CIRJE-F-34

# Learning as the Dynamics of Emotions and Strength of Confidence

Hitoshi Matsushima
University of Tokyo

December 1998

Discussion Papers are a series of manuscripts in their draft form. They are not intended for circulation or distribution except as indicated by the author. For that reason Discussion Papers may not be reproduced or distributed without the written consent of the author.

## Abstract

This paper investigate a Bandit problem in which a decision maker chooses between the safe action and the uncertain action infinitely many times. She does not know the expected payoff for the uncertain action. Her decision on whether to experiment is influenced by the atmosphere in the society. An adaptive learning rule is introduced which regulates her unconscious dynamics of emotions. We argue that realistic decision makers mostly fail to choose the efficient action in the long run, even though they experiment infinitely often. We present a necessary and sufficient condition for efficient learning which requires a very restrictive psychological nature that the confidence which a decision maker has in the uncertain action being better is the strongest.



**JEL Classification Numbers:** D80, D81, D83.

**Key Words:** Uncertainty, Learning, Experimentation, Atmosphere, Strength of Confidence.

# Learning as the Dynamics of Emotions and Strength of Confidence

Hitoshi Matsushima+

Faculty of Economics, University of Tokyo

December 15, 1998

+ Faculty of Economics, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan. Fax: (03)3818-7082. e-mail: hitoshi@e.u-tokyo.ac.jp

# 1. Introduction

This paper investigates a Bandit problem in which a decision maker chooses between the safe action and the uncertain action infinitely many times. She knows the payoff for the safe action, but does not know the expected payoff for the uncertain action. The purpose of this paper is to clarify the possibility that the decision maker comes to choose the efficient action in the long run. We argue that realistic decision makers mostly fail to choose the efficient action in the long run, even though they experiment infinitely often.

In a daily life, a person sometimes encounters into the opportunity to decide whether to challenge something special. A scientist bothers deciding whether to try to discover a breakthrough or be engaged in a follow-up. An entrepreneur decides whether to devote herself to a routine work or look for a new enterprise whose consequence can not be anticipated. As Knight (1921) has emphasized, whenever an entrepreneur looks for a profitable enterprise then it is inevitable that she faces "radical uncertainty" in which probabilities are unknown and not well-estimated.

When radical uncertainty is present, the basis of rational calculation for decision is greatly weakened. Instead, psychological forces play the central role. Of particular importance, a decision maker is influenced by the atmosphere in the society. The more merry the atmosphere is, the more likely it is in the decision maker's mind that the uncertain action yields the higher payoff. The atmosphere emerges from the mass psychology as being outside her control, and randomly fluctuates due to payoff-irrelevant factors.[1]

Based on these observations, we model a decision maker by an adaptive learning rule which regulates the decision maker's unconscious dynamics of emotions. The decision maker decides to choose the uncertain action if and only if her "state of mind" suggests that the atmosphere is merry enough to make her optimistic. The higher the state of mind is, the wider the range of atmosphere which urges the decision maker to choose or experiment with the uncertain action extends. Because of the dissonance between her subjective belief and the experiences, the state of mind is updated in ways that the state of mind becomes higher (respectively, lower) when she obtains a higher

[1] Keynes (1936, p.154) has argued: "The market will be subject to waves of optimistic and pessimistic sentiment, which are unreasoning and yet in a sense legitimate where no solid basis exists for a reasonable calculation".

(respectively, lower) payoff than the payoff for the safe action.

We characterize learning rules according to which the decision maker always comes to choose the efficient action in the long run. When the likelihood that the merry atmosphere occurs is the *highest*, the decision maker is always urged to experiment with the uncertain action most often, and therefore, comes to choose the efficient action in the long run irrespective of how to specify a learning rule.

It is, however, very rare in real environments that this likelihood is the highest. When this likelihood is *not* the highest, the consequences change drastically: We introduce an index accompanied with a learning rule which expresses the *confidence* that the decision maker has in the uncertain action being better than the safe action. Roughly speaking, this index is the proportion of the number of possible states of mind which urge the decision maker to choose the uncertain action to the number of all possible states of mind. The larger this index is, the stronger this confidence is. We show as the main statement of this paper that a learning rule is efficient if and only if this confidence is the *strongest*. This necessary and sufficient condition for efficient learning implies the very restrictive psychological nature of human being. Typical learning rules do not satisfy this condition, and therefore, are never efficient.

We also shows that the set of efficient learning rules is *not* empty, though it is indeed very narrow. Importantly, such special learning rules always lead the decision maker to choose the efficient action *irrespective* of how often the merry atmosphere occurs.

The previous works related to this paper are Sarin and Vahid (1997) and Matsushima (1997,1998a), which showed that a decision maker never chooses the uncertain action in the long run. They assumed that the decision maker is not influenced by the atmosphere, and never experiments with the uncertain action once she believes it less profitable than the safe action. In the present paper, on the contrary to these works, decision makers always experiment infinitely often, but, nevertheless, mostly fail to achieve long-run efficiency except the decision makers who have a very special psychological nature such as the strongest confidence.

This paper assumes that the decision maker is well-motivated and never "mistakes". By contrast, if a decision maker always makes a mistake to choose the uncertain action with at least some positive probability, any perturbation would avoid the safe action being "absorbing" and make it possible to achieve long-run efficiency. Needless to say, such an analogy to a "mutation" is not compelling when a decision is a well-laid plan and divides into several sub-decisions which are carefully made step by step.

At the present time of academic circumstance, it must be admitted that we still have a serious lack of a theoretical foundation on how a real individual learns from past experiences and makes a decision under uncertainty.[2] Hence, it might be inevitable to ask ad hoc assumptions for help to a certain extent, such as the assumptions that a learning rule is a Markov chain with reflecting barriers, there are only two alternative actions, the rule of updating is time-independent, and so on. Despite of this difficulty, we should make our best endeavors to bring in a "reality" into a model of learning. This paper shows that long-run efficiency may be trivially achieved by an artificially designed learning rule, but, once such unrealistic learning rules are excluded, it would turn out to be far from triviality.

Finally, we must mention that this paper stands against Bayesian learning, which is at present regarded as a well-developed theory with the solid foundation of Bayesian statistical methods. A Bayesian learner summarizes her uncertainty with some simplified subjective prior distribution, and makes her view of the world more accurate according to a process of Bayesian updating on the presumption that the experiences can never contradict her fixing subjective prior. Rothschild (1974), for example, is a related paper on the basis of the Bayesian framework. The present paper will not follow the Bayesian framework, because when radical uncertainty are present Bayesian learning is quite unrealistic and even has little ability to converge to the right predictions.[3]

The organization of this paper is as follows. Section 2 presents the model of Bandit problem. Section 3 introduces a learning rule defined by a Markov chain with reflecting barriers. We will investigate a learning rule which is not restricted by the reflecting barriers too much, and therefore, is approximated by the limit of an infinite sequence of learning rules. Section 4 presents the main theorem, and section 5 gives the complete proof of it. Finally, section 6 generalizes learning rules and investigates sophisticated learners.

---

[2] For the recent progress of learning theory, see Marimon (1997) and its references.

[3] For more detailed discussions, see Marimon (1997) and its references again.

## 2. Bandit Problem

We consider a Bandit problem in which a decision maker chooses between the *safe* action and the *uncertain* action infinitely many times. At the beginning of every period $t \geq 1$, the decision maker observes a *context* denoted by an integer $c(t)$. The decision maker interprets:

*The smaller the context is, the more merry the atmosphere in the society is.*

The decision maker then chooses an action $a(t)$ and obtains a payoff $v(t)$. The payoff for the safe action is equal to $0$ and she a priori knows this payoff. The payoff for the uncertain action is randomly determined. It is equal to $1$ with probability $p \in (0,1)$, whereas it is equal to $-1$ with probability $1-p$. Hence, the uncertain action is more (respectively, less) profitable than the safe action if $p > \frac{1}{2}$ (respectively, $p < \frac{1}{2}$). We assume that the decision maker does not know this probability $p$.

The context is randomly determined according to a probability function $f$ on the set of all integers: For every integer $c$, context $c$ occurs with probability $f(c)$. The context is payoff-irrelevant in the sense that the payoff for the uncertain action is determined independently of the context.

Define the cumulative distribution by $F(c) \equiv \sum_{c' \leq c} f(c')$. We assume

$$f(c) > 0 \text{ for all integers } c,$$

which guarantees that the decision maker always experiments with the uncertain action *infinitely often*.

## 3. Learning Rule

A *learning rule* is defined by a Markov chain with reflecting barriers $\Gamma \equiv (\underline{x}, \overline{x})$, where $\underline{x}$ and $\overline{x}$ are integers and $\underline{x} < 0 < \overline{x}$. In every period $t$, the decision maker's *state of mind* is denoted by an integer $x(t)$ in the set $\{\underline{x}, \ldots, \overline{x}\}$ which is regarded as being the maximal context in which she chooses the uncertain action. The decision maker chooses the uncertain action if and only if

$$c(t) \leq x(t),$$

i.e., if and only if the current atmosphere is at least as merry as the atmosphere in the context $c = x(t)$. Given a state of mind $x$, the probability that the decision maker chooses the uncertain action is equal to the cumulative distribution $F(x)$.

The state of mind $x(t)$ is updated in ways that the state of mind becomes higher (respectively, lower) when she obtains the positive (respectively, negative) payoff. When the decision maker chooses the uncertain action in period $t$,

$$x(t+1) = x(t) + 1 \text{ if } v(t) = 1 \text{ and } x(t) < \overline{x},$$
$$x(t+1) = x(t) - 1 \text{ if } v(t) = -1 \text{ and } x(t) > \underline{x},$$

and

$$x(t+1) = x(t) \text{ otherwise.}$$

When she chooses the safe action, the state of mind does not change, i.e., $x(t+1) = x(t)$. Hence, the probability that the state of mind changes from $x \neq \overline{x}$ into $x+1$ is $pF(x)$, and the probability that the state of mind changes from $x \neq \underline{x}$ into $x-1$ is $(1-p)F(x)$.

The stationary distribution of state of mind in the long run, $g(x) = g(x, p, f, \Gamma)$, is defined by the following inequalities:

$$g(x) = \{1 - F(x)\}g(x) + pF(x-1)g(x-1)$$
$$+ (1-p)F(x+1)g(x+1) \text{ for all } x \notin \{\underline{x}, \overline{x}\}, \quad (1)$$
$$g(\underline{x}) = \{1 - F(\underline{x}) + (1-p)F(\underline{x})\}g(\underline{x})$$
$$+ (1-p)F(\underline{x}+1)g(\underline{x}+1), \quad (2)$$

and

$$g(\overline{x}) = \{1 - F(\overline{x}) + pF(\overline{x})\}g(\overline{x}) + pF(\overline{x}-1)g(\overline{x}-1).$$

Here, $g(x)$ is the stationary probability that the state of mind is equal to $x$ in the long run. Finally, the stationary probability that the decision maker chooses the uncertain action in the long run is defined by

$$\Lambda = \Lambda(p, f, \Gamma) \equiv \sum_{x=\underline{x}}^{\overline{x}} F(x)g(x).$$

Our purpose is to clarify the possibility that a learning rule is (approximately) efficient in the sense that $\Lambda$ is close to 1 if $p > \frac{1}{2}$, whereas $\Lambda$ is close to 0 if $p < \frac{1}{2}$. In order for a learning rule to be efficient, it is necessary that the transition of state of mind is not restricted by the reflecting barriers too much. Hence, we shall confine our attention to a learning rule which is approximated by the limit of a *sequence of learning rules* $\Gamma^\infty \equiv (\Gamma_m)_{m=1}^{+\infty}$ such that

$$\Gamma_m = (\underline{x}_m, \overline{x}_m),$$
$$\lim_{m \to +\infty} \underline{x}_m = -\infty \text{ and } \lim_{m \to +\infty} \overline{x}_m = +\infty.$$

The stationary probability that the decision maker chooses the uncertain action in the long run is approximated by

$$\Lambda^\infty = \Lambda^\infty(p, f, \Gamma^\infty) \equiv \lim_{m \to +\infty} \Lambda(p, f, \Gamma_m).$$

**Definition:** A sequence of learning rules $\Gamma^\infty$ is *efficient* with respect to $f$ if for every $p \in (0,1)$,

$$\Lambda^\infty(p, f, \Gamma^\infty) = 1 \text{ if } p > \frac{1}{2},$$

and

$$\Lambda^\infty(p, f, \Gamma^\infty) = 0 \text{ if } p < \frac{1}{2}.$$

## 4. The Main Theorem

We introduce two indices $\gamma$ and $\beta$ which are relevant to a sequence of learning rule $\Gamma^{\infty}$ and a probability distribution of context $f$, respectively.

First, we define

$$\gamma = \gamma(\Gamma^{\infty}) \equiv \lim_{m \to +\infty} \left( \frac{\overline{x}_m}{\overline{x}_m - \underline{x}_m} \right).$$

The index $\gamma$ expresses the proportion of the number of states of mind which urge the decision maker to choose the uncertain action to the number of possible states of mind in the following sense. Fix $c$ and $m$ arbitrarily. When the decision maker conforms to $\Gamma_m$, she chooses the uncertain action in context $c$ if and only if the state of mind $x$ is larger than or equal to $c$. Hence, the number of states of mind which urge the decision maker to choose the uncertain action is $\overline{x}_m - c + 1$. The proportion of it to the number of possible states of mind $\overline{x}_m - \underline{x}_m$ is $\frac{\overline{x}_m - c + 1}{\overline{x}_m - \underline{x}_m}$ which converges to $\gamma$ as $m$ increases, whichever context $c$ is fixed.

The index $\gamma$ implies the *strength of confidence* that the decision maker has in the uncertain action being better than the safe action: The higher the index $\gamma$ is, the stronger the confidence is. The confidence is the strongest when $\gamma = 1$, whereas it is the weakest when $\gamma = 0$.

Next, we define

$$\beta = \beta(f) \equiv \lim_{c \to -\infty} \frac{F(c-1)}{F(c)}.$$

**Lemma 1:** *If* $\beta(f) > \beta(\tilde{f})$, *then for every small enough* $c$,

$$F(c) > \tilde{F}(c),$$

*where* $\tilde{F}(c)$ *is the cumulative distribution associated with* $\tilde{f}$.

**Proof:** Choose positive real numbers $b$ and $\tilde{b}$ such that

$$\beta(f) > b > \tilde{b} > \beta(\tilde{f}).$$

From the definition of $\beta(\cdot)$, there exists an integer $\hat{c}$ such that

$$\frac{F(c-1)}{F(c)} > b \text{ and } \frac{\tilde{F}(c-1)}{\tilde{F}(c)} < \tilde{b} \text{ for all } c \leq \hat{c}.$$

Hence,

$$F(c) > b^{\hat{c}-c} F(\hat{c}) \text{ and } \tilde{F}(c) < (\tilde{b})^{\hat{c}-c} \tilde{F}(\hat{c}) \text{ for all } c \leq \hat{c}.$$

Since $b^{\hat{c}-c}F(\hat{c}) > (\widetilde{b})^{\hat{c}-c}\widetilde{F}(\hat{c})$ for every small enough $c$, one gets

$$F(c) > \widetilde{F}(c) \text{ for every small enough } c.$$

**Q.E.D.**

From Lemma 1, we can regard the index $\beta$ as expressing the *likelihood* that the merry atmosphere occurs. The higher $\beta$ is, the more likely it is that the merry atmosphere occurs. This likelihood is the highest when $\beta = 1$, whereas it is the lowest when $\beta = 0$.

Based on these indices, we present the main theorem of this paper as follows.

**Theorem 2:**

(i) *Suppose* $\beta(f) = 1$. *Then, any sequence of learning rules* $\Gamma^{\infty}$ *is efficient with respect to* $f$.

(ii) *Suppose* $0 < \beta(f) < 1$. *Then, a sequence of learning rules* $\Gamma^{\infty}$ *is efficient if and only if*

$$\gamma(\Gamma^{\infty}) = 1.$$

The implication of part (i) is as follows: Equality $\beta(f) = 1$ implies the highest likelihood that the merry atmosphere occurs, which guarantees that the decision maker experiments with the uncertain action most often, and therefore, succeeds to choose the efficient action in the long run, irrespective of how a sequence of learning rules is specified.

The drawback of part (i) is that it is very rare that the exogenous distribution $f$ possesses the highest likelihood, and it is rather typical that this likelihood is neither the highest nor the lowest, i.e., $0 < \beta(f) < 1$. Part (ii) of Theorem 2 says that in this typical case, the consequences change drastically: A sequence of learning rules is efficient if and only if it possesses the *strongest* confidence that the decision maker has in the uncertain action being better than the safe action. This necessary and sufficient condition implies the very restrictive psychological nature of human being. Most sequences of learning rules do not satisfy this condition, and therefore, are never efficient.

Part (ii) of Theorem 2 also says that the set of efficient sequences of learning rules is *non-empty*, though it is very narrow. The following corollary is straightforward from Theorem 2.

**Corollary 3:** *If a sequence of learning rules is efficient with respect to some* $f$ *such* *that* $0 < \beta(f) < 1$, *then it is efficient with respect to every* $f'$ *such that* $0 < \beta(f') < 1$.

Corollary 3 says that any sequence of learning rule $\Gamma^{\infty}$ such that the confidence is the strongest is efficient with respect to almost all $f$. According to such a sequence of learning rule, the decision maker always comes to choose the efficient action irrespective of how often the merry atmosphere occurs, as well as irrespective of how often payoff 1 is realized.

In the next section, we present the complete proof of Theorem 2.

## 5. Proof of the Main Theorem

We denote by $\lambda(x) = \lambda(x, p, f)$ the proportion of the probability of transiting from state of mind $x-1$ into $x$ to the probability of transiting from $x$ into $x-1$. We must note

$$\lambda(x) = \frac{pF(x-1)}{(1-p)F(x)}.$$

The proportion of the probability of transiting from the lowest state of mind $\underline{x}_m$ into the highest state of mind $\bar{x}_m$ during $\bar{x}_m - \underline{x}_m$ periods to the probability of transiting from $\bar{x}_m$ into $\underline{x}_m$ during $\bar{x}_m - \underline{x}_m$ periods is equal to $\prod_{x=\underline{x}_m+1}^{\bar{x}_m} \lambda(x)$.

**Lemma 4:**

$$\lim_{x \to -\infty} \lambda(x) = \left(\frac{p}{1-p}\right)\beta, \tag{3}$$

$$\lim_{x \to +\infty} \lambda(x) = \frac{p}{1-p}, \tag{4}$$

and

$$\lim_{m \to +\infty} \left( \prod_{x=\underline{x}_m+1}^{\bar{x}_m} \lambda(x) \right)^{\frac{1}{\bar{x}_m - \underline{x}_m}} = \left(\frac{p}{1-p}\right)\beta^{1-\gamma}. \tag{5}$$

**Proof:** It is straightforward from the definition of $\beta(\cdot)$ that equalities (3) and (4) hold. Equalities (3) and (4) say

$$\lim_{m \to +\infty} \left( \prod_{x=\underline{x}_m+1}^{0} \lambda(x) \right)^{-\frac{1}{\underline{x}_m}} = \left(\frac{p}{1-p}\right)\beta \text{ and } \lim_{m \to +\infty} \left( \prod_{x=1}^{\bar{x}_m} \lambda(x) \right)^{\frac{1}{\bar{x}_m}} = \frac{p}{1-p},$$

and therefore, one gets

$$\begin{aligned}
&\lim_{m \to +\infty} \left( \prod_{x=\underline{x}_m+1}^{\bar{x}_m} \lambda(x) \right)^{\frac{1}{\bar{x}_m - \underline{x}_m}} \\
&= \lim_{m \to +\infty} \left\{ \left( \prod_{x=\underline{x}_m+1}^{0} \lambda(x) \right)^{-\frac{1}{\underline{x}_m}} \right\}^{\frac{-\underline{x}_m}{\bar{x}_m - \underline{x}_m}} \times \lim_{m \to +\infty} \left\{ \left( \prod_{x=1}^{\bar{x}_m} \lambda(x) \right)^{\frac{1}{\bar{x}_m}} \right\}^{\frac{\bar{x}_m}{\bar{x}_m - \underline{x}_m}} \\
&= \left(\frac{p}{1-p}\right)^{1-\gamma} \beta^{1-\gamma} \times \left(\frac{p}{1-p}\right)^{\gamma} = \left(\frac{p}{1-p}\right)\beta^{1-\gamma}.
\end{aligned}$$

Hence, equality (5) holds.

**Q.E.D.**

The following proposition is substantial for the proof of Theorem 2.

**Proposition 5:** *For every* $p \in (0,1)$,

(iii) $$\Lambda^{\infty} = 1 \text{ if } \left(\frac{p}{1-p}\right)\beta^{1-\gamma} > 1,$$

*and*

(iv) $$\Lambda^{\infty} = 0 \text{ if } \left(\frac{p}{1-p}\right)\beta^{1-\gamma} < 1.$$

Proposition 5 says that:

(iii) The decision maker chooses the uncertain action in the long run, if the probability of transiting from the lowest state of mind into the highest during $\bar{x}_m - \underline{x}_m$ periods is larger than the probability of transiting from the highest into the lowest during $\bar{x}_m - \underline{x}_m$ periods.

(iv) She chooses the safe action in the long run, if the probability of transiting from the lowest into the highest during $\bar{x}_m - \underline{x}_m$ periods is smaller than the probability of transiting from the highest into the lowest during $\bar{x}_m - \underline{x}_m$ periods.

Proposition 5 implies that:

(iii) $$\Lambda^{\infty} = 1 \text{ if } p > \frac{1}{1+\beta^{1-\gamma}},$$

whereas

(iv) $$\Lambda^{\infty} = 0 \text{ if } p < \frac{1}{1+\beta^{1-\gamma}}.$$

Hence, a sequence of learning rules $\Gamma^{\infty}$ is efficient with respect to $f$ if and only if

$$\beta(f)^{1-\gamma(\Gamma^{\infty})} = 1.$$

This equality holds if and only if

$$\text{either } \beta(f) = 1, \text{ or, } \gamma(\Gamma^{\infty}) = 1 \text{ and } \beta(f) > 0.$$

Hence, we have completed the proof of Theorem 2.

All we have to do is to prove Proposition 5. We denote

$$g_m(x) \equiv g(x, p, \Gamma_m).$$

The following lemma says that the proportion of the probability of emergence of state of mind $x$ to the probability of emergence of state of mind $x+1$ is approximated by the proportion of the probability of transiting from state of mind $x-1$ into $x$ to the probability of transiting from $x$ into $x-1$.

**Lemma 6:** For every $m$ and every $x \in \{\underline{x}_m + 1, \ldots, \bar{x}_m\}$,

$$\frac{g_m(x)}{g_m(x-1)} = \lambda(x), \qquad (6)$$

and

$$\lim_{m\to+\infty} \left(\frac{g_m(\bar{x}_m)}{g_m(\underline{x}_m)}\right)^{\frac{1}{\bar{x}_m-\underline{x}_m}} = \left(\frac{p}{1-p}\right)\beta^{1-\gamma}. \qquad (7)$$

**Proof:** Equality (2) says

$$\frac{g_m(\underline{x}+1)}{g_m(\underline{x})} = \lambda(\underline{x}+1),$$

which, together with equality (1), means that for every $x' \in \{\underline{x}_m + 2, \ldots, \bar{x}_m\}$, if equality (6) holds for all $x < x'$, then

$$\begin{aligned}
&F(x'-1)g_m(x'-1) - pF(x'-2)g_m(x'-2) + (1-p)F(x')g_m(x') \\
&= F(x'-1)g_m(x'-1) - (1-p)F(x'-1)g_m(x'-1) \\
&+ (1-p)F(x')g_m(x') \\
&= -pF(x'-1)g_m(x'-1) + (1-p)F(x')g_m(x') \\
&= 0,
\end{aligned}$$

that is,

$$\frac{g_m(x')}{g_m(x'-1)} = \frac{pF(x'-1)}{(1-p)F(x')} = \lambda(x').$$

Hence, we have proved that equality (6) holds. Moreover, it is straightforward from equalities (5) and (6) that

$$\lim_{m\to+\infty} \left(\frac{g_m(\bar{x}_m)}{g_m(\underline{x}_m)}\right)^{\frac{1}{\bar{x}_m-\underline{x}_m}} = \lim_{m\to+\infty} \left(\prod_{x=\underline{x}_m+1}^{\bar{x}_m} \lambda(x)\right)^{\frac{1}{\bar{x}_m-\underline{x}_m}} = \left(\frac{p}{1-p}\right)\beta^{1-\gamma}.$$

**Q.E.D.**

The following lemma says that the probability of emergence of low states of mind is near zero, provided that the uncertain action is more profitable than the safe action and the inequality in part (iii) holds.

**Lemma 7:** *If* $\left(\frac{p}{1-p}\right)\beta^{1-\gamma} > 1$ *and* $\frac{p}{1-p} > 1$, *then for every large enough positive integer* $x$ *and every small enough negative integer* $\tilde{x}$,

$$\lim_{m\to+\infty} \frac{\sum_{x'=x+1}^{\bar{x}_m} g_m(x')}{\sum_{x'=\underline{x}_m}^{\tilde{x}-1} g_m(x')} = +\infty.$$

**Proof:** We can choose positive real numbers $\varepsilon$ and $\widetilde{\varepsilon}$ such that

$$\frac{p}{1-p} > \varepsilon, \quad (\frac{p}{1-p})\beta > \widetilde{\varepsilon},$$

$$\varepsilon \neq 1, \quad \widetilde{\varepsilon} < 1,$$

and

$$\varepsilon^{\gamma}(\widetilde{\varepsilon})^{1-\gamma} > 1.$$

Equalities (3) and (4) say that for every large enough positive integer $x$ and every small enough negative integer $\widetilde{x}$,

$$\lambda(x') \geq \varepsilon \text{ for all } x' \geq x,$$

and

$$\lambda(x') \geq \widetilde{\varepsilon} \text{ for all } x' \leq \widetilde{x}.$$

Equality (6) says

$$g_m(x') = \{\prod_{x''=x+1}^{x'} \lambda(x'')\} g_m(x) \geq \varepsilon^{x'-x} g_m(x) \text{ for all } x' > x,$$

and therefore,

$$\sum_{x'=x+1}^{\bar{x}_m} g_m(x') \geq \varepsilon(\frac{\varepsilon^{\bar{x}_m - x} - 1}{\varepsilon - 1}) g_m(x).$$

Equality (6) says also

$$g_m(x') = \{\prod_{x''=x'}^{\widetilde{x}-1} \frac{1}{\lambda(x''+1)}\} g_m(\widetilde{x}) \leq (\widetilde{\varepsilon})^{-(\widetilde{x}-x')} g_m(\widetilde{x}) \text{ for all } x' < \widetilde{x},$$

and therefore,

$$\sum_{x'=\underline{x}_m}^{\widetilde{x}-1} g_m(x') \leq (\frac{(\widetilde{\varepsilon})^{\underline{x}_m - \widetilde{x}} - 1}{1 - \widetilde{\varepsilon}}) g_m(\widetilde{x}).$$

Hence, one gets

$$\frac{\sum_{x'=x+1}^{\bar{x}_m} g_m(x')}{\sum_{x'=\underline{x}_m}^{\widetilde{x}-1} g_m(x')} \geq (\frac{\varepsilon(1-\widetilde{\varepsilon})}{\varepsilon - 1})(\frac{\varepsilon^{\bar{x}_m - x} - 1}{(\widetilde{\varepsilon})^{\underline{x}_m - \widetilde{x}} - 1}) \frac{g_m(x)}{g_m(\widetilde{x})}$$

$$= (\frac{\varepsilon(\widetilde{\varepsilon} - 1)}{1 - \varepsilon})(\frac{((\widetilde{\varepsilon})^{(\frac{-\underline{x}_m + \widetilde{x}}{\bar{x}_m - \underline{x}_m})} \varepsilon^{(\frac{\bar{x}_m - x}{\bar{x}_m - \underline{x}_m})})^{\bar{x}_m - \underline{x}_m} - (\widetilde{\varepsilon})^{-\underline{x}_m + \widetilde{x}}}{1 - (\widetilde{\varepsilon})^{-\underline{x}_m + \widetilde{x}}}) \prod_{x''=x'+1}^{x} \lambda(x''),$$

which approaches $+\infty$ as $m \to +\infty$, because

$$\lim_{m \to +\infty} (\widetilde{\varepsilon})^{-\underline{x}_m + \widetilde{x}} = 0, \quad \lim_{m \to +\infty} \frac{-\underline{x}_m + \widetilde{x}}{\bar{x}_m - \underline{x}_m} = 1 - \gamma, \quad \lim_{m \to +\infty} \frac{\bar{x}_m - x}{\bar{x}_m - \underline{x}_m} = \gamma,$$

and

$$\lim_{m \to +\infty} ((\widetilde{\varepsilon})^{1-\gamma} \varepsilon^{\gamma})^{\bar{x}_m - \underline{x}_m} = +\infty.$$

**Q.E.D.**

The following lemma says that the probability of emergence of medium states of mind is near zero, provided that the uncertain action is more profitable than the safe action.

**Lemma 8:** *If* $\frac{p}{1-p} > 1$, *then for every large enough positive integer* $x$ *and every integer* $\tilde{x} < x$,

$$\lim_{m \to +\infty} \frac{\sum_{x'=x+1}^{\bar{x}_m} g_m(x')}{\sum_{x'=\tilde{x}}^{x} g_m(x')} = +\infty .$$

**Proof:** Choose $\varepsilon$ such that

$$\frac{p}{1-p} > \varepsilon > 1 .$$

Equality (4) says that for every large enough positive integer $x$,

$$\lambda(x') \geq \varepsilon \text{ for all } x' \geq x .$$

In the same way as Lemma 7, one gets

$$\sum_{x'=x+1}^{\bar{x}_m} g_m(x') \geq \varepsilon(\frac{\varepsilon^{\bar{x}_m - x} - 1}{\varepsilon - 1}) g_m(x).$$

Fix an integer $\tilde{x} < x$ arbitrarily. Equality (6) says

$$\sum_{x'=\tilde{x}}^{x} g_m(x') = \sum_{x'=\tilde{x}}^{x} \{ \prod_{x''=x'+1}^{x} \frac{1}{\lambda(x'')} \} g_m(x),$$

and therefore,

$$\frac{\sum_{x'=x+1}^{\bar{x}_m} g_m(x')}{\sum_{x'=\tilde{x}}^{x} g_m(x')} \geq \varepsilon(\frac{\varepsilon^{\bar{x}_m - x} - 1}{\varepsilon - 1}) [\sum_{x'=\tilde{x}}^{x} \{ \prod_{x''=x'+1}^{x} \frac{1}{\lambda(x'')} \}]^{-1} ,$$

which approaches $+\infty$ as $m \to +\infty$, because $\lim_{m \to +\infty} \varepsilon^{\bar{x}_m - x} = +\infty$.

**Q.E.D.**

The following lemma says that the probability of emergence of high states of mind is near zero, provided that the inequality in part (iv) and inequality $(\frac{p}{1-p})\beta < 1$ hold.

**Lemma 9:** *If* $(\frac{p}{1-p})\beta^{1-\gamma} < 1$ *and* $(\frac{p}{1-p})\beta < 1$, *then for every large enough positive integer* $x$ *and every small enough negative integer* $\tilde{x}$,

$$\lim_{m\to+\infty} \frac{\sum_{x'=x+1}^{\bar{x}_m} g_m(x')}{\sum_{x'=\underline{x}_m}^{\tilde{x}-1} g_m(x')} = 0.$$

**Proof:** We can choose positive real numbers $\varepsilon$ and $\tilde{\varepsilon}$ such that

$$\frac{p}{1-p} \le \varepsilon,\ (\frac{p}{1-p})\beta \le \tilde{\varepsilon},$$

$$\varepsilon \ne 1,\ \tilde{\varepsilon} < 1,$$

and

$$\varepsilon^{\gamma}(\tilde{\varepsilon})^{1-\gamma} < 1.$$

Equalities (3) and (4) say that for every large enough positive integer $x$ and every small enough negative integer $\tilde{x}$,

$$\lambda(x') \le \varepsilon \text{ for all } x' \ge x,$$

and

$$\lambda(x') \le \tilde{\varepsilon} \text{ for all } x' \le \tilde{x}.$$

Equality (6) says

$$g_m(x') = \{\prod_{x''=x+1}^{x'} \lambda(x'')\} g_m(x) \le \varepsilon^{x'-x} g_m(x) \text{ for all } x' > x,$$

and therefore,

$$\sum_{x'=x+1}^{\bar{x}_m} g_m(x') \le \varepsilon(\frac{\varepsilon^{\bar{x}_m - x} - 1}{\varepsilon - 1}) g_m(x).$$

Equality (6) says also

$$g_m(x') = \{\prod_{x''=x'}^{\tilde{x}-1} \frac{1}{\lambda(x''+1)}\} g_m(\tilde{x}) \ge (\tilde{\varepsilon})^{-(\tilde{x}-x')} g_m(\tilde{x}) \text{ for all } x' < \tilde{x},$$

and therefore,

$$\sum_{x'=\underline{x}_m}^{\tilde{x}-1} g_m(x') \ge (\frac{(\tilde{\varepsilon})^{\underline{x}_m - \tilde{x}} - 1}{1-\tilde{\varepsilon}}) g_m(\tilde{x}).$$

Hence, one gets

$$\frac{\sum_{x'=x+1}^{\bar{x}_m} g_m(x')}{\sum_{x'=\underline{x}_m}^{\tilde{x}-1} g_m(x')} \le (\frac{\varepsilon(1-\tilde{\varepsilon})}{\varepsilon - 1})(\frac{\varepsilon^{\bar{x}_m - x} - 1}{(\tilde{\varepsilon})^{\underline{x}_m - \tilde{x}} - 1}) \frac{g_m(x)}{g_m(\tilde{x})}$$

$$= \left(\frac{\varepsilon(\tilde{\varepsilon}-1)}{1-\varepsilon}\right)\left(\frac{\left((\tilde{\varepsilon})^{\left(\frac{-\underline{x}_m+\tilde{x}}{\bar{x}_m-\underline{x}_m}\right)}\varepsilon^{\left(\frac{\bar{x}_m-x}{\bar{x}_m-\underline{x}_m}\right)}\right)^{\bar{x}_m-\underline{x}_m}-(\tilde{\varepsilon})^{-\underline{x}_m+\tilde{x}}}{1-(\tilde{\varepsilon})^{-\underline{x}_m+\tilde{x}}}\right)\prod_{x''=x'+1}^{x}\lambda(x''),$$

which approaches zero as $m \to +\infty$, because

$$\lim_{m\to+\infty}(\tilde{\varepsilon})^{-\underline{x}_m+\tilde{x}} = 0, \quad \lim_{m\to+\infty}\frac{-\underline{x}_m+\tilde{x}}{\bar{x}_m-\underline{x}_m} = 1-\gamma, \quad \lim_{m\to+\infty}\frac{\bar{x}_m-x}{\bar{x}_m-\underline{x}_m} = \gamma,$$

and

$$\lim_{m\to+\infty}\left((\tilde{\varepsilon})^{1-\gamma}\varepsilon^{\gamma}\right)^{\bar{x}_m-\underline{x}_m} = 0.$$

**Q.E.D.**

The following lemma says that the probability of emergence of medium states of mind is near zero, provided that inequality $(\frac{p}{1-p})\beta < 1$ holds.

**Lemma 10:** *If* $(\frac{p}{1-p})\beta < 1$, *then for every small enough negative integer* $\tilde{x}$ *and every integer* $x > \tilde{x}$,

$$\lim_{m\to+\infty}\frac{\sum_{x'=\underline{x}_m}^{\tilde{x}-1} g_m(x')}{\sum_{x'=\tilde{x}}^{x} g_m(x')} = +\infty.$$

**Proof:** Choose $\varepsilon$ such that

$$(\frac{p}{1-p})\beta < \varepsilon < 1.$$

Equality (3) says that for every small enough negative integer $\tilde{x}$,

$$\lambda(x') \le \varepsilon \text{ for all } x' \le \tilde{x}.$$

In the same way as Lemma 9, one gets

$$\sum_{x'=\underline{x}_m}^{\tilde{x}-1} g_m(x') \ge \left(\frac{(\tilde{\varepsilon})^{\underline{x}_m-\tilde{x}}-1}{1-\tilde{\varepsilon}}\right) g_m(\tilde{x}).$$

Fix an integer $x < \tilde{x}$ arbitrarily. Equality (3) says

$$\sum_{x'=\tilde{x}}^{x} g_m(x') = \sum_{x'=\tilde{x}}^{x}\left\{\prod_{x''=\tilde{x}+1}^{x'}\lambda(x'')\right\} g_m(\tilde{x}),$$

and therefore,

$$\frac{\sum_{x'=\underline{x}_m}^{\tilde{x}-1} g_m(x')}{\sum_{x'=\tilde{x}}^{x} g_m(x')} \geq \left(\frac{(\tilde{\varepsilon})^{\underline{x}_m-\tilde{x}}-1}{1-\tilde{\varepsilon}}\right)\left[\sum_{x'=\tilde{x}}^{x}\left\{\prod_{x''=\tilde{x}+1}^{x'}\lambda(x'')\right\}\right]^{-1},$$

which approaches $+\infty$ as $m \to +\infty$, because $\lim_{m\to+\infty}(\tilde{\varepsilon})^{\underline{x}_m-\tilde{x}} = +\infty$.

**Q.E.D.**

Based on these lemmata, we can complete the proof of Proposition 5 as follows.

**Proof of Proposition 5:** Suppose $(\frac{p}{1-p})\beta^{1-\gamma} > 1$. Since $\beta^{1-\gamma} \leq 1$, one gets $\frac{p}{1-p} > 1$. Lemmata 7 and 8 say that for every large enough positive integer $x$ and every integer $\tilde{x} < x$,

$$\lim_{m\to+\infty}\frac{\sum_{x'=x+1}^{\bar{x}_m} g_m(x')}{\sum_{x'=\tilde{x}}^{x} g_m(x')} = +\infty \text{ and } \lim_{m\to+\infty}\frac{\sum_{x'=x+1}^{\bar{x}_m} g_m(x')}{\sum_{x'=\tilde{x}}^{x} g_m(x')} = +\infty.$$

Hence, the stationary probability that the state of mind is larger than $x$ in the long run converges to unity as $m$ increases. Since we can choose $x$ as large as possible, we have proved $\Lambda^{\infty} = 1$, i.e., part (iii).

Suppose $(\frac{p}{1-p})\beta^{1-\gamma} < 1$. Since $\beta^{1-\gamma} \geq \beta$, one gets $(\frac{p}{1-p})\beta < 1$. Lemmata 9 and 10 say that for every large enough positive integer $x$ and every integer $\tilde{x} < x$,

$$\lim_{m\to+\infty}\frac{\sum_{x'=x+1}^{\bar{x}_m} g_m(x')}{\sum_{x'=\underline{x}_m}^{\tilde{x}-1} g_m(x')} = 0 \text{ and } \lim_{m\to+\infty}\frac{\sum_{x'=\underline{x}_m}^{\tilde{x}-1} g_m(x')}{\sum_{x'=\tilde{x}}^{x} g_m(x')} = +\infty.$$

Hence, the stationary probability that the state of mind is smaller than $\tilde{x}$ in the long run converges to unity as $m$ increases. Since we can choose $\tilde{x}$ as small as possible, we have proved $\Lambda^{\infty} = 0$, i.e., part (iv).

**Q.E.D.**

# 6. Discussion: Sophisticated Learner

In the previous sections, we have shown that it is difficult for a decision maker to choose the uncertain action in the long run, even though it is actually more profitable than the safe action. Hence, it is natural as the next step to shed light on a wider class of learning rules.

We generalize learning rules by modifying the rule of updating in ways that the state of mind is less likely to decline: Fix $q \in (0,1]$ arbitrarily. When the current state of mind $x(t)$ is smaller than $0$,

$$x(t+1) = x(t)+1 \text{ if } v(t)=1 \text{ and } x(t) < \bar{x},$$

and

$$x(t+1) = \begin{cases} x(t)-1 & \text{with probability } q \\ x(t) & \text{with probability } 1-q \end{cases} \text{ if } v(t)=-1 \text{ and } x(t) > \underline{x}.$$

Even if the decision maker obtains the positive payoff, the state of mind remains unchanged with probability $1-q$. On the other hand, when $x(t)$ is larger than or equal to $0$,

$$x(t+1) = x(t)+1 \text{ if } v(t)=1 \text{ and } x(t) < \bar{x},$$

and

$$x(t+1) = x(t)-1 \text{ if } v(t)=-1 \text{ and } x(t) > \underline{x}.$$

Similarly to Section 5, the proportion of the probability of transiting from $x-1$ into $x$ to the probability of transiting from $x$ into $x-1$ is given by

$$\lambda(x) = \frac{pF(x-1)}{(1-p)F(x)} \text{ for all } x > 0,$$

and

$$\lambda(x) = \frac{pF(x-1)}{(1-p)qF(x)} \text{ for all } x \leq 0.$$

Similarly to Section 5 also, one gets

$$\lim_{x \to -\infty} \lambda(x) = \left(\frac{p}{1-p}\right)\left(\frac{\beta}{q}\right),$$

$$\lim_{x \to +\infty} \lambda(x) = \frac{p}{1-p},$$

and

$$\lim_{m \to +\infty} \left( \prod_{x=\underline{x}_m+1}^{\bar{x}_m} \lambda(x) \right)^{\frac{1}{\bar{x}_m - \underline{x}_m}} = \left(\frac{p}{1-p}\right)\left(\frac{\beta}{q}\right)^{1-\gamma}.$$

**Proposition 11:** *Suppose* $q \geq \beta$. *Then, for every* $p \in (0,1)$,

(v) $$\Lambda^{\infty} = 1 \ \textit{if} \ p > \frac{1}{1+(\frac{\beta}{q})^{1-\gamma}},$$

*and*

(vi) $$\Lambda^{\infty} = 0 \ \textit{if} \ p < \frac{1}{1+(\frac{\beta}{q})^{1-\gamma}}.$$

**Proof:** By replacing $\beta$ with $\frac{\beta}{q}$ and using inequality $\frac{\beta}{q} < 1$, we can prove parts (v) and (vi) in the same way as part (iii) and (iv), respectively.

**Q.E.D.**

Proposition 11 implies that given $\frac{\beta(f)}{q} \leq 1$, a sequence of learning rules $\Gamma^{\infty}$ is efficient with respect to $f$ if and only if

either $q = \beta(f)$ or $\gamma(\Gamma^{\infty}) = 1$.

**Proposition 12:** *Suppose* $\frac{\beta}{q} > 1$. *Then, for every* $p \in (0,1)$,

(vii) $$\Lambda^{\infty} = 1 \ \textit{if} \ p > \frac{1}{2},$$

(viii) $$\Lambda^{\infty} = 0 \ \textit{if} \ p < \frac{1}{1+\frac{\beta}{q}},$$

*and*

(ix) $$0 < \Lambda^{\infty} < 1 \ \textit{if} \ \frac{1}{1+\frac{\beta}{q}} < p < \frac{1}{2}.$$

**Proof:** Suppose $p > \frac{1}{2}$. Since $(\frac{\beta}{q})^{1-\gamma} \geq 1$, one gets $p > \frac{1}{1+(\frac{\beta}{q})^{1-\gamma}}$. Hence, by replacing $\beta$ with $\frac{\beta}{q}$, we can prove part (vii) in the same way as part (iii).

Suppose $p < \frac{1}{1+\frac{\beta}{q}}$. Since $(\frac{\beta}{q})^{1-\gamma} \leq \frac{\beta}{q}$, one gets $p < \frac{1}{1+(\frac{\beta}{q})^{1-\gamma}}$. Hence, by replacing $\beta$ with $\frac{\beta}{q}$, we can prove part (viii) in the same way as part (iv).

Finally, we prove part (ix) as follows. Suppose $\frac{p}{1-p} < 1$. Choose $\varepsilon$ such that

$$\frac{p}{1-p} < \varepsilon < 1.$$

Equality (4) says that for every large enough positive integer $x$,

$$\lambda(x') \le \varepsilon \text{ for all } x' \ge x.$$

In the same way as Lemma 9, one gets

$$\sum_{x'=x+1}^{\bar{x}_m} g_m(x') \le \varepsilon\left(\frac{\varepsilon^{\bar{x}_m - x} - 1}{\varepsilon - 1}\right) g_m(x).$$

Fix an integer $\tilde{x} < x$ arbitrarily. Equality (6) says

$$\sum_{x'=\tilde{x}}^{x} g_m(x') = \sum_{x'=\tilde{x}}^{x} \{ \prod_{x''=x'+1}^{x} \frac{1}{\lambda(x'')} \} g_m(x),$$

and therefore,

$$\frac{\sum_{x'=x+1}^{\bar{x}_m} g_m(x')}{\sum_{x'=\tilde{x}}^{x} g_m(x')} \le \varepsilon\left(\frac{\varepsilon^{\bar{x}_m - x} - 1}{\varepsilon - 1}\right) [\sum_{x'=\tilde{x}}^{x} \{ \prod_{x''=x'+1}^{x} \frac{1}{\lambda(x'')} \}]^{-1},$$

which approaches

$$\varepsilon\left(\frac{1}{1-\varepsilon}\right) [\sum_{x'=\tilde{x}}^{x} \{ \prod_{x''=x'+1}^{x} \frac{1}{\lambda(x'')} \}]^{-1} < +\infty$$

as $m \to +\infty$, because $\lim_{m \to +\infty} \varepsilon^{\bar{x}_m - x} = 0$. Hence, one gets $\Lambda^{\infty} < 1$.

Next, suppose $p < \frac{1}{1+\beta}$. Choose $\varepsilon$ such that

$$\left(\frac{p}{1-p}\right)\beta > \varepsilon > 1.$$

Equality (3) says that for every small enough negative integer $\tilde{x}$,

$$\lambda(x') \ge \varepsilon \text{ for all } x' \le \tilde{x}.$$

In the same way as Lemma 6, one gets

$$\sum_{x'=\underline{x}_m}^{\tilde{x}-1} g_m(x') \le \left(\frac{(\tilde{\varepsilon})^{\underline{x}_m - \tilde{x}} - 1}{1 - \tilde{\varepsilon}}\right) g_m(\tilde{x}).$$

Fix an integer $x < \tilde{x}$ arbitrarily. Equality (6) says

$$\sum_{x'=\tilde{x}}^{x} g_m(x') = \sum_{x'=\tilde{x}}^{x} \{ \prod_{x''=\tilde{x}+1}^{x'} \lambda(x'') \} g_m(\tilde{x}),$$

and therefore,

$$\frac{\sum_{x'=\underline{x}_m}^{\widetilde{x}-1} g_m(x')}{\sum_{x'=\widetilde{x}}^{x} g_m(x')} \leq \left(\frac{(\widetilde{\varepsilon})^{\underline{x}_m-\widetilde{x}}-1}{1-\widetilde{\varepsilon}}\right)\left[\sum_{x'=\widetilde{x}}^{x}\left\{\prod_{x''=\widetilde{x}+1}^{x'}\lambda(x'')\right\}\right]^{-1},$$

which approaches

$$\left(\frac{1}{\widetilde{\varepsilon}-1}\right)\left[\sum_{x'=\widetilde{x}}^{x}\left\{\prod_{x''=\widetilde{x}+1}^{x'}\lambda(x'')\right\}\right]^{-1} < +\infty$$

as $m \to +\infty$, because $\lim_{m\to+\infty}(\widetilde{\varepsilon})^{\underline{x}_m-\widetilde{x}} = +\infty$. Hence, one gets $\Lambda^{\infty} > 0$.

From these arguments, we have completed the proof of part (ix), and therefore, the proof of Proposition 12.

**Q.E.D.**

Proposition 12 implies that given $\frac{\beta(f)}{q} > 1$, there exists no sequence of learning rules which is efficient with respect to $f$.

From the above arguments, one gets that a sequence of learning rules $\Gamma^{\infty}$ is efficient with respect to every $f$ such that $\beta(f) \in (0,1]$, if and only if

$$q = 1 \text{ and } \gamma(\Gamma^{\infty}) = 1.$$

Hence, even if the class of possible learning rules extends wider, we can not find another sequence of learning rules which is efficient with respect to every $f$ such that $\beta(f) \in (0,1]$.

We will devote the rest of this section to seeking furthermore for another possibility of efficient learning, by investigating a decision maker who is so sophisticated as to control the probability $q$ deliberately. Assume that at the very beginning of the Bandit problem, a decision maker is informed of $f$, and then equalizes $q$ to $\beta(f)$. Straightforwardly from the above arguments, such a sophisticated learner succeeds to achieve long-run efficiency, irrespective of how $f$ is given.

I think, however, it is quite implausible in most real environments that a decision maker is sophisticated in this way. Learning rules are for the most parts regarded as being exogenous like preferences, technological conditions, initial endowments, and so on. Many substantial elements of a learning rule such as the probability $q$ belong to the realm of the unconsciousness, and therefore, should not be treated indiscriminately as the decision maker's control variables.

## References


Keynes, J. M. (1936): *The General Theory of Employment, Interest, and Money*, London: Macmillan.

Knight, F. (1921): *Risk, Uncertainty and Profit*, Houghton Mifflin Co.

Marimon, R. (1997): "Learning from Learning in Economics," in Kreps, D. and K. Wallis (eds.), *Advances in Economics and Econometrics: Theory and Applications Vol. 1*, Cambridge University Press.

Matsushima, H. (1997): "Procedural Rationality and Inductive Learning I: Towards a Theory of Subjective Games," Discussion Paper 97-F-21, Faculty of Economics, University of Tokyo.

Matsushima, H. (1998a): "Learning about Stochastic Payoff Structure," Discussion Paper CIRJE-F-7, Faculty of Economics, University of Tokyo.

Matsushima, H. (1998b): "Efficient Entrepreneurship," Discussion Paper CIRJE-F-13, Faculty of Economics, University of Tokyo.

Rothschild, M. (1974): "A Two-Armed Bandit Theory of Market Pricing," *Journal of Economic Theory* 9, 185-202.

Sarin, R. and F. Vahid (1997): "Payoff Assessments without Probabilities: A Simple Dynamic Model of Choice," mimeo.