}(a)$ and $f(a)$ to coincide.

**Definition 3** *The* certainty equivalent *of a learning rule $L$ for action $a$ with distribution $F_a$ is the payoff $c_a(F_a)$ that action $a$ in environment $\tilde{F}$ would have to pay with certainty for $\tilde{f}(a) = f(a)$. The* risk premium *of an action $a$ with distribution $F_a$ is the amount of money $r_a(F_a)$ such that if $a$ in $\tilde{F}$ offers $\pi_a - r_a(F_a)$ with certainty then $\tilde{f}(a) = f(a)$.*

**Proposition 2** *A learning rule $L$ is risk averse $\Longleftrightarrow$ $c_a(F_a) \leq \pi_a$ for all actions $a$ and distributions $F_a \Longleftrightarrow r_a(F_a) \geq 0$ for all actions $a$ and distributions $F_a$.*

**Proof.** Suppose $L$ is risk averse. Consider any environment $F$ in which $a$ has a distribution $F_a$ and the associated environment $\tilde{F}$ in which $a$ has the degenerate distribution $\tilde{F}_a$ that places all probability on the payoff $\pi_a$. All other aspects of the two environments are identical. Then,

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

$\Longleftrightarrow$

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

$\Longleftrightarrow$

$$L_{(a,\pi_a)}(a) \geq L_{(a,c_a(F_a))}(a)$$

$\iff$

$$\pi_a \geq c_a(F_a)$$

as we assumed that $L_{(a,x)}(a)$ is strictly increasing in $x$.

The equivalence of the last statement in the Proposition and the others follows immediately from the fact that $r_a(F_a) = \pi_a - c_a(F_a)$. ■

The following example illustrates some aspects of the above definitions and results.

**Example 1** *Square-root-Cross Learning Rule*

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

*where* $x \in [0,1]$. *This learning rule is own-concave and hence by Proposition 1 is risk averse. To illustrate the notion of certainty equivalence consider an environment* $F = (F_a, F_{a'}) = \{(.09, .25; .5, .5), (.36; 1)\}$. *That is, a gives .09 with probability .5 and .25 with probability .5 and* $a'$ *gives .36 with probability 1. Furthermore, suppose* $\sigma = (\sigma_a, \sigma_{a'}) = (.5, .5)$. *Then, it is easy to show that* $f(a) = -.05$. *The certainty equivalent of* $F_a$ *is* $c_a(F_a) = .16$, *whereas the expected payoff of* $F_a$ *is* $\pi_a = .17$. *Hence, the risk premium of* $F_a$ *is* $r_a(F_a) = .01$.

To further develop the analogy between risk averse learning rules and risk averse expected utility maximizers consider the case when $L_{(a,x)}(a)$ is a twice differentiable function of $x$. We may then adapt the well known Arrow-Pratt measure of absolute risk aversion 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

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

For the above example we find that $ar_{L_a}(x) = 1/(2x)$ for all $a$.

We may now turn to consider when to call one learning rule more risk averse than another. We follow Pratt (1964) closely in this regard who defines the more risk averse relation in terms of the risk premiums. For any two learning rules $L$ and $L'$ denote their risk premiums by $r_{La}(F_a)$ and

$r_{L_a'}(F_a)$, respectively. We follow the same notational convention when we refer to the certainty equivalents of two learning rules.

**Definition 4** *A learning rule $L$ is* more risk averse *than another $L'$ if for all distributions $F_a$ we have that $r_{La}(F_a) \geq r_{L_a'}(F_a)$ for all $a$.*

The following Proposition gives us four different ways of saying when a learning rule is more risk averse than another.

**Proposition 3** *Suppose $L_{(a,x)}(a)$ is twice differentiable for all $a$. Then, the following statements are equivalent:*

1. $r_{L_a}(F_a) \geq r_{L_a'}(F_a)$ *for all $a$ and $F_a$*
2. $c_{L_a}(F_a) \leq c_{L_a'}(F_a)$ *for all $a$ and $F_a$*
3. *There exist concave functions $\beta_a$ such that, $L_{(a,x)}(a) = \beta_a\left(L'_{(a,x)}(a)\right)$ for all $a$.*
4. $ar_{L_a} \geq ar_{L_a'}$ *for all $a$ and $x$.*

**Proof.** This involves a straightforward application of the result of Pratt (1964, Theorem 1). ■

Expected utility theory provides an attractive way of saying when one distribution is more risky than another. Specifically, it describes a distribution $F_a$ as more risky than another $\tilde{F}_a$ if both have the same means 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) $F_a$. Our next result provides an analogous result from the viewpoint of the theory of risk averse learning rules. Specifically, it shows that a distribution $\tilde{F}_a$ sosd $F_a$ if and only if every risk averse learning rule is expected to add more probability mass to the action that gives $\tilde{F}_a$ than to an action which gives $F_a$, when other aspects of the environment are held fixed.

**Proposition 4** *$\tilde{F}_a$ 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.** $\tilde{f}(a) \geq f(a)$

$\iff$

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

$\iff$

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

It may be observed that risk aversion of a learning rule does not require the learning rule to map points in the simplex into itself. 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 properties we discuss in the next section provides such restrictions.

# 4 Two Sufficient Conditions

Our definition of when to call a learning rule risk averse was inspired by standard decision theory. This had the consequence that we were able to uncover results for risk averse learning rules analogous to the well known results for risk averse expected utility maximizers. Learning, however, differs in many respects from choice. Important among these is that behavior in learning models is described as stochastic which results in a probability distribution being "selected" whereas in choice theory behavior is assumed to be deterministic and a single action is chosen.

For learning rules we may, hence, consider alternative ideas of when to call them risk averse. In this section, we introduce two such properties (in Definition 5 and Definition 6, respectively). Both pay particular attention to the fact that the individual learner will be choosing a probability distribution in the next period. The first property looks at the entire (expected) behavior tomorrow but restricts attention to a subset of environments. Roughly speaking, a learning rule will be called super risk averse if in every environment that is completely ordered by the sosd relation[7] the expected behavior

[7] An environment $F$ is *completely ordered by the sosd relation* if for all $F_a$, $F_{a'} \in F$ either $F_a$ sosd $F_{a'}$ or $F_{a'}$ sosd $F_a$ or both. In the last case, we have $F_a = F_{a'}$.

tomorrow is such that the distribution over payoffs the individual faces tomorrow second order stochastically dominates the distribution of today.

**Definition 5** *A learning rule L is* super risk averse *if in every environment that is completely ordered by the sosd relation we have that* $\sum_a (\sigma_a + f(a)) F_a$ *sosd* $\sum_a \sigma_a F_a$.

Even though the definition of super risk aversion calls for improved performance only on environments which are completely ordered by the sosd relation, this property imposes a lot of structure on the learning rule (see Lemma 2 and Proposition 5, below). If we required that $\sum_a (\sigma_a + f(a)) F_a$ sosd $\sum_a \sigma_a F_a$ in every environment then 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 et al (2004).[8][9] Restricting the set of environments on which the improvement is required leads us to identify a larger class of learning rules.

Our next result shows that every super risk averse learning rule is risk averse. The proof of the result reveals that super risk averse learning rules have the property that if all actions have the same distribution of payoffs then there is no expected movement in probability mass on any action. This is shown in Lemma 1 of the proof. A second lemma then characterizes all learning rules that satisfy this property.

**Proposition 5** *Every super risk averse learning rule L is a risk averse learning rule.*

The following two Lemmas help us in proving the result. 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 $a, a' \in A$.

[8]Unbiased rules are those which exhibit zero expected movement in probability mass when all actions have the same expected payoffs. Such rules satisfy Definition 5 in a trivial manner because the expected distribution tomorrow is the same as today.

[9]It can also be shown that 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 are the unbiased learning rules.

**Lemma 1** *If the learning rule $L$ is super risk averse then in every environment with $A^* = A$ we have that $f(a) = 0$ for all $a$.*

**Proof.** The proof is by contradiction. Suppose $L$ is super risk averse but that there exist an environment $F$ with $A^* = A$ and $f(a) > 0$ for some $a$. To begin with consider the case in which $F_a$ places strictly positive weight in the interior of $[x_{\min}, x_{\max}]$. We now construct the associated environment $\tilde{F}$ in which $\tilde{F}_a$ is a mean preserving spread of $F_a$ and $\tilde{F}_{a'} = F_{a'}$ for all $a' \neq a$. Specifically, suppose that $\tilde{F}_a$ is obtained by taking out $\varepsilon$ probability uniformly[10] from $[x_{\min}, x_{\max}]$ and placing $(\pi_a - x_{\min})\,\varepsilon / (x_{\max} - x_{\min})$ on $x_{\max}$ and $(x_{\max} - \pi_a)\,\varepsilon / (x_{\max} - x_{\min})$ on $x_{\min}$. By construction, the $\tilde{F}_{a'}$ sosd $\tilde{F}_a$ for all $a' \neq a$. Since $\tilde{f}(a)$ is continuous in $\varepsilon$, there exists a small enough $\varepsilon$ such that $\tilde{f}(a) > 0$. This contradicts that $L$ is super risk averse. Next consider the case in which $F_a$ places all probability mass on $x_{\max}$. In this case, all distributions $F_{a'}$ must also place all probability mass on $x_{\max}$. Consider another environment $\hat{F}$ in which each action $a \in A$ assigns probability $(1-\varepsilon)$ to $x_{\max}$ and $\varepsilon$ to $y \in (x_{\min}, x_{\max})$. Now consider an environment $\tilde{F}$ in which all actions except $a$ have the same payoff distributions as in $\hat{F}$ and $a$ removes probability $\varepsilon$ from $y$ and assigns $\varepsilon/2$ to $y - \delta$ and $\varepsilon/2$ to $y + \delta$ where $0 < \delta < \min\{x_{\max} - y, y\}$. By construction, the distribution of $a$ in $\tilde{F}$ is a mean preserving spread of the distribution of $a$ in $\hat{F}$. As $\tilde{f}(a)$ is continuous in $\varepsilon$ there is a small enough $\varepsilon$ such that $\tilde{f}(a) > 0$. This contradicts the assumption that $L$ is super risk averse. The argument for the case in which $F_a$ places all probability mass on $x_{\min}$ is analogous. The last case we need to consider is one in which $F_a$ places all probability mass on $x_{\min}$ and $x_{\max}$ with strictly positive weight on each. In this case, $\pi_a \in (x_{\min}, x_{\max})$. Now consider an environment $\hat{F}$ in which each $a \in A$ has $\varepsilon$ probability taken away proportionally from each of the extreme points and this is placed on $\pi_a$. Now consider an environment $\tilde{F}$ in which all actions except $a$ have the same payoff distributions as in $\hat{F}$ and the distribution of $a$ removes probability $\varepsilon$ from $\pi_a$ and assigns $\varepsilon/2$ to $\pi_a + \delta$ and $\varepsilon/2$ to $\pi_a - \delta$ for some $\delta \in (0, \min\{\pi_a, (x_{\max} - \pi_a)\})$. By construction, the distribution of $a$ in $\tilde{F}$ is a mean preserving spread of the distribution of $a$ in $\hat{F}$. Since $f(a) > 0$ by hypothesis, and because $\tilde{f}(a)$ is continuous in $\varepsilon$ we have that for small $\varepsilon$, $\tilde{f}(a) > 0$. This contradicts the assumption that $L$ is super risk averse. ■

[10] By which we mean that for every interval $D$ contained in $[x_{\min}, x_{\max}]$, the probability that $\tilde{F}_a$ places on $D$ is $(1-\varepsilon)$ that which $F_a$ places on $D$.

**Lemma 2** *A learning rule $L$ has $f(a) = 0$ for all $a$ whenever $A = A^*$ if and only if there exist functions $u_{aa'}(x)$ such that for all $a, a' \in A$ and all $x \in X$ and $\sigma \in int(\Delta(A))$*

$$\begin{aligned} L_{(a,x)}(a) &= \sigma_a + (1-\sigma_a) u_{aa}(x) \\ L_{(a',x)}(a) &= \sigma_a - \sigma_a u_{a'a}(x) \qquad \forall a' \neq a \\ u_{aa}(x) &= \sum_{a'} \sigma_{a'} u_{a'a}(x) \end{aligned}$$

**Proof.** To show sufficiency suppose that $A = A^*$ and hence $F_a = F_{a'}$ for all $a, a' \in A$. Then

$$\begin{aligned} f(a) &= \sigma_a \left( \sigma_a + (1-\sigma_a) \int u_{aa}(x)\, dF_a(x) \right) + \sum_{a' \neq a} \sigma_{a'} \left( \sigma_a - \sigma_a \int u_{a'a}(x)\, dF_{a'}(x) \right) - \sigma_a \\ &= \sigma_a \left( \int u_{aa}(x)\, dF_a(x) - \sum_{a'} \sigma_{a'} \int u_{a'a}(x)\, dF_{a'}(x) \right) \\ &= \sigma_a \left( \sum_{a'} \sigma_{a'} \int u_{a'a}(x)\, dF_a(x) - \sum_{a'} \sigma_{a'} \int u_{a'a}(x)\, dF_{a'}(x) \right) \\ &= 0 \end{aligned}$$

To show necessity consider an environment in which $A = A^*$ and $F_a$ places all probability mass on $x \in [x_{\min}, x_{\max}]$. Hence, all $F_{a'}$ place all probability mass on $x$ and

$$f(a) = \sigma_a L_{(a,x)}(a) + \sum_{a' \neq a} \sigma_{a'} L_{(a',x)}(a) - \sigma_a = 0$$

Define,

$$\begin{aligned} u_{aa}(x) &\equiv \left( L_{(a,x)}(a) - \sigma_a \right) / (1-\sigma_a) \\ u_{a'a}(x) &\equiv \left( \sigma_a - L_{(a',x)}(a) \right) / \sigma_a \end{aligned}$$

Then,

$$f(a) = \sigma_a (\sigma_a + (1-\sigma_a) u_{aa}(x)) + \sum_{a' \neq a} \sigma_{a'} (\sigma_a - \sigma_a u_{a'a}(x)) - \sigma_a = 0$$

$\iff$

$$\sigma_a \left( u_{aa}(x) - \sum_{a'} \sigma_{a'} u_{a'a}(x) \right) = 0$$

$\Longleftrightarrow$

$$u_{aa}(x) = \sum_{a'} \sigma_{a'} u_{a'a}(x)$$

which gives us the result. ■

**Remark 1** *Note that for all the rules satisfying Lemma 1 and Lemma 2,*

$$f(a) = \sigma_a \left( \sum_{a'} \sigma_{a'} \left( \int u_{a'a}(x)\, dF_a(x) - \int u_{a'a}(x)\, dF_{a'}(x) \right) \right)$$

**Proof.** To complete the proof of Proposition 5, consider any two payoff $x', x'' \in [x_{\min}, x_{\max}]$ with $x' \neq x''$ and any $\lambda \in (0,1)$. Let $x \equiv \lambda x' + (1-\lambda)\, x''$. Consider an environment in which $F_a$ gives $x$ with probability 1 and all other actions $a' \neq a$ give $x'$ with $\lambda$ probability and $x''$ with $(1-\lambda)$ probability. Clearly, $A^* = \{a\}$ and

$$\begin{aligned} f(a) &= \sigma_a \left( \sum_{a' \neq a} \sigma_{a'} \left( u_{a'a}(x) - \lambda u_{a'a}(x') - (1-\lambda)\, u_{a'a}(x'') \right) \right) \\ &= \sigma_a (1-\sigma_a) \left( u_{aa}(x) - \left( \lambda u_{aa}(x') + (1-\lambda)\, u_{aa}(x'') \right) \right) \end{aligned}$$

Super risk aversion requires that $f(a) \geq 0$ which requires that

$$u_{aa}(x) \geq \lambda u_{aa}(x') + (1-\lambda)\, u_{aa}(x'')$$

Hence, $u_{aa}$ must be concave. ■

The property referred to in Lemma 1 is similar to the unbiasedness property studied in Börgers et al (2004). Whereas unbiased requires zero expected motion when all actions have the same expected payoffs, the property in Lemma 1 requires zero expected movement when all actions have the same distribution of payoffs. Clearly, a larger set of learning rules satisfy the property in Lemma 1. Börgers et al characterize all unbiased learning rules and show that all absolutely expedient rules are unbiased. In contrast, Lemma 2 characterizes all learning rules that satisfy $f(a) = 0$ for all $a$ when all actions have the same distribution and Lemma 1 shows that all super risk averse learning rules satisfy this property.[11]

Our next definition introduces a property of learning rules related to risk aversion and super risk aversion of learning rules. In contrast to the latter it looks at expected behavior only with regard to the (expected) probability with which the best actions (in a sosd sense) are chosen. 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). For any subset $\hat{A} \subset A$, let $f\left(\hat{A}\right) \equiv \Sigma_{a\in\hat{A}} f(a)$.

**Definition 6** *A learning rule $L$ is* monotonically risk averse *if in all environments we have that $f(A^*) \geq 0$.*

The next result characterizes the set of monotonically risk averse learning rules.

**Proposition 6** *A learning rule $L$ is monotonically risk averse if and only if there exist functions $u_{aa'}(x)$ such that for all $a, a' \in A$ and all $x \in X$ and $\sigma \in int(\Delta(A))$*

$$
\begin{aligned}
L_{(a,x)}(a) &= \sigma_a + (1-\sigma_a)\, u_{aa}(x) \\
L_{(a',x)}(a) &= \sigma_a - \sigma_a u_{a'a}(x) \qquad \forall a' \neq a \\
u_{aa}(x) &= \sum_{a'} \sigma_{a'} u_{a'a}(x)
\end{aligned}
$$

*and $u_{a'a}(x)$ are concave functions for all $a' \neq a$.*

**Proof.** Sufficiency:

[11] Note that absolute expediency requires the expected movement of probability mass to satisfy a strict inequality whereas super risk aversion only requires the expected movement to satisfy a weak inequality. Our choice of a weak inequality allows us to develop a theory of risk averse learning rules that is isomorphic to the theory of risk averse expected utility maximizers.

Suppose $A^* \neq A$ and $A^* \neq \emptyset$, otherwise $f(A^*) = 0$ immediately. Let $a \in A^*$. We know that (see Remark 1)

$$\begin{aligned} f(a) &= \sigma_a \left( \sum_{a'} \sigma_{a'} \left( \int u_{a'a}(x)\, dF_a(x) - \int u_{a'a}(x)\, dF_{a'}(x) \right) \right) \\ &= \sigma_a \left( \sum_{a' \notin A^*} \sigma_{a'} \left( \int u_{a'a}(x)\, dF_a(x) - \int u_{a'a}(x)\, dF_{a'}(x) \right) \right) \end{aligned}$$

because $F_a = F_{a'}$ for all $a, a' \in A^*$. Since $F_a$ sosd $F_{a'}$ for all $a' \notin A^*$, each term in the sum is non-negative. Therefore, $f(a) \geq 0$ for all $a \in A^*$ and, hence, $f(A^*) \geq 0$.

Necessity:

It is easily checked that if $L$ is a monotonically risk averse learning rule then in every environment with $A^* = A$ we have that $f(a) = 0$ for all $a$. To see this we need only see that the argument given in Lemma 1 applies when we replace super risk averse with monotonically risk averse. Hence, the conditions derived in Lemma 2 need to be satisfied. Hence, we only need to show that $u_{a'a}(x)$ are concave functions for all $a' \neq a$. We argue by contradiction. Suppose there exist $a'$ and $a$ , with $a' \neq a$, for which $u_{a'a}(x)$ is not concave. This implies there exist payoffs $x'$ and $x''$ and some $\lambda \in (0,1)$ for which

$$u_{a'a}(x) < \lambda u_{a'a}(x') + (1-\lambda) u_{a'a}(x'')$$

where $x \equiv \lambda x' + (1-\lambda) x''$. Now consider an environment $F$ in which $F_a$ gives $x$ with certainty, $F_{a'}$ gives $x'$ with probability $\lambda$ and $x''$ with probability $(1-\lambda)$. All other actions $a''$, if any, give $x$ with probability $(1-\varepsilon)$, $x'$ with probability $\varepsilon\lambda$, and $x''$ with probability $\varepsilon(1-\lambda)$, where $\varepsilon \in (0,1)$. Clearly, $A^* = \{a\}$, and using the expression for $f(a)$ provided in Remark 1 we have

$$\begin{aligned} f(a) &= \sigma_a \{ \sigma_{a'} \left( u_{a'a}(x) - [\lambda u_{a'a}(x') + (1-\lambda) u_{a'a}(x'')] \right) \\ &\quad + \Sigma_{a'' \neq a, a'} \sigma_{a''} \left( u_{a''a}(x) - [(1-\varepsilon) u_{a''a}(x) + \varepsilon\lambda u_{a''a}(x') + \varepsilon(1-\lambda) u_{a''a}(x'')] \right) \} \\ &= \sigma_a \{ \sigma_{a'} \left( u_{a'a}(x) - [\lambda u_{a'a}(x') + (1-\lambda) u_{a'a}(x'')] \right) \\ &\quad + \varepsilon \Sigma_{a'' \neq a, a'} \sigma_{a''} \left( u_{a''a}(x) - [\lambda u_{a''a}(x') + (1-\lambda) u_{a''a}(x'')] \right) \} \end{aligned}$$

The term $u_{a'a}(x) - [\lambda u_{a'a}(x') + (1-\lambda) u_{a'a}(x'')]$ in the last expression is negative because of our working hypothesis. Therefore, for small enough $\varepsilon$, we have $f(a) = f(A^*) < 0$. This contradicts our hypothesis that $L$ was a monotonically risk averse learning rule. ■

**Remark 2** *Monotonic risk aversion does not imply any meaningful restrictions in environments in which $A^*$ is empty. In such environments it may seem desirable that the subset of actions whose distributions are not second order stochastically dominated by any others are expected to have probability mass added to them. Formally, let $\hat{A} = \{a \in A : \{a' \in A : F_{a'} \neq F_a \text{ and } F_{a'} \text{ sosd } F_a\} = \emptyset\}$. Note that $\hat{A}$ is always non-empty and $\hat{A} = A^*$ whenever $A^* \neq \emptyset$. It is easily seen that every monotonically risk averse learning rule that satisfies $u_{a'a} = u_{aa'}$ for all $a, a' \in A$ has $f\left(\hat{A}\right) \geq 0$.*

Observe that every monotonically risk averse learning rule is risk averse because $u_{aa}(x)$ is just a weighted average of concave functions and so it must be concave. It can also be shown that every monotonically risk averse learning rule is super risk averse. A formal proof of this is contained in the Appendix. The monotone learning rules studied by Börgers et al (2004) are related. Their notion requires that probability mass be expected to strictly increase on the set of expected payoff maximizing actions in all environments in which this is possible. Our proof that monotonely risk averse learning rules are super risk averse closely follows the proof of Börgers et al that shows that monotone learning rules are absolutely expedient.

# 5 Examples

Several learning rules have been considered in the literature. In this section we discuss the risk attitudes of the Cross (1973) and the Roth and Erev (1995) learning rules. We also discuss the risk attitudes of a learning rule derived from the work of Fudenberg and Levine (1998).

The Cross learning rule is given by the following equations:

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

where $x \in [0,1]$. As $L_{(a,x)}(a)$ is a linear function of $x$ this rule is both risk averse and risk seeking and hence is risk neutral. It is equally easily seen that the Cross learning rule is monotonically risk neutral and hence it is also super risk neutral.

The Roth and Erev learning rule describes an agent by a vector $v \in R_{++}^{|A|}$. The vector $v$ describes the decision makers "propensity" 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 propensity to play $a$, leaving all other propensities unchanged. Hence, the Roth and Erev learning rule is given by

$$\begin{aligned} L_{(a,x)}^{v}(a) &= \sigma_a + \frac{1}{\sum_{a'} v_{a'} + x}(1-\sigma_a)x \\ L_{(a'',x)}^{v}(a) &= \sigma_a - \frac{1}{\sum_{a'} v_{a'} + x}\sigma_a x \qquad \forall a'' \neq a \end{aligned}$$

where $x \in [0, x_{\max}]$. It is easily shown that $L_{(a,x)}^{v}(a)$ is a concave function of $x$ and that the coefficient of absolute risk aversion for the Roth and Erev learning rule is positive and hence this rule is risk averse. It is similarly easy to see that this learning rule is monotonically risk averse and hence is also super risk averse and satisfies the condition in Lemma 1. Note that this rule satisfies none of the properties studied by Börgers et al (2004) who have shown that this rule is neither monotone, absolutely expedient or unbiased.

Another learning rule in which agents only respond to payoff information from the chosen action has been studied by Fudenberg and Levine (1998, section 4.8.4). The agent is described by the $|A| \times 2$ matrix $(y, \kappa)$ where $\kappa_a$ denotes the number of times action $a$ was chosen in the past, $\kappa = (\kappa_a)_{a \in A}$, and $y = (y_a)_{a \in A}$ gives the vector of attractions. The next period attraction of an action that was chosen today is its current attraction plus $(x - y_a)/\kappa_a$. The attractions of unchosen actions are not updated. The learning rule is specified as

$$\begin{aligned} L_{(a,x)}^{y,\kappa}(a) &= \frac{e^{y_a + (x-y_a)/\kappa_a}}{e^{y_a + (x-y_a)/\kappa_a} + \sum_{a' \neq a} e^{y_{a'}}} \\ L_{(a',x)}^{y,\kappa}(a) &= \frac{e^{y_a}}{e^{y_{a'} + (x-y_{a'})/\kappa_{a'}} + \sum_{a'' \neq a'} e^{y_{a''}}} \qquad \forall a' \neq a \end{aligned}$$

It is easily seen that this rule is risk averse if $L_{(a,x)}(a) \geq 1/2$ and is risk-seeking otherwise. Equivalently, the learning rule is risk averse (at $\sigma$) if $x \geq y_a + \kappa_a [\ln(1-\sigma_a) - \ln \sigma_a]$. However, this rule is not super-risk averse (or seeking) as it fails to satisfy the condition in Lemma 1.[12]

[12] The analysis of the risk attitudes of several other learning rules (e.g. minimal infor-

# 6 Conclusion

The expected movement of a learning rule has been studied on many previous occasions. Finding the long term properties of a learning rule typically involves studying the expected movement. As is well known under conditions of slow learning the actual movement of a learning rule closely approximate the expected movement.[13] Combining properties of the expected movement and of the speed of learning inform us about the long term properties of learning rules. Recently, Börgers et al (2004) have investigated short term properties of learning rules that also pertain to its expected movement. The properties they study relate to expected payoffs.[14]

In this paper we have shown that defining the risk attitudes of a learning rule in terms of how its expected movement responds to alternative distributions allows us to develop a theory of the risk attitudes of learning rules that is isomorphic to the well known theory of the risk attitudes of expected utility maximizers. This reveals a close relation between the expected utility theory of risk and the risk attitudes of learning rules and allows learning theory to draw upon the many results available for risk averse expected utility maximizers.

Risk averse learning rules need not satisfy what may be considered an elementary consistency condition: That the learning rule not be expected to move probability mass when all actions have the same distribution of payoffs. We characterize all learning rules that satisfy this property. The two sufficient conditions we provide for a learning rule to be risk averse satisfy this property. Lastly, we have provided necessary and sufficient conditions for a learning rule to satisfy one of those sufficient conditions.

The risk attitudes of agents differ from situation to situation. For example, they tend to insure their houses while still buying lotteries. If such behavior is learnt it could be the case that the agents use different learning rules in different circumstances. Empirical work is needed to assess if this is,

mation versions of Camerer and Ho (1999) and Rustichini (1999)), which use a logistic transform of the attractions to obtain the probabilities with which each action is chosen, is closely related.

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

[14]All the learning rules they study are, implicitly, risk neutral. Section 4 of our paper extends their results to allow learning rules to respond to risk.

in fact, the case. Our characterizations of learning rules that have different risk attitudes should facilitate this enterprise. Experimental work could also be conducted to determine the risk attitudes of the learning rules people use.

## APPENDIX

In this Appendix we prove that every monotonely risk averse learning rule is super risk averse. Suppose $L$ is a monotonely risk averse learning rule and consider an environment $F$ that is completely ordered by the sosd relation. The set $F$ can be partitioned so that $F = \cup_{i=1}^{m} F_i$, where all the distributions in $F_i$ are the same and there are $m$ distinct distributions in $F$, i.e. there are $m$ elements in this partition. The proof is by induction on $m$.

Let $w$ be any concave function and let

$$g = \sum_{F_a \in F} (\sigma_a + f(a)) \int w(x)\, dF_a(x) - \sum_{F_a \in F} \sigma_a \int w(x)\, dF_a(x)$$

We need to show that $g \geq 0$ for all $m$. Suppose $m = 1$. Then, $A = A^*$ and hence $g = 0$. Now suppose $g \geq 0$ for $m = b - 1$ and consider an environment with $m = b$. Let $A^{**} = \{a : F_{a'} \text{ sosd } F_a \text{ for all } a' \in A\}$. That is, $A^{**}$ is the set of least preferred actions for any risk averse agent. Let $\pi_a^{**} = \int w(x)\, dF_a(x)$ for all $a \in A^{**}$. Let $A^{***} = \{a : F_{a'} \text{ sosd } F_a \text{ for all } a' \notin A^{**}\}$. That is, $A^{***}$ is the set of second least preferred actions for any risk averse agent. Let $\pi_a^{***} = \int w(x)\, dF_a(x)$ for all $a \in A^{***}$. Observe that $k \equiv \pi^{***} - \pi^{**} \geq 0$.

Now consider another environment $\tilde{F}$ in which the distribution of lotteries in $A^{**}$ is changed to the distribution of lotteries in $A^{***}$ (and all other distributions are as in $F$). By the inductive hypothesis we know that $\tilde{g} \geq 0$ and hence showing $g - \tilde{g} \geq 0$ completes the proof.

$$\begin{aligned} g - \tilde{g} &= \sum_{a \notin A^{**}} f(a)\, \pi_a + \sum_{a \in A^{**}} f(a)\, \pi_a^{**} - \left[ \sum_{a \notin A^{**}} \tilde{f}(a)\, \pi_a + \sum_{a \in A^{**}} \tilde{f}(a)\, (\pi_a^{**} + k) \right] \\ &= \sum_{a \notin A^{**}} \left( f(a) - \tilde{f}(a) \right) \pi_a + \sum_{a \in A^{**}} \left( f(a) - \tilde{f}(a) \right) \pi_a^{**} - \sum_{a \in A^{**}} \tilde{f}(a)\, k \qquad (1) \end{aligned}$$

From Remark 1, we know that for all $a \notin A^{**}$ we have

$$\begin{aligned}
f(a)-\tilde{f}(a) &= \sigma_a \sum_{a' \in A} \sigma_{a'} \left[ \int u_{a'a}(x)\, d\tilde{F}_{a'}(x) - \int u_{a'a}(x)\, dF_{a'}(x) \right] \\
&= \sigma_a \sum_{a' \in A^{**}} \sigma_{a'} \left[ \int u_{a'a}(x)\, d\tilde{F}_{a'}(x) - \int u_{a'a}(x)\, dF_{a'}(x) \right] \\
&\quad -\sigma_a \sum_{a' \notin A^{**}} \sigma_{a'} \left[ \int u_{a'a}(x)\, d\tilde{F}_{a'}(x) - \int u_{a'a}(x)\, dF_{a'}(x) \right] \\
&= \sigma_a \sum_{a' \in A^{**}} \sigma_{a'} \left[ \int u_{a'a}(x)\, d\tilde{F}_{a'}(x) - \int u_{a'a}(x)\, dF_{a'}(x) \right] \quad (2)
\end{aligned}$$

Note that each term in the square brackets is non-negative since $\tilde{F}_{a'}$ sosd $F_{a'}$ for all $a' \in A^{**}$. Since, $\sum_{a \in A} f(a) = \sum_{a \in A} \tilde{f}(a) = 0$ we get

$$\begin{aligned}
\sum_{a \in A^{**}} \left( f(a) - \tilde{f}(a) \right) &= -\sum_{a \notin A^{**}} \left( f(a) - \tilde{f}(a) \right) \\
&= -\sum_{a \notin A^{**}} \sum_{a' \in A^{**}} \sigma_a \sigma_{a'} \left[ u_{a'a}(x)\, d\tilde{F}_{a'}(x) - \int u_{a'a}(x)\, dF_{a'}(x) \right]
\end{aligned}$$

Using this and (2) in (1) we obtain

$$\begin{aligned}
g - \tilde{g} &= \sum_{a \notin A^{**}} \sum_{a' \in A^{**}} \sigma_a \sigma_{a'} \left[ u_{a'a}(x)\, d\tilde{F}_{a'}(x) - \int u_{a'a}(x)\, dF_{a'}(x) \right] \pi_a \\
&\quad - \sum_{a \notin A^{**}} \sum_{a' \in A^{**}} \sigma_a \sigma_{a'} \left[ u_{a'a}(x)\, d\tilde{F}_{a'}(x) - \int u_{a'a}(x)\, dF_{a'}(x) \right] \pi_a^{**} - \sum_{a \in A^{**}} \tilde{f}(a)\, k \\
&= \sum_{a \notin A^{**}} \sum_{a' \in A^{**}} \sigma_a \sigma_{a'} \left[ u_{a'a}(x)\, d\tilde{F}_{a'}(x) - \int u_{a'a}(x)\, dF_{a'}(x) \right] (\pi_a - \pi_a^{**}) - k \sum_{a \in A^{**}} \tilde{f}(a)
\end{aligned}$$

The first term on the RHS is non-negative because $\pi_a \geq \pi_a^{**}$ for all $a$ and the term in the squared brackets is also non-negative as explained above. The second term non-negative because $\tilde{f}(a) \leq 0$ for all $a \in A^{**}$. It follows from Remark 1 and the concavity of $u_{a'a}(x)$ and the fact that $F_{a'}$ sosd $F_a$ for all $a' \in A$ and $a \in A^{**}$. It follows that $g - \tilde{g} \geq 0$.

## REFERENCES


1. Arrow, K.J. (1971): *Essays on the theory of risk bearing.* Chicago: Markham.

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. Della Vigna, S. and M. Li Calzi (2001): "Learning to make risk-neutral choices in a symmetric world," *Mathematical Social Sciences*, 41, 19-37.

8. Fudenberg, D. and D. Levine (1998): *The theory of learning in games*, Cambridge: MIT Press.

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

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

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

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

13. 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.

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

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