TO APPEAR IN
Advances in Applied Probability
(1980) 12 No. 3.

TWO-ARMED BANDITS WITH A GOAL;
I: ONE ARM KNOWN

by

Donald A. Berry*

and

Bert Fristedt**

University of Minnesota

Technical Report No. 344

August, 1979

Corrections: October 1979

* This author's research sponsored by the NSF under Grant No. 78-02694.

** This author's research sponsored by the NSF under Grant No. 74-05786 A02.

Two-armed bandits with a goal;
I: One arm known

by

Donald A. Berry and Bert Fristedt

ABSTRACT

One of two random variables, X and Y, can be selected at each of a possibly infinite number of stages. Depending on the outcome, one's fortune is either increased or decreased by one. The probability of increase may not be known for either X or Y. The objective is to increase one's fortune to G before it decreases to g, for some integral g and G; either may be infinite.

In the current part of the paper, the distribution of X is unknown and that of Y is known. We characterize the situations in which optimal strategies exist and, for certain kinds of information concerning X and Y, we characterize optimal sequential strategies for choosing to observe X and Y.

In Part II (Berry and Fristedt 1980), it is known that either X or Y has probability α of increasing the current fortune by one and the other has probability β of increasing the fortune by one, where α and β are known, but which goes with X is not known.

Key words and phrases: Achieving a goal, two-armed bandits, how to gamble if you must, gambler's ruin, sequential decisions, Bayesian decision making, sequential medical treatments, stochastic control, optimal dynamic designs.

Acknowledgement: This problem and its applicability to certain medical treatment problems were suggested to us by Professor William D. Sudderth.

1. Introduction. A gambler has an integral fortune k and is allowed to make a sequence of unit bets. He is to continue betting until his fortune becomes either of the integral goals G or g, with $g < k < G$. Each bet is made by choosing to observe one of two random variables, X and Y, where

$$P(X = 1) = 1-P(X = -1) = \rho,$$

$$P(Y = 1) = 1-P(Y = -1) = \lambda,$$

and ρ and λ may be unknown, and so we regard them as random variables.

After observing X or Y he may choose to observe the same variable for the next bet or switch and observe the other variable. Which variable he chooses at the $(n + 1)$st stage may depend on the first n choices and observations. Using traditional bandit problem terminology, an observation of X is "a pull of the right arm $\mathfrak{R}$" and of Y is "a pull of the left arm $\mathfrak{L}$."

DEFINITION 1.1. A strategy, sometimes denoted by τ, is a function that assigns the symbol $\mathfrak{R}$ or $\mathfrak{L}$ to each finite history of bets and resulting observations.

The variable assigned by a strategy denotes the bet to be made next when following that strategy. Let Z_n denote the result of the nth bet and let $S_n = \sum_1^n Z_i$. The gambler's fortune at the $(n + 1)$st stage is

$$k_n = k + S_n . \tag{1.1}$$

If $G < \infty$, then the objective is to reach G before g. Any strategy that maximizes the probability that there exists an n such that

$$k_1 > g, \ldots, k_{n-1} > g, k_n = G$$

will be called optimal. If $G = \infty$ then any strategy that maximizes the

probability that $k_n > g$ for all n and $k_n \to \infty$ will be called optimal. Any strategy that comes within ε of the maximum, or supremum in case the maximum does not exist, will be called ε-optimal. Evidently, if $k_n = g$ or $k_n = G$ for some n, the same is true for all larger n.

Let X_i and Y_j denote, respectively, the observations on the ith bet on $\mathcal{R}$ and the jth bet on $\mathcal{L}$. For convenience we assume that the sequences $(X_1, X_2, \ldots)$ and $(Y_1, Y_2, \ldots)$ are non-terminating whether or not infinitely many bets are made on each variable.

The two sequences $(X_1, X_2, \ldots)$ and $(Y_1, Y_2, \ldots)$ are assumed to be independent of each other. Also, given the random variables ρ and λ, $(X_1, X_2, \ldots)$ and $(Y_1, Y_2, \ldots)$ are sequences of independent random variables. Therefore, the unconditional finite-dimensional distributions of $(X_1, X_2, \ldots)$ and $(Y_1, Y_2, \ldots)$ are invariant under permutations of the subscripts; that is, each is a sequence of exchangeable random variables. In the present part, Part I, of this paper, λ will be assumed known with $0 < \lambda < 1$. Under this assumption, $(Y_1, Y_2, \ldots)$ is a sequence of independent random variables. The distribution R of ρ will be arbitrary. In Part II of this paper (Berry and Fristedt 1980), (ρ,λ) is known to be either (α,β) or (β,α), where α and β are arbitrary but known.

When possible, the notation used in this paper is consistent with that of Berry and Fristedt (1979), which deals with classical bandit problems. While many of the aspects of the current problem are different from those of classical bandit problems, some are the same. In particular, in both problems the desire to pull $\mathcal{R}$ or $\mathcal{L}$ depends on the probability of immediate success and on the possibility of gaining information about the variables X and Y which will aid future decisions.

In both problems these criteria may be conflicting, that is, suggesting different strategies. While it is not possible to consider these two criteria separately in either problem, the interrelationship is more complicated in the current problem.

For fortune $k \in (g,G)$, distribution R on $[0,1]$, and $\lambda \in (0,1)$, let $U(k; R,\lambda)$ denote the maximum over the set of strategies, or supremum in case the maximum does not exist, of the probability of approaching fortune G (thereby reaching G when G is finite) before approaching g. The maximal probability of approaching G is at least that of the strategy which observes the same variable at each stage, and is no larger than if ρ is also known at the outset. That is,

$$(1.2) \qquad u(k,\lambda) \vee Eu(k,\rho) \leq U(k; R,\lambda) \leq E[u(k,\rho) \vee u(k,\lambda)],$$

where u is the probability of success in the gambler's ruin problem. Expressions for this latter probability are well-known and are given below.

Goals	$u(k,p)$	Restrictions
$-\infty < g < G < \infty$	$\dfrac{1 - (\frac{1-p}{p})^{k-g}}{1 - (\frac{1-p}{p})^{G-g}}$	$p \neq 1/2$
	$\dfrac{k-g}{G-g}$	$p = 1/2$
$-\infty < g < G = \infty$	0	$p \leq 1/2$
	$1 - (\frac{1-p}{p})^{k-g}$	$p > 1/2$
$-\infty = g < G < \infty$	$(\frac{p}{1-p})^{G-k}$	$p < 1/2$
	1	$p \geq 1/2$

We cannot find optimal strategies for the problem as stated in its full generality. We confine our attention mainly to special cases of (R,λ). As we consider different cases, or even different instances within a single case, the reader will notice major qualitative differences in the optimal strategies. Some optimal strategies are counterintuitive. For example, it may be uniquely optimal to bet on Y (i.e., to pull $\mathfrak{L}$) even though it is practically certain that $\rho > \lambda$. The next example is especially easy and gives one instance of this phenomenon.

EXAMPLE 1.1. Suppose λ is .5 and

$$1-R(\{0\}) = R(\{1\}) = r > 0.$$

Suppose both goals g and G are finite so we can take $g = 0$. We will abbreviate $U(k; R,\lambda)$, the maximal probability of reaching G, to $U(k,r)$. Again, there is no information to be gained by pulling $\mathfrak{L}$ since λ is known; that is, r is not affected by pulling $\mathfrak{L}$. On the other hand, complete information is gained with probability one by a single pull of $\mathfrak{R}$. It is reasonable, therefore, to expect that if r is large enough it is optimal to pull $\mathfrak{R}$ first and then proceed with complete knowledge. Actually, the optimal strategy is quite the opposite. If $k = 2, \ldots, G-1$, it is optimal (uniquely unless $r = 1$) to pull $\mathfrak{L}$ for all r, and it is optimal to pull $\mathfrak{R}$ if $k = 1$, for $r \geq 1/G$ and $\mathfrak{L}$ for $r < 1/G$. Of course, if $\mathfrak{R}$ is ever pulled, then it is optimal (though not necessarily uniquely) to pull $\mathfrak{R}$ forever (until reaching G) if it succeeds. Let τ_0 denote the strategy claimed to be optimal in these statements.

Let $U_0(k,r)$ denote the probability of reaching G using τ_0 when the current fortune is k. Then

$$U_0(k,r) = \frac{k-1}{G-1} + \frac{G-k}{G-1}\, r \,, \qquad \text{if} \quad r \geq 1/G$$

$$= k/G, \qquad \text{if} \quad r < 1/G.$$

We show that $U = U_0$ by showing that U_0 is excessive (Corollary 2.1 of Section 2, or Dubins and Savage 1976, Theorem 2.12.1); that is, the expected value of U_0 under either initial choice is not greater than U_0. It follows (Theorem 2.1 of Section 2) that the initial choice in question is optimal when and only when this expectation actually equals U_0. We prove that U_0 is excessive by considering two cases that are determined by the arm pulled initially when following τ_0.

According to τ_0, $\mathfrak{L}$ should be pulled when $r < 1/G$ or both $k > 1$ and $r < 1$. For such a fortune k and probability r that $\rho = 1$, the expected value of U_0 under a pull of $\mathfrak{L}$ is simply $U_0(k,r)$ from the definition of τ_0. The expected value of U_0 under a single pull of $\mathfrak{A}$ equals

$$rU_0(k+1,\, 1) + (1-r)U_0(k-1,\, 0)$$

$$= r + (1-r)(k-1)/G \,.$$

Subtracting this from $U_0(k,r)$ gives

$$\begin{cases} (1-r)\,\dfrac{k-1}{G(G-1)} & \text{if} \quad r \geq 1/G, \\ \dfrac{1}{G} - r + \dfrac{r}{G}(k-1) & \text{if} \quad r < 1/G \,; \end{cases}$$

both of which are ≥ 0 (in fact > 0 in the case under consideration) as desired.

In the remaining cases, $\mathfrak{A}$ is pulled when following τ_0; accordingly, we suppose that either $k = 1$ and $r \geq 1/G$ or $k > 1$ and $r = 1$.

For such a fortune and probability that $\rho = 1$ the expected value of U_0 under a pull of $\mathfrak{R}$ is $U_0(k,r)$ from the definition of τ_0. The expected value of U_0 under a single pull of $\mathfrak{L}$ equals

$$\frac{1}{2} U_0(2,r) = \frac{1}{2(G-1)} + \frac{G-2}{2(G-1)} r$$

if $k = 1$, and it equals 1 if $k > 1$ and $r = 1$. Subtracting this from $U_0(k,r)$ gives $(rG-1)/(2G-2)$ if $k = 1$ and 0 if $k > 1$ and $r = 1$; both of which are ≥ 0 for the cases at hand (in fact > 0 unless $k = 1$ and $r = 1/G$ or $k > 1$ and $r = 1$) as desired.

The claim that τ_0 is optimal is established. Further, the only deviations from τ_0 that are optimal occur when equality holds in the excessivity calculations. So, pulling $\mathfrak{L}$ and pulling $\mathfrak{R}$ are both optimal when $k = 1$ and $r = 1/G$ or when $k > 1$ and $r = 1$; otherwise, the arm indicated by τ_0 is uniquely optimal. ∎

Two-armed bandits with a goal are motivated by the following medical treatment decision problem. Two treatments are available for use on a patient with a known level of a disease; one of the treatments may be "no treatment". Between applications the patient's condition either improves a little or worsens a little, independently of previous treatments and responses. Assume that levels can be quantified so that each new level is the old level plus or minus one. Furthermore, there is a level g, which is below the patient's current level, at which he can no longer be cured, and a higher level G, at which the patient no longer has the disease. Then the problem we have posed may be applicable. Perhaps the least realistic assumption we make is that it is only eventual cure or not that matters, and not the length of time that the patient is in a diseased state. Still, there are diseases

for which standard medical practice is similar to an optimal strategy in our problem for certain kinds of information. In particular, there are diseases in which treatments that are risky, having unknown effectiveness, are used only when the disease is in very advanced stages. As we saw in Example 1.1, some optimal strategies have this property. In that example, the right arm $\mathcal{Q}$ is a "risky treatment." On the other hand, there are diseases in which a risky treatment is only used when the patient is relatively healthy. We shall encounter situations in Section 5 in which optimal strategies have this property.

As far as we know, the problem considered here has not been studied previously, but it is a variation of a number of problems. Articles dealing with classical bandit problems--in which the expected number of successes (possibly discounted) is to be maximized--include those by Berry (1972) and Berry and Fristedt (1979). Related work in the theory of gambling includes the fundamental book of Dubins and Savage (1976), and (Berry, Heath, and Sudderth 1974). In a problem used as motivation by Dubins and Savage (1976), bets with known win probability are made with the objective of reaching a goal before going broke. The problem is to decide what sequence of stakes to use, depending on the sequence of fortunes. Berry, Heath, and Sudderth (1974) consider a generalization in which the win probability is itself a random variable. In the current problem there is a second sequence of bets available, but now only unit bets are allowed.

In Section 2 we introduce some necessary notation and terminology. In Section 3 we develop some general theory, considering the question of existence and, to a certain extent, the nature of optimal strategies.

In Section 4 we find optimal strategies when ρ is known to be 1 if it is greater than λ. In Section 5 we find optimal strategies when ρ is known to be 0 if it is less than λ. Finally, in Section 6 we consider the case in which $G = \infty$ and R is a two-point distribution, the two points being symmetric about .5.

2. Preliminaries. The discussion in the previous section focusses on distinct initial fortunes k and distributions R of ρ without regard to their interconnections. In this section we knit these individual problems into one. Our purposes will be well-served by the following definition.

DEFINITION 2.1. For fixed λ, g, and G, a scheme Ψ is a function $\Psi(k,R)$ which assigns either of the symbols $\mathcal{R}$ or $\mathcal{L}$ to each $k \in (g,G)$ and distribution R of ρ.

To connect the concepts of scheme and strategy, as defined in Definition 1.1, we need some further notation.

Just as we defined k_n in (1.1) to be the fortune at stage $n+1$, we let R_n be the conditional distribution of ρ at stage $n+1$. Specifically, if s successes and f failures have been obtained on $\mathcal{R}$ in the first n stages, then $R_n = \sigma^s\varphi^f R$, where each $\sigma^s\varphi^f R$ is absolutely continuous with respect to R and

$$\frac{d\sigma^s\varphi^f R}{dR}(x) = \frac{x^s(1-x)^f}{\int u^s(1-u)^f R(du)} .$$

(In case the support of R is contained in $\{0,1\}$, written supp $R \subset \{0,1\}$, some values of (s,f) are not possible.) Stage by stage considerations yield the characterization: $R_0 = R$ and

$$\begin{aligned} R_n &= \sigma R_{n-1} \quad \text{if } \mathcal{R} \text{ is pulled at stage } n \text{ and } Z_n = 1 \\ &= \varphi R_{n-1} \quad \text{if } \mathcal{R} \text{ is pulled at stage } n \text{ and } Z_n = -1 \\ &= R_{n-1} \quad \text{if } \mathcal{L} \text{ is pulled at stage } n. \end{aligned}$$

DEFINITION 2.2. The strategies induced by the scheme Ψ are those that require a pull of $\Psi(k_n,R_n)$ at the $(n+1)$st stage, for $n = 0,1,\ldots$.

In Example 1.1 we restricted attention to strategies induced by a scheme, that is, strategies that specify the arm to be pulled at each stage depending only on the current fortune and current distribution of ρ, and not otherwise on the history of observations. While there may be optimal strategies outside this class, we shall see in Theorems 3.1 and 3.2 that there are always strategies inside the class that perform at least as well as those outside.

For the consideration of schemes we shall require a notation that reflects the simultaneous consideration of a variety of initial fortunes k and distributions R. For a strategy τ, let $E_\tau^{(k,R)}$ and $P_\tau^{(k,R)}$ (or $P_\psi^{(k,R)}$ when τ is induced by the scheme Ψ) denote the expectation and probability starting at (k,R). For instance,

$$U(k;R,\lambda) = \sup_\tau P_\tau^{(k,R)}(k_n \to G).$$

For situations in which expectation depends on the strategy τ only through the arm pulled initially, we shall use the name of that arm as the subscript. In this notation, we can express what is called the "functional equation" in sequential decision theory:

$$U(k;R,\lambda) = E_R^{(k,R)} U(k_1;R_1,\lambda) \vee E_{\mathfrak{L}}^{(k,R)} U(k_1;R_1,\lambda) \quad . \tag{2.1}$$

When convenient, we shall drop any of the superscripts or subscripts from the E and P notations. For instance, the probability of a second success with R given one success can be written in a variety of ways:

$$E^{\sigma R}(\rho) = \int x\ (\sigma R)(dx) = \frac{\int x^2\ R(dx)}{\int x\ R(dx)}$$

$$= \frac{E^R(\rho^2)}{E^R(\rho)} = \frac{E\rho^2}{E\rho} \quad .$$

An arm G, where G equals R or L, is called conserving at (k,R) if

$$E_G^{(k,R)} U(k_1;R_1,\lambda) = U(k;R,\lambda);$$

that is, if the maximum in (2.1) is taken on at arm G. Clearly, it is never optimal to pull an arm that is not conserving. However, as we shall see, it may not be optimal to always pull a conserving arm.

To carry over the concepts of conserving and optimal to schemes, we first need the following definition.

DEFINITION 2.3. Let $\underset{\sim}{R}$ be a family of distributions of ρ. We say that $\underset{\sim}{R}$ is closed under sampling whenever $R \in \underset{\sim}{R}$ implies $\sigma^s\varphi^f R \in \underset{\sim}{R}$ for all s and f that are possible under R. If $R((0,1)) > 0$, then all pairs of nonnegative integers (s,f) are possible.

DEFINITION 2.4. A scheme Ψ is called conserving for $\underset{\sim}{R}$, a class of distributions that is closed under sampling, if $\Psi(k,R)$ is conserving at (k,R) for each $k \in (g,G)$ and $R \in \underset{\sim}{R}$.

DEFINITION 2.5. Let $\varepsilon \geq 0$. A scheme Ψ for which the induced strategy for each initial $k \in (g,G)$ and each initial $R \in \underset{\sim}{R}$, a class of distributions closed under sampling, is ε-optimal is called an ε-optimal scheme for $\underset{\sim}{R}$. A 0-optimal scheme for $\underset{\sim}{R}$ is called an optimal scheme for $\underset{\sim}{R}$.

An arm G is called optimal at (k,R) if $G = \Psi(k,R)$ for some scheme Ψ that is optimal for some class containing R.

While we give an elementary proof of the following theorem, it can be viewed as a special case of Theorem 14 of Sudderth (1972), which in turn is a generalization of Theorem 3.5.1 of Dubins and Savage (1976). In case $G < \infty$, our theorem is also a special case of the Dubins and Savage result.

THEOREM 2.1. A conserving scheme Ψ for a class $\underset{\sim}{R}$, closed under sampling, is optimal for $\underset{\sim}{R}$ if, for each $k \in (g,G)$ and $R \in \underset{\sim}{R}$,

$$P_{\Psi}^{(k,R)}(U(k_n;R_n,\lambda) \to 0 \mid k_n \to G) = 1 \quad . \tag{2.2}$$

PROOF. Suppose Ψ is conserving for $\underset{\sim}{R}$. By induction,

$$U(k;R,\lambda) = E_{\Psi}^{(k,R)} U(k_n;R_n,\lambda) \quad .$$

Let $n \to \infty$ to obtain

$$U(k;R,\lambda) \leq P_{\Psi}^{(k,R)}(k_n \to G) + P_{\Psi}^{(k,R)}(k_n \not\to G) \operatorname*{ess\,sup}_{[k_n \not\to G]} \limsup_{n \to \infty} U(k_n;R_n,\lambda),$$

where $\operatorname*{ess\,sup}_{[k_n \not\to G]}$ denotes the essential supremum for the measure $P_{\Psi}^{(k,R)}$ over the event $[k_n \not\to G]$. ∎

We now make precise the concept of excessivity used in Example 1.1.

DEFINITION 2.6. Let $W(k,R)$ be a real-valued function defined for $g-1 < k < G+1$ and $R \in \underset{\sim}{R}$, a class closed under sampling. The function W is <u>excessive for</u> $\underset{\sim}{R}$ if, for each arm G and each $(k,R) \in (g-1,G+1) \times \underset{\sim}{R}$,

$$W(k,R) \geq E_{G}^{(k,R)} W(k_1,R_1) \quad .$$

The following theorem enables one to check that a particular scheme is optimal. While we give an elementary proof of the theorem, it can be viewed as a special case of Theorem 1 of Dubins and Sudderth (1977), which in turn is a generalization of Theorem 2.12.1 of Dubins and Savage (1976).

THEOREM 2.2. Suppose that, for some scheme Ψ, the function

$$(k,R) \mapsto P_{\Psi}^{(k,R)}(k_n \to G)$$

is excessive for some class $\underset{\sim}{R}$ that is closed under sampling. If, in addition,

(2.3) $\quad P_{\psi}^{(k,R)}(k_n \to G) \to 1$ as $k \to G$, uniformly for $R \in \underset{\sim}{R}$,

then Ψ is optimal for $\underset{\sim}{R}$.

PROOF. Fix $(k,R) \in (g,G) \times \underset{\sim}{R}$ and let τ be an arbitrary strategy for starting at (k,R). Let τ_m denote the strategy: use τ at stages $1, \ldots, m$ and thereafter use the strategy induced by Ψ. By induction and the assumed excessivity,

$$P_{\tau_m}^{(k,R)}(k_n \to G) \leq P_{\psi}^{(k,R)}(k_n \to G) \ . \tag{2.4}$$

On the other hand,

$$\begin{aligned} &\liminf_{m \to \infty} P_{\tau_m}^{(k,R)}(k_n \to G) \\ &= \liminf_{m \to \infty} P_{\tau}^{(k,R)}(P_{\psi}^{(k_m,R_m)}(k_n \to G) \,|\, (k_m,R_m)) \qquad (2.5) \\ &\geq P_{\tau}^{(k,R)}(k_m \to G) \inf_{[k_m \to G]} \liminf_{m \to \infty} P_{\psi}^{(k_m,R_m)}(k_n \to \infty). \end{aligned}$$

By (2.3) the right side of (2.5) equals $P_{\tau}(k_m \to G)$ which, by (2.4) and (2.5) is less than or equal to $P_{\psi}(k_n \to G)$. ∎

For $G < \infty$, Theorem 2.2 is a special case of Theorem 2.12.1 of Dubins and Savage (1976), and since (2.3) is trivially satisfied, simplifies as follows.

COROLLARY 2.1. Suppose $G < \infty$. If $(k,R) \mapsto P_{\psi}^{(k,R)}(k_n \to G)$ is excessive for $\underset{\sim}{R}$, then Ψ is optimal for $\underset{\sim}{R}$.

3. Optimal Schemes; General Theory. In this section we prove two theorems dealing with the existence of optimal schemes and one giving a qualitative property of optimal schemes.

THEOREM 3.1. There is an optimal scheme for the class of all distributions of ρ if $g > -\infty$ or $G = \infty$ or $\lambda \geq .5$.

PROOF. (i) Suppose both $G = \infty$ and $\lambda \leq .5$. Then the only chance of approaching ∞ is if $\rho > .5$; therefore, it is optimal to pull $\mathfrak{R}$ at each stage.

(ii) Suppose $g = -\infty$, $\lambda \geq .5$, and either $G < \infty$ or $\lambda \neq .5$. Then it is clearly optimal to pull $\mathfrak{L}$ at each stage.

(iii) Finally, suppose $g > -\infty$ and either $G < \infty$ or $\lambda > .5$. Since we are assuming $\lambda < 1$ throughout (thereby ruling out the possibility of oscillating forever), the event $[k_n \neq G]$ differs by a null event from the event $[k_n \to g]$ (that is, the event $[k_n = g$ for some $n]$). Thus, since $g > -\infty$, condition (2.2) is satisfied for every scheme Ψ. It follows from Theorem 2.1 that every conserving scheme is optimal. Since there exists at least one conserving scheme, there is at least one optimal scheme. ∎

The only case not covered in Theorem 3.1 will be considered in Theorem 3.2. For one circumstance arising in that theorem we shall need the following lemma.

LEMMA 3.1. Suppose that $\lambda < \lambda + \delta < .5 - \delta$ and $g = -\infty$. Then there exists $\gamma > 0$ such that

$$\frac{R([0,\lambda+\delta))}{R([\lambda+\delta,.5-\delta])} < \gamma$$

implies $E_{\mathfrak{L}}^{(k,R)}(U(k_1;R_1,\lambda) < E_{\mathfrak{R}}^{(k,R)}(U(k_1;R_1,\lambda)$ for every k.

PROOF. From (1.2), we obtain

$$E_{\mathfrak{L}}^{(k,R)}U(k_1;R_1,\lambda) \le \lambda E^R[u(k+1,\rho)\vee u(k+1,\lambda)]+(1-\lambda)E^R[u(k-1,\rho)\vee u(k-1,\lambda)]$$

and

$$E_{\mathfrak{R}}^{(k,R)}U(k_1;R_1,\lambda) \ge E^R u(k,\rho)$$

$$= E[\rho u(k+1,\rho)+(1-\rho)u(k-1,\rho)].$$

Therefore

$$E_{\mathfrak{R}}^{(k,R)}U(k_1;R_1,\lambda) - E_{\mathfrak{L}}^{(k,R)}U(k_1;R_1,\lambda)$$

$$\ge E^R\Big[\{1_{\sim[0,\lambda+\delta)}(\rho)+1_{\sim[\lambda+\delta,.5-\delta]}(\rho)+1_{\sim(.5-\delta,1]}(\rho)\}\{\rho u(k+1,\rho)$$

$$+ (1-\rho)u(k-1,\rho)-\lambda[u(k+1,\rho)\vee u(k+1,\lambda)]-(1-\lambda)[u(k-1,\rho)\vee u(k-1,\lambda)]\}$$

$$= I_1 + I_2 + I_3,$$

where I_1, I_2, and I_3 denote the terms involving the respective indicator functions.

Clearly,

$$I_1 \ge E^R\Big[1_{\sim[0,\lambda+\delta)}(\rho)\{-\lambda u(k+1,\lambda+\delta)-(1-\lambda)u(k-1,\lambda+\delta)\}\Big] \ge -R([0,\lambda+\delta))u(k+1,\lambda+\delta).$$

We also have

$$I_2 = E^R\Big[1_{\sim[\lambda+\delta,.5-\delta]}(\rho)(\rho-\lambda)\{u(k+1,\rho)-u(k-1,\rho)\}\Big]$$

$$\ge \delta R([\lambda+\delta,.5-\delta])\inf\{u(k+1,p)-u(k-1,p):\lambda+\delta \le p \le .5-\delta\}$$

and, by a similar calculation, $I_3 \ge 0$. Since, for $p \in [\lambda+\delta,.5-\delta]$,

$$u(k+1,p)-u(k-1,p) = u(k+1,p)-\left(\frac{p}{1-p}\right)^2 u(k+1,p)$$

$$\ge u(k+1,p)\left[1-\left(\frac{.5-\delta}{.5+\delta}\right)^2\right] \ge u(k+1,\lambda+\delta)\left[1-\left(\frac{.5-\delta}{.5+\delta}\right)^2\right],$$

$I_1 + I_2 + I_3 > 0$ for $R([0,\lambda+\delta])/R([\lambda+\delta,.5-\delta]) < \gamma$, where $\gamma = \delta[1-(.5-\delta)^2/(.5+\delta)^2]$. ∎

THEOREM 3.2. Suppose $g = -\infty$, $G < \infty$, and $\lambda < .5$. For each $\varepsilon > 0$ there is an ε-optimal scheme for the class of all distributions of ρ. There is an optimal scheme for $\underset{\sim}{R}_3 \cup \underset{\sim}{R}_4 \cup \underset{\sim}{R}_5$ where

$$\underset{\sim}{R}_3 = \{R: R((\lambda,.5)) > 0\} ,$$

$$\underset{\sim}{R}_4 = \{R: R([0,\lambda]) = 1\} .$$

$$\underset{\sim}{R}_5 = \{R: R([\lambda,1]) = 1\} .$$

There is no optimal strategy for any initial (k,R) for R belonging to the complement of $\underset{\sim}{R}_3 \cup \underset{\sim}{R}_4 \cup \underset{\sim}{R}_5$.

REMARKS. Each of the classes $\underset{\sim}{R}_3$, $\underset{\sim}{R}_4$, and $\underset{\sim}{R}_5$ is closed under sampling, and therefore, so is their union. The last sentence in the theorem implies that $\underset{\sim}{R}_3 \cup \underset{\sim}{R}_4 \cup \underset{\sim}{R}_5$ is the largest class closed under sampling for which there is an optimal scheme.

PROOF OF THEOREM 3.2. Let $\underset{\sim}{R}_6$ denote the complement of $\underset{\sim}{R}_3 \cup \underset{\sim}{R}_4 \cup \underset{\sim}{R}_5$, so

$$\underset{\sim}{R}_6 = \{R: R((\lambda,.5)) = 0,\ R([0,\lambda)) > 0,\ R([.5,1]) > 0\} .$$

We shall define a scheme Ψ and show that it induces an optimal strategy for any $R \in \underset{\sim}{R}_3 \cup \underset{\sim}{R}_4 \cup \underset{\sim}{R}_5$ and an ε-optimal strategy for any $R \in \underset{\sim}{R}_6$.

Define $\Psi(k,R) = \mathfrak{L}$ for $R \in \underset{\sim}{R}_4$ and all k. Define $\Psi(k,R) = \mathfrak{R}$ for $R \in \underset{\sim}{R}_5 - \underset{\sim}{R}_4$ and all k. Clearly, as so far defined, Ψ is an optimal scheme for $\underset{\sim}{R}_4 \cup \underset{\sim}{R}_5$.

For $R \in \underset{\sim}{R}_6$ define

$$\Psi(k,R) = \begin{cases} \mathfrak{L} \\ \mathfrak{R} \end{cases} \text{ for } k \begin{cases} > \\ \leq \end{cases} k^* = \left[G - \frac{\log \varepsilon}{\log[\lambda/(1-\lambda)]} \right] ,$$

the largest integer less than or equal to $G - \log \varepsilon / \log[\lambda/(1-\lambda)]$.

Since $\underset{\sim}{R}_4 \cup \underset{\sim}{R}_5 \cup \underset{\sim}{R}_6$ is closed under sampling, we can calculate, for $R \in \underset{\sim}{R}_6$, that the probability of reaching G when starting at k and following Ψ is at least

$$1 - \frac{1 - [\lambda/(1-\lambda)]^{G-(k \vee k^*)}}{1 - [\lambda/(1-\lambda)]^{G-k^*}} R([0,\lambda])$$

$$\geq 1 - [1 - (\frac{\lambda}{1-\lambda})^{G-(k \vee k^*)} + (\frac{\lambda}{1-\lambda})^{G-k^*}]R([0,\lambda])$$

$$\geq 1 - [1 - (\frac{\lambda}{1-\lambda})^{G-k} + \varepsilon]R([0,\lambda]) \ .$$

Therefore,

$$(3.1) \qquad U(k; R,\lambda) \geq 1 - R([0,\lambda])(1 - [\lambda/(1-\lambda)]^{G-k}).$$

The lower bound in (3.1) is precisely the upper bound given in (1.2), so that for $R \in \underset{\sim}{R}_6$,

$$U(k; R,\lambda) = 1 - R([0,\lambda])(1 - [\lambda/(1-\lambda)]^{G-k}),$$

which means that Ψ is ε-optimal. The supremum U can only be attained if R is pulled eventually (and often) when $\rho \geq .5$ and if $\mathfrak{L}$ is pulled exclusively when $\rho < \lambda$. Since, for $R \in \underset{\sim}{R}_6$, no strategy fulfills both these requirements with probability one, there is no optimal strategy for $R \in \underset{\sim}{R}_6$, the complement of $\underset{\sim}{R}_3 \cup \underset{\sim}{R}_4 \cup \underset{\sim}{R}_5$.

It remains to consider $\underset{\sim}{R}_3$, a class closed under sampling. Define Ψ to be conserving for $\underset{\sim}{R}_3$. For each $R \in \underset{\sim}{R}_3$ there exists a k^* such that: for $k < k^*$

$$E_{\mathfrak{L}}^{(k,R)} U(k_1; R_1, \lambda)$$

$$\leq \lambda E^R[u(k+1,\lambda) \vee u(k+1,\rho)] + (1-\lambda)E^R[u(k-1,\lambda) \vee u(k-1,\rho)]$$

$$< E^R[\rho u(k+1,\rho) + (1-\rho)u(k-1,c)]$$

$$= E^{R}u(k,\rho)$$

$$\leq E_{\mathcal{R}}^{(k,R)}U(k_1;R_1,\lambda)$$

and, hence, $\Psi(k,R) = \mathcal{R}$. Therefore, for any $(k,R) \in (-\infty,G)\times R_3$,

$$(3.2) \qquad P_{\Psi}^{(k,R)}(\Psi(k_n,R_n) = \mathcal{R} \text{ i.o.} \mid k_n \to -\infty) = 1 .$$

By the strong law of large numbers

$$(3.3) \qquad P_{\Psi}(s_n/(s_n+f_n) \to \rho \mid k_n \to -\infty) = 1,$$

where s_n and f_n denote the numbers of successes and failures on $\mathcal{R}$ for stages 1 through n.

Choose $\delta > 0$, depending on $R \in R_3$, so that $R([\lambda+\delta,.5-\delta]) > 0$. An easy calculation starting with (3.3) gives

$$P_{\Psi}\left(\frac{R_n([0,\lambda+\delta))}{R_n([\lambda+\delta,.5-\delta])} \to 0 \mid k_n \to -\infty, \rho \geq .5 \right) = 1.$$

Now, by Lemma 3.1,

$$P_{\Psi}(\Psi(k_n,R_n) = \mathcal{L} \text{ only finitely often} \mid k_n \to -\infty, \rho \geq .5) = 1$$

which implies

$$P_{\Psi}(k_n \to G \mid k_n \to -\infty, \rho \geq .5) = 1$$

and, thus,

$$(3.4) \qquad P_{\Psi}(\rho \geq .5 \mid k_n \to -\infty) = 0.$$

From (3.2) and the weak law of large numbers we obtain

$$P_{\Psi}(R_n \to \delta(\rho) \mid k_n \to -\infty) = 1,$$

where $\delta(x)$ denotes the probability measure with mass 1 at x. Hence, using (1.2),

$$P_\psi(U(k_n;R_n,\lambda) \to 0 | k_n \to -\infty, \rho < .5) = 1.$$

By (3.4) we can drop the condition $\rho < .5$ and because we are assuming $\lambda > 0$ throughout we can replace $k_n \to -\infty$ by $k_n \not\to G$. Hence, (2.2) holds, and so Theorem 2.1 implies that Ψ is optimal for R_3. ∎

Because there is no gain in information when $\mathcal{L}$ is pulled, it seems unlikely for there to exist a fortune k at which $\mathcal{L}$ is uniquely optimal and yet $\mathcal{R}$ is optimal at fortunes $k-1$ and $k+1$. The next theorem verifies this, saying that the set of fortunes for which $\mathcal{R}$ is optimal is contiguous.

THEOREM 3.3. The optimal and ε-optimal schemes of Theorems 3.1 and 3.2 may be chosen from among the set of schemes Ψ for which $\{k: \Psi(k,R) = \mathcal{R}\}$ is an interval. Furthermore, if $G - g = \infty$, the schemes of Theorems 3.1 and 3.2 may be chosen from among the set of schemes Ψ for which $\{k: \Psi(k,R) = \mathcal{R}\}$ is empty or an unbounded interval.

PROOF. Let $\Psi(k,R) = \mathcal{R}$ if $\mathcal{R}$ is conserving at (k,R) and $\mathcal{L}$ otherwise. By Definition 2.4, Ψ is conserving for the class of all distributions of ρ. Using a well-known argument, we shall show by contradiction that $\{k: \Psi(k,R) = \mathcal{R}\}$ is an interval.

Suppose for a certain R and $k' \leq k'' - 2$ that $\Psi(k',R) = \mathcal{R}$, $\Psi(k'',R) = \mathcal{R}$, and $\Psi(k,R) = \mathcal{L}$ for all k with $k' < k < k''$. For this certain R and some $k \in (k',k'')$ let τ denote the strategy: pull $\mathcal{R}$ at stage 1, then pull $\mathcal{L}$ until either fortune $k' + 1$ or fortune $k'' + 1$ is reached in case $Z_1 = 1$ or until either fortune $k' - 1$ or fortune $k'' - 1$ is reached in case $Z_1 = -1$, and thereafter follow Ψ. Then

$$P_\tau^{(k,R)}(k_n \to G) = P_\psi^{(k,R)}(k_n \to G).$$

Therefore, $\mathcal{R}$ is conserving at (k,R) and, hence, $\Psi(k,R) = \mathcal{R}$, the desired contradiction. When $G - g = \infty$, similar reasoning shows that $\{k: \Psi(k,R) = \mathcal{R}\}$ is either empty or an unbounded interval. Therefore, Ψ is a conserving scheme with the desired characteristic.

For $g > -\infty$ and either $G < \infty$ or $\lambda > .5$, Theorem 2.1 together with the fact that the difference between $[k_n \to G]$ and $[k_n \to g]$ is a null set implies that Ψ is optimal for the class of all distributions of ρ. For $g = -\infty$, $G < \infty$, $\lambda < .5$, the proof of Theorem 3.2 shows every conserving scheme for $\underset{\sim}{R}_3$, in particular Ψ, to be optimal for $\underset{\sim}{R}_3$.

For the remaining cases the schemes explicitly constructed in the proofs of Theorems 3.1 and 3.2 satisfy Theorem 3.3. ■

Suppose that we were to replace the objective we have used throughout by the objective: maximize the probability that $k_n \to g$. None of the conclusions in this paper would change except in case $g = -\infty$, $G = \infty$, and $\lambda = .5$. With the change in objective, $\Psi(k,R) \equiv \mathcal{L}$ would become optimal in this case.

The remainder of this paper considers particular subclasses of the class of all distributions of ρ. The reader familiar with the emphasis on "stay with the winner" strategies in the literature will notice the various situations in which it is not optimal to pull an arm immediately after having obtained a success with that arm when following an optimal strategy.

4. Optimal Schemes for the Class $\underset{\sim}{R}_1$. In this section we find all optimal schemes when ρ is known to be either 1 or no greater than λ, thereby generalizing Example 1.1. Define

$$\underset{\sim}{R}_1 = \{R: \text{supp } R \subset [0,\lambda] \cup \{1\}\},$$

a class of distributions that is closed under sampling. If R is in this class, then supp $\mathcal{R} \subset [0,\lambda]$; so the effectiveness of $\mathcal{L}$ is known to be at least that of $\mathcal{R}$ if ever $\mathcal{R}$ yields a failure. Of course, no finite number of successes on $\mathcal{R}$ is as revealing in this regard as a single failure.

For reasons discussed in the previous section, the case $g = -\infty$ is rather trivial, so we exclude it from consideration. Also, if $G = \infty$ and $\lambda \leq .5$ then the problem is trivial; the only chance of winning is that $\rho = 1$, so U is just $R(\{1\})$. We therefore exclude this case as well.

Let $v(\rho)$ denote the probability, given ρ, of approaching G rather than reaching g when starting at $g+1$ and using the strategy: pull $\mathcal{R}$ at $g+1$ and $\mathcal{L}$ at all other fortunes. Standard calculations yield:

$$v(\rho) = \frac{\rho[1-(1-\lambda)/\lambda]}{1-\rho+\rho[1-(1-\lambda)/\lambda]} \quad \text{if} \quad \lambda > .5 \quad \text{and} \quad G = \infty$$

$$= \frac{\rho[1-(1-\lambda)/\lambda]}{(1-\rho)[1-((1-\lambda)/\lambda)^{G-g-1}]+\rho[1-(1-\lambda)/\lambda]} \quad \text{if} \quad \lambda \neq .5 \quad \text{and} \quad G < \infty$$

$$= \frac{\rho}{(1-\rho)(G-g-1)+\rho} \quad \text{if} \quad \lambda = .5 \quad \text{and} \quad G < \infty.$$

The next theorem generalizes Example 1.1, which concerns the case $\text{supp } R \subset \{0,1\}$ and $G < \infty$.

THEOREM 4.1. Assume that $g > -\infty$ and either $G < \infty$ or $\lambda > .5$. Then

$$\Psi_1(k,R) = \begin{cases} R & \text{if } k = g+1 \text{ and } E^R v(\rho) > v(\lambda) \\ f & \text{otherwise} \end{cases}$$

is an optimal scheme for $\underset{\sim}{R}_1$.

REMARK. From the definition of u in Section 1, it is evident that

$$v(\lambda) = u(g+1, \lambda).$$

PROOF OF THEOREM 4.1. By calculations, which we omit, that are similar to those used in Example 1.1,

$$V_1(k,R) = P_{\Psi_1}^{(k,R)}(k_n \to G)$$

can be shown to be excessive. For $G < \infty$ the optimality of Ψ_1 follows from Corollary 2.1. For $G = \infty$, Theorem 2.2 applies; for, (2.3) is satisfied by Ψ_1 since $\lambda > .5$ and $\Psi_1(k,R) = f$ for $k > g+1$. ∎

One of the calculations mentioned in the above proof shows that

$$V_1(k,R) \geq E(\rho)V_1(k+1,\sigma R) + E(1-\rho)V_1(k-1,\varphi R)$$

for $k > g+1$, with equality if and only if $G = \infty$ and $Ev(\rho) \geq v(\lambda)$. Keeping track of this and other situations in which equality holds when checking the excessivity of V_1 allows us to list every conserving scheme. By Theorem 2.1 these are exactly the optimal schemes. Accordingly, we have the following theorem.

THEOREM 4.2. Assume that $g > -\infty$ and either $G < \infty$ or $\lambda > .5$. A scheme Ψ is optimal for $\underset{\sim}{R}_1$ if and only if

$$\Psi(k,R) = \mathfrak{R} \text{ if } k = g+1 \text{ and } E^R v(\rho) > v(\lambda)$$
$$= \mathfrak{L} \text{ if } k > g+1,\ R(\{\lambda,1\}) < 1, \text{ and } G < \infty$$
$$= \mathfrak{L} \text{ if } E^R v(\rho) < v(\lambda).$$

REMARKS. According to this theorem, there are many cases in which either arm can be pulled without losing optimality. For example, if $k > g+1$, $Ev(\rho) \geq v(\lambda)$, and $G = \infty$, then either arm can be pulled. This makes it clear that there are many optimal strategies that are not induced by any optimal scheme. One such strategy in the case $G = \infty$ makes the following minor modification of some strategy induced by Ψ_1: arm $\mathfrak{L}$ is used when (and if) fortune $g+2$ is reached for the second time provided that $Ev(\rho) \geq v(\lambda)$.

In the next section we consider a class of distributions which is at the other extreme from the class considered in this section.

5. Optimal Schemes for the Class $\underset{\sim}{R}_2$. In this section we investigate optimal schemes when ρ is known to be either 0 or at least as large as λ. Define

$$\underset{\sim}{R}_2 = \{R: \text{supp } R \subset \{0\} \cup [\lambda,1] \ , \ R((\lambda,1)) > 0\} \ ,$$

a class of distributions that is closed under sampling. We have introduced the condition $R((\lambda,1)) > 0$ since R's for which $R(\{0,\lambda,1\}) = 1$ are contained in $\underset{\sim}{R}_1$ and so are covered in the previous section.

For a distribution $R \in \underset{\sim}{R}_2$, supp $cR \subset [\lambda,1]$, so the effectiveness of $\mathcal{R}$ is known to be at least that of $\mathcal{L}$ if ever $\mathcal{R}$ yields a success. In a sense $\underset{\sim}{R}_2$ is at the opposite extreme from $\underset{\sim}{R}_1$ since it is now a single success rather than a failure that settles the issue of which is the better arm. While the classes $\underset{\sim}{R}_1$ and $\underset{\sim}{R}_2$ are comparable in this regard, the matter of finding optimal schemes for $\underset{\sim}{R}_2$ is much more complicated than for $\underset{\sim}{R}_1$.

We shall construct a scheme Ψ_2 only on $(g,G)\times\underset{\sim}{R}_2$ and prove it to be optimal for $\underset{\sim}{R}_2$, checking the ways a scheme may differ from Ψ_2 and still be optimal for $\underset{\sim}{R}_2$. For "typical" R's belonging to $\underset{\sim}{R}_2$ it will develop that $\Psi(k,R)$ must equal $\Psi_2(k,R)$ for all k in order that Ψ be optimal. Throughout we exclude the case $g = -\infty$; our methods do not work in that case.

We first consider the relatively simple case $G = \infty$ and then examine the modifications necessary when $G < \infty$.

Suppose $G = \infty$ and, to exclude a trivial case discussed in Section 3, assume $\lambda > .5$. Propositions 5.1 and 5.2 allow us to construct the scheme Ψ_2 recursively.

Proposition 5.3 makes it clear that, for any fixed (k,R), $\Psi_2(k,R)$ can be calculated in a finite number of steps. Then Proposition 5.2 and Theorem 5.1 imply that, for any fixed initial (k,R), an optimal strategy (for all stages) and

$$P_{\Psi_2}^{(k,R)}(k_n \to G)$$

both can be calculated in a finite number of steps. All optimal strategies for $\underset{\sim}{R}_2$ are characterized in Theorem 5.1.

To define Ψ_2 we need the following notation:

$$\underset{\sim}{S}_0 = \{R \in \underset{\sim}{R}_2 : E^R(\rho|\rho\neq\lambda) \leq (2\lambda-1)/\lambda\} ; \tag{5.1}$$

$$\underset{\sim}{S}_m = \{R \in \underset{\sim}{R}_2 : \varphi R \in \underset{\sim}{S}_{m-1},\ R \notin \underset{\sim}{S}_{m-1}\},\ m = 1,2,\ldots; \tag{5.2}$$

$$\underset{\sim}{S}_\infty = \{R \in \underset{\sim}{R}_2 : R(\{0\}) = 0\} . \tag{5.3}$$

If $R \in \underset{\sim}{S}_\infty$, then

$$E^{\varphi^m R}(\rho|\rho\neq\lambda) > \lambda > (2\lambda-1)/\lambda$$

for all m, and if $R \in \underset{\sim}{R}_2 - \underset{\sim}{S}_\infty$, then

$$E^{\varphi^m R}(\rho|\rho\neq\lambda) \to 0 \quad \text{as} \quad m \to \infty.$$

Therefore, we have the next result.

PROPOSITION 5.1. $\{\underset{\sim}{S}_0, \underset{\sim}{S}_1, \ldots, \underset{\sim}{S}_\infty\}$ is a partition of $\underset{\sim}{R}_2$.

For easy reference we formally state the following obvious facts.

PROPOSITION 5.2. The classes $\underset{\sim}{S}_\infty$ and $\underset{\sim}{S}_\infty \cup \bigcup_{i=0}^{m} \underset{\sim}{S}_i$, $m = 0,1,2,\ldots$, are closed under sampling. Moreover, for $R \in \underset{\sim}{S}_m$, $1 \leq m < \infty$, $\sigma R \in \underset{\sim}{S}_\infty$ and $\varphi R \in \underset{\sim}{S}_{m-1}$.

Let

(5.4) $\Psi_2(k,R) = \mathcal{R}$ if $(k,R) \in (g,G)\times S_\infty$

$= \mathcal{L}$ if $(k,R) \in (g,G)\times S_0$.

Suppose $\Psi_2(k,R)$ has been defined for $R \in S_\infty \cup \bigcup_{i=0}^{m} S_i$ and all k. Consider an R belonging to S_{m+1}. For starting at $(g+1,R)$ let τ_j denote the strategy: pull $\mathcal{L}$ until fortune j is reached (if ever), then pull $\mathcal{R}$ once, and thereafter follow Ψ_2. For $j < \infty$, let

(5.5) $$q_j(R) = P_{\tau_j}^{(g+1,R)}(k_n \to G)$$

which can be calculated using the portion of Ψ_2 already constructed. Let

(5.6) $$a(R) = \inf\{k: q_k(R) = \sup_j q_j(R)\},$$

(5.7) $\Psi_2(k,R) = \mathcal{L}$ if $k < a(R)$

$= \mathcal{R}$ if $k \geq a(R)$.

From Propositions 5.1 and 5.2 we see that we have completed a recursive definition of a scheme Ψ_2. The next result makes it clear that we can use (5.6) to calculate $a(R)$ in a finite number of steps.

PROPOSITION 5.3. Suppose $G = \infty$. For $R \in \bigcup_{m=1}^{\infty} S_m$, the sequence $j \mapsto q_j(R)$ defined by (5.5) satisfies:

(i) $q_j(R) \to (2\lambda-1)/\lambda$ as $j \to \infty$;

(ii) $q_j(R) > (2\lambda-1)/\lambda$ for some j;

(iii) $$q_j(R) < \frac{(2\lambda-1)/\lambda}{1 - (\frac{1-\lambda}{\lambda})^{j-g}} \quad \text{for all } j.$$

REMARKS. By calculating, for $R \in \bigcup_{m=1}^{\infty} S_m$, the numbers $q_1(R), q_2(R), \ldots$ in order, one will eventually find some j and

some $y > (2\lambda-1)/\lambda$ for which $y = q_j(R)$. For j greater than or equal to some J it will be true that

$$\frac{(2\lambda-1)/\lambda}{1 - (\frac{1-\lambda}{\lambda})^{j-g}} < y .$$

By Proposition 5.3, only $q_j(R)$ for $j < J$ need be examined in order to calculate $a(R)$.

PROOF OF PROPOSITION 5.3 (i), (iii). Assertion (iii) follows from the fact that $q_j(R)$ is less than the probability of getting to j with $\mathfrak{L}$ starting at $g+1$. Combine this with the fact that

$$P_{\tau_j}^{(g+1,R)}(k_n \to G | k_n = j \text{ for some } n) \to 1$$

as $j \to \infty$ to obtain (i). ∎

We must delay the proof of (ii) until we have proved the forthcoming Theorem 5.1.

For $R \in \bigcup_{m=1}^{\infty} S_m$, let

$$b(R) = a(R) \text{ if } q_{a(R)+1}(R) < q_{a(R)}(R) \tag{5.8}$$

$$= a(R) + 1 \text{ if } q_{a(R)+1}(R) = q_{a(R)}(R) .$$

THEOREM 5.1. Suppose that $G = \infty$, $g > -\infty$, and $\lambda > .5$. Let $a(R) = b(R) = g+1$ for $R \in S_\infty$, $a(R) = b(R) = \infty$ for $R \in S_0$, and a and b be given by (5.6) and (5.8) on $\bigcup_{m=1}^{\infty} S_m$. A scheme Ψ is optimal for R_2 if and only if

$$\Psi(k,R) = \mathfrak{L} \text{ if } k < a(R)$$

$$= \mathfrak{R} \text{ if } k \geq b(R).$$

In particular, Ψ_2, given via (5.4) and (5.7), is optimal for R_2. The functions a and b are finite on $S_\infty \cup \bigcup_{m=1}^{\infty} S_m$.

PROOF. With no loss of generality we take $g = 0$. Let

$$V_2(k,R) = P_{\Psi_2}^{(k,R)}(k_n \to \infty).$$

Clearly Ψ_2 is optimal for $S_{\sim\infty}$ and, hence,

$U(k;R,\lambda) = V_2(k,R) = E^R\,[1 - (\frac{1-\rho}{\rho})^k]$ for $R \in S_{\sim\infty}$ and all k.

The reader may formally check or argue directly that $\mathcal{L}$ is not conserving, and therefore not optimal, at any (k,R) for which $R \in S_{\sim\infty}$.

Clearly

$$V_2(k,R) = 1 - (\frac{1-\lambda}{\lambda})^k \quad \text{for} \quad R \in S_{\sim 0} \quad \text{and all} \quad k.$$

Condition (2.3) of Theorem 2.2 is obviously satisfied; the excessivity of V_2, and optimality of Ψ_2, for $S_{\sim\infty} \cup S_{\sim 0}$ is equivalent to

$$1 - (\frac{1-\lambda}{\lambda})^k \geq E^R(\rho)E^{\sigma R}[1-(\frac{1-\rho}{\rho})^{k+1}]+E^R(1-\rho)[1-(\frac{1-\lambda}{\lambda})^{k-1}]$$

for $R \in S_{\sim 0}$ and all k, that is,

$$E^R(\rho) \leq \frac{2\lambda-1}{\lambda}\,(\frac{1-\lambda}{\lambda})^{k-1}/E^{\sigma R}[(\frac{1-\lambda}{\lambda})^{k-1} - (\frac{1-\rho}{\rho})^{k+1}] \tag{5.9}$$

for $R \in S_{\sim 0}$ and all k. We omit the straightforward calculation showing (5.9) to hold with strict inequality, and, therefore, $\mathcal{L}$ to be uniquely optimal, at each (k,R), $R \in S_{\sim 0}$.

Suppose, as an induction hypothesis, that Ψ_2 is optimal for $S_{\sim\infty} \cup \bigcup_{i=0}^{m} S_{\sim i}$. For $R \in S_{\sim m+1}$, we have, by Proposition 5.2, that $\varphi R \in S_{\sim m}$ and $\sigma R \in S_{\sim\infty}$. By the definition (5.6) of $a(R)$ it is clear that $\mathcal{L}$, and only $\mathcal{L}$, is conserving at (k,R) for $k < a(R)$ and that φ is conserving at $(a(R), R)$. By the proof of Theorem 3.3, $\mathcal{R}$ is conserving at (k,R) for $k > a(R)$. An appeal to Theorem 2.1 completes the proof by induction that Ψ_2 is optimal.

Clearly $\mathcal{L}$ is conserving at $(a(R), R)$ if $b(R) = a(R) + 1$. On the other hand, suppose $\mathcal{L}$ is conserving at $(a(R),R)$. Then, since $\mathcal{R}$ is conserving at $(a(R) + 1, R)$ and Theorem 2.1 is applicable, $\tau_{a(R)+1}$ is optimal. Hence $b(R) = a(R) + 1$.

Suppose, for a proof by contradiction which is very similar to the proof of Theorem 3.3, that $\mathcal{L}$ is optimal at some (k,R) for $k > a(R)$. Then an optimal strategy for that initial (k,R) is to pull $\mathcal{L}$ at stage 1, pull $\mathcal{R}$ at stage 2, and thereafter follow Ψ_2. An equally good, and therefore optimal, strategy is to pull $\mathcal{R}$ at stage 1, pull $\mathcal{L}$ at stage 2, and thereafter follow Ψ_2. But this latter strategy is not optimal since it can be improved by pulling $\mathcal{R}$ at stage 2 if $Z_1 = 1$.

That $b(R) < \infty$ for $R \in \underset{\sim}{R}_2 - \underset{\sim}{S}_0$ follows from Proposition 5.3. Even though the proof of Proposition 5.3 has not yet been completed, no circular reasoning is involved; for, only the optimality of Ψ_2 from Theorem 5.1 will be needed for the completion of the proof of Proposition 5.3. ■

PROOF OF PROPOSITION 5.3 (ii). Since Ψ_2 is optimal from Theorem 5.1, $q_j(R)$ is at least as large as the probability of approaching ∞ by pulling $\mathcal{L}$ until the fortune j is reached, then pulling $\mathcal{R}$ once, and thereafter pulling $\mathcal{R}$ or $\mathcal{L}$ accordingly as $X_1 = 1$ or $X_1 = -1$; that is (in case $g = 0$),

$$q_j(R) \geq \frac{(2\lambda-1)/\lambda}{1-[(1-\lambda)/\lambda]^j} \{E^R(\rho)E^{\sigma R}[1 - (\tfrac{1-\rho}{\rho})^{j+1}] + E^R(1-\rho)[1-(\tfrac{1-\lambda}{\lambda})^{j-1}]\}$$

which, for $R \in \underset{\sim}{R}_2 - \underset{\sim}{S}_0$, can be shown to be larger than $(2\lambda-1)/\lambda$ for sufficiently large j. ■

EXAMPLE 5.1. Suppose $G = \infty$, $g = 0$, $\lambda = .6$, and ρ is known to be either 0 or .625. The class

$$\underset{\sim}{R}_2' = \{R: \text{supp } R \subset \{0, .625\}\}$$

is a subset of $\underset{\sim}{R}_2$ and is closed under sampling. Let $\underset{\sim}{S}_m' = \underset{\sim}{S}_m \cap \underset{\sim}{R}_2'$, $m = 0, 1, \ldots, \infty$. Since $(2\lambda-1)/\lambda = 1/3$,

$$\underset{\sim}{S}_0' = \{R \in \underset{\sim}{R}_2', R(\{.625\}) \leq 8/15\} ;$$

$$\underset{\sim}{S}_m' = \{R \in \underset{\sim}{R}_2': \frac{8^m}{8^m+7\cdot 3^{m-1}} < R(\{.625\}) \leq \frac{8^{m+1}}{8^{m+1}+7\cdot 3^m}\}, \quad m = 1,2,\ldots;$$

$$\underset{\sim}{S}_\infty' = \{R \in \underset{\sim}{R}_2': R(\{.625\}) = 1\} .$$

Suppose $R(\{.625\}) = 32/35$. Then $R \in \underset{\sim}{S}_3'$, $\varphi R \in \underset{\sim}{S}_2'$, $\varphi^2 R \in \underset{\sim}{S}_1'$, and $\varphi^3 R \in \underset{\sim}{S}_0'$. Also $\varphi R(\{.625\}) = 4/5$, $\varphi^2 R(\{.625\}) = 3/5$, and $\varphi^3 R(\{.625\}) = 9/25$. Straightforward calculations using (5.6) and (5.8) yield $a(R) = b(R) = 1$, $a(\varphi R) = b(\varphi R) = 3$, and $a(\varphi^2 R) = b(\varphi^2 R) = 14$.

If the initial fortune is $k = 2$, then $\mathfrak{R}$ should be pulled (since $k \geq b(R)$). If $X_1 = Z_1 = -1$, so that $k_1 = 1$ and $R_1 = \varphi R$, then $\mathfrak{L}$ should be pulled until fortune 3 is reached (if ever), at which time $\mathfrak{R}$ should be pulled again. If $X_2 = -1$, then $\mathfrak{L}$ should be pulled until fortune 14 is reached, at which time $\mathfrak{R}$ should be pulled again. If $X_3 = -1$, then $\mathfrak{R}$ is never pulled again, no matter how large the fortune becomes. Of course, $\mathfrak{R}$ is used exclusively if it ever gives a success.

If, instead, the initial fortune is $k \geq 16$, then $\mathfrak{R}$ should be pulled the first three times. The modifications for other possible initial fortunes are now obvious.

In order to approach infinity from any starting point, R should always be pulled at least three times. Interestingly however, if k is small these pulls should be delayed until failures on R can be more easily withstood. ∎

We turn to the case in which g and G are both finite. When $G = \infty$, the gambler, knowing that he wants to play forever, is quite interested in gathering information about R. Now that $G < \infty$, the gambler, especially when his fortune is close to G, may take a more immediate view. It is, therefore, not surprising to find that the gambler may now be wise to pull ℓ when his fortune is close to G, thus eschewing the opportunity to learn about ρ that is provided by having a fortune far from g.

The definition (5.1) of S_0 must be replaced by

$$S_0 = \{R \in R_2 : E^R(\rho)E^{\sigma R}[u(k+1,\rho)-u(k-1,\lambda)] < u(k,\lambda)-u(k-1,\lambda) \text{ for all } k \in (g,G)\}, \tag{5.10}$$

which is equivalent to (5.1) when $G = \infty$. We define S_m for $m = 1,\ldots,\infty$ via (5.2) and (5.3). Propositions 5.1 and 5.2 remain true for $G < \infty$. We continue to use (5.5) for $g < j < G$. We can also view τ_j as used in (5.5) as a strategy for starting at $(G-1, R)$, as well as a strategy for starting at $(g+1, R)$. In analogy with (5.5), let

$$Q_j(R) = P^{(G-1,R)}_{\tau_j}(k_n \to G).$$

In analogy with (5.6), let

$$A(R) = \max\{k \in (g,G): Q_k(R) = \max_{g<j<G} Q_j(R)\}. \tag{5.11}$$

Of course, in the present context, "max" and "$g < j < G$" appear in (5.6) also. Replace (5.7) by

$$(5.12)\qquad \begin{aligned} \Psi_2(k,R) &= \mathcal{L} \text{ if } k < a(R) \\ &= \mathcal{R} \text{ if } a(R) \le k \le A(R) \\ &= \mathcal{L} \text{ if } A(R) < k . \end{aligned}$$

For $G < \infty$ there is no need for an analogue of Proposition 5.3. In analogy with (5.8), let

$$(5.13)\qquad \begin{aligned} B(R) &= A(R) \text{ if } Q_{A(R)-1}(R) < Q_{A(R)}(R) \\ &= A(R)-1 \text{ if } Q_{A(R)-1}(R) = Q_{A(R)}(R) \end{aligned}$$

for $R \in \bigcup_{m=1}^{\infty} \underset{\sim}{S}_m$.

For $G = \infty$ and any $R \in \underset{\sim}{R}_2$ we saw that either $\mathcal{L}$ is uniquely optimal at (k,R) for all k or else there is some k such that $\mathcal{L}$ is not conserving at (k,R). That this dichotomy does not hold when $G < \infty$ is the reason for one further notation. Let

$$(5.14)\qquad \begin{aligned} \overline{\underset{\sim}{S}}_0 = \{R \in \underset{\sim}{R}_2 : & E^R(\rho)E^{\sigma R}[u(k+1,\rho)-u(k-1,\lambda)] \\ & \le u(k,\lambda)-u(k-1,\lambda) \text{ for all } k \in (g,G)\} . \end{aligned}$$

Clearly, $\underset{\sim}{S}_0 \subset \overline{\underset{\sim}{S}}_0 \subset \underset{\sim}{S}_0 \cup \underset{\sim}{S}_1$.

THEOREM 5.2. Suppose g and G are finite. Let $a(R) = b(R) = g+1$ and $A(R) = B(R) = G-1$ for $R \in \underset{\sim}{S}_\infty$, $a(R) = G$ and $A(R) = g$ for $R \in \underset{\sim}{S}_0$, and a, b, A, and B be defined by (5.6), (5.8), (5.11), and (5.13) on $\bigcup_{m=1}^{\infty} \underset{\sim}{S}_m$. A scheme Ψ is optimal for $\underset{\sim}{R}_2$ if and only if

$$\begin{aligned} \Psi(k,R) &= \mathcal{L} \text{ if } k < a(R) \\ &= \mathcal{R} \text{ if } b(R) \le k \le B(R) \text{ and } R \in \underset{\sim}{R}_2 - \overline{\underset{\sim}{S}}_0 \\ &= \mathcal{L} \text{ if } A(R) < k . \end{aligned}$$

In particular, Ψ_2, given via (5.4) and (5.12), is optimal for $\underset{\sim}{R}_2$. For $R \in \underset{\sim}{R}_2 - \overline{\underset{\sim}{S}}_0$, $b(R) \leq B(R)$.

We omit the proof which is very similar to the proof of Theorem 5.1.

EXAMPLE 5.2. Suppose $g = 0$, $G = 4$, $\lambda = .74$, and $R(\{.75\}) = 1 - R(\{0\}) = .97554$. Straightforward calculations show that $R \in \underset{\sim}{S}_1 - \overline{\underset{\sim}{S}}_0$ and that $a(R) = b(R) = A(R) = B(R) = 2$. According to Theorem 5.2, in order for Ψ to be optimal it must satisfy $\Psi(1,R) = \Psi(3,R) = \mathfrak{L}$ and $\Psi(2,R) = \mathfrak{R}$. This is a situation in which the set of fortunes where it is optimal to pull $\mathfrak{L}$ is not an interval. ∎

Suppose that $g = -\infty$, $\lambda < .5$, and $R((\lambda,.5) > 0$. Then, according to the proof of Theorem 3.2, $\{k: \Psi(k,R) = \mathfrak{R}\} \neq \emptyset$ for every optimal scheme. Therefore, there is no analogue of the set $\underset{\sim}{S}_0$ and, as mentioned earlier, we pursue this case no further.

6. Optimal Schemes for the Class $\underset{\sim}{R}_\beta$. In this final section we consider the class

$$\underset{\sim}{R}_\beta = \{R: \text{supp } R \subset \{1-\beta,\beta\}\}$$

for the case $G = \infty$, $g > -\infty$. The class $\underset{\sim}{R}_\beta$ is closed under sampling and has the pleasant property that, for $R \in \underset{\sim}{R}_\beta$, $\sigma^s\varphi^f R$ depends on s and f only through $s-f$.

We carry over a notational convention from Example 1.1--if $r = R(\{\beta\})$, we not only use r as a number but we replace R by r in various symbols. For instance, we write $U(k;r,\lambda)$ for $U(k;R,\lambda)$ and $\sigma^s\varphi^f r$ for $(\sigma^s\varphi^f R)(\{\beta\})$.

Since the case $\beta = 1$ is discussed in Section 4 we take $\beta \in (.5,1)$ with no loss. Under the condition $\lambda \leq .5$ or the condition $\beta \leq \lambda$ the problem is trivial: $\mathcal{R}$ is optimal at every stage for the former and $\mathcal{L}$ is optimal at every stage for the latter. We therefore assume $.5 < \lambda < \beta < 1$. It seems reasonable to pull $\mathcal{R}$ if r is large enough and $\mathcal{L}$ otherwise. The next theorem says that some, but not all, optimal schemes have this characteristic. It indicates that, while a certain amount of flexibility is available for k large, $\mathcal{R}$ should never be pulled if r is small enough. The critical value of r is

$$r^* = \frac{(2\lambda-1)/\lambda}{(2\beta-1)/\beta} ,$$

which is the ratio $u(g+1,\lambda)/u(g+1,\beta)$ where u is defined in Section 1.

THEOREM 6.1. Suppose $G = \infty$, $g > -\infty$, and $.5 < \lambda < \beta < 1$. Then a scheme Ψ is optimal for $\underset{\sim}{R}_\beta$ if and only if

$$\Psi(k,r) = \mathcal{R} \text{ if } r > \left[1+\left(\frac{1-\beta}{\beta}\right)^{k-g-1}\left(\frac{1-r^*}{r^*}\right)\right]^{-1}$$

$$= \mathcal{L} \text{ if } r < r^* .$$

REMARK. The quantity

$$[1 + (\frac{1-\beta}{\beta})^{k-g-1}(\frac{1-r^*}{r^*})]^{-1}$$

is equal to $\sigma^{k-g-1}r^*$; it is the probability that $\rho = \beta$ given $k-g-1$ successes on $\mathcal{R}$ when the initial probability that $\rho = \beta$ is r^*.

PROOF. With no loss of generality we may assume $g = 0$.

One of the schemes given by the theorem is

$$\Psi_\beta(k,r) = \mathcal{R} \text{ if } r \geq r^*$$
$$= \mathcal{L} \text{ if } r < r^* .$$

Let

$$k^*(r) = \min\{j:\varphi^j r < r^*\} .$$

Since $\sigma^s\varphi^f r$ depends only on $s-f$ and r and, in view of the definition of k^*, the strategy induced by Ψ_β for starting at (k,r) specifies pulls of $\mathcal{R}$ until (if ever) reaching fortune $0 \vee [k-k^*(r)]$, and thereafter (if $k > k^*(r)$) it specifies pulls of $\mathcal{L}$ indefinitely. With the notation

$$V_\beta(k,r) = P_{\Psi_\beta}^{(k,r)}(k_n \to \infty)$$

we obtain

$$V_\beta(k,r) = ru(k,\beta) \quad \text{if } k \leq k^*(r)$$
$$= 1 - [1 - ru(k^*(r),\beta)][1-u(k-k^*(r),\lambda)] \text{ if } k > k^*(r).$$

The remaining steps are: (i) to prove that $V_\beta(k,r) \to 1$ as $k \to \infty$ uniformly in r, that is, that (2.3) holds; (ii) to show that V_β is excessive for $\underset{\sim}{R}_\beta$ and, thus, by Theorem 2.2, that Ψ_β is optimal for $\underset{\sim}{R}_\beta$; (iii) to observe where equality holds in the calculations

that show V_β to be excessive for $R_{\sim\beta}$ and, thus, conclude which schemes are conserving, and optimal by Theorem 2.1, for $R_{\sim\beta}$. We shall carry out step (i) and leave the straightforward steps (ii) and (iii) to the reader.

Let $\varepsilon > 0$. Choose K so that $u(k,\lambda) > 1-\varepsilon$ for $k \geq K/2$, $u(k,\beta) > 1-\varepsilon/2$ for $k \geq K/2$, and $r > 1-\varepsilon/2$ if $k^*(r) \geq K/2$. Suppose $k \geq K$. Either $k-k^*(r) \geq K/2$ in which case $1-u(k-k^*(r),\lambda) < \varepsilon$ or $k^*(r) \geq K/2$ in which case $1-ru(k \wedge k^*(r),\beta) < \varepsilon$. In either case $V_\beta(k,r) > 1-\varepsilon$. □

References

Berry, Donald A. (1972). A Bernoulli two-armed bandit. Annals of Mathematical Statistics 43 871-897.

Berry, Donald A. and Bert Fristedt (1979). Bernoulli one-armed bandits--arbitrary discount sequences, Annals of Statistics 7 1086-1105.

Berry, Donald A. and Bert Fristedt (1980). Two-armed bandits with a goal; II: Dependent arms. Advances in Applied Probability 12 No. 4.

Berry, Donald A., David C. Heath, and William D. Sudderth (1976). Red-and-black with unknown win probability. Annals of Statistics 2 602-608.

Dubins, Lester E. and Leonard J. Savage (1976). Inequalities for Stochastic Processes: How to Gamble If You Must. Dover, New York.

Dubins, Lester E. and William D. Sudderth (1977). Countably additive gambling and optimal stopping. Z. Wahrscheinlichkeitstheorie 41 59-72.

Sudderth, William (1972). On the Dubins and Savage characterization of optimal strategies. Annals of Mathematical Statistics 43 498-507.