Mean and Variance Responsive Learning

Carlos Oyarzun, University of Queensland
Rajiv Sarin, University of Birmingham
Mean and Variance Responsive Learning*

Carlos Oyarzun†  Rajiv Sarin
University of Alicante  Texas A&M University

December 31, 2009

Abstract

Decision makers are often described as seeking higher expected payoffs and avoiding higher variance in payoffs. We provide some necessary and some sufficient conditions for learning rules, that assume the agent has little prior and feedback information about the environment, to reflect such preferences. We adopt the framework of Börgers, Morales and Sarin (2004, Econometrica) who provide similar results for learning rules that seek higher expected payoffs. Our analysis reveals that a concern for variance leads to quadratic transformations of payoffs to appear in the learning rule.

1 Introduction

A popular and commonly used measure of the “risk” of a distribution is its variance. Following the seminal papers by Markowitz (1952) and Tobin (1958) many studies in finance and monetary economics have analyzed alternative distributions of payoffs according to their mean and variance. This literature describes people as seeking higher means and lower variances. If

---

*We are grateful for the comments we received from two anonymous referees. All errors belong to us.

†Financial support from the Ministry of Science and Technology of Spain (Project SEJ 2007-62656) is gratefully appreciated.
behavior is, to some extent, learnt then it is likely that the learning rules people use lead them to choose actions with higher means and lower variances. In this paper we study such learning rules.\textsuperscript{1}

We consider learning rules in which the agent has very little prior and feedback information about her environment. She knows the set of actions and chooses each with positive probability.\textsuperscript{2} We refer to the probability with which the agent chooses the different actions as her behavior. The agent does not know the distribution of payoffs from any action. In any period, information is obtained about the action chosen and the payoff received. We focus on how this information is used to modify behavior from one period to the next. Since the change in behavior depends on the payoff obtained, the payoff distributions associated with the different actions will determine the expected change in behavior. We refer to this set of distributions as the environment faced by the agent.

Our analysis begins by considering environments in which all actions have the same mean but could differ in other respects. We refer to such environments as flat environments. We call a learning rule variance averse if it is expected to result in behavior tomorrow that induces a reduced distribution over payoffs that has a lower variance in every flat environment.\textsuperscript{3} We call a learning rule monotonically variance averse if the probability with which it chooses the variance minimizing actions is expected to increase in every flat environment. We show that all monotonically variance averse learning rules are variance averse. Note that variance aversion, and monotonic variance aversion, in flat environments coincides with aversion to the second moment.

Variance averse rules satisfy a consistency condition that we call weak unbiasedness which requires that there is no expected movement in probability on any action if all actions have the same variance (and the same

\textsuperscript{1}At the time of writing of the first version of this paper we knew of no paper that studies learning rules that respond to both the mean and the variance of a distribution. Recently, Denrell (2008) has investigated this subject. We discuss his paper further later in this paper. See Fudenberg and Levine (1998) for a comprehensive study of learning in games.

\textsuperscript{2}This is a common assumption in learning models. Gul and Pesendorfer (2006) provide an axiomatic foundation for such probabilistic choice behavior.

\textsuperscript{3}The reduced distribution over payoffs arises by compounding (the distribution that describes) behavior and the payoff distributions associated with the actions.
expected payoff). We characterize all weakly unbiased learning rules. All such learning rules allow for a polynomial of order two transformation of payoffs, in which the coefficients of the transformation are allowed to depend on the action chosen and the action whose probability is being updated, before applying Cross’ learning rule (Cross (1973)). We characterize all monotonically variance averse learning rules. Such rules place restrictions on the coefficients of the quadratic transformation of payoffs. We also prove that all monotonically variance averse learning rules are variance averse, which in turn we show to be a strict subset of the set of weakly unbiased learning rules.

Our main result considers learning rules in environments in which actions may have different expected payoffs and different variances of payoffs. We now study monotonically mean variance averse rules which require that the probability of choosing actions which have the highest means and lowest variances, provided they exist, is expected to increase. We characterize monotonically mean variance averse rules. Such rules have to be monotonically variance averse. Additionally, the characterization reveals the restrictions on the relative sizes of the coefficients of the linear and quadratic terms that appear in weakly unbiased learning rules. Intuitively, both linear and quadratic terms arise because the learning rule has to respond to both the mean and the variance of the distribution.

We show that monotonically mean variance averse rules are expected to move behavior toward lower variance actions in environments in which all actions have the same expected payoff. Furthermore, they move behavior toward higher expected payoffs in environments in which all actions have the same variance. Hence, they capture certain features of rules that respond only to the first moment of payoffs and those that respond only to the second moment of payoffs.

The analysis in this paper complements that in Börgers, Morales and Sarin (2004), henceforth BMS, and Oyarzun and Sarin (2006), henceforth OS. BMS characterize monotonic learning rules which are expected to increase probability mass on the expected payoff maximizing actions. OS characterize monotonically risk averse learning rules which are expected to

\footnote{In such environments, we show that a monotonically variance averse learning rule may lead to lower expected payoffs.}
increase probability mass on the actions whose distributions second order stochastically dominate (sosd) all other actions. In contrast to BMS which is concerned with the mean of a distribution and OS which is concerned with its risk as measured by sosd, this paper considers learning rules that respond to both the mean of the distribution and its variance, where the latter is used as a measure of risk. BMS (respectively, OS) find that a learning rule is monotonic (respectively, monotonically risk averse) if and only if the agent first transforms the payoff in an affine (respectively, concave) manner and then apply the Cross rule (Cross (1973)). The transformation of payoffs is allowed to depend on the action chosen and the action whose probability is being updated.

As in BMS our analysis takes the initial behavior of the individual as given and fixed. Furthermore, like that paper, our analysis is best viewed as a “reduced form” analysis as it focuses on an individual’s behavior. The “true” learning rule may specify the agent’s beliefs, how these beliefs change in response to experience and how these are transformed into behavior via some “choice rule”. Analyzing the reduced form allows us to focus directly on the agent’s behavior which is more readily observed and usually of more economic interest than her beliefs. Alternate belief formation procedures combined with different choice rules might yield the same learning rule as we study in this paper. Hence, a reduced form approach may not be suitable if we want to study an agent’s beliefs.

Given the results in BMS and in OS, the main result in this paper may not seem surprising. The arguments involved in it, however, need to be considerably different. This is especially true in the characterization of weakly unbiased learning rules. More importantly, our results allow us to describe agents whose learning responds to two different features of a distribution. Responding to the first two moments of a distribution seems both to be intuitive and what is observed in experimental data (see, e.g., Erev, Bereby-Meyer and Roth (1999), Haruvy and Erev (2001) and Haruvy, Erev and Sonsino (2001)).

March (1996), Denrell and March (2001) and Burgos (2002) investigate how specific learning rules respond to risk by way of simulations. Denrell (2007) analytically considers the long term properties of some learning rules in the same informational setting as this paper. He finds a class of adaptive
learning rules that lead to more risk averse choices in the long run. DellaVigna and LiCalzi (1999) also pay close attention to the long run risk attitudes of a specific learning rule. The learning rule they consider, however, assumes more knowledge about the environment than the rules we consider. Finally, the work of Karni and Schmeidler (1986) and Robson (1996a, 1996b) is related. In their work current preferences over gambles, i.e., risk attitudes, are the result of an evolutionary process in which either the expected number of offspring or the probability of survival is maximized.

This paper is structured as follows. In the next section we introduce the framework of the paper. Section 3 and 4 consider the case in which all actions have the same expected payoff. Section 3 characterizes learning rules which satisfy a weak consistency requirement that all variance averse and monotonically variance averse learning rules need to satisfy. Section 4 characterizes monotonically variance averse rules and provides some necessary and some sufficient conditions for variance averse rules. In section 5 we consider environments in which actions have different means and variances of payoff and provide results concerning monotonically mean variance averse rules. Section 6 provides some concluding comments.

2 Framework

The decision maker has a finite number of actions \( a \in A \). Payoffs are interpreted as monetary magnitudes that lie in some bounded interval known to the decision maker which we normalize to be \([0, 1]\).\(^5\) For any Borel subset \( D \subset [0, 1] \) the probability that \( a \) gives a payoff in \( D \) is \( \mu_a (D) \).\(^6\) Let \( \mu = (\mu_a)_{a \in A} \) denote the vector of countably additive probability measures associated with each action. We refer to \( E \equiv (A, \mu) \) as an environment. The expected payoff from action \( a \) is denoted \( \pi_a = \int_0^1 x d\mu_a \) and \( A^* = \{ a : \pi_a \geq \pi_{a'} \text{ for all } a' \in A \} \) denotes the set of expected payoff maximizing actions. Let \( s_a = \int_0^1 x^2 d\mu_a \) denote the second moment of payoffs of action \( a \). The variance of payoffs of action \( a \) is given by \( v_a = s_a - \pi_a^2 \). Let \( A_* = \{ a : v_a \leq v_{a'} \text{ for all } a' \in A \} \) denote the set of variance minimizing actions.

\(^5\)This normalization is without loss of generality for our results.

\(^6\)The probability that action \( a \) gives a payoff \( x \) is denoted by \( \mu_a (x) \).
The individual knows the set of actions $A$ but does not know $\mu$. She chooses among her actions today according to the mixed action vector $\sigma \in \Delta (A)$, where $\Delta (A)$ is the set of all probability distributions over $A$. $\sigma$ describes the behavior of the individual today. We assume that $\sigma$ lies in the interior of the simplex $\Delta (A)$. The expected payoff associated with $\sigma$ is $\pi_\sigma = \sum a \sigma_a \pi_a$ and the variance associated with it is $\nu_\sigma = \sum a \sigma_a s_a - (\pi_\sigma)^2$.

Given that we assume behavior today is fixed and given, a learning rule specifies the individual’s behavior tomorrow as a function of the action she chooses and the payoff she obtains. The learning rule should be viewed as a “reduced form” of the true learning process. The true learning rule may, for example, specify how the decision maker updates her beliefs and how these beliefs are transformed into behavior. Combining the two steps of belief updating and behavior change we get the learning rule we specify. We shall assume the behavior today is given and fixed and hence we may write a learning rule as $L : A \times [0, 1] \rightarrow \Delta (A)$.

Denote by $L_{(a', x)} (a)$ the probability that the learning rule $L$ assigns to action $a$ tomorrow if action $a'$ was chosen today and a payoff of $x$ is received. Fix an environment and a learning rule. Then, the expected change in the probability of action $a$ is given by $f (a) = \sum_{a' \in A} \sigma_{a'} \int_0^1 L_{(a', x)} (a) \, d\mu_{a'} - \sigma_a$. For any subset of actions, $\tilde{A} \subset A$, let $\left( \tilde{A} \right)$ denote the expected change in probability mass on that subset. That is, $f \left( \tilde{A} \right) = \sum_{a \in \tilde{A}} f (a)$. The expected change in expected payoffs is given by $g_1 = \sum_a f (a) \, \pi_a$. The expected change in the variance of the payoff is given by $g_2 = \sum_{a \in \tilde{A}} f (a) \, s_a - \left[ \mathbb{E} \left( (\pi_{a'})^2 \right) - (\pi_\sigma)^2 \right]$, where $\sigma'$ is the next period behavior and the expectation $\mathbb{E}$ is taken with regard to its value.

We shall begin our analysis by studying environments in which each action gives the same expected payoff. We refer to such environments as flat environments.

**Definition 1** An environment is flat if $\pi_a = \pi_{a'}$ for all $a, a' \in A$.

We begin our analysis with flat environments because this allows us to isolate the effect of (payoff) variance on the learning rule. From the point

---

7The properties of learning rules investigated in this paper cannot be satisfied if this condition is violated.
of view of the analysis, working in flat environments reduces the analysis of variance of the payoff distributions to its second moment. Formally, in flat environments $A = A^*$ and $g_2 = \Sigma_a f (a) s_a$. We now introduce two properties of learning rules which pertain to their behavior in flat environments. Hence, in each of these definitions, when we refer to “variance” we are actually concerned with only the second moment.

**Definition 2** A learning rule $L$ is variance averse [neutral, seeking] if for all flat environments with $A \neq A_*$, $g_2 < [=, >] 0$.

In words, a learning rule is said to be variance averse [neutral, seeking] if for all environments in which each action has the same expected payoff, but at least two actions have a different variance of payoff, the expected change in the variance of payoff is strictly negative [zero, strictly positive]. Hence, a learning rule is variance averse if it is expected to lead to a strict reduction in the variance of payoff from one period to the next, provided the environment stays the same.

**Definition 3** A learning rule $L$ is monotonically variance averse [neutral, seeking] if for all flat environments with $A \neq A_*$, $f (A_*) > [=, <] 0$.

In words, a learning rule is monotonically variance averse [neutral, seeking] if for all environments in which each action has the same expected payoff, but at least two actions have a different variance of payoff, the expected change in the probability with which the variance minimizing actions are chosen is strictly positive [zero, strictly negative].

**Remark 1** If $|A| = 2$ then a learning rule is variance averse [neutral, seeking] if and only if the learning rule is monotonically variance averse [neutral, seeking].

---

8 A Referee observed that any generic environment is flat if we consider learning rules that depend upon deviations from the mean. Note, however that the agent, or learning rule, in this paper is unable to perform this renormalization of payoffs given our assumptions about her knowledge of her environment.

9 An alternate, and perhaps more intuitive way to define monotonically variance seeking rules, could be as those learning rules that satisfy $f (A_{a-}) > 0$, where $A_{a-} \equiv \{a : v_a \geq v_{a'}$ for all $a' \in A\}$ in every environment with $A \neq A_*$. It turns out that the learning rules for which $f (A_{a-}) > 0$ are precisely those for which $f (A_*) < 0$ for all flat environments with $A \neq A_*$. 

---
We end this section with an example of a learning rule due to Cross (1973).

**Example 1** Cross Learning rule

\[
L_{(a,x)}(a) = \sigma_a + (1 - \sigma_a) x \\
L_{(a,x)}(a') = \sigma_{a'} - \sigma_{a'} x \quad \forall a' \neq a
\]

In the Cross learning rule,

\[
f(a) = \sigma_a (\pi_a - \Sigma_{a'} \sigma_{a'} \pi_{a'})
\]

for all \(a \in A\). In flat environments, \(f(a) = 0\) for all \(a\) and

\[
g_2 = \Sigma_{a \in A} \sigma_a s_a (\pi_a - \Sigma_{a'} \sigma_{a'} \pi_{a'}) = 0.
\]

Hence, this rule is variance neutral and monotonically variance neutral.

### 3 Weak Unbiasedness

We begin this section by defining the concept of unbiasedness which was introduced in BMS. Unbiasedness requires that for all flat environments there is no expected movement in probability mass.

**Definition 4** A learning rule \(L\) is said to be **unbiased** if for all flat environments \(f(a) = 0\) for each \(a \in A\).

The following remark, which is a straightforward consequence of the definitions, reveals a connection between unbiased learning rules and those that are variance neutral or monotonically variance neutral. It also generalizes the observation, following Example 1 in the preceding section, that the Cross learning rule, which is unbiased, is variance neutral and monotonically variance neutral.

**Remark 2** Every unbiased learning rule is variance neutral and monotonically variance neutral.

Next we introduce a property related to unbiasedness which we call weak unbiasedness. This requires that in environments in which all actions have the same expected payoff and the same variance of payoff there is no expected movement in probability mass. Lemma 1 below shows why we are interested in weakly unbiased learning rules.
Definition 5 A learning rule $L$ is weakly unbiased if for all flat environments with $A = A_s$, $f(a) = 0$ for all $a \in A$.

Lemma 1 Every learning rule that is variance averse [neutral, seeking] or monotonically variance averse [neutral, seeking] is weakly unbiased.

Proof. We first prove this result for the variance averse case in Step 1. The proof is by contradiction. Suppose the learning rule $L$ is variance averse and monotonically variance averse but not weakly unbiased. Hence, there exist a flat environment $E$ with $A = A_s$ and $f(a) > 0$ for some $a \in A$.

Step 1: We consider two cases:

Case 1: $\mu_a(0, 1) > 0$. Consider another environment $\tilde{E}$ in which action $a$ removes $\varepsilon$ probability uniformly from $[0, 1]^{10}$ and adds $(1 - \pi) \varepsilon$ to the probability of 0 and $\pi \varepsilon$ to the probability of 1. Other than this difference both environments offer the same distributions of payoffs for all actions. By construction, the payoff distribution of $a$ in $\tilde{E}$ is a mean preserving spread of $a$ in $E$. Let $\tilde{f}(a)$ denote the expected change in probability on $a$ in environment $\tilde{E}$. Since $\tilde{f}(a)$ is continuous in $\varepsilon$, there is a small enough $\varepsilon$ such that $\tilde{f}(a) > 0$. This contradicts the assumption that $L$ is variance averse and monotonically variance averse.

Case 2: $\mu_a(0, 1) = 0$. We break down the argument into three subcases.

Subcase 2.1: $\mu_a(1) = 1$. Since the environment is flat all actions $a' \in A$ must also have $\mu_{a'}(1) = 1$. Consider another environment $\tilde{E}$ in which each action $a' \in A$ assigns probability $(1 - \varepsilon)$ to 1 and $\varepsilon$ to $y \in (0, 1)$. Now consider an environment $\tilde{E}$ in which all actions except $a$ have the same payoff distributions as in $\tilde{E}$ and $a$ removes probability $\varepsilon$ from $y$ and assigns $\varepsilon/2$ to $y - \delta$ and $\varepsilon/2$ to $y + \delta$ where $0 < \delta < \min \{1 - y, y\}$. By construction, $a$ in $\tilde{E}$ is a mean preserving spread of $a$ in $\tilde{E}$. 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 variance averse and monotonically variance averse.

Subcase 2.2: $\mu_a(0) = 1$. The argument is analogous to Subcase 2.1.

Subcase 2.3: $\mu_a(1) \neq 1$, $\mu_a(0) \neq 1$, $\mu_a(0) + \mu_a(1) = 1$. In this case, $\pi_a \in (0, 1)$. Now consider an environment $\tilde{E}$ in which each $a' \in A$ has $\tilde{\mu}_{a'}(1) = (1 - \varepsilon) \mu_{a'}(1)$ and $\tilde{\mu}_{a'}(0) = (1 - \varepsilon) \mu_{a'}(0)$ and $\tilde{\mu}_{a'}(\pi_a) = \varepsilon$. Now consider an environment $\tilde{E}$ in which all actions except $a$ have the same payoff distributions as in $\tilde{E}$ and $a$ removes probability $\varepsilon$ from $\pi_a$ and assigns $\varepsilon/2$ to $\pi_a + \delta$ and $\varepsilon/2$ to $\pi_a - \delta$ for some $\delta \in (0, \min \{\pi, (1 - \pi)\})$. By construction,

---

10By which we mean that for every Borel subset $D \subset [0, 1]$, $\tilde{\mu}(D) = (1 - \varepsilon) \mu(D)$.\)
the distribution of $a$ in $\tilde{E}$ is a mean preserving spread of $a$ in $\tilde{E}$. 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 variance averse and monotonically variance averse.

**Step 2:** The above argument suffices to show that every variance neutral and monotonically variance neutral rule is weakly unbiased. One simply has to replace “variance averse” with “variance neutral” in the above argument.

**Step 3:** The proof when a learning rule is variance seeking and monotonically variance seeking is analogous to that provided in Step 1. In this case we begin by considering a flat environment $E$ with $A = A_*$ and supposing that $\tilde{f}(a) < 0$. We can then construct a new environment $\tilde{E}$ in which all actions $a' \neq a$ have the same distribution as in $E$ but action $a$ in $\tilde{E}$ has a distribution which is a mean preserving spread of the distribution of $a$ in $E$. The continuity of $\tilde{f}(a)$ will ensure that for a small enough change in the the distribution of $a$ in $E$, we will have $\tilde{f}(a) < 0$ contradicting that the rule is variance seeking and monotonically variance seeking.

The next result characterizes the set of weakly unbiased learning rules.

**Proposition 1** A learning rule is weakly unbiased if and only if there exist matrices $(A_{aa'})_{a,a'=1,...,|A|}$, $(B_{aa'})_{a,a'=1,...,|A|}$ and $(C_{aa'})_{a,a'=1,...,|A|}$ such that for every $(a, x) \in A \times [0, 1]$

\begin{align}
L_{(a,x)}(a) &= \sigma_a + (1 - \sigma_a) (A_{aa} + B_{aa}x - C_{aa}x^2) \\
L_{(a,x)}(a') &= \sigma_{a'} - \sigma_a' (A_{aa'} + B_{aa'}x - C_{aa'}x^2) \quad \forall a' \neq a
\end{align}

and for all $a$,

\begin{align}
A_{aa} &= \sum_{a'} \sigma_{a'} A_{a'a} \\
B_{aa} &= \sum_{a'} \sigma_{a'} B_{a'a} \\
C_{aa} &= \sum_{a'} \sigma_{a'} C_{a'a}
\end{align}

**Proof. Sufficiency**

Consider a flat environment $E$ with $A = A_*$ and suppose the learning rule satisfies the properties stated in Proposition 1. Then (1) – (2) imply
that, for all $a$,
\[
    f(a) = \sigma_a \int_0^1 (1 - \sigma_a) \left( A_{aa} + B_{aa}x - C_{aa}x^2 \right) d\mu_a \\
    - \sum_{a' \neq a} \sigma_{a'} \int_0^1 \sigma_a \left( A_{a'a} + B_{a'a}x - C_{a'a}x^2 \right) d\mu_{a'} \\
    = \sigma_a \left[ A_{aa} + B_{aa}\pi_a - C_{aa}s_a - \sum_{a'} \sigma_{a'} (A_{a'a} + B_{a'a}\pi_{a'} - C_{a'a}s_{a'}) \right].
\]

Equations (3) – (5) and the fact that $A = A^* = A_*$ imply $f(a) = 0$ for all $a$.

**Necessity:**

Let $L$ be a weakly unbiased learning rule and let $E$ and $\tilde{E}$ be two flat environments with $A = A_*$. Denote the common mean of the actions by $\pi$ and their common variance by $\nu$. The distribution for all actions $a$ in $E$ is given by $(0, x, \pi, 1; \mu_a(0), \mu_a(x), \mu_a(\pi), \mu_a(1))$.\(^{11}\) Note that the the variance of payoff $\nu$ lies between 0 and $\pi(1 - \pi)$.\(^{12}\) Suppose $\nu \in (0, \pi(1 - \pi))$. In environment $\tilde{E}$ all actions $a' \neq a$ have the same distribution as in $E$, while action $a$ has a distribution given by $(0, \pi, 1; \tilde{\mu}_a(0), \tilde{\mu}_a(\pi), \tilde{\mu}_a(1))$. By assumption, $\tilde{\pi}_a = \pi_a$ and $\tilde{\nu}_a = \nu_a$.

To see that for any mean $\pi$, the variance $\nu$ (with $0 < \nu < \pi(1 - \pi)$) can be achieved by a suitably chosen distribution with support on the payoffs 0, $\pi$ and 1, we need to observe that the probabilities $\tilde{\mu}_a(0), \tilde{\mu}_a(\pi)$, and $\tilde{\mu}_a(1)$ have to satisfy the three equations $\tilde{\mu}_a(0) + \tilde{\mu}_a(\pi) + \tilde{\mu}_a(1) = \pi$, $\tilde{\mu}_a(0) 0 + \tilde{\mu}_a(\pi) \pi + \tilde{\mu}_a(1) 1 = \pi$ and $\tilde{\mu}_a(0) 0^2 + \tilde{\mu}_a(\pi) \pi^2 + \tilde{\mu}_a(1) 1^2 = \nu$. These three equations can be solved to give
\[
    \tilde{\mu}_a(0) = \frac{s - \pi^2}{\pi}, \quad \tilde{\mu}_a(\pi) = \frac{\pi - s}{\pi - \pi^2}, \quad \tilde{\mu}_a(1) = \frac{s - \pi^2}{1 - \pi}.
\]

\(^{11}\)The first four terms in the brackets are payoffs and the last four terms are their respective probabilities.

\(^{12}\)The variance of a distribution with mean $\pi$ whose support is contained in $[0, 1]$ cannot be greater than the variance of a Bernoulli distribution which gives 1 with probability $\pi$ and 0 otherwise.
which have a unique interior solution between 0 and 1 given $0 < v < \pi (1 - \pi)$.

Next we solve the three equations $\mu_a(0) + \mu_a(x) + \mu_a(\pi) + \mu_a(1) = 1$, $\mu_a(0) + \mu_a(x)x + \mu_a(\pi)\pi + \mu_a(1)\pi = 1$ and $\mu_a(0) + \mu_a(x)x^2 + \mu_a(\pi)\pi^2 + \mu_a(1) = \pi^2 = s$ in terms of $\mu_a(x)$ to obtain

$$\mu_a(0) = \frac{s - \pi^2}{\pi} - \mu_a(x) \left(1 - \frac{x - x^2}{\pi - \pi^2} - \frac{\pi x^2 - \pi^2 x}{\pi - \pi^2}\right)$$  (9)

$$\mu_a(\pi) = \frac{\pi - s}{\pi - \pi^2} - \mu_a(x) \frac{x - x^2}{\pi - \pi^2}$$  (10)

$$\mu_a(1) = \frac{s - \pi^2}{1 - \pi} - \mu_a(x) \frac{\pi x^2 - \pi^2 x}{\pi - \pi^2}.$$  (11)

Substituting (6) in (9), (7) in (10) and (8) in (11) we get,

$$\mu_a(0) = \tilde{\mu}_a(0) - \mu_a(x) \left(1 - \frac{x - x^2}{\pi - \pi^2} - \frac{\pi x^2 - \pi^2 x}{\pi - \pi^2}\right)$$  (12)

$$\mu_a(\pi) = \tilde{\mu}_a(\pi) - \mu_a(x) \frac{x - x^2}{\pi - \pi^2}$$  (13)

$$\mu_a(1) = \tilde{\mu}_a(1) - \mu_a(x) \frac{\pi x^2 - \pi^2 x}{\pi - \pi^2}.$$  (14)

Let $\tilde{L}_{(a,\cdot)}(a')$ denote the expected value of $\sigma_{a'}$ tomorrow given that $a$ was chosen today. That is, $\tilde{L}_{(a,\cdot)}(a') = \int_0^1 L_{(a,x)}(a') d\mu_a$. We now compute the expected change in the probability of an arbitrary action $a'$ in the two environments,

$$f(a') = \sigma_a \left[ \mu_a(0) L_{(a,0)}(a') + \mu_a(x) L_{(a,x)}(a') + \mu_a(\pi) L_{(a,\pi)}(a') + \mu_a(1) L_{(a,1)}(a') \right]$$

$$+ \sum_{a'' \neq a} \sigma_{a''} \tilde{L}_{(a'',\cdot)}(a') - \sigma_{a'} = 0$$

$$\tilde{f}(a') = \sigma_a \left[ \tilde{\mu}_a(0) L_{(a,0)}(a') + \tilde{\mu}_a(x) L_{(a,x)}(a') + \tilde{\mu}_a(\pi) L_{(a,\pi)}(a') + \tilde{\mu}_a(1) L_{(a,1)}(a') \right]$$

$$+ \sum_{a'' \neq a} \sigma_{a''} \tilde{L}_{(a'',\cdot)}(a') - \sigma_{a'} = 0.$$  

Subtracting these two equations and slightly rearranging we get,

$$\mu_a(x) L_{(a,x)}(a') = [\tilde{\mu}_a(0) - \mu_a(0)] L_{(a,0)}(a')$$

$$+ [\tilde{\mu}_a(\pi) - \mu_a(\pi)] L_{(a,\pi)}(a')$$

$$+ [\tilde{\mu}_a(1) - \mu_a(1)] L_{(a,1)}(a')$$  (15)
Substituting (12) – (14) in (15) and slightly re-arranging we obtain,

$$L(a;x) = \left[ 1 - \frac{x - x^2}{\pi - \pi^2} - \frac{\pi x^2 - \pi^2 x}{\pi - \pi^2} \right] L_{(a,0)}(a') + \frac{x - x^2}{\pi - \pi^2} L_{(a,\pi)}(a') + \frac{\pi x^2 - \pi^2 x}{\pi - \pi^2} L_{(a,1)}(a'),$$

which is a polynomial of order no greater than two in $x$. Hence, $L_{(a,x)}(a')$ can be written as

$$L_{(a,x)}(a') = \hat{A}_{aa'} + \hat{B}_{aa'} x + \hat{C}_{aa'} x^2.$$ 

Straightforward algebra reveals that these conditions impose no extra restrictions on $\hat{A}_{aa'}, \hat{B}_{aa'}, \hat{C}_{aa'}$ beyond $\hat{A}_{aa'} = L_{(a,0)}(a')$. In particular, these coefficients do not depend on $\pi$.

Equating,

$$L_{(a,x)}(a) = \hat{A}_{aa} + \hat{B}_{aa} x + \hat{C}_{aa} x^2$$

with (1), we set the coefficients of $A_{aa}, B_{aa}$ and $C_{aa}$ in such a way that they satisfy

$$\hat{A}_{aa} = \sigma_a + (1 - \sigma_a) A_{aa}$$
$$\hat{B}_{aa} = (1 - \sigma_a) B_{aa}$$
$$\hat{C}_{aa} = - (1 - \sigma_a) C_{aa}.$$ 

Similarly, we set the coefficients of $A'_{aa'}, B'_{aa'}$ and $C'_{aa'}$ in such a way that they satisfy

$$\hat{A}'_{aa'} = \sigma_{a'} - \sigma_a A'_{aa'}$$
$$\hat{B}'_{aa'} = - \sigma_{a'} B'_{aa'}$$
$$\hat{C}'_{aa'} = \sigma_a C'_{aa'}.$$ 

Finally, to prove the necessity of conditions (3) – (5) consider an environment in which every action pays $x$ with probability 1. In this environment

$$f_a = \sigma_a \left[ (A_{aa} + B_{aa} x + C_{aa} x^2) - \sum_{a'} \sigma_{a'} \left( A'_{a'a} + B'_{a'a} x - C'_{a'a} x^2 \right)\right]$$
$$= \sigma_a \left[ (A_{aa} - \sum_{a'} \sigma_{a'} A_{a'a}) + (B_{aa} - \sum_{a'} \sigma_{a'} B_{a'a}) x - (C_{aa} - \sum_{a'} \sigma_{a'} C_{a'a}) x^2 \right].$$
This expression has to be zero for all $x \in [0, 1]$. This requires that conditions (3) – (5) are satisfied.

Conditions (1) and (2) of Proposition 1 reveal that all weakly unbiased rules apply Cross’ rule after transforming payoffs by a polynomial of order two in which the coefficients of the transformation are allowed to depend on the chosen action and the action whose probability is being updated. Conditions (3)-(5) restrict the coefficients of this transformation.

**Remark 3** The set of unbiased learning rules identified by BMS is contained in the set of weakly unbiased learning rules. They are obtained by setting $C_{a'a} = 0$ for all $a, a' \in A$.

**Remark 4** The expected change in probability for an action $a$ for all weakly unbiased learning rules is given by

$$f(a) = \sigma_a \sum_{a'} \sigma_{a'} [B_{a'a} (\pi_a - \pi_{a'}) - C_{a'a} (s_a - s_{a'})].$$

which in a flat environment reduces to

$$f(a) = \sigma_a \sum_{a'} \sigma_{a'} C_{a'a} (s_{a'} - s_a).$$

As Remark 4 shows, the coefficients $C_{a'a}$ play a fundamental role in determining the expected movement of probability in flat environments. The precise relation between these coefficients and attitudes towards variance of learning rules is studied in the next section.

Several learning rules, in which agents observe information only about the payoff they observe from the chosen action, have been getting attention recently. Prominent amongst these is the Roth and Erev (1995) learning rule. It is readily seen from the above results that this learning rule is not variance averse [neutral, seeking]. Indeed, it is not weakly unbiased. The same applies for logit learning rules (e.g., Denrell (2008), Fudenberg and Levine (1998), Chapter 4.8.4).

\[13\] In this learning rule the agent is described by a vector $v \in \mathbb{R}_{++}^{|A|}$. The vector $v$ describes the decision maker’s “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, this rule can be written as $L^v_{a,x} (a) = (v_a + x) / (\Sigma_{a'} v_{a' + x})$, $\forall a$, and $L^v_{a',x} (a) = v_a / (\Sigma_{a'} v_{a'} + x)$, $\forall a' \neq a$, where $x \in [0, x_{\text{max}}]$ and the superscript $v$ on the learning rule defines it at that state of learning.
4 Learning and behavior toward variance

The next Lemma provides restrictions on the $C_{aa}$ coefficients for a learning rule to be variance averse [neutral, seeking] or monotonically variance averse [neutral, seeking].

**Lemma 2** Every variance averse [neutral, seeking] and every monotonically variance averse [neutral, seeking] learning rule has $C_{aa} > [=, <] 0$.

**Proof.** Consider a flat environment in which $\mu_a(x) = 1$ for some $x \in (0, 1)$, and all other actions $a' \neq a$ have $\mu_{a'}(x) = 1 - \varepsilon$, and $\mu_{a'}(x - \delta) = \varepsilon / 2$ and $\mu_{a'}(x + \delta) = \varepsilon / 2$ (for $0 < \delta < \min \{x, 1 - x\}$). Clearly, $s_a = x^2$ and $s_{a'} = x^2 + \varepsilon \delta^2$ and $A_\ast = \{a\}$. Furthermore,

$$f(a) = \sigma_a \left[ \sum_{a' \neq a} \sigma_{a'} C_{a'a} (s_{a'} - s_a) \right]$$

$$= \sigma_a \left[ \sum_{a' \neq a} \sigma_{a'} C_{a'a} \varepsilon \delta^2 \right]$$

$$= \sigma_a (1 - \sigma_a) C_{aa} \varepsilon \delta^2.$$

Hence, for a learning rule to be variance averse [neutral, seeking] or monotonically variance averse [neutral, seeking] we require $C_{aa} > [=, <] 0$.

From Lemma 2 it follows that no variance averse or monotonically variance averse learning rule is BMS monotonic because such learning rules cannot be BMS unbiased (see Remark 3).

**Proposition 2** A learning rule is monotonically variance averse [neutral, seeking] if and only if

1. $C_{a'a} \geq [=, \leq] 0$ for all $a' \neq a$,

2. if $D$ is a non-empty proper subset of $A$ then there are actions $a \in D$ and $a' \in A - D$ such that $C_{a'a} > [=, <] 0$.

**Proof.** Sufficiency:
From Lemma 1 and Proposition 1 we know that every monotonically variance averse [neutral, seeking] learning rule has to satisfy equations (1) – (5). (1) and (2) imply

\[ f(a) = \sigma_a \int_0^1 (1 - \sigma_a) (A_{aa} + B_{aa}x - C_{aa}x^2) \, d\mu_a \]

\[ - \sum_{a' \neq a} \sigma_{a'} \int_0^1 \sigma_a (A_{a'a} + B_{a'a}x - C_{a'a}x^2) \, d\mu_{a'} \]

\[ = \sigma_a \left( A_{aa} + B_{aa} \pi_a - C_{aa}s_a \right) - \sum_{a'} \sigma_{a'} \left( A_{a'a} + B_{a'a} \pi_{a'} - C_{a'a}s_{a'} \right) \]

Since the environment is flat, and using (3) – (5), we get

\[ f(a) = \sigma_a \left( \sum_{a'} \sigma_{a'} C_{a'a} (s_{a'} - s_a) \right). \]

Since \( s_{a'} - s_a > 0 \) for every pair of actions \( a \in A_s \) and \( a' \in A - A_s \), \( f(A_s) > [\geq, <] 0 \) by the restrictions on \( C \) stated in the Proposition.

**Necessity:**

We focus on the case of variance aversion (the arguments for the other cases are analogous). The proof is by contradiction.

Suppose condition 1 is not satisfied so that \( C_{a'a} < 0 \) for some \( a' \neq a \). Consider a flat environment in which \( a \) gives a payoff of \( x \in (0, 1) \) with probability one, and \( a' \) gives \( x \) with probability \( 1 - \varepsilon \) and \( x + \delta \) with probability \( \varepsilon / 2 \) and gives \( x - \delta \) with probability \( \varepsilon / 2 \), where \( \delta \in (0, \min \{x, 1 - x\}) \) and \( \varepsilon \in (0, 1] \). All other actions \( a'' \), if any, give a payoff of \( x \in (0, 1) \) with probability \( 1 - \varepsilon' \) and \( x + \delta \) with probability \( \varepsilon'/2 \) and gives \( x - \delta \) with probability \( \varepsilon'/2 \) where \( \varepsilon' \in (0, 1] \). It is easily seen that \( s_a = x^2 \), \( s_{a'} = x^2 + \varepsilon \delta^2 \), \( s_{a''} = x^2 + \varepsilon' \delta^2 \) and \( A_s = \{a\} \). Hence, the conditions on \( L \) derived in Proposition 1 imply that

\[ f(a) = \sigma_a \left\{ \sigma_a C_{a'a} (s_{a'} - s_a) + \sum_{a'' \neq a, a'} \sigma_{a''} C_{a''a} (s_{a''} - s_a) \right\} \]

\[ = \sigma_a \left\{ \sigma_a C_{a'a} \varepsilon^2 + \sum_{a'' \neq a, a'} \sigma_{a''} C_{a''a} \varepsilon' \delta^2 \right\}. \]

16
For $\varepsilon'$ small enough $f(a) = f(A_*) < 0$ which contradicts monotonicity. Therefore, $C_{a'a}$ cannot be less than zero for any $a, a'$ with $a \neq a'$.

Next, suppose condition 2 is not satisfied. Then there exists a non-empty subset of $D \subset A$, such that $A - D \neq \emptyset$ and $C_{a'a} = 0$ for all pairs of lotteries $a \in D$ and $a' \in A - D$. Now consider an environment in which $a \in D$ receives a payoff of $x$ with probability 1 and each action $a' \in A - D$ receives $x$ with probability $1 - \varepsilon$ and receives $x - \delta$ with probability $\varepsilon/2$ and receives $x + \delta$ with probability $\varepsilon/2$ where $\varepsilon \in (0, 1]$ and $\delta \in (0, \min \{x, 1 - x\})$. Clearly, $A_* = D$. Also, for any $a \in D$, the restrictions of Proposition 1 imply

$$f(a) = \sigma_a \sum_{a' \in A - D} \sigma_{a'} C_{a'a} (s_{a'} - s_a).$$

The RHS vanishes because of our assumption that $C_{a'a} = 0$ for all actions $a' \in A - D$ and $a \in D$. Therefore, $f(a) = f(A_*) = 0$ which contradicts the hypothesis that the rule is monotonically variance averse. ■

Our next result shows that every learning rule that is monotonically variance averse [neutral, seeking] is variance averse [neutral, seeking].

**Proposition 3** Every learning rule that is monotonically variance averse [neutral, seeking] is variance averse [neutral, seeking].

**Proof.** The proof is analogous to the one provided in BMS for their Proposition 3(ii). In our case, we have to perform the induction on the second moments rather than on the first moments. We omit the details. ■

Monotonic variance averse rules are a very special case of the OS monotonically risk averse rules. This specific class is the one in which the response towards the risk of a distribution is completely determined by the response toward its variance.\(^\text{14}\) The relationship between monotonically variance averse rules and the monotonically risk averse learning rules studied in OS is analogous to the relation between quadratic utility functions and risk averse utility functions in expected utility theory (Meyer (1987)).

The next example motivates us to consider learning rules that respond in desirable ways to both higher mean and lower variance.

\(^{14}\)From Remark 4 it follows that in flat environments $f(a) = \sigma_a \sum_{a'} \sigma_{a'} C_{a'a} (v_{a'} - v_a)$. So, only the variance matters in determining expected movement.
Example 2 Negative Cross-Square Learning Rule

\[ L_{(a,x)}(a) = \sigma_a + (1 - \sigma_a)(1 - x^2) \]
\[ L_{(a,x)}(a') = \sigma_{a'} - \sigma_{a'}(1 - x^2) \quad \forall a' \neq a. \]

The negative Cross-Square rule has \( A_{a' a} = C_{a' a} = 1 \) and \( B_{a' a} = 0 \) for all \( a, a' \in A \). Hence, this learning rule is variance averse and monotonically variance averse. Note, however, that the higher the payoff obtained the more probability mass that is taken away from the chosen action. This has the consequence that this rule may not have desirable properties on non-flat environments. In particular, it is not difficult to construct an example in which each action gives a unique payoff (with probability one) and the action with the highest payoff has a negative expected movement.

5 Monotone Mean-Variance Averse Rules

In environments that are not necessarily flat, we now define the most desirable set of actions for an agent who prefers higher means and lower variances. Specifically, we now consider \( A^*_a = \{ a : \pi_a \geq \pi_{a'} \text{ and } v_a \leq v_{a'} \text{ for all } a' \in A \} \) which defines the set of actions that have the maximum expected payoff and the minimum variance of payoff.

Definition 6 A learning rule is monotonically mean-variance averse if \( f(A^*_a) > 0 \) for all environments with \( A^*_a \neq A \) and \( A^*_a \neq \emptyset \).

Every monotonically mean-variance averse learning rule is monotonically variance averse. Hence, every monotonically mean-variance averse learning rule satisfies the conditions we derived on monotonically variance averse rules. The next Proposition provides the necessary and sufficient conditions for a learning rule to be monotonically mean-variance averse. As it will be seen this provides some additional restrictions on the relative magnitudes of the \( B \) and \( C \) coefficients.

Proposition 4 A learning rule is monotonically mean-variance averse if and only if

1. The learning rule is monotonically variance averse.
2. \( B_{a' a} - 2C_{a' a}x \geq 0 \) for all \( x \in (0, 1) \) and for all \( a, a' \in A \).
3. If \( D \) is a nonempty subset of \( A \), then for all \( x \in (0,1) \) there exists an action \( a \in D \) and an action \( a' \in A - D \) such that \( B_{a'a} - 2C_{a'a}x > 0 \).

**Proof. Sufficiency:**

Let \( a \in A^*_x \). Then,

\[
f(a) = \sigma_a \left\{ (B_{aa} \pi_a - C_{aa} s_a) - \sum_{a'} \sigma_{a'} (B_{a'a} \pi_{a'} - C_{a'a} s_{a'}) \right\}
\]

\[
= \sigma_a \left\{ (B_{aa} \pi_a - C_{aa} (\pi_a^2 + v_a)) - \sum_{a'} \sigma_{a'} (B_{a'a} \pi_{a'} - C_{a'a} (\pi_{a'}^2 + v_{a'})) \right\}
\]

\[
= \sigma_a \left\{ \sum_{a'} \sigma_{a'} \left[ (B_{a'a} \pi_a - C_{a'a} \pi_a^2) - (B_{a'a} \pi_{a'} - C_{a'a} \pi_{a'}^2) \right] + \sum_{a'} \sigma_{a'} C_{a'a} (v_{a'} - v_a) \right\}.
\]

The sign of each term in the first sum is given by the sign of

\[
\left[ (B_{a'a} \pi_a - C_{a'a} \pi_a^2) - (B_{a'a} \pi_{a'} - C_{a'a} \pi_{a'}^2) \right].
\]

Since \( a \in A^*_x \), \( \pi_a \geq \pi_{a'} \) for all \( a' \in A \). It follows that if the expression \( (B_{a'a} \pi_a - C_{a'a} \pi_a^2) \) is increasing in \( \pi_a \), then each term of the first sum is non-negative (or, equivalently \( (B_{a'a} \pi_a - C_{a'a} \pi_a^2) \geq (B_{a'a} \pi_{a'} - C_{a'a} \pi_{a'}^2) \) for all \( a' \)). The first derivative of \( (B_{a'a} \pi_a - C_{a'a} \pi_a^2) \) with respect to \( \pi_a \) is given by \( B_{a'a} - 2C_{a'a} \pi_a \). Hence Condition 2 guarantees that \( (B_{a'a} \pi_a - C_{a'a} \pi_a^2) \) is increasing in \( \pi_a \), and it follows that each term of the first sum is non-negative. Furthermore, each term in the second sum is non-negative because of condition 1 in Proposition 2. Since, \( A^*_x \neq A \), condition 3 of the hypotheses and condition 2 of Proposition 2 ensure that at least one of the terms of the first sum or one of the terms of the second sum is strictly positive for some \( a \in A^*_x \).

**Necessity:**

The necessity of condition 1 in this Proposition is obvious. To see that condition 2 is necessary suppose by way of contradiction that \( B_{a'a} - 2C_{a'a}x < 0 \) for some \( a' \neq a \) and some \( x \in (0,1) \). Consider an environment in which \( \mu_a(x) = 1 \) so that \( \pi_a = x \) and \( v_a = 0 \) for some \( x \in (0,1) \). Furthermore, suppose that \( \mu_{a'}(x - \delta) = 1 \) for \( \delta \in (0,x) \) so that \( \pi_{a'} = x - \delta \) and \( v_{a'} = 0 \) for some \( x \in (0,1) \). Lastly, suppose that for any \( a'' \in A \setminus \{a, a'\} \) we have
\( \mu_{x^*}(x - \varepsilon) = 1 \) for \( \varepsilon \in (0, x) \) so that \( \pi_{x^*} = x - \varepsilon \) and \( v_{x^*} = 0 \) for some \( x \in (0, 1) \). Clearly, in this environment \( A^*_x = \{a\} \neq A \). We have

\[
\begin{align*}
f(a) &= \sigma_a \left\{ \sum_{a'} \sigma_{a'} \left[ (B_{a'a}x - C_{a'a}x^2) - (B_{a'a}(x - \delta) - C_{a'a}(x - \delta)^2) \right] \right. \\
&\quad + \sum_{a''} \sigma_{a''} \left[ (B_{a''a}x - C_{a''a}x^2) - (B_{a''a}(x - \varepsilon) - C_{a''a}(x - \varepsilon)^2) \right] \right\}
\end{align*}
\]

For small enough \( \delta \) the first term in the first square bracket is negative. Hence, for small enough \( \varepsilon \), we have \( f(a) < 0 \) which contradicts the learning rule being monotonically mean variance averse.

Finally, to prove the necessity of condition 3, suppose by way of contradiction that for some \( x \in (0, 1) \) we have \( B_{a'a} = 2C_{a'a}x = 0 \) for all \( a, a' \) with \( a \in D \) and \( a' \in A - D \). We know from the necessity of condition 2 of Proposition 2 that \( C_{a'a} > 0 \) for some \( a \in D \) and \( a' \in A - D \). This implies that there exists an \( x' \in (x, 1) \) such that \( B_{a'a} = 2C_{a'a}x < 0 \) which violates condition 2 and therefore violates mean variance aversion. \closedbox

The class of monotonically mean variance averse rules identified in Proposition 4 excludes the class of monotonic learning rules studied in Börgers, Morales and Sarin (2004). Consequently, there must exist environments in which the concern for variance leads to no improvement in terms of means (in the next period).\textsuperscript{15} The next result reveals that monotonically mean variance averse rules are expected to result in behavior that reduces variance in flat environments. In environments in which all actions have the same variance such rules tend to increase expected payoffs.

**Proposition 5** Suppose \( L \) is a monotonically mean-variance averse rule. Then:

1. In every flat environment with \( A^*_x \neq A \), we have \( g_2 < 0 \).

\textsuperscript{15}One could, therefore, investigate the “minimal” modification required of any monotonically mean-variance learning rule that ensures that it is increasing in mean in a “maximally” large class of environments. In a context in which the agent is assumed to have beliefs about the distribution she faces, Maccheroni et al. (2009) analyze a modified version of mean variance preferences, which they call monotone mean variance preferences.
2. In every environment with \( v_a = v_{a'} \) for all \( a, a' \in A \) and \( A^*_a \neq A \), we have \( g_1 > 0 \).

**Proof.** The argument follows from an inductive argument similar to that provided in Proposition 3(ii) in BMS. We omit the details.

We end this section with an example of a learning rule that is monotonically mean-variance averse.

**Example 3** Mean-Variance Cross Learning rule

\[
L_{(a, x)}(a) = \sigma_a + (1 - \sigma_a) \left( x - \frac{1}{2} x^2 \right)
\]

\[
L_{(a, x)}(a') = \sigma_{a'} - \sigma_{a'} \left( x - \frac{1}{2} x^2 \right) \quad \forall a' \neq a.
\]

Comparing Example 2 with Example 3 we see that in the Mean-Variance Cross learning rule the updated probability of the chosen action responds positively to the obtained payoff whereas in the Negative Cross learning rule this response is negative. Intuitively, this difference allows the Mean-Variance Cross learning rule to respond positively to the expected payoff of the chosen action.

**6 Discussion**

In the class of learning rules in which people learn from their experience, we have characterized those that would be expected to lead them to choose actions with higher means and lower variances in all environments. In our analysis the current behavior of the agent is taken as given. It is also short run, being concerned with the change in behavior from one period to the next. Our short run analysis could be extended to include the long run properties of learning. Such an extension would require studying learning at different possible behaviors. Future work could explore the long run behavior of individuals whose learning rules satisfy the characterizations we provide at all behaviors.

There exist learning rules that are expected to increase probability mass on the safer actions in some environments and for some given behaviors. In
particular, Denrell (2008) studies a class of learning rules in which agents summarize the quality of an action according to its assessment or score and choose probabilistically. This class includes many rules that are well known in the experimental literature. He shows that, under certain conditions, these rules are expected to increase probability mass on the safer actions over time. These rules, however, are not variance averse in the sense of this paper.

REFERENCES


---


