Research Article
Rational Probabilistic Deciders—Part I: Individual Behavior

P. T. Kabamba, W.-C. Lin, and S. M. Meerkov

This paper is intended to model a decision maker as a rational probabilistic decider (RPD) and to investigate its behavior in stationary and symmetric Markov switch environments. RPDs take their decisions based on penalty functions defined by the environment. The quality of decision making depends on a parameter referred to as level of rationality. The dynamic behavior of RPDs is described by an ergodic Markov chain. Two classes of RPDs are considered—local and global. The former take their decisions based on the penalty in the current state while the latter consider all states. It is shown that asymptotically (in time and in the level of rationality) both classes behave quite similarly. However, the second largest eigenvalue of Markov transition matrices for global RPDs is smaller than that for local ones, indicating faster convergence to the optimal state. As an illustration, the behavior of a chief executive officer, modeled as a global RPD, is considered, and it is shown that the company performance may or may not be optimized—depending on the pay structure employed. While the current paper investigates individual RPDs, a companion paper will address collective behavior.

1. Introduction

1.1. Motivation. The theory of rational behavior (TRB) is a set of models intended to capture one of the main features of living organisms' behavior—the possibility of selecting the most favorable decisions among all possible options in a decision space. TRB involves two major components: dynamical systems, which model rational behavior, and rules of interaction among them and with the environment. Analysis of the resulting complex dynamics reveals fundamental properties of rational behavior.

TRB emerged in the early '60s in the work of mathematicians and physicists, primarily in Russia. A summary of their results can be found in [1, 2] while the earliest publications are [3, 4]. In [1–4], the decision makers were modeled as automata, the states of which represent various decisions. The transitions among the states were driven by inputs (penalties or rewards) generated either by the environment or by the decisions of other automata. The transition diagrams of the automata were designed so that the steady state probabilities corresponding to the decisions with the largest rewards were maximized. This was interpreted as rational behavior of an individual decision maker or a collective of decision makers.

In [5], this approach was generalized by assuming that the decision makers were not necessarily automata but general dynamical systems in certain decision spaces. The trajectories of these systems were assumed to satisfy two axioms: ergodicity and retardation. Roughly speaking, the ergodicity axiom implied that all possible decisions were considered within the process of decision making, while the retardation axiom required that the trajectories slow down (i.e., retard) in the vicinity of the most advantageous states. Along with enlarging the set of possible decision makers, this framework exhibited additional properties of rational behavior, such as the possibility of rapid convergence to the optimal state, which was impossible in the framework of automata.

These two modeling approaches involved not only issues of rationality but also complex dynamic behavior, which, on one hand, made the analysis of the resulting systems difficult and, on the other hand, obscured the issue of rational behavior. As a result, the steady state probabilities of various decisions as functions of the parameters of the decision makers and rules of interactions were all but impossible to analyze, especially when the environment was time-varying and/or more than one decision makers were involved.

The main purpose of the present work is to develop a purely probabilistic modeling approach to rational behavior, which does not lead to complicated dynamics and which provides a more complete and transparent analysis of main features of rational behavior. To accomplish this, we introduce the notion of *rational probabilistic deciders* (RPDs), which select various states in their decision spaces with certain probabilities. How this selection takes place is omitted; it could be accomplished by either automata, or the dynamical systems of [5], or any other, perhaps unknown, mechanism. But, as long as this selection takes place, the approach of this work, being divorced from the issues of dynamics of the decision making, leads to a simple Markov chain analysis of fundamental properties of rational behavior. This paper addresses the issues of individual behavior, while in a forthcoming paper collective behavior is analyzed.

1.2. Brief review of existing results. The work on using automata to model the simplest form of rational behavior first appeared in [3], where the so-called finite automata with linear tactics were constructed and their individual behavior in stationary and nonstationary media was investigated. Based on this work, [6–10] developed other types of rational automata, and investigated their individual behavior. All these automata were shown to behave optimally with arbitrarily large memory in stationary media. In nonstationary media, it was shown that there exists a finite memory size for the automata to behave optimally.

The collective behavior of the automata mentioned above was investigated [4, 11–21]. In particular, [14–16] analyzed the collective behavior of asymptotically optimal automata as players in zero-sum matrix games. The results showed that automata with arbitrarily large memory converge to the saddle point in matrix games with pure optimal strategy and do not converge to the saddle point in mixed strategies. References [17–19] investigated the collective behavior of automata in the so-called Gur game. Conditions for the group to behave optimally were derived. Specifically, it was shown that as the number of automata, M, and their memory capacity, N, become arbitrarily large, the ratio N/M must exceed some constant for the collective to behave optimally.

Similar results, in the framework of general dynamical systems, were obtained in [5]. In addition, [5] provided many examples of dynamical systems that exhibit rational behavior. The main among them was the so-called ring-element, which could be viewed as the simplest search algorithm for global extrema. Unlike the automata, where the convergence to the optimal decision is exponentially slow (with the exponent being the memory capacity), ring-elements could converge arbitrarily fast. Also, [5] provided a detailed study of collective behavior under homogeneous and nonhomogeneous fractional interactions and showed that even if N/M tends to 0 (as $N, M \to \infty$), the convergence to the optimal state still may take place if the interaction was nonhomogeneous.

The earliest western work on rational behavior appeared in [22–24]. Specifically, [22] introduced a new type of rational automata and [23, 24] initiated the application to learning systems. The learning system approach, along with applications to telephone routing problems, has continued in [25–28]. The results of this research were summarized in [29].

A number of applications of rational behavior to various practical problems have also been reported. Namely, [30, 31] applied rational automata to distributed systems, and [32–34] discussed the applications to cooperative mobile robots, to quality of service for sensor networks, and to control of flapping wings in micro air vehicles, respectively. Recently, TRB was applied to power-efficient operation of wireless personal area networks [35].

1.3. Goals of the study.   The main goals of this work are as follows.

(α) Contributing to TRB by introducing a purely probabilistic modeling framework for rational behavior.

(β) Using this framework, it develops methods for *analysis* of rational behavior, that is, methods, which allow to investigate the steady state probabilities of various decisions as functions of system parameters.

(γ) Addressing the issues of *synthesis* of desirable rational behavior, that is, methods for selecting parameters of RPDs and rules of interactions, which lead to desired individual and collective behavior.

This paper pursues these goals for the case of individual behavior; a forthcoming paper will address the collective one.

The outline of this paper is as follows. In Section 2, we define and characterize RPDs, and state analysis and synthesis problems. The individual behavior of RPDs in stationary and symmetric Markov switch environments is investigated in Sections 3 and 4, respectively. An application of RPDs to a pay and incentive system is described in Section 5. Finally, in Section 6, we state the conclusions. All proofs are given in the appendices.

2. Probabilistic modeling framework

2.1. Probabilistic deciders.

A *probabilistic decider* (PD) is defined as a quadruple,

$$(\mathcal{S}, \Phi, N, \mathcal{P}), \tag{2.1}$$

where

(a) $\mathcal{S} = \{1, 2, \ldots, s\}$ is the *decision space* of the PD. At each time moment, $n = 0, 1, 2, \ldots$, the PD is in one of the states of $\mathcal{S}$;

(b) $\Phi = [\varphi_1\ \varphi_2\ \cdots\ \varphi_s]$ is a vector, which characterizes the environment surrounding the PD. The entry $0 < \varphi_i < \infty$, $i = 1, 2, \ldots, s$, denotes the *penalty* associated with state i;

(c) $N \in (0, \infty)$ is a positive real number, which will be used below to characterize the *level of rationality* of the PD;

(d) $\mathcal{P} = \{P_1(\Phi; N), P_2(\Phi; N), \ldots, P_s(\Phi; N)\}$ is a set of *transition probabilities* depending on Φ and N. If at time n the PD is in state i, the probability that it makes a state transition at time $n + 1$ is

$$P[x(n+1) \neq i \mid x(n) = i] = P_i(\varphi_1, \varphi_2, \ldots, \varphi_s; N) = P_i(\Phi; N), \tag{2.2}$$

$$0 < P_i(\Phi; N) < 1; \tag{2.3}$$

(e) when a state transition occurs, the PD selects any other state with equal probability

$$P[x(n+1) = j \mid x(n) = i] = \frac{P_i(\Phi; N)}{s - 1} \quad \text{for } j \neq i. \tag{2.4}$$

2.2. Steady state probabilities.

Due to (2.2)–(2.4), the dynamics of PD in $\mathcal{S}$ are described by an ergodic Markov chain with transition matrix

$$A = \begin{bmatrix} 1 - (P_1(\Phi; N)) & \dfrac{P_1(\Phi; N)}{s - 1} & \cdots & \dfrac{P_1(\Phi; N)}{s - 1} \\ \dfrac{P_2(\Phi; N)}{s - 1} & 1 - (P_2(\Phi; N)) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{P_s(\Phi; N)}{s - 1} & \cdots & \cdots & 1 - P_s(\Phi; N) \end{bmatrix}. \tag{2.5}$$

Therefore, the vector of steady state probabilities,

$$\kappa = \begin{bmatrix} \kappa_1 & \kappa_2 & \cdots & \kappa_s \end{bmatrix}, \quad \kappa_i = \kappa_i(\Phi; N), \tag{2.6}$$

where κ_i is the steady state probability of the PD choosing state i, can be calculated from

$$\kappa = \kappa A, \quad \sum_{i=1}^{s} \kappa_i = 1, \tag{2.7}$$

which has the solution

$$\kappa_i(\Phi;N) = \frac{1/P_i(\Phi;N)}{\sum_{j=1}^{s}\left(1/P_j(\Phi;N)\right)} \quad \forall i \in \{1,2,\dots,s\}. \tag{2.8}$$

Due to (2.3),

$$0 < \kappa_i(\Phi;N) < 1. \tag{2.9}$$

2.3. Rational probabilistic deciders. A probabilistic decider $(\mathcal{S},\Phi,N,\mathcal{P})$ is referred to as a *rational probabilistic decider* (RPD) if

$$\varphi_i < \varphi_j, \quad i,j \in \mathcal{S}, \ i \neq j, \tag{2.10}$$

implies

$$\frac{\kappa_i(\Phi;N)}{\kappa_j(\Phi;N)} > 1 \quad \forall N \in (0,\infty), \tag{2.11}$$

and, moreover,

$$\lim_{N\to\infty} \frac{\kappa_i(\Phi;N)}{\kappa_j(\Phi;N)} = \infty. \tag{2.12}$$

Equation (2.11) means that, at steady state, the probability of an RPD being in a less penalized state is higher than the probability of it being in a more penalized state. Equation (2.12) means that as N increases, the RPD becomes more selective, preferring states that are least penalized. As before, the parameter, N, is referred to as the *level of rationality* (LR) of the RPD.

2.4. Analytical characterization of RPDs. It turns out that there are several types of PDs satisfying (2.12) and (2.11). Two of them—local and global RPDs—are introduced below.

2.4.1. Local RPDs. Consider PDs such that

$$P_i(\Phi;N) = P(\varphi_i,N), \quad i \in \mathcal{S},$$
$$P(\varphi,N) : (0,\infty) \times (0,\infty) \longrightarrow (0,1), \tag{2.13}$$

that is, the probability of leaving a state depends only on the penalty associated with this state. For this reason, these PDs are referred to as *local probabilistic deciders* (L-PDs).

The properties of function P, which guarantee that an L-PD is, in fact, an RPD, that is, (2.11) and (2.12) are satisfied, are analyzed next.

Assume the following.

(P.1) The relation, $\varphi_1 < \varphi_2$, implies

$$P(\varphi_1,N) < P(\varphi_2,N) \quad \forall N \in (0,\infty), \tag{2.14}$$

and, moreover,

$$\lim_{N \to \infty} \frac{P(\varphi_2, N)}{P(\varphi_1, N)} = \infty. \tag{2.15}$$

The class of functions satisfying (P.1) is denoted as Π.

THEOREM 2.1. *An L-PD is, in fact, an RPD if and only if $P \in \Pi$.*

We refer to L-PDs that are RPDs as *local RPDs* (L-RPDs). Clearly, an L-RPD is characterized by the function P, $P \in \Pi$. An example of a function P satisfying (P.1) is

$$P(\varphi, N) = p(\varphi)^N, \tag{2.16}$$

where

$$p : (0, \infty) \longrightarrow (0, 1) \tag{2.17}$$

is a strictly increasing function, for example,

$$p(\varphi) = \frac{\varphi}{1 + \varphi}, \tag{2.18}$$

or

$$p(\varphi) = \frac{2}{1 + e^{-\varphi}} - 1, \tag{2.19}$$

or

$$p(\varphi) = \frac{2}{\pi} \tan^{-1}(\varphi). \tag{2.20}$$

Finally, we state an analytic property of P, which will be useful later.

LEMMA 2.2. *If $P \in \Pi$, then $\lim_{N \to \infty} P(\varphi, N) = 0$, for all $\varphi \in (0, \infty)$.*

2.4.2. Global RPDs. Consider PDs, which satisfy the following.
   (a) The decision space contains two states

$$\mathcal{S} = \{1, 2\}. \tag{2.21}$$

   (b) The steady state probabilities, $\kappa_i(\varphi_1, \varphi_2; N)$, $i = 1, 2$, can be written in the form of a composite function,

$$\kappa_i(\varphi_1, \varphi_2; N) = \begin{cases} F\left(G\left(N, \dfrac{\varphi_2}{\varphi_1}\right)\right) & \text{if } i = 1, \\ F\left(G\left(N, \dfrac{\varphi_1}{\varphi_2}\right)\right) & \text{if } i = 2, \end{cases} \tag{2.22}$$

$$F : (0, \infty) \longrightarrow (0, 1),$$
$$G : (0, \infty) \times (0, \infty) \longrightarrow (0, \infty). \tag{2.23}$$

The purpose of (2.21) is to simplify the characterization. The function G in (2.22) is intended to model how the PD perceives the ratio of the penalties associated with the states, while the function F in (2.22) models how the PD makes decisions upon this perception. In addition, (2.22) implies

$$\kappa_1(\varphi_1, \varphi_2; N) = \kappa_2(\varphi_2, \varphi_1; N), \tag{2.24}$$

which indicates that the decisions of the PD are not prejudiced. We specify what functions in $\mathscr{P}$ of the PD give rise to (2.22).

THEOREM 2.3. *A necessary and sufficient condition for $\mathscr{P}$ to lead to (2.22) is*

$$\frac{P_1(\Phi; N)}{P_2(\Phi; N)} = \frac{F(G(N, \varphi_1/\varphi_2))}{F(G(N, \varphi_2/\varphi_1))}, \tag{2.25}$$

where $F(G(N, \varphi_1/\varphi_2)) + F(G(N, \varphi_2/\varphi_1)) = 1$.

Note that, according to (2.25), the probability of leaving a state depends not only on the penalty of this particular state, but also on the penalty of the other state. For this reason, PDs, satisfying (a) and (b), are called *global probabilistic deciders* (G-PDs).

We investigate the properties of F and G, which guarantee that a G-PD is an RPD, that is, (2.11) and (2.12) are satisfied. Assume the following.

(F.1) F is a continuous, strictly increasing function, and $\lim_{x \to \infty} F(x) = 1$, $\lim_{x \to 0} F(x) = 0$.

(F.2) F satisfies the functional equation, $F(x) + F(x^{-1}) = 1$ for $x \in (0, \infty)$.

(G.1) For any $y \in (0, 1)$, $G(N, y)$ is a continuous, strictly decreasing function of $N \in (0, \infty)$, and $\lim_{N \to \infty} G(N, y) = 0$.

(G.2) For any $y \in (1, \infty)$, $G(N, y)$ is a continuous, strictly increasing function of $N \in (0, \infty)$, and $\lim_{N \to \infty} G(N, y) = \infty$.

(G.3) For any $y \in (0, \infty)$, $\lim_{N \to 0} G(N, y) = 1$.

(G.4) For any $N \in (0, \infty)$, $G(N, y)$ is a continuous, strictly increasing function of $y \in (0, \infty)$, and $\lim_{y \to \infty} G(N, y) = \infty$, $\lim_{y \to 0} G(N, y) = 0$.

(G.5) For any $N \in (0, \infty)$, $G(N, y)$ satisfies the functional equation, $G(N, y^{-1}) = G(N, y)^{-1}$ for $y \in (0, \infty)$.

Assumptions (G.1) and (G.2) imply that when $N \to \infty$, the PD becomes arbitrarily perceptive to the ratio of the penalties, while (G.3) means that when $N \to 0$, this perception diminishes. Moreover, (F.1) means that, as this perception becomes arbitrarily good, the probability of the PD making the right decision becomes arbitrarily close to one. Assumptions (F.2), (G.4), and (G.5) are introduced for technical reasons.

Denote the class of functions F satisfying (F.1) and (F.2) as $\mathscr{F}$, and the class of functions G satisfying (G.1)–(G.5) as $\mathscr{G}$.

THEOREM 2.4. *A G-PD is, in fact, an RPD if $F \in \mathscr{F}$ and $G \in \mathscr{G}$.*

We refer to G-PDs that are RPDs as *global RPDs* (G-RPDs). A G-RPD is characterized by the pair (F, G), $F \in \mathscr{F}$, $G \in \mathscr{G}$. The functions in $\mathscr{P}$ of the G-RPD can be reconstructed by Theorem 2.3. For example, consider the G-RPD characterized by the pair (F_1, G_1),

where

$$F_1(x) = \frac{x}{1+x}, \qquad G_1(N, y) = y^N. \qquad (2.26)$$

The functions in $\mathscr{P}$ can be either

$$P_1(\Phi, N) = \begin{cases} \dfrac{1}{2}\left(\dfrac{\varphi_1}{\varphi_2}\right)^N & \text{if } \varphi_1 < \varphi_2, \\ \dfrac{1}{2} & \text{if } \varphi_1 > \varphi_2, \end{cases} \qquad P_2(\Phi, N) = \begin{cases} \dfrac{1}{2} & \text{if } \varphi_1 < \varphi_2, \\ \dfrac{1}{2}\left(\dfrac{\varphi_2}{\varphi_1}\right)^N & \text{if } \varphi_1 > \varphi_2, \end{cases} \qquad (2.27)$$

or

$$P_1(\Phi, N) = \frac{\varphi_1^N}{\varphi_1^N + \varphi_2^N}, \qquad P_2(\Phi, N) = \frac{\varphi_2^N}{\varphi_1^N + \varphi_2^N}. \qquad (2.28)$$

We note that Krinskiy's automata [8] with two actions can be characterized by the pair (F_1, G_1), defined in (2.26). Furthermore, new RPDs can be found by using other functions in $\mathscr{F}$ and $\mathscr{G}$. For example, a G-RPD characterized by (F_2, G_2), where

$$F_2(x) = \frac{x-1}{2(x+1)} + \frac{1}{2}, \qquad G_2(N, y) = \begin{cases} Ny - N + 1 & \text{if } y \geq 1, \\ \dfrac{1}{N/y - N + 1} & \text{if } y \leq 1, \end{cases} \qquad (2.29)$$

has not been found before.

Remark 2.5. Two methods can be used to extend the characterization of G-RPDs discussed above to a G-RPD with more than two states in the decision space. The first method is to characterize the G-RPD iteratively: in the first step of the iteration, the decision space is partitioned into two subspaces. The G-RPD then selects one of the two subspaces in a probabilistic way. The probability of this selection is modeled in a form similar to (2.22). In the next iteration, the selected subspace in the previous step is partitioned into two subspaces and the G-RPD proceeds as before. The second method is to characterize the G-RPD in a pairwise fashion. Specifically, for each pair of the states, we characterize the probabilities of the decisions of the G-RPD in the form of (2.22). This method is used in Section 5 below.

In conclusion of this subsection, we formulate a lemma, which will be useful in later sections.

LEMMA 2.6. *Suppose $F(x) \in \mathscr{F}$ and $G(N, y) \in \mathscr{G}$. Then,*
   (i) *$F(x)$ has a continuous, strictly increasing inverse, $F^{-1} : (0, 1) \to (0, \infty)$;*
   (ii) *given $y > 1$, $G(N, y)$ has a continuous, strictly increasing inverse with respect to the variable N, $G_y^{-1} : (1, \infty) \to (0, \infty)$;*
   (iii) *given $N \in (0, \infty)$, $G(N, y)$ has a continuous, strictly increasing inverse with respect to the variable y, $G_N^{-1} : (0, \infty) \to (0, \infty)$.*

2.5. RPDs' analysis and synthesis problems. The environment, modeled in Section 2.1 by a constant vector Φ, is referred to as *stationary*. RPDs may operate also in nonstationary environments, modeled by time varying Φ's. In Section 4, we introduce a specific case of such situation referred to as the *symmetric Markov switch environment*.

This paper investigates the individual behavior of RPDs in both stationary and symmetric Markov switch environments. In each of these environments, the following analysis and synthesis problems are addressed.

Analysis. Given an RPD and an environment, analyze the probability of various decisions as a function of the level of rationality and the parameters of the environment.

Synthesis. Given an RPD and an environment, calculate the level of rationality and/or parameters of the environment, which lead to various types of RPD's behavior.

Exact formulations of these problems along with appropriate answers are given in Sections 3 and 4.

3. RPDs in stationary environment

3.1. Environment. In this section, we consider the decision space $\mathcal{S} = \{1, 2\}$ and the penalty vector $\Phi = [\varphi_1 \ \varphi_2]$. Without loss of generality, assume that $\varphi_1 < \varphi_2$, that is, state 1 is associated with less penalty than state 2.

3.2. Analysis. In order to characterize the behavior of RPDs qualitatively, introduce the following definition.

Definition 3.1 (Asymptotically optimal behavior). The behavior of a PD is *asymptotically optimal* if for all $0 < \epsilon < 1/2$ and for all Φ with $\varphi_1 < \varphi_2$, there exists N^* such that $N \geq N^* \Rightarrow 1 \geq \kappa_1(\Phi, N) \geq 1 - \epsilon$.

Clearly, asymptotically optimal behavior means that no matter how close the penalties associated with the two states are, there is an LR large enough so that the RPD selects the state with less penalty reliably.

We have the following qualitative results.

THEOREM 3.2. *Both L-RPDs and G-RPDs exhibit asymptotically optimal behavior.*

Although, as it is stated in Theorem 3.2, L-RPDs and G-RPDs asymptotically behave qualitatively similar, their behavior for fixed N and $\varphi_2 - \varphi_1$ might be different. To illustrate this, consider L-RPD defined by (2.16), (2.18) and G-RPD defined by (2.26), respectively. Figures 3.1(a) and 3.1(b) show the probability of selecting state 1 as a function of N and φ_1, respectively. Clearly, G-RPD, having more information, outperforms L-RPD.

In addition, G-RPDs have a faster rate of convergence to the steady state probabilities than L-RPDs. Indeed, when $\mathcal{S} = \{1, 2\}$, the second largest eigenvalue, λ_2, of the transition matrix (2.5) for the case of L-RPDs is

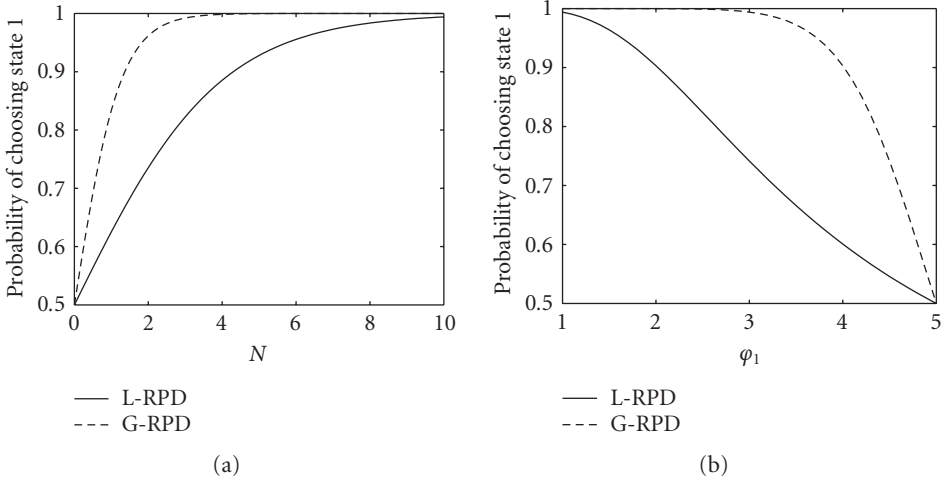$$\lambda_2 = 1 - P(\varphi_1, N) - P(\varphi_2, N). \tag{3.1}$$

Figure 3.1. Probability of selecting state 1. (a) $\varphi_1 = 1$, $\varphi_2 = 5$. (b)$N = 10$, $\varphi_2 = 5$.

Due to Lemma 2.2, this implies that $\lambda_2 \to 1$ as $N \to \infty$, that is, the transient period tends to infinity when LR becomes large.

This phenomenon does not take place for G-RPDs. Indeed, for the case of G-RPD defined by (2.26) and (2.27),

$$\lambda_2 = 1 - \frac{1}{2} - \frac{1}{2}\left(\frac{\varphi_1}{\varphi_2}\right)^N, \tag{3.2}$$

which tends to 1/2 as $N \to \infty$. This qualitatively different behavior is illustrated in Figure 3.2 for L-RPD defined by (2.16), (2.18) and G-RPD defined by (2.26), (2.27).

3.3. Synthesis. We address the following synthesis question.

   S: given $\varphi_1 < \varphi_2$ and $0 < \epsilon < 1/2$, how large should N be so that

$$1 \geq \kappa_1(\varphi_1, \varphi_2; N) \geq 1 - \epsilon \tag{3.3}$$

   holds?

The following theorem give answers to these synthesis questions.

THEOREM 3.3. *Given (F, G), $F \in \mathscr{F}$, $G \in \mathscr{G}$, $\Phi = [\varphi_1 \ \varphi_2]$, $\varphi_1 < \varphi_2$, and $0 < \epsilon < 1/2$, the value of N^* introduced in Definition 3.1 is given by*

$$N^*(\epsilon, y) = G_y^{-1}(F^{-1}(1 - \epsilon)), \tag{3.4}$$

where $y = \varphi_2/\varphi_1$, and F^{-1} and G_y^{-1} are defined in Lemma 2.6.

To illustrate the behavior of N^*, consider two G-RPDs, characterized by (F_1, G_1) and (F_2, G_2) given in (2.26) and (2.29), respectively. Let $N_1^*(\epsilon, y) = G_{1,y}^{-1}(F_1^{-1}(1 - \epsilon))$ and
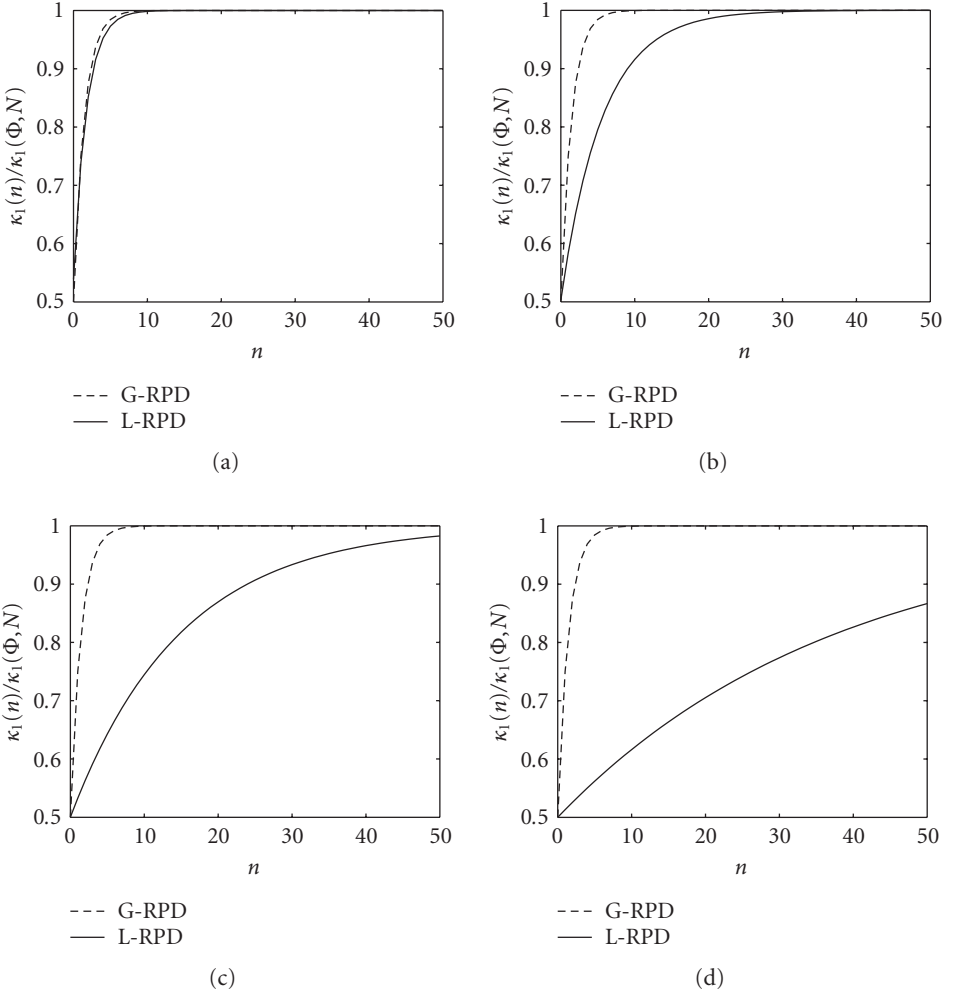
Figure 3.2. Behavior of $\kappa_1(n)/\kappa(\Phi;N)$ for various N ($\varphi_1 = 1$, $\varphi_2 = 5$). (a)$N = 5$. (b) $N = 10$. (c) $N = 15$. (d) $N = 20$.

$N_2^*(\epsilon, y) = G_{2,y}^{-1}(F_2^{-1}(1 - \epsilon))$. Figure 3.3 shows $N_1^*(\epsilon, y)$ and $N_2^*(\epsilon, y)$ as functions of ϵ when $y = 2$ and $y = 5$. As expected, N_i^*, $i = 1,2$, is monotonically decreasing in ϵ and y. In addition, we observe that the G-RPD defined by (F_1, G_1) is more efficient than that defined by (F_2, G_2) since it requires a smaller LR for the same probability of making the right decision.

4. RPDs in symmetric Markov switch environment

4.1. Environment. In this section, we consider RPDs in an environment defined by
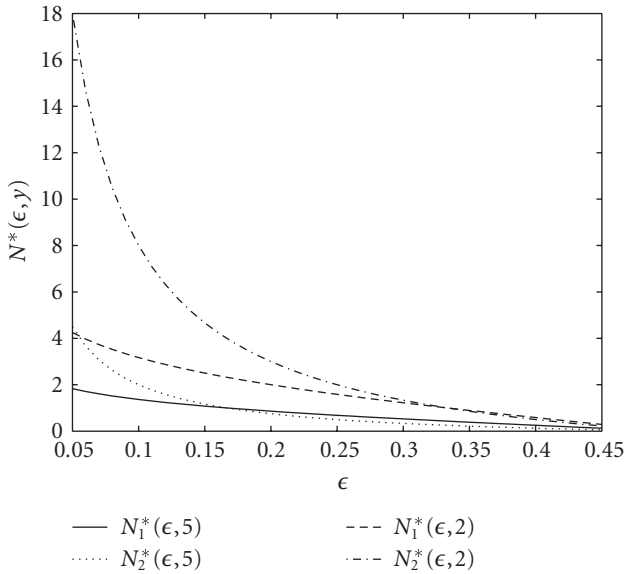
(a) $\mathcal{S} = \{1, 2\}$;

Figure 3.3. N_1^* and N_2^* versus ϵ when $y = 2$ and $y = 5$.

(b) $\Phi(n) = [\varphi_1(n)\ \varphi_2(n)]$, where $\varphi_1(n)$ and $\varphi_2(n)$ are the penalties associated with states 1 and 2, respectively, at time n;

(c) the dynamics of $\varphi_i(n)$, $i = 1, 2$, are defined by a symmetric Markov chain with two states, E_1 and E_2, and the state transition matrix given by

$$\begin{bmatrix} 1 - \epsilon & \epsilon \\ \epsilon & 1 - \epsilon \end{bmatrix}, \quad 0 < \epsilon < 1. \tag{4.1}$$

At time n, the penalties, $\varphi_i(n)$, $i = 1, 2$, are

$$\begin{aligned} \varphi_1(n) = \phi_1, \quad \varphi_2(n) = \phi_2 \quad &\text{if the environment is in state } E_1, \\ \varphi_1(n) = \phi_2, \quad \varphi_2(n) = \phi_1 \quad &\text{if the environment is in state } E_2, \end{aligned} \tag{4.2}$$

where $0 < \phi_i < \infty$, $i = 1, 2$, and without loss of generality, $\phi_1 < \phi_2$.

The environment defined by (a)–(c) is referred to as symmetric Markov switch environment. In this section, we consider the behavior of L-RPDs and G-RPDs.

4.2. Analysis

4.2.1. L-RPDs. The dynamics of the system consisting of an L-RPD characterized by the function P in (2.16) and the symmetric Markov switch environment (a)–(c) are described

by a four-state ergodic Markov chain with the state transition Matrix

$$
A = \begin{bmatrix}
(1-a^N)(1-\epsilon) & a^N(1-\epsilon) & (1-a^N)\epsilon & a^N\epsilon \\
b^N(1-\epsilon) & (1-b^N)(1-\epsilon) & b^N\epsilon & (1-b^N)\epsilon \\
(1-b^N)\epsilon & b^N\epsilon & (1-b^N)(1-\epsilon) & b^N(1-\epsilon) \\
a^N\epsilon & (1-a^N)\epsilon & a^N(1-\epsilon) & (1-a^N)(1-\epsilon)
\end{bmatrix}, \quad (4.3)
$$

where $a = p(\phi_1)$, $b = p(\phi_2)$, $0 < a < b < 1$, and $0 < \epsilon < 1$. Denote the steady state probability vector of this chain by

$$
\kappa = \begin{bmatrix} \kappa_{11} & \kappa_{12} & \kappa_{21} & \kappa_{22} \end{bmatrix}, \quad (4.4)
$$

where κ_{ij} is the steady state probability of the environment being in state E_i and the L-RPD being in state j. Solving the equations

$$
\kappa = \kappa A, \quad \sum_{i,j}\kappa_{ij} = 1, \quad (4.5)
$$

we obtain

$$
\kappa_{11} = \kappa_{22} = \frac{-\epsilon + 2b^N\epsilon - b^N}{2(2a^N\epsilon - a^N - \epsilon + 2b^N\epsilon - b^N - \epsilon)},
$$
$$
\kappa_{12} = \kappa_{21} = \frac{-\epsilon + 2a^N\epsilon - a^N}{2(2a^N\epsilon - a^N - \epsilon + 2b^N\epsilon - b^N - \epsilon)}. \quad (4.6)
$$

Hence, the average penalty incurred by the L-RPD at steady state is

$$
\Psi_\epsilon(N) = \phi_1\kappa_{11} + \phi_2\kappa_{12} + \phi_2\kappa_{21} + \phi_1\kappa_{22} = \phi_1 + (\phi_2 - \phi_1)f_\epsilon(N), \quad (4.7)
$$

where

$$
f_\epsilon(N) = \frac{a^N(2\epsilon - 1) - \epsilon}{(a^N + b^N)(2\epsilon - 1) - 2\epsilon}. \quad (4.8)
$$

In Section 3, the RPDs are in a stationary environment, and the steady state probability of the RPDs to be in the less penalized state is a measure of performance. However, in a symmetric Markov switch environment, the less penalized state changes from time to time. Hence, the steady state probability of being in a certain state is not a suitable measure of performance. Instead, we use $\Psi_\epsilon(N)$ as a measure of performance. The analysis considered here is

A: given an L-RPD in a symmetric Markov switch environment, how does the value of $\Psi_\epsilon(N)$ behave as a function of N and ϵ?

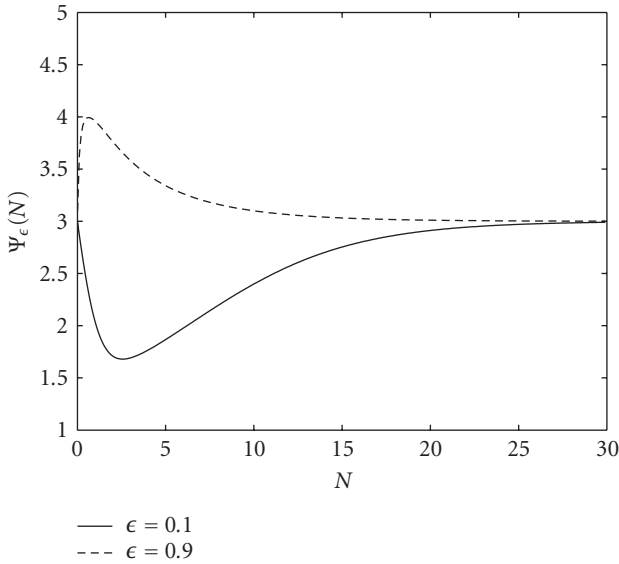A partial answer to this question is provided by the following theorem.

Figure 4.1. $\Psi_\epsilon(N)$ versus N when $\epsilon = 0.1$ and $\epsilon = 0.9$.

THEOREM 4.1. *For $\Psi_\epsilon(N)$ defined in (4.7), the following four properties hold:*

(i)

$$\Psi_\epsilon(N) < \frac{\phi_1 + \phi_2}{2} \quad \text{if } 0 < \epsilon < \frac{1}{2},$$

$$\Psi_\epsilon(N) > \frac{\phi_1 + \phi_2}{2} \quad \text{if } 1 > \epsilon > \frac{1}{2};$$

(4.9)

(ii)

$$\lim_{N \to 0} \Psi_\epsilon(N) = \lim_{N \to \infty} \Psi_\epsilon(N) = \frac{\phi_1 + \phi_2}{2};$$

(4.10)

(iii) *if $0 < \epsilon < 1/2$, then $\Psi_\epsilon(N)$ has a unique minimum on $N \in (0, \infty)$;*
(iv) *if $1/2 < \epsilon < 1$, then $\Psi_\epsilon(N)$ has a unique maximum on $N \in (0, \infty)$.*

Thus, for $1 > \epsilon > 1/2$, $\Psi_\epsilon(N)$ is a function of N with a unique maximum that is larger than $(\phi_1 + \phi_2)/2$. This is clearly not an interesting case. For $0 < \epsilon < 1/2$, $\Psi_\epsilon(N)$ is a function of N with a unique minimum that is smaller than $(\phi_1 + \phi_2)/2$. The behavior of $\Psi_\epsilon(N)$ is illustrated in Figure 4.1 for $\phi_1 = 1$, $\phi_2 = 5$, $a = 0.2$, and $b = 0.8$.

Let N^* be the minimizer of $\Psi_\epsilon(N)$ for $0 < \epsilon < 1/2$. Then, we have the following theorem.

THEOREM 4.2. *Function $\Psi_\epsilon(N)$ with $0 < \epsilon < 1/2$ satisfies the following property:*

$$\epsilon \text{ decreases} \implies N^* \text{ increases and } \Psi_\epsilon(N^*) \text{ decreases.}$$
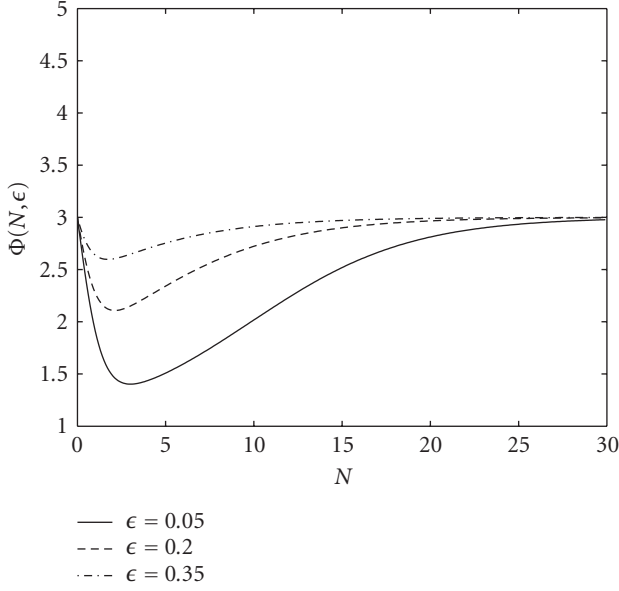
(4.11)

Figure 4.2. $\Psi_\epsilon(N)$ versus N for $\epsilon = 0.05$, 0.2, and 0.35.

Thus, as the environment is switching slower, the L-RPD can use a higher LR, which takes more time for the L-RPD to settle at the least penalized state. As a result, the L-RPD has better ability to discriminate the two states in between switches of the environment. Hence, the average penalty incurred at steady state is less. Figure 4.2 illustrates Theorem 4.2 for $\phi_1 = 1$, $\phi_2 = 5$, $a = 0.2$, $b = 0.8$.

4.2.2. G-RPDs. Consider a G-RPD characterized by (F_1, G_1) defined in (2.26) with $\mathscr{P}$ given by (2.27). The dynamics of a system consisting of this G-RPD and the symmetric Markov switch environment (a)–(c) is described by a four state ergodic Markov chain with transition matrix

A

$$
= \begin{bmatrix} \left(1 - \frac{1}{2}\left(\frac{\phi_1}{\phi_2}\right)^N\right)(1-\epsilon) & \frac{1}{2}\left(\frac{\phi_1}{\phi_2}\right)^N(1-\epsilon) & \left(1 - \frac{1}{2}\left(\frac{\phi_1}{\phi_2}\right)^N\right)\epsilon & \frac{1}{2}\left(\frac{\phi_1}{\phi_2}\right)^N\epsilon \\[2ex] \frac{1}{2}(1-\epsilon) & \frac{1}{2}(1-\epsilon) & \frac{1}{2}\epsilon & \frac{1}{2}\epsilon \\[2ex] \frac{1}{2}\epsilon & \frac{1}{2}\epsilon & \frac{1}{2}(1-\epsilon) & \frac{1}{2}(1-\epsilon) \\[2ex] \frac{1}{2}\left(\frac{\phi_1}{\phi_2}\right)^N\epsilon & \left(1 - \frac{1}{2}\left(\frac{\phi_1}{\phi_2}\right)^N\right)\epsilon & \frac{1}{2}\left(\frac{\phi_1}{\phi_2}\right)^N(1-\epsilon) & \left(1 - \frac{1}{2}\left(\frac{\phi_1}{\phi_2}\right)^N\right)(1-\epsilon) \end{bmatrix}.
$$

$$(4.12)$$

The average penalty incurred at steady state is

$$\Psi'_\epsilon(N) = \phi_1 + (\phi_2 - \phi_1)f'_\epsilon(N), \tag{4.13}$$

where

$$f'_\epsilon(N) = \frac{(\phi_1/\phi_2)^N(\epsilon - 1/2) - \epsilon}{(\phi_1/\phi_2)^N(\epsilon - 1/2) - \epsilon - 1/2}. \tag{4.14}$$

The analysis question considered here is as follows
   A: given an G-RPD in a symmetric Markov switch environment, how does the value
   of $\Psi'_\epsilon(N)$ behave as a function of N and ϵ?
A partial answer to this question is given by the following.

THEOREM 4.3. *The function $\Psi'_\epsilon(N)$ satisfies the following properties:*
   (i)

$$\Psi'_\epsilon(N) < \frac{\phi_1 + \phi_2}{2} \quad if\, 0 < \epsilon < \frac{1}{2},$$

$$\Psi'_\epsilon(N) > \frac{\phi_1 + \phi_2}{2} \quad if\, 1 > \epsilon > \frac{1}{2}; \tag{4.15}$$

   (ii)

$$\lim_{N\to 0}\Psi'_\epsilon(N) = \frac{\phi_1 + \phi_2}{2}, \qquad \lim_{N\to\infty}\Psi'_\epsilon(N) = \phi_1 + (\phi_2 - \phi_1)\frac{\epsilon}{\epsilon + 1/2}; \tag{4.16}$$

   (iii) *if $0 < \epsilon < 1/2$, then $\Psi'_\epsilon(N)$ is a strictly decreasing function of N;*
   (iv) *if $1/2 < \epsilon < 1$, then $\Psi'_\epsilon(N)$ is a strictly increasing function of N.*

Thus, unlike L-RPDs, the larger LR, the less penalty incurred by the G-RPD. When the LR is arbitrarily large, the G-RPD is least penalized. Figure 4.3 illustrates Theorem 4.3 for $\phi_1 = 1$ and $\phi_2 = 5$.

4.3. Synthesis. In this section, we ask the synthesis question.
      S: given an L-RPD in a symmetric Markov switch environment with $0 < \epsilon < 1/2$,
      what is the optimal LR, N^*, that minimizes $\Psi_\epsilon(N)$?
The answer to S can be obtained by solving for N in

$$\frac{d}{dN}\Psi_\epsilon(N) = 0. \tag{4.17}$$

However, this involves solving for the transcendental equation

$$\frac{(\ln a)a^N}{a^N + \epsilon - 2a^N\epsilon} = \frac{(\ln b)b^N}{b^N + \epsilon - 2b^N\epsilon}. \tag{4.18}$$
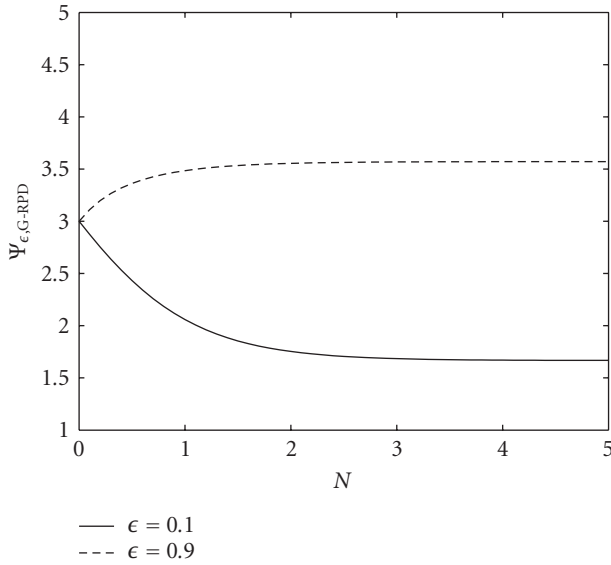
Figure 4.3. $\Psi'_\epsilon(N)$ versus N for $\phi_1 = 1$ and $\phi_2 = 5$.

Let functions $Z_a(N)$ and $Z_b(N)$ be the left-hand side and right-hand side of (4.18), respectively. From plots of functions $Z_a(N)$ and $Z_b(N)$ for various values of a and b, we note that the solution to (4.18) happens near the maximum of

$$\frac{d}{dN} Z_a(N). \tag{4.19}$$

An example is shown in Figure 4.4, where $a = 0.2$, $b = 0.6$, and $\epsilon = 0.1$. An initial approximation of N^*, $\widehat{N^*}$, can then be obtained by solving

$$\frac{d^2}{dN^2} Z_a(N) = 0, \tag{4.20}$$

which gives

$$\widehat{N^*} = \frac{\ln\left(\epsilon/(1 - 2\epsilon)\right)}{\ln a}. \tag{4.21}$$

Note that $\widehat{N^*}$ is a very rough approximation of N^* since this approximation only depends on a, but not on b. An example is shown in Figure 4.5 where the percentage errors

$$\Delta_N = \frac{\widehat{N^*} - N^*}{N^*} \times 100\%,$$

$$\Delta_\Psi = \frac{\Psi_\epsilon\left(\widehat{N^*}\right) - \Psi_\epsilon\left(N^*\right)}{\Psi_\epsilon\left(N^*\right)} \times 100\%, \tag{4.22}$$
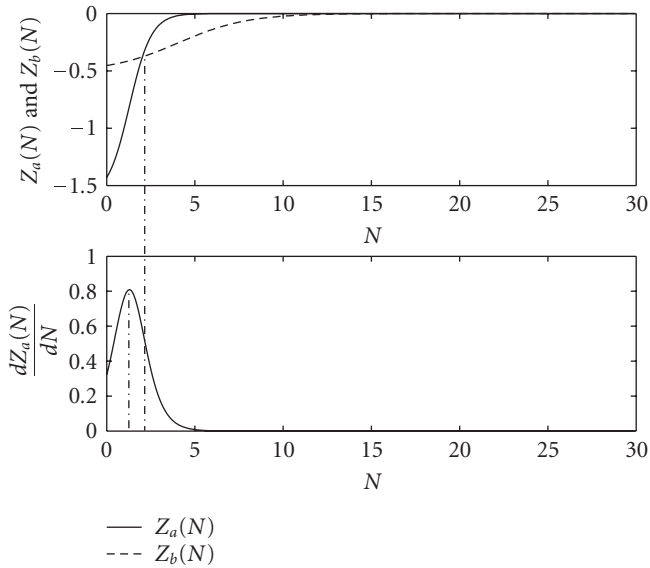
Figure 4.4. Solution to (4.18) happens near maximum of $(d/dN)Z_a(N)$.
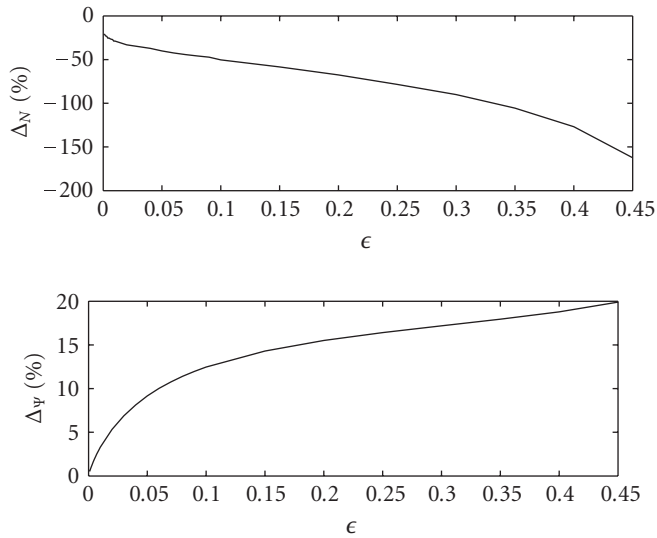
Figure 4.5. Percentage errors Δ_N and Δ_Ψ with respect to varying ϵ.

for given $\phi_1 = 1$, $\phi_2 = 5$, $a = 0.2$, and $b = 0.8$, are plotted as functions of ϵ. The percentage errors, Δ_N and Δ_Ψ, go as high as -160% and 20%, respectively, as ϵ approaches 0.45.
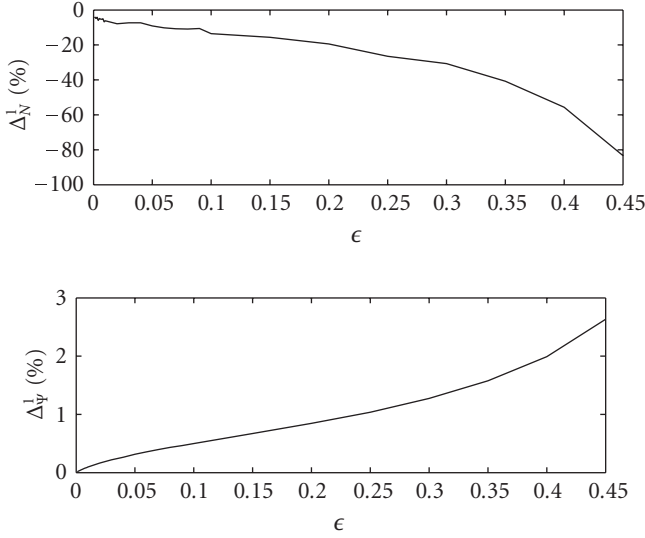
Figure 4.6. Percentage errors Δ_N^1 and Δ_Ψ^1 with respect to varying ϵ.

The approximation of N^* can be improved by iterating once in Newton's method for solving (4.18) using $\widehat{N^*}$ as initial guess. This improved approximation is then

$$\widehat{N^{*1}} = \widehat{N^*} - \frac{Z_a(\widehat{N^*}) - Z_b(\widehat{N^*})}{(d/dN)(Z_a(\widehat{N^*}) - Z_b(\widehat{N^*}))}, \tag{4.23}$$

where $\widehat{N^*}$ is given in (4.21). For the same values of ϕ_1, ϕ_2, a, and b as in Figure 4.5, Figure 4.6 plots the percentage errors

$$\Delta_N^1 = \frac{\widehat{N^{*1}} - N^*}{N^*} \times 100\%,$$

$$\Delta_\Psi^1 = \frac{\Psi_\epsilon(\widehat{N^{*1}}) - \Psi_\epsilon(N^*)}{\Psi_\epsilon(N^*)} \times 100\%, \tag{4.24}$$

as functions of ϵ. As compared to Δ_N and Δ_Ψ, Δ_N^1 and Δ_Ψ^1 improve to -80% and 3%, respectively, as ϵ approaches 0.45. Further improvement can be expected if Newton's method for solving (4.18) is carried out with more than one iteration.

5. Application

As an application, we use G-RPDs to model the behavior of a Chief Executive Officer (CEO) of a company to provide insights on how to induce the CEO, through his own self-interest, to act for the best benefits of the company.
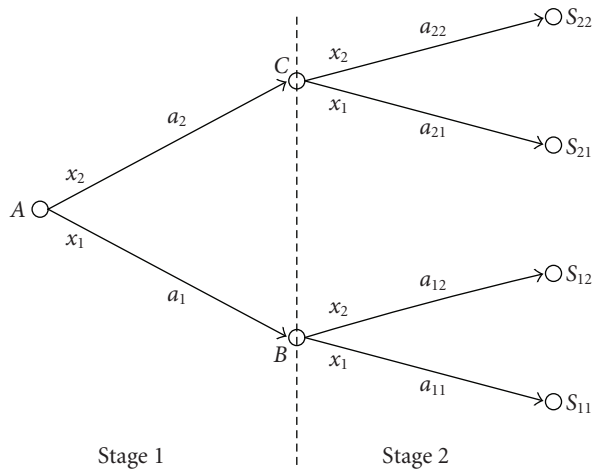
Figure 5.1. Information structure of the two-stage decision process.

5.1. Environment. Assume that the CEO is making decisions in a two-stage decision process. The information structure of the decision process is depicted by the graph in Figure 5.1. At Stage 1, the company is at node A and the CEO can choose between two decisions, x_1 or x_2, which lead the company to nodes B or C, respectively. Similarly, at Stage 2, whether the company is at node B or C, the CEO can choose between two decisions, x_1 or x_2. At the end of the process, the company is at node S_{ij} if the CEO takes the sequence of decisions, x_i at Stage 1 and x_j at Stage 2. We denote the sequence of decisions by $x_i x_j$.

The numbers a_1, a_2, a_{11}, a_{12}, a_{21}, and a_{22} on the edges of the graph denote the reward received by the CEO for each decision made. These numbers are assumed to be functions of the company's stock prices and reflect the situation of the company due to the decisions of the CEO. The larger the number, the better the situation the company is in. The information structure is assumed to be "probabilistically" known to the CEO in the sense that he can take the decision in the form of a G-RPD, that is,

$$P[\text{decision} = x_i] = \begin{cases} F_1\left(G_1\left(N, \dfrac{\varphi_2}{\varphi_1}\right)\right) & \text{for } i = 1, \\ F_1\left(G_1\left(N, \dfrac{\varphi_1}{\varphi_2}\right)\right) & \text{for } i = 2, \end{cases} \tag{5.1}$$

where φ_1 and φ_2 are the penalties associated with states x_1 and x_2, respectively, and F_1 and G_1 are defined in (2.26). The decision process at Stages 1 and 2 are described below.

Stage 1. At node A, the CEO considers all four possible sequences of decisions, $x_i x_j$, $i, j = 1, 2$. The objective reward for the sequence of decisions $x_i x_j$ is $a_i + a_{ij}$. However, subjectively, the CEO views his reward for $x_i x_j$ as $a_i + \alpha a_{ij}$, $0 \le \alpha \le 1$, where the value of α is defined by the contractual relation between the CEO and the company. When the

employment of the CEO is long term, encompassing the two stages, α is large. Otherwise, α can be small. Thus, parameter α models how much the rewards at Stage 2 are taken into account by the CEO at Stage 1.

Let κ_{ij} denote the probability that the CEO favors the sequence of decisions $x_i x_j$. Then, the probability that the CEO chooses x_i, $i = 1, 2$, at Stage 1 is

$$\kappa_i^A = \kappa_{i1} + \kappa_{i2}. \tag{5.2}$$

The probabilities, κ_{ij}, are determined by pairwise comparison of the sequence of decisions. To be more specific, taking the reciprocal of rewards as penalties,

$$\frac{\kappa_{ij}}{\kappa_{kl}} = \frac{F_1\left(G_1\left(N, \left(1/(a_k + \alpha a_{kl})\right)/\left(1/(a_i + \alpha a_{ij})\right)\right)\right)}{F_1\left(G_1\left(N, \left(1/(a_i + \alpha a_{ij})\right)/\left(1/(a_k + \alpha a_{kl})\right)\right)\right)}, \quad 1 \le i,\ j \le 2,\ (i,j) \ne (k,l), \tag{5.3}$$

where the numerator and denominator on the right-hand side of the ratio are the probabilities of the CEO favoring $x_i x_j$ and $x_k x_l$, respectively, if there were only these two choices. Since F_1 and G_1 are defined in (2.26), (5.3) implies

$$\frac{\kappa_{ij}}{\kappa_{kl}} = \left(\frac{a_i + \alpha a_{ij}}{a_k + \alpha a_{kl}}\right)^N. \tag{5.4}$$

Since $\sum_{ij} \kappa_{ij} = 1$, we have

$$\kappa_{ij} = \frac{(a_i + \alpha a_{ij})^N}{(a_1 + \alpha a_{11})^N + (a_1 + \alpha a_{12})^N + (a_2 + \alpha a_{21})^N + (a_2 + \alpha a_{22})^N}. \tag{5.5}$$

Hence, by (5.2), we obtain

$$\kappa_i^A = \frac{(a_i + \alpha a_{i1})^N + (a_i + \alpha a_{i2})^N}{(a_1 + \alpha a_{11})^N + (a_1 + \alpha a_{12})^N + (a_2 + \alpha a_{21})^N + (a_2 + \alpha a_{22})^N}. \tag{5.6}$$

Stage 2. If the company is at node B, the CEO considers the two decisions, x_1 and x_2, with rewards a_{11} and a_{12}, respectively. Taking the reciprocals of rewards as penalties, the probability of selecting x_i is

$$\kappa_i^B = F_1\left(G_1\left(N, \frac{1/a_{1j}}{1/a_{1i}}\right)\right) = \frac{a_{1i}^N}{a_{1i}^N + a_{1j}^N}, \tag{5.7}$$

where $i, j = 1, 2$, $i \ne j$. Similarly, if the company is at node C, the probability of selecting x_i is

$$\kappa_i^C = F_1\left(G_1\left(N, \frac{1/a_{2j}}{1/a_{2i}}\right)\right) = \frac{a_{2i}^N}{a_{2i}^N + a_{2j}^N}, \tag{5.8}$$

where $i, j = 1, 2$, $i \ne j$.

Given the above discussion, at the end of the decision process, the probability that the company is at S_{ij} is

$$
P_{ij} = \begin{cases}
\kappa_1^A \times \kappa_1^B & \text{for } i = 1, \ j = 1, \\
\kappa_1^A \times \kappa_2^B & \text{for } i = 1, \ j = 2, \\
\kappa_2^A \times \kappa_1^C & \text{for } i = 2, \ j = 1, \\
\kappa_2^A \times \kappa_2^C & \text{for } i = 2, \ j = 2.
\end{cases}
\tag{5.9}
$$

5.2. Analysis. We next analyze how the probabilities, P_{ij}, change as functions of α and N. For this purpose, we assume $a_1 = 1$, $a_2 = 5$, $a_{11} = 10$, $a_{12} = 15$, $a_{21} = 2$, $a_{22} = 3$. Since $a_1 + a_{12}$ has the largest value, the sequence of decisions, $x_1 x_2$, is the best for the company. Figure 5.2 shows P_{ij}, calculated according to (5.6)–(5.9), as a function of N for various values of α, and Figure 5.3 shows P_{ij} as a function of α for various N's.

Based on these figures, we observe the following.

(i) For a fixed α, the probability of the CEO making the best decision for the company, P_{12}, does not necessary become large as N becomes large. For small α (0.2 and 0.3), the CEO does not make the best decision with high probability as N becomes large. Moreover, it tends to 0 as N becomes large. However, for large α (0.6 and 1), the CEO does make the best decision with high probability if N is large enough. This means that for the CEO to make the best decision for the company, having a high LR is not enough. The CEO must also take into account the future rewards.

(ii) For a fixed N, P_{12} increases when α becomes larger. This means that for a CEO with small LR, the probability of making the best decision can be improved by increasing his ability to take into account future rewards.

5.3. Synthesis. From the observations in Section 5.2, it follows that it is best for a company to have a CEO who takes into account the future rewards when he is making decisions, that is, a CEO with large α. One way to ensure a large α is to guarantee a relatively long-term contractual relationship between the CEO and the company.

6. Conclusions

This paper shows that rational behavior can be modeled in a purely probabilistic way, thus avoiding complex dynamics associated with other approaches. The rational probabilistic deciders, introduced in this work, both in the so-called local and global implementations, allow us to quantitatively investigate individual rational behavior in stationary and symmetric Markov switch environments. Although both G-RPD and L-RPD perform well, there are qualitative differences between them. Specifically,

(1) in the stationary environment, with the same LR and penalties associated with their states, G-RPD selects its least penalized state with higher steady state probability than L-RPD;
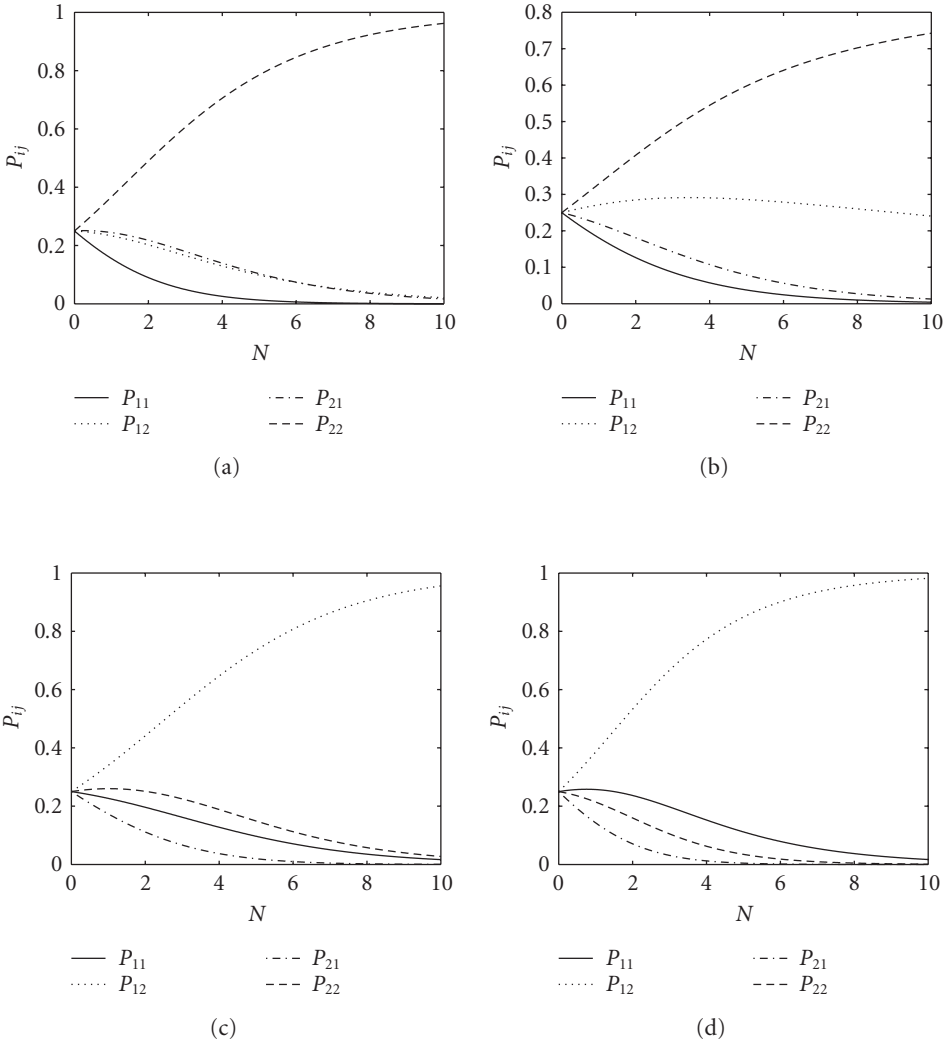
Figure 5.2. P_{ij} versus N for various fixed α. (a) $\alpha = 0.2$. (b) $\alpha = 0.3$. (c) $\alpha = 0.6$. (d) $\alpha = 1$.

(2) as the LR becomes large, the rate of convergence for G-RPD is much faster than that for L-RPD in stationary environments;

(3) in a slowly switching symmetric Markov switch environment, the average penalty incurred by G-RPD is a monotonically decreasing function of the LR, while for L-RPD, the average penalty is minimized by a finite optimal LR and increases as the LR deviates from this optimal value.

The results of this paper, as it is shown in the application, can be used as a mathematical model for investigating efficacy of various pay and incentive systems.
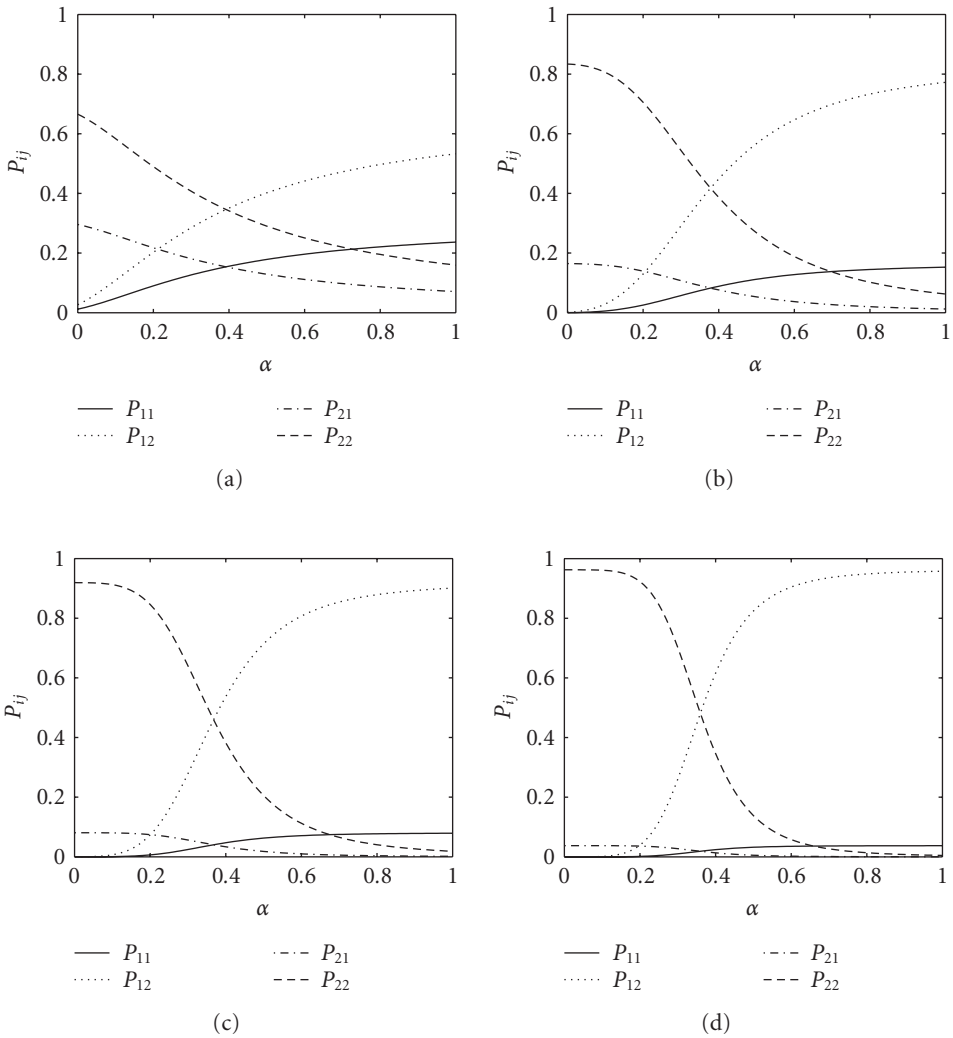
Figure 5.3. P_{ij} versus α for various fixed N. (a) $N = 2$. (b) $N = 4$. (c) $N = 6$. (d) $N = 8$.

Appendices

A. Proofs of Section 2

Proof of Theorem 2.1. (Sufficiency). Suppose an L-PD is an RPD, that is, (2.11) and (2.12) hold. Then,

$$\kappa_i(\Phi, N) = \frac{1/P(\varphi_i, N)}{1/P(\varphi_1, N) + 1/P(\varphi_2, N)}, \quad i = 1, 2, \tag{A.1}$$

implies that (2.14) and (2.15) hold.

(Necessity). Suppose the function P in an L-PD satisfies (2.14) and (2.15). Then, (A.1) implies that (2.11) and (2.12) hold.   □

Proof of Lemma 2.2. Suppose that there exists φ' such that $\lim_{N\to\infty} P(\varphi', N) = a \neq 0$. Then, for all $\varphi'' > \varphi'$,

$$\lim_{N\to\infty} \frac{P(\varphi'', N)}{P(\varphi', N)} \neq \infty, \tag{A.2}$$

which is a contradiction to (2.15). Hence, Lemma 2.2 must be true.   □

Proof of Theorem 2.3. (a) Suppose (2.22) holds. Then,

$$\kappa_i(\varphi_1, \varphi_2; N) = \frac{1/P_i(\Phi; N)}{1/P_1(\Phi; N) + 1/P_2(\Phi; N)}, \quad i = 1, 2, \tag{A.3}$$

implies (2.25).

(b) Suppose (2.25) holds. Then, (A.3) implies (2.22).

By (a) and (b), the theorem is proved.   □

Proof of Theorem 2.4. (a) Suppose $\varphi_1 < \varphi_2$. Then, by (G.4), we have $G(N, \varphi_2/\varphi_1) > G(N, \varphi_1/\varphi_2)$. Therefore, by assumption (F.1), we have $F(G(N, \varphi_2/\varphi_1)) > F(G(N, \varphi_1/\varphi_2))$. Hence, by (2.22),

$$\frac{\kappa_1(\varphi_1, \varphi_2; N)}{\kappa_2(\varphi_1, \varphi_2; N)} = \frac{F(G(N, \varphi_2/\varphi_1))}{F(G(N, \varphi_1/\varphi_2))} > 1. \tag{A.4}$$

Furthermore, as $N \to \infty$, $G(N, \varphi_2/\varphi_1) \to \infty$ and $G(N, \varphi_1/\varphi_2) \to 0$. Hence, by assumption (F.1), as $N \to \infty$, $F(G(N, \varphi_2/\varphi_1)) \to 1$ and $F(G(N, \varphi_1/\varphi_2)) \to 0$. This implies

$$\lim_{N\to\infty} \frac{\kappa_1(\varphi_1, \varphi_2; N)}{\kappa_2(\varphi_1, \varphi_2; N)} = \infty. \tag{A.5}$$

(b) Suppose $\varphi_1 > \varphi_2$. Similar to (a), we have

$$\frac{\kappa_2(\varphi_1, \varphi_2; N)}{\kappa_1(\varphi_1, \varphi_2; N)} = \frac{F(G(N, \varphi_1/\varphi_2))}{F(G(N, \varphi_2/\varphi_1))} > 1,$$

$$\lim_{N\to\infty} \frac{\kappa_2(\varphi_1, \varphi_2; N)}{\kappa_1(\varphi_1, \varphi_2; N)} = \infty. \tag{A.6}$$

From (a) and (b), we note that (2.11) and (2.12) are satisfied and the theorem is proved.   □

Proof of Lemma 2.6. The inverse function theorem and assumption (F.1) imply part (i). Similarly, the inverse function theorem along with assumptions (G.2) and (G.3) imply part (ii). Finally, the inverse function theorem and assumption (G.4) imply part (iii).   □

B. Proofs of Section 3

Proof of Theorem 3.2. Suppose the PD is an RPD. Then,

    (i) for all $0 < \epsilon < 1/2$, $\exists R^* = (1 - \epsilon)/\epsilon$ such that $\kappa_1(\Phi; N)/\kappa_2(\Phi; N) \geq R^* \Rightarrow 1 \geq \kappa_1(\Phi; N) \geq 1 - \epsilon$;

    (ii) from (2.12), for all R and for all Φ with $\varphi_1 < \varphi_2$, there exists $N^*(R, \Phi)$ such that $N \geq N^*(R, \Phi) \Rightarrow \kappa_1(\Phi; N)/\kappa_2(\Phi; N) \geq R$.

Observations (i) and (ii) imply the theorem. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Proof of Theorem 3.3. First, note that assumptions (F.1) and (F.2) imply

$$F^{-1}(z) > 1 \quad \text{if } z \in \left(\frac{1}{2}, 1\right), \tag{B.1}$$

where F^{-1} is defined in Lemma 2.6. Equation (B.1) implies that, for $0 < \epsilon < 1/2$, the number, $G_y^{-1}(F^{-1}(1 - \epsilon))$, where $y = \varphi_2/\varphi_1$ and G_y^{-1} is defined in Lemma 2.6, exists. Furthermore,

$$\infty > N \geq G_y^{-1}\left(F^{-1}(1 - \epsilon)\right) \implies \infty > G(N, y) \geq F^{-1}(1 - \epsilon)$$

$$\implies 1 \geq F\left(G(N, y)\right) \geq 1 - \epsilon \tag{B.2}$$

$$\implies 1 \geq \kappa_1\left(\varphi_1, \varphi_2; N\right) \geq 1 - \epsilon.$$

Let $N^*(\epsilon, y) = G_y^{-1}(F^{-1}(1 - \epsilon))$, and the theorem is proved. $\qquad\qquad\qquad\square$

C. Proofs of Section 4

Proof of parts (i) and (ii) of Theorem 4.1. We first prove part (i). The function $f_\epsilon(N)$ defined in (4.8) satisfies

$$f_\epsilon(N) - \frac{1}{2} = \frac{(b^N - a^N)(1 - 2\epsilon)}{2\left[2\epsilon(a^N - 1) + 2\epsilon(b^N - 1) - a^N - b^N\right]}. \tag{C.1}$$

From the discussions in Section 4.2, we have

$$0 < a < b < 1. \tag{C.2}$$

Hence, $b^N - a^N > 0$ and $(2\epsilon(a^N - 1) + 2\epsilon(b^N - 1) - a^N - b^N) < 0$. Therefore,

$$f_\epsilon(N) - \frac{1}{2} < 0 \quad \text{if } 0 < \epsilon < \frac{1}{2},$$

$$f_\epsilon(N) - \frac{1}{2} > 0 \quad \text{if } 1 > \epsilon > \frac{1}{2}. \tag{C.3}$$

Equations (4.7) and (C.3) give (4.9).

Next, we prove part (ii). Note that from (4.8) and (C.2),

$$\lim_{N \to 0} f_\epsilon(N) = \frac{1}{2}, \qquad \lim_{N \to \infty} f_\epsilon(N) = \frac{1}{2}. \tag{C.4}$$

Equations (4.7) and (C.4) give (4.10). □

To prove parts (iii) and (iv), we need the following lemmas.

LEMMA C.1. *Suppose $N' \in (0, \infty)$ is a solution to*

$$\frac{d\Psi_\epsilon(N)}{dN} = 0. \tag{C.5}$$

Then,

$$\left. \frac{d^2\Psi_\epsilon(N)}{dN^2} \right|_{N=N'} > 0 \quad if \, 0 < \epsilon < \frac{1}{2},$$

$$\left. \frac{d^2\Psi_\epsilon(N)}{dN^2} \right|_{N=N'} < 0 \quad if \, \frac{1}{2} < \epsilon < 1. \tag{C.6}$$

Proof. Note that

$$\frac{d\Psi_\epsilon(N)}{dN} = (\varphi_2 - \varphi_1) \frac{(2\epsilon - 1)\left(\left(a^N \ln a(2b^N \epsilon - b^N - \epsilon)\right) - \left(b^N \ln b(2a^N \epsilon - a^N - \epsilon)\right)\right)}{\left(2a^N \epsilon - a^N - \epsilon + 2b^N \epsilon - b^N - \epsilon\right)^2}. \tag{C.7}$$

Hence, N' must satisfy

$$(\ln a)a^{N'} \left(b^{N'} + \epsilon - 2b^{N'} \epsilon\right) = (\ln b)b^{N'} \left(a^{N'} + \epsilon - 2a^{N'} \epsilon\right). \tag{C.8}$$

Furthermore,

$$\left. \frac{d^2\Psi_\epsilon(N)}{dN^2} \right|_{N=N'} = (2\epsilon - 1) \frac{\ln a(\ln a)a^{N'} \left(b^{N'} + \epsilon - 2b^{N'} \epsilon\right) - \ln b(\ln b)b^{N'} \left(a^{N'} + \epsilon - 2a^{N'} \epsilon\right)}{\left(2a^{N'} \epsilon - a^{N'} - \epsilon + 2b^{N'} \epsilon - b^{N'} - \epsilon\right)^2}. \tag{C.9}$$

Equations (C.8) and (C.2) imply

$$\ln a(\ln a)a^{N'} \left(b^{N'} + \epsilon - 2b^{N'} \epsilon\right) - \ln b(\ln b)b^{N'} \left(a^{N'} + \epsilon - 2a^{N'} \epsilon\right) < 0 \tag{C.10}$$

for all $\epsilon \in (0, 1)$. Equations (C.9) and (C.10) imply the lemma. □

LEMMA C.2. *If $0 < \epsilon < 1$ and $\epsilon \neq 1/2$, then (C.5) has a unique real positive solution.*

Proof. We note that part (ii) of Theorem 4.1 and Rolle's theorem imply that there must be at least one real positive solution to (C.5). Suppose that $0 < \epsilon < 1$ and $\epsilon \neq 1/2$ and that there are more than two solutions to (C.5). Then, Lemma C.1 implies that these solutions are isolated. Hence, we can find two solutions, N_1 and N_2, to (C.5) such that $N_1 < N_2$ and there are no other solutions to (C.5) in between them.

(a) If $0 < \epsilon < 1/2$, then there exists a δ_1 small enough such that

$$\frac{d\Psi_\epsilon(N)}{dN}\bigg|_{N=N_1+\delta_1} > 0, \qquad \frac{d\Psi_\epsilon(N)}{dN}\bigg|_{N=N_2-\delta_1} < 0. \tag{C.11}$$

By the intermediate value theorem, there exists N_3 such that $N_1 < N_3 < N_2$ and $d\Psi_\epsilon(N)/dN|_{N=N_3} = 0$, which is a contradiction.

(b) If $1/2 < \epsilon < 1$, then there exists a δ_2 small enough such that

$$\frac{d\Psi_\epsilon(N)}{dN}\bigg|_{N=N_1+\delta_2} < 0, \qquad \frac{d\Psi_\epsilon(N)}{dN}\bigg|_{N=N_2-\delta_2} > 0. \tag{C.12}$$

By the intermediate value theorem, there exists N_3 such that $N_1 < N_3 < N_2$ and $d\Psi_\epsilon(N)/dN|_{N=N_3} = 0$, which is a contradiction.

By (a) and (b), there must be only one real positive solution to (C.5). $\qquad\square$

Proof of parts (iii) and (iv) of Theorem 4.1. Since $\Psi_\epsilon(N)$ is a differentiable function of $N \in (0, \infty)$, its extreme values happen at the end points, $N \to 0$ and $N \to \infty$, and the real positive solution to (C.5). Parts (i) and (ii) of Theorem 4.1 and Lemmas C.1 and C.2 imply the following.

(a) If $0 < \epsilon < 1/2$, then $\Psi_\epsilon(N)$ has a unique minimum at the real positive solution to (C.5).

(b) If $1/2 < \epsilon < 1$, then $\Psi_\epsilon(N)$ has a unique maximum at the real positive solution to (C.5).

Parts (iii) and (iv) of Theorem 4.1 follow from (a) and (b), respectively. $\qquad\square$

Proof of Theorem 4.2. From Section 4.3, N^* satisfies (4.18), which can be re-written as

$$Z_a(N) - Z_b(N) = 0, \tag{C.13}$$

where $Z_a(N)$ and $Z_b(N)$ are the left-hand side and right-hand side of (4.18), respectively. By the implicit function theorem,

$$\begin{aligned}
\frac{dN^*}{d\epsilon} &= -\frac{\partial(Z_a(N) - Z_b(N))/\partial\epsilon}{\partial(Z_a(N) - Z_b(N))/\partial N}\bigg|_{N=N^*} \\
&= -\frac{2b^{N^*}(\ln b/\ln a - 1) + (-b^{N^*}\ln b/a^{N^*}\ln a + 1)}{\epsilon \ln b (b^{N^*}/a^{N^*} - 1)}.
\end{aligned} \tag{C.14}$$

From (C.2) and (4.18), we have

$$\frac{\ln b}{\ln a} - 1 < 0, \quad \frac{b^{N^*}}{a^{N^*}} - 1 > 0, \quad -\frac{b^{N^*}\ln b}{a^{N^*}\ln a} + 1 < 0, \quad \ln b < 0. \tag{C.15}$$

Hence, $dN^*/d\epsilon < 0$, and hence, ϵ decreases implies N^* increases.

Next, note that

$$\Psi_\epsilon(N^*) = \varphi_1 + (\varphi_2 - \varphi_1) f_\epsilon(N^*)$$

$$= \varphi_1 + (\varphi_2 - \varphi_1) \frac{1}{1 + (\ln b/\ln a)(b/a)^{N^*}}. \tag{C.16}$$

Consider ϵ_1 and ϵ_2 such that $0 < \epsilon_1 < \epsilon_2 < 1/2$. Let N_1^* and N_2^* be the optimal LRs when the symmetric Markov switch environment is such that $\epsilon = \epsilon_1$ and $\epsilon = \epsilon_2$, respectively. By the above discussions, we have $N_1^* > N_2^*$. By (C.2) and (C.16), we have $\Psi_{\epsilon_1}(N_1^*) < \Psi_{\epsilon_2}(N_2^*)$. Hence, ϵ decreases implies $\Psi_\epsilon(N^*)$ decreases. $\qquad\square$

Proof of Theorem 4.3. We first prove part (ii). We note that

$$\lim_{N \to 0} f_\epsilon'(N) = f_\epsilon'(N)\big|_{N=0} = \frac{1}{2}. \tag{C.17}$$

Furthermore, the relation

$$0 < \phi_1 < \phi_2 < \infty \tag{C.18}$$

implies

$$\lim_{N \to \infty} f_\epsilon'(N) = \frac{\epsilon}{\epsilon + 1/2}. \tag{C.19}$$

Equations (C.17), (C.19), and (4.13) imply part (ii).

Next, we prove parts (iii) and (iv). Note that

$$\frac{df_\epsilon'(N)}{dN} = \frac{-(1/2)\big(\ln(\phi_1/\phi_2)\big)(\phi_1/\phi_2)^N(\epsilon - 1/2)}{\big((\phi_1/\phi_2)^N \epsilon - (1/2)(\phi_1/\phi_2)^N - \epsilon - 1/2\big)^2}. \tag{C.20}$$

Equation (C.18) implies that

$$\begin{aligned}
\frac{df_\epsilon'(N)}{dN} &< 0 \quad \text{if } 0 < \epsilon < \frac{1}{2}, \\
\frac{df_\epsilon'(N)}{dN} &> 0 \quad \text{if } \frac{1}{2} < \epsilon < 1.
\end{aligned} \tag{C.21}$$

Equations (C.21) and (4.13) imply parts (iii) and (iv).

Finally, parts (ii), (iii), and (iv) imply part (i). $\qquad\square$

References

[1] M. L. Tsetlin, *Automaton Theory and Modeling of Biological Systems*, Academic Press, New York, NY, USA, 1973.

[2] V. I. Varshavskiĭ, *Collective Behavior of Automata*, "Nauka", Moscow, 1973, (in Russian).

[3] M. L. Tsetlin, "On the behavior of finite automata in random media," *Automation and Remote Control*, vol. 22, no. 10, pp. 1210–1219, 1961.

[4] I. M. Gel'fand, I. I. Pjateckiĭ-Šapiro, and M. L. Tsetlin, "On certain classes of games and robot games," *Doklady Akademii Nauk SSSR*, vol. 152, no. 4, pp. 845–848, 1963.

[5] S. M. Meerkov, "Mathematical theory of behavior—individual and collective behavior of retardable elements," *Mathematical Biosciences*, vol. 43, no. 1-2, pp. 41–106, 1979.

[6] V. Y. Krylov, "On one stochastic automata which is asymptotically optimal in a random medium," *Automation and Remote Control*, vol. 24, no. 9, pp. 1226–1228, 1963.

[7] V. A. Ponomarev, "A construction of an automaton which is asymptotically-optimal in a stationary random medium," *Biophysics*, vol. 9, no. 1, pp. 104–110, 1964.

[8] V. I. Krinskiy, "An asymptotically-optimal automata with an exponential rate of convergence," *Biophysics*, vol. 9, no. 4, pp. 484–487, 1964.

[9] A. A. Miljutin, "On automata with an optimal behavior in random media," *Automation and Remote Control*, vol. 26, no. 1, pp. 117–131, 1965.

[10] V. I. Varshavskiĭ and I. P. Vorontsova, "On the behavior of stochastic automata with a variable structure," *Automation and Remote Control*, vol. 24, no. 3, pp. 327–333, 1963.

[11] V. I. Bryzgalov, I. M. Gel'fand, I. I. Pjateckiĭ-Šapiro, and M. L. Tsetlin, "Homogeneous games of automata and their simulation on digital computers," *Automation and Remote Control*, vol. 25, no. 11, pp. 1572–1580, 1964.

[12] S. L. Ginsburg and M. L. Tsetlin, "On modeling of collective behavior of automata," *Problems of Information Transmission*, vol. 1, no. 2, pp. 54–62, 1965.

[13] V. I. Varshavskiĭ, M. I. Meleshina, and M. L. Tsetlin, "The behavior of automata in periodic random media and the problem of synchronization in the presence of noise," *Problems of Information Transmission*, vol. 1, no. 1, pp. 65–71, 1965.

[14] V. Y. Krylov and M. L. Tsetlin, "On automaton games," *Automation and Remote Control*, vol. 24, no. 7, pp. 975–987, 1963.

[15] V. I. Krinskiy, "Zero-sum games for two asymptotically optimal sequences of automata," *Problems of Information Transmission*, vol. 2, no. 2, pp. 43–53, 1966.

[16] E. T. Gurvich, "Method for asymptotic investigation of games between automata," *Automation and Remote Control*, vol. 36, no. 2, part 1, pp. 257–270, 1975.

[17] V. A. Borovikov and V. I. Bryzgalov, "The simplest symmetric game of many automata," *Automation and Remote Control*, vol. 26, no. 4, pp. 681–684, 1965.

[18] V. A. Volkonskiĭ, "Asymptotic properties of the behavior of simple automata in a game," *Problems of Information Transmission*, vol. 1, no. 2, pp. 36–53, 1965.

[19] B. G. Pittel, "The asymptotic properties of a version of the gur game," *Problems of Information Transmission*, vol. 1, no. 3, pp. 99–112, 1965.

[20] B. G. Pittel, "A simple probability model of collective behavior," *Problems of Information Transmission*, vol. 3, no. 3, pp. 30–44, 1967.

[21] Yu. D. Tenisberg, "Some models of collective behavior in dynamic processes of market price formation," *Automation and Remote Control*, vol. 30, no. 7, pp. 1140–1148, 1969.

[22] G. J. McMurtry and K.-S. Fu, "A variable structure automaton used as a multi-modal searching technique," *IEEE Transactions on Automatic Control*, vol. 11, no. 3, pp. 379–387, 1966.

[23] K. S. Fu, "Stochastic automata as models of learning systems," in *A Prelude to Neural Networks: Adaptive and Learning Systems*, pp. 393–431, Prentice-Hall Press, Englewood Cliffs, NJ, USA, 1994.

[24] K.-S. Fu, "Learning control systems—review and outlook," *IEEE Transactions on Automatic Control*, vol. 15, no. 2, pp. 210–221, 1970.

[25] K. S. Narendra and R. Viswanathan, "A two-level system of stochastic automata for periodic random environments," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 2, no. 2, pp. 285–289, 1972.

[26] K. S. Narendra and M. A. L. Thathachar, "Learning automata—a survey," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 4, no. 4, pp. 323–334, 1974.

[27] K. S. Narendra, E. A. Wright, and L. G. Mason, "Application of learning automata to teletraffic routing and management," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 7, no. 11, pp. 785–792, 1977.

[28] K. S. Narendra and M. A. L. Thathachar, "On the behavior of learning automata in a changing environment with application to telephone traffic routing," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 10, no. 5, pp. 262–269, 1980.

[29] K. S. Narendra and M. A. L. Thathachar, *Learning Automata: An Introduction*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1989.

[30] B. Tung and L. Kleinrock, "Distributed control methods," in *Proceedings of the 2nd International Symposium on High Performance Distributed Computing*, pp. 206–215, Spokane, Wash, USA, July 1993.

[31] Y.-C. Tung, "Distributed control using finite state automata," Ph.D. dissertation, Computer Science Department, UCLA, Los Angeles, Calif, USA, 1994.

[32] Y. U. Cao, A. S. Fukunaga, and A. B. Kahng, "Cooperative mobile robotics: antecedents and directions," *Autonomous Robots*, vol. 4, no. 1, pp. 7–27, 1997.

[33] R. Iyer and L. Kleinrock, "QoS control for sensor networks," in *Proceedings of IEEE International Conference on Communications (ICC '03)*, vol. 1, pp. 517–521, Seattle, Wash, USA, May 2003.

[34] S. Ho, H. Nassef, N. Pornsinsirirak, Y.-C. Tai, and C.-M. Ho, "Unsteady aerodynamics and flow control for flapping wing flyers," *Progress in Aerospace Sciences*, vol. 39, no. 8, pp. 635–681, 2003.

[35] J. Flinn, P. T. Kabamba, W.-C. Lin, S. M. Meerkov, and C. Y. Tang, "Collaborative power management in WPANs using rational controllers: a case study," *Mathematical Problems in Engineering*, vol. 2005, no. 5, pp. 491–501, 2005.

P. T. Kabamba: Department of Electrical Engineering and Computer Science,
University of Michigan, Ann Arbor, MI 48109-2122, USA
Email address: kabamba@umich.edu

W.-C. Lin: Department of Electrical Engineering and Computer Science,
University of Michigan, Ann Arbor, MI 48109-2122, USA
Email address: wenchiao@umich.edu

S. M. Meerkov: Department of Electrical Engineering and Computer Science,
University of Michigan, Ann Arbor, MI 48109-2122, USA
Email address: smm@eecs.umich.edu