Rochester Center for Economic Research

Switching Costs and the Gittins Index

Banks, Jeffrey S. and Rangarajan K. Sundaram

Working Paper No. 353
July 1993

University of
Rochester

Switching Costs and the Gittins Index

Jeffrey S. Banks and Rangarajan K. Sundaram

Rochester Center for Economic Research
Working Paper No. 353

Switching Costs and the Gittins Index[1]

Jeffrey S. Banks and Rangarajan K. Sundaram
Department of Economics, Harkness Hall
University of Rochester
Rochester, NY 14627

July 8, 1993

[1] We are very grateful to Andy McLennan for several helpful conversations. We would also like to thank Martin Hellwig and two referees for their comments. The first author gratefully acknowledges financial support provided by the Sloan Foundation and the NSF. An earlier version of this paper was written during the second author's sabbatical at the California Institute of Technology, and he would like to thank them for their hospitality.

Abstract

We study independent-armed Bandit problems with geometric discounting over an infinite horizon. When there is no cost of switching between arms, it is well known that the "Gittins Index" completely characterizes the set of optimal strategies in such problems. In contrast, we show that if switching costs are are allowed to be non-zero, optimal index strategies no longer exist. This is true even if attention is restricted to that family of Bandit problems in which the cost of switching between any two arms is fixed *a priori* at some constant (non-zero) level.

Keywords: independent-armed bandit problems, switching costs, Gittins index, index strategies.

1 Introduction

The Theorem of Gittins and Jones (1974) is, perhaps, the single most powerful result in the literature on Bandit problems. This result establishes that in independent-armed Bandit problems with geometric discounting over an infinite horizon, *all* optimal strategies may be obtained by solving a family of simple optimal stopping problems that associate with each arm an index known as the *Dynamic Allocation Index* or, more popularly, as the *Gittins index.* Importantly, the Gittins index of an arm depends solely on the characteristics of that arm and the rate of discounting, and is otherwise completely independent of the problem under consideration. These features simplify significantly the task of characterizing optimal strategies in this class of problems.[1]

The purpose of this paper is to examine the extent to which the Gittins-Jones Theorem remains valid when the cost of switching between arms is possibly non-zero, *i.e.,* to determine whether suitably defined index strategies continue to remain optimal in this case. The need to include switching costs arises primarily from economic considerations. Such an extension is, perhaps, of especial interest in a labor market setting where the Bandit framework has found wide applicability,[2] but appears more generally important. Indeed, it is difficult to imagine a relevant economic decision problem in which the decision-maker may costlessly move between alternatives.[3]

As our framework of analysis, we use a generalized version of Whittle (1982) with arm-specific costs of switching. It is obvious that in the most inclusive case where there is a cost c_{ij} for switching from an arm i to another arm j, there cannot exist an optimal index strategy with the index on an arm depending solely on the arm's characteristics. We consider therefore, the more restrictive case where the cost of switching away from an arm (resp. to an arm) are independent of the arm to which (resp. from which) the switch is made. In principle, at least, this leaves open the possibility that an optimal index strategy may exist.

Unfortunately, our main result is negative. We show that, in general, it is not possible to define indices which have the property that the resulting index strategy is optimal on the domain of all Bandit problems with switching costs. Indeed, this remains true even if attention is restricted to that subset of the domain in which the cost of switching is a given (non-zero) constant. From one point of view, this non-existence may not appear very surprising, for the Gittins Index is known to be non-robust in some (other) directions. For instance, the stationarity of the underlying problem is very important: Berry and Fristedt (1985) show that geometric discounting is a *necessary* condition for the validity of the Gittins-

[1] See, *e.g.,* Whittle (1982), or Banks and Sundaram (1992).

[2] See, *e.g.,* Mortensen (1985). See also Banks and Sundaram (1992), who provide a list of other applications in economics and political science.

[3] It is somewhat surprising to note, therefore, that the literature on switching costs consists only of a few stray papers. Examples include Kolonko and Benzing (1983) who study the special case of a two-armed Bandit with one known arm, where the other arm generates rewards according to a Bernoulli distribution with unknown parameters; and Agrawal, *et al* (1988) who study the existence of asymptotically efficient adaptive allocation rules (in the sense of Lai and Robbins, 1985) in the presence of switching costs. See also the recent contribution of Feldman and Spagat (1993) on the impact of switching costs in a general model of optimal Bayesian learning.

Jones Theorem. On the other hand, the addition of switching costs does not affect the model's stationarity, and certainly, there is no *a priori* reason to expect lack of robustness in this direction. In particular, Weitzman (1979) shows that optimal index strategies do exist in the closely related "Pandora's Box" problem, where there is a non-zero cost to be paid the *first* time an arm is used, but subsequent visits to the arm are free, even if one has switched away to another arm in the interim.

Our proof of the non-existence of an optimal index uses a *reductio ad absurdum* approach: we assume that an optimal index strategy does exist, derive some of the properties it must satisfy, and show that these properties are not mutually consistent. The intuition underlying our construction is quite straightforward. Consider an n-armed Bandit problem, and suppose the decision maker is currently on some arm (the *incumbent arm*). If, in the optimal continuation, there is any possibility of switching back to the incumbent arm after leaving it (depending on, say, the realizations from the other arms), then the index on the incumbent arm must depend non-trivially on the cost of switching (back) to it, since a higher cost of coming back should make the decision-maker more reluctant to leave the arm. If, on the other hand, coming back to this arm is a zero-probability event (say, because the worst realizations on the other arms would still dominate the present incumbent) the arm's index must be independent of the cost of switching back to it.[4] Together these statements furnish the required contradiction, since (by definition) the index on an arm cannot depend on features extraneous to the arm, such as the payoff prospects from other arms.

The remainder of this paper is organized as follows. Section 2 provides a description of Bandit problems, introduces switching costs, and defines the notion of an optimal index in this case. Section 3 formalizes the intuition of the previous paragraph in proving the non-existence of an optimal index under switching costs, both in general and in the restricted case where attention is limited to problems having an *a priori* given and fixed switching cost.

2 Bandit Problems

2.1 The Standard Framework

Our description of the standard Bandit framework in this paper is, of necessity, terse. We also keep the technical exposition at a relatively informal level. For omitted details, we refer the reader to Whittle (1982).

An independent-armed Bandit problem with geometric discounting (hereafter, simply *Bandit problem*) is defined by the following objects:[5]

[4] Alternatively put, the index on an incumbent arm changes depending on whether we compare it to an arm whose payoff prospects are known for certain, or to an arm whose payoffs involve some uncertainty.

[5] Our description follows Whittle (1982). In the "classical" version of the Bandit problem, as used for instance, by Berry and Fristedt (1985), the states of arm i would correspond to the set of possible beliefs the decision-maker may have regarding the "true" distribution of rewards from arm i; the transition probabilities are implicitly defined by the map taking prior beliefs and observed rewards into posterior beliefs.

1. A set $N = \{1, \ldots, n\}$ of *arms* of the Bandit, where n is a positive integer.
2. A tuple $F_i = (X_i, r_i, Q_i)$ for each arm i where
   (a) X_i, a subset of some Polish (*i.e.*, complete, separable, metric) space, describes the set of possible *states* of arm i, with generic element x_i;
   (b) $r_i : X_i \rightarrow \Re$ is a bounded measurable function describing the instantaneous *reward* from arm i; and
   (c) Q_i represents a family of *transition probabilities* on X_i, i.e., for each $x_i \in X_i$, $Q_i(.|x_i)$ is a probability distribution on X_i, and for each fixed Borel subset D of X_i, $Q_i(D|.)$ is a measurable mapping from X_i into [0,1].
3. A *discount factor* $\rho \in [0, 1)$.

The Bandit problem has the following interpretation. In each period $t = 0, 1, 2, \ldots,$ of an infinite horizon, a decision-maker must decide which arm of the Bandit is to be employed in that period, given the vector of states $(x_1^t, ..., x_n^t)$ at the begining of that period. This decision is made with full knowledge of the history of the problem to date. If arm i is chosen in period t, two things happen. First, the decision-maker receives a reward of $r_i(x_i^t)$. Second, the state of arm i transits to its period $(t+1)$-value x_i^{t+1} according to the (conditional) probability distribution $Q_i(.|x_i^t)$. The states of all other arms remain frozen, so that we have $x_j^{t+1} = x_j^t$ for all $j \neq i$. The decision-maker discounts future rewards by the factor $\rho \in [0, 1)$, and aims to maximize total discounted expected reward over the infinite horizon.

More formally, for any $t \geq 0$, a *t-history* h_t for the problem is a description of the state of each arm in each period upto t, the action taken in each of those periods, and the period-t state. Let $H_0 = X_1 \times \cdots \times X_n$, and let H_t be the set of all possible histories upto t. A *strategy* σ for the decision-maker is a rule that recommends the arm to be played at any point in time as a function of the history upto that point, i.e., it is a sequence of maps $\{\sigma_t\}$, where for each $t \geq 0$, σ_t is a measurable map from H_t into N.

Each strategy σ defines in the obvious way an expected t-th period reward, denoted $r_t[x]$, from each initial state $x = (x_1, \ldots, x_n)$ and for each t. The total *worth* of σ from x, denoted $W(\sigma)(x)$, is then defined as $W(\sigma)(x) = \sum_{t=0}^{\infty} \rho^t r_t[x]$. A strategy σ^* is an *optimal* strategy if its worth is maximal amongst all strategies, i.e., if $W(\sigma^*)(x) = \sup_\sigma W(\sigma)(x)$ for all x.

Standard arguments from dynamic programming (see for instance, Whittle, 1982) establish that optimal strategies exist in this problem, and, indeed, that *stationary Markovian* optimal strategies[6] exist. The breakthrough achieved by Gittins and Jones (1974) lies in showing that a particularly simple class of strategies—those defined through the Gittins index—actually suffice to obtain *all* optimal strategies. We turn now to a brief description of this result. In the sequel, ρ is assumed fixed at some level in [0,1), and all dependence on ρ is suppressed.

[6] A stationary Markovian strategy σ is a strategy under which the period-t action depends solely on the period-t state vector $x^t = (x_1^t, \ldots, x_n^t)$, but not on how or when that state was reached. Such a strategy can evidently be represented by a measurable function $g: X_1 \times \cdots \times X_n \rightarrow N$, with the interpretation that $g(x)$ is the action recommended by the strategy when the state is x.

2.2 The Gittins Index

The Gittins index on an arm i, whose characteristics are given by $F_i = (X_i, r_i, Q_i)$, is obtained by the following procedure. Let $m \in \Re$ be given. Consider the stopping problem in which in each period (given that the terminal reward m has not yet been accepted) the decision-maker must choose between playing arm i for one more period, and stopping and accepting the terminal reward m. Routine arguments show that the value $V(x_i, F_i; m)$ of this problem is well-defined and finite from any initial state $x_i \in X_i$. The *Gittins index* on arm i, denoted by $\mu(x_i, F_i)$ is then defined by

$$\mu(x_i, F_i) = \inf\{m | \, V(x_i, F_i; m) = m\} \tag{2.1}$$

Since r_i is bounded by assumption, it follows that for large m, we have $V(.; F_i, m) = m$, while for $-m$ large, $V(x_i, F_i; m)$ is independent of m. Thus, the Gittins index is well-defined. The importance of this index lies in the following result. Let $\{N, \, (F_i)_{i\in N}\}$ be an arbitrary Bandit problem.

Theorem 1 (Gittins and Jones (1974)) *The optimal selections at the state* $(x_1, \ldots, x_n)$ *in the Bandit* $\{N, (F_i)_{i\in N}\}$ *are those arms* i *for which* $\mu(x_i, F_i) = \max\{\mu(x_j, F_j) | j \in N\}$.

Equivalently, the Gittins-Jones Theorem may be stated as follows: a strategy σ for a Bandit problem $\{N, \, (F_i)_{i\in N}\}$ is an optimal strategy if, and only if, the set of histories on which the recommendations of σ differ from the Gittins index-maximal arms after that history has probability zero.

2.3 Switching Costs in the Bandit Framework

The most general way to introduce switching costs in the Bandit framework is to assume that there exists a cost c_{ij} for switching from arm i to arm j, $i, j \in \{1, ..., N\}$. It is clear, however, that no index can then be defined which is such that the resulting strategy is optimal, if the index for an arm is to depend on that arm alone. The formulation we use here, therefore, is more specialized, and one that leaves open the possibility, at least at the intuitive level, that optimal index strategies may exist.

Specifically, we associate with each arm a pair (c_i, d_i) of real numbers where (i) c_i is the cost of switching *to* arm i (from any arm), and (ii) d_i is the cost of switching *away from* arm i (to any arm). Thus, if a switch occurs from arm i to arm j, the total cost paid is $d_i + c_j$.

To avoid further complicating notation, in the sequel the tuple F_i describing arm i is to be understood as including the vector (c_i, d_i) also.

When switching costs are admitted, the state of the Bandit problem in any period cannot, except at the very begining, be adequately described by just the vector $(x_1, ..., x_n)$. Rather, it is also important to know the arm that was in use in the period immediately preceding. (We will henceforth refer to this arm as the arm "currently in use.") Defining $\Delta = X_1 \times$

$\cdots \times X_n \times N$, and letting x denote the vector $(x_1, ..., x_n)$, routine arguments now show that the value function $V: \Delta \rightarrow \Re$ for this problem satisfies the Bellman optimality equation at each $(x, j) \in \Delta$:

$$V(x,j) = \max_{i \in N} L_i V(x,j) \tag{2.2}$$

where, for $i \neq j$,

$$L_i V(x,j) = r_i(x_i) - c_i - d_j + \rho \int V(x_{-i}, \hat{x}_i, i) Q_i(d\hat{x}_i | x_i) \tag{2.3}$$

while

$$L_j V(x,j) = r_j(x_j) + \rho \int V(x_{-j}, \hat{x}_j, j) Q_j(d\hat{x}_j | x_j); \tag{2.4}$$

and that any measurable selection from the correspondence of maximizers of (2.2) constitutes a stationary Markovian optimal strategy. In order, however, to examine the existence of optimal index strategies, we must first define the notion of an index for this problem. We turn to this now.

2.4 The Index with Switching Costs

As with the Gittins Index, we shall define an index on a generic arm i to be any function obtainable solely from the characteristics F_i of arm i. However, if we require the index to depend on F_i alone, then simple examples show that index strategies cannot be optimal. Consider the following:

> **Example 1:** Let $N = \{1,2\}$; $X_1 = X_2 = [0,1]$; $r_1 = r_2 = r$, where $r(x) = x$, for all $x \in [0,1]$; $Q_1 = Q_2 = Q$, where $Q(1|x) = 1 - Q(0|x) = x$, for all $x \in [0,1]$; $d_1 = d_2 = d > 0$, while $c_1 = c_2 = 0$; and, finally, let ρ be any value in (0,1).
>
> Consider any indices $\lambda(x_i, F_i)$ for the arms. Since the arms are identical upto the initial state, we must have $\lambda(., F_1) = \lambda(., F_2) = \lambda(., F)$, say, where $F_1 = F_2 = F$. It is evident that the attractiveness of an arm is increasing in the value of the initial state, so that, if at all λ is to be optimal, it must be increasing on [0,1]. In particular, we must have $\lambda(x, F) > \lambda(0, F)$ for any $x > 0$.
>
> But $\lambda(x, F) > \lambda(0, F)$ for all $x > 0$ is inconsistent with the optimality of λ. For, suppose we had $0 = x_1 < x_2$, and $d > x_2/(1 - \delta)$. Suppose further that the decision-maker is currently on arm 1. It is clear then that the unique optimal policy is simply to stay with arm 1 forever, but λ recommends a shift to arm 2, which is strictly suboptimal.

The reason this example "works" is that in requiring the index to depend on F_i alone, we have omitted the crucial bit of information about whether arm i was the arm that was in use in the previous period. For, it is obvious that in comparing two otherwise identical arms, one of which was used in the previous period, the one which was in use must necessarily be more attractive than the one which was idle. This motivates the following:

Definition: *An index in the presence of switching costs is any function λ which specifies for a generic arm i, a value $\lambda(x_i, F_i, s_i)$, where F_i denotes the characteristics of arm i, x_i is the current state of arm i, and $s_i \in \{0,1\}$ is a variable that specifies whether $(s_i = 1)$, or not $(s_i = 0)$, i is the arm currently in use.*

An index λ induces in each Bandit problem $\{N, (F_i)_{i \in N}\}$, a strategy $\sigma(\lambda)$ in the obvious manner: let x be the vector of initial states. In period 0, $\sigma(\lambda)$ plays any of the arms i for which $\lambda(x_i, F_i, 0) = \max\{\lambda(x_j, F_j, 0) | j \in N\}$. For each subsequent period t, let x^t denote the vector of states at the begining of period t, and $i(t-1)$ the arm that was used in period $(t-1)$. Then, in period t, $\sigma(\lambda)$ plays any of the arms i for which $\lambda(x_i, F_i, s_i) = \max\{\lambda(x_j, F_j, s_j) | j \in N\}$, where $s_i = 1$ iff $i = i(t-1)$.

Finally, an index λ is said to be an *optimal index* in the presence of switching costs if $\sigma(\lambda)$ is optimal in *every* Bandit problem $\{N, (F_i)_{i \in N}\}$.

3 The Non-Existence of an Optimal Index

We show in this section that an optimal index does not exist in the presence of switching costs. Our argument consists of two parts. First, we will show that any Bandit problem with costs of switching "from" (and, possibly, also costs of switching "to"), is equivalent to another problem in which there are only costs of switching "to." We will then consider the case where the only switching costs are costs of switching "to," and show that in a series of steps, that if an optimal index does exist a contradiction must result, completing the proof.

So let a Bandit $B = \{N, (F_i)_{i \in N}\}$ be given. Define another Bandit $B^* = \{N, (F_i^*)_{i \in N}\}$ from B as follows: for each i, let $X_i^* = X_i$; $r_i^*(x_i) = r_i(x_i) + (1-\rho)d_i$; $Q_i^* = Q_i$; $c_i^* = c_i + d_i$; and, finally, $d_i^* = 0$. We will show that the bandits B and $B*$ are equivalent.

Indeed, this is almost immediate. Viewed as dynamic programming problems, B and B^* have the same state and action spaces, hence the same strategy spaces. Moreover, since the transition probabilities also coincide, a given strategy induces the same distribution on infinite histories in either problem. Thus, it suffices to show that a given history yields the same reward in either problem, or more specifically, that the net reward from a given arm over the periods of its contiguous use is the same in either problem. To see that this is true, observe that, in essence, the only difference between Bandits B and B^* is that in bandit B, a cost of d_i is paid every time a switch away from arm i occurs, whereas in Bandit B^*, d_i is paid "in advance" when the switch to arm i occurs, but an additional reward of $(1-\rho)d_i$ is received every period arm i is in use. If, therefore, arm i is used for t contiguous periods before a switch to another arm occurs, the present value of the total switching cost paid in the Bandit B will be $c_i + \rho^t d_i$. In Bandit B^*, the total cost will be $c_i + d_i$; but an additional reward of $(1-\rho)d_i$ is received in each of t periods, so that the *net* cost is $c_i + d_i(1 - \sum_{s=0}^{t-1} \rho^s(1-\rho)) = c_i + d_i\rho^t$, which is exactly the same as in the Bandit B. Thus, the sums of discounted rewards in the two problems differ only by the cost of switching away from the initial arm, and the equivalence of B and B^* follows.

We now proceed to the second part of our proof, which consists essentially of formalizing the arguments given in the Introduction. We suppose from now on that there are only costs of switching "to," and that an optimal index, denoted λ, does exist in this case. Since all the Bandits we shall consider from this point on involve arms with similar characteristics, we simplify notation as follows. For any $a, b \in \Re$, $x \in [0,1]$ and $c \geq 0$, let $[x\delta_a + (1-x)\delta_b, c]$ denote an arm with state space [0,1] and initial state x; reward function $r(x) = xa + (1-x)b$; transition probabilities $Q(1|x) = 1 - Q(0|x) = x$ for all $x \in [0,1]$; and switching cost c. In particular, $[\delta_a, c]$ will denote an arm with switching cost c, that pays a reward of a in each period with certainty.

In this notation, $\lambda([x\delta_a + (1-x)\delta_b, c]; s)$ will denote the value of the optimal index on the arm $[x\delta_a + (1-x)\delta_b, c]$ at the state $s \in \{0,1\}$. Recall that $s = 1$ denotes that the arm is currently in use.

Claim 1 *$\lambda([\delta_a, c]; 1)$ is independent of c, while $\lambda([\delta_a, c]; 0)$ is strictly decreasing in c.*

Proof: Consider the situation where there are only two arms. Suppose arm i (= 1,2) pays a_i for certain, and that the cost of switching to arm i is c_i. Finally, suppose that the decision-maker is currently on the first arm.

It is trivial to see that the uniquely optimal strategy is to stay with arm 1 forever if $a_1 > a_2 - c_2(1-\rho)$; that picking either arm initially and staying with it forever is optimal if $a_1 = a_2 - c_2(1-\rho)$; and that switching to arm 2 and staying there forever is uniquely optimal if $a_1 < a_2 - c_2(1-\rho)$. It follows immediately that the only indices λ that can be invariably optimal are those that are strict monotone transformations of the index $\hat{\lambda}$, defined by

$$\begin{aligned} \hat{\lambda}([\delta_a, c]; 1) &= a, \text{ and} && (3.1) \\ \hat{\lambda}([\delta_a, c]; 0) &= a - c(1-\rho), && (3.2) \end{aligned}$$

establishing the claim. ◊

On the other hand, the following two claims together establish that $\lambda([\delta_a, c]; 1)$ must be strictly increasing in c:

Claim 2 *For $a > b$, $\lambda([x\delta_a + (1-x)\delta_b, c]; 0)$ is increasing in x.*

Proof: Let $x_1, x_2 \in [0,1]$, with $x_1 > x_2$. Pick $\alpha \in \Re$ so that

$$[ax_1 + b(1-x_1) - c(1-\rho)] > \alpha > [ax_2 + b(1-x_2) - c(1-\rho)]. \quad (3.3)$$

Now consider a two-armed Bandit in which the first arm is given by $[\delta_\alpha, c^*]$ for some c^*, and the second arm by $[x_i\delta_a + (1 - x_i)\delta_b, c]$. Suppose further, that the decision maker is currently on the first arm. When c^* is sufficiently large, a simple calculation shows that it is uniquely optimal to switch to the second arm and stay there forever if $i = 1$; and to continue indefinitely with the first arm if $i = 2$. Since λ is optimal by hypothesis, we must have

$$\lambda([x_1\delta_a + (1 - x_1)\delta_b, c]; 0) \; > \; \lambda([\delta_\alpha, c^*]; 1) \; > \; \lambda([x_2\delta_a + (1 - x_2)\delta_b, c]; 0), \tag{3.4}$$

establishing Claim 2. ◇

Claim 3 *For any $a \in \Re$, $\lambda([\delta_a, c]; 1]$ must be increasing in c.*

Proof: We derive this as a consequence of Claim 2. Let $c_1 > c_2$. Pick $\alpha, \beta \in \Re$ such that

(i) $\alpha \; > \; a$, and

(ii) $\beta \; < \; [a - (1 - \rho)c_1] \; < \; [a - (1 - \rho)c_2)]$.

Pick $x_1 \in [0, 1]$ such that in the two-armed Bandit problem where the first arm is given by $[\delta_a, c_1]$, and the second arm is given by $[x_1\delta_\alpha + (1 - x_1)\delta_\beta, 0]$, either arm is an optimal initial choice. Define x_2 analogously. Some calculation shows that the desired values are[7]

$$x_i = \frac{(1 - \rho)(a - \beta + \rho c_i)}{[\alpha - (1 - \rho)\beta - \rho a + \rho(1 - \rho)c_i]}. \tag{3.5}$$

Note that $1 > x_1 > x_2 > 0$. It follows, by the presumed optimality of λ, that

$$\lambda([\delta_a, c_i]; 1) = \lambda([x_i\delta_\alpha + (1 - x_i)\delta\beta, 0]; 0). \tag{3.6}$$

Claim 3 is now an immediate consequence of Claim 2, since $x_1 > x_2$. ◇

Claims 1 and 3 are in obvious contradiction, establishing the impossibility of consistently defining an optimal index λ on the domain of all Bandit problems with switching costs.

On the other hand, claims 1-3 do not rule out the possibility that an optimal index λ may still exist in the restricted subset of Bandit problems where costs of switching are fixed *a priori* at some constant level, and, in particular, are not allowed to vary across arms.[8]

[7] To check that these calculated values are correct, note that if the decision-maker sticks to the first arm forever, then the total discounted reward is $a/(1 - \rho)$. Suppose, on the other hand, that the decision-maker switches to the second arm at the outset. By choice of α and β, it is optimal to switch back to arm 1 if, and only if, the state on the second arm moves to 0, *i.e.*, the second arm yields a continuation reward of β in every period. Thus, the total discounted reward in this case is $x_i\alpha/(1 - \rho) + (1 - x_i)(\beta + \rho a/(1 - \rho) - \rho c_i)$, which by choice of x_i is simply $a/(1 - \rho)$.

[8] The possibility of existence of an optimal index in this case was raised by a referee.

Nonetheless, it is easily shown that existence on this limited subset of the overall domain is also impossible.

For, suppose some (non-zero) switching cost c is fixed and given as the cost of switching to any arm. By claim 1, we may assume (without any loss of generality) that the values of the index $\hat{\lambda}$ defined there are in fact the values taken on by our hypothetical optimal index λ when the arms are of the form $[\delta_a, c]$. Consider, first, a two-armed Bandit where the first arm is of type $[\delta_a, c]$ and the second is of type $[x\delta_1 + (1-x)\delta_0, c]$.

Suppose that the decision-maker is on the second arm. Simple calculation reveals that when

$$a = \mu(x) := \frac{x}{(1-\rho(1-x))} + c(1-\rho), \tag{3.7}$$

either arm is an optimal initial selection. Thus, we must have $\lambda([x\delta_1 + (1-x)\delta_0, c]; 1) = \lambda([\delta_{\mu(x)}, c]; 0)$, and so, by the presumed optimality of λ,

$$\lambda([x\delta_1 + (1-x)\delta_0, c]; 1) = \frac{x}{1-\rho(1-x)}. \tag{3.8}$$

Similarly, by supposing that the decision-maker is initially on the first arm, and calculating optimal continuations, we obtain the following (the details are omitted): if $x \in [2c(1-\rho), 1]$, then

$$\lambda([x\delta_1 + (1-x)\delta_0, c]; 0) = \frac{x - c(1-\rho)(1+\rho(1-x))}{1-\rho(1-x)}, \tag{3.9}$$

while, if $x \in [0, 2c(1-\rho))$, then

$$\lambda(x, 0) = x - c(1-\rho). \tag{3.10}$$

We will use equations (3.8)-(3.10) to derive a contradiction. To this end, consider a two-armed Bandit, where the first arm is specified by $[x\delta_1 + (1-x)\delta_0, c]$, and the second arm by $[y\delta_1 + (1-y)\delta_0, c]$. Assume that the decision-maker is currently on the first arm. Simple calculation reveals that the uniquely optimal strategy is to pick the first arm (and then to switch to the second iff the state of the first moves to 0) when the following conditions are met:

$$y > c(1-\rho) > x \tag{3.11}$$

$$\frac{x}{(1-\rho)(1-\rho(1-x))} > \frac{y}{(1-\rho)} - c \tag{3.12}$$

Thus, we must have $\lambda(x, 1) > \lambda(y, 0)$ whenever these two conditions are met.

But it is easy to construct examples where these inequalities are inconsistent with the definition of λ from (3.8)-(3.10). For instance , let $\rho = c = 1/2$, $x = 3/17$, and $y = 27/50$. Then, $y > 1/4 = c(1-\rho) > x$, so (3.11) is met. (Note that, in fact, $y > 1/2 = 2c(1-\rho)$.) Moreover, the LHS of (3.12) reduces to $4x/(1+x) = 3/5$, while the RHS of (3.12) is $29/50$, so (3.12) is also met. However, $\lambda(x,1) = 3/5$, while $\lambda(y,0) = (9y-3)/(2+2y) = 93/154 > 3/5$. $\diamondsuit$

References

Agrawal, R.; M.V. Hegde, and D. Teneketzis (1988): "Asymptotically Efficient Adaptive Allocation Rules for the Multiarmed Bandit Problem with Switching Costs," *IEEE Transactions on Optimal Control* 33(10), 899-906.

Banks, J.S. and R.K. Sundaram (1992): "Denumerable-Armed Bandits,"*Econometrica* 60(5), 1071-1096.

Berry, D.A. and B. Fristedt (1985): *Bandit Problems: Sequential Allocation of Experiments,* London: Chapman and Hall.

Feldman, M. and M. Spagat (1993): Optimal Learning with Costly Adjustment, Working Paper No. 93-10, Department of Economics, Brown University.

Gittins, J.C. and D.M. Jones (1974): "A Dynamic Allocation Index for the Sequential Allocation of Experiments," in *Progress in Statistics* (J. Gani, et al, Eds.), Amsterdam: North Holland, pp.241-266.

Kolonko, M. and H. Benzing (1983): "The Sequential Design of Bernoulli Experiments including Switching Costs," unpublished manuscript; cited in Berry and Fristedt (1985).

Lai, T.L. and H. Robbins (1985): "Asymptotically Efficient Adaptive Allocation Rules," *Advances in Applied Mathematics* 6, 4-22.

Mortensen, D. (1985): "Job Search and Labor Market Analysis," in *Handbook of Labor Economics,* Vol.II (O. Ashenfelter and J. Layard, Eds.), New York: North Holland, pp. 849-919.

Weitzman, M. (1979): "Optimal Search for the Best Alternative," *Econometrica* 47, 641-654.

Whittle, P. (1982) *Optimization over Time: Dynamic Programming and Stochastic Control,* Vol I, New York: Wiley.