INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

DYNAMIC DECISION BEHAVIOR: COMPETITIVE TESTS OF DECISION POLICIES IN A CLASS OF TWO-ARMED BANDIT PROBLEMS

by

Elizabeth Verghese Abraham

A Dissertation Submitted to the Faculty of the

COMMITTEE ON BUSINESS ADMINISTRATION

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY
WITH A MAJOR IN MANAGEMENT

In the Graduate College

THE UNIVERSITY OF ARIZONA

2000

UMI Number: 9992129

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

THE UNIVERSITY OF ARIZONA ®
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have read the dissertation prepared by Elizabeth Verghese Abraham

entitled DYNAMIC DECISION BEHAVIOR: COMPETITIVE TESTS OF DECISION POLICIES IN A CLASS OF TWO-ARMED BANDIT PROBLEMS

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy

Amnon Rapoport	Date: Sept. 27, 2000
Ken Koput	Date: 9/28/00
Lisa Ordonez	Date: 9/28/00
	Date
	Date

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director: Amnon Rapoport — Date: Sept. 27, 2000

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under the rule of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.

Signed: Elizabeth Abraham

ACKNOWLEDGMENTS

I would first and foremost like to thank Professor Amnon Rapoport for his unwavering support over the past several years. Your guidance, patience and encouragement were invaluable and will never be forgotten.

I would also like to thank my family (Paul, Yohan and Stefan) for their years of patience and encouragement during this entire endeavor. A special word of thanks to my extended family and friends who are too numerous to name. Please know that your love, prayers, and friendship have sustained me through this program.

Last, but not least, my friend Judith. I cannot tell you how much your friendship and hospitality have meant to me.

Partial support for the research summarized in this dissertation was provided by grant CA 98/99. BMOI "Bargaining, coordination and public good provision in experimental markets" to the Hong Kong University of Science and Technology.

DEDICATION

To my wonderful parents

Ivan and Mary,

(for teaching me to aim for the stars)

My beloved husband

Paul,

(for being my pillar of support)

and my dear children

Yohan and Stefan

(for making this all worthwhile).

TABLE OF CONTENTS

TABLE OF CONTENTS - Continued

LIST OF TABLES

LIST OF TABLES - Continued

LIST OF FIGURES

ABSTRACT

The Two-Armed Bandit Problem (TAB) is an individual decision making problem that is dynamic in nature. In a dynamic task, stage-to-stage changes in the state of the system are affected by the decision-maker's (DM) previous decisions as well as by the states of the system at the preceding stages. Unlike most of the dynamic tasks that are very complex, the TAB is much simpler with respect to the way it is presented to the DM and the range of decisions on each trial (simple binary choice). Thus it is an excellent choice for the study of behavior in dynamic tasks and is the focus of this thesis.

Most of the earlier research has focused on developing theoretical solutions to the problem and variations of the problem. Very little effort has been directed to examining the performance of naïve subjects. Hence this dissertation tries to present subjects with a series of realistic and consequently progressively more difficult tasks, and to account for their behavior. We looked at both individual and aggregate behavior. Given that the problems we tackled were so ill defined, it precluded us from developing normative solutions. Hence we took a descriptive approach, sacrificing mathematical tractability for realism.

Our research focused on the classic TAB problem and three variations of it (namely, the one armed bandit problem (OAB), TAB problem with increasing probabilities and the TAB problem with one arm with increasing probabilities and the other arm with decreasing probabilities. Instead of just capturing aggregate behavior we paid particular attention to individual decision makers comparing their decisions in the

classic TAB case to the optimal policy and to two degradations of the optimal policy. In the other three experiments subjects' decisions were compare to heuristics that were developed and actual earnings were compared to potential earnings.

Results were mixed in the experiments. None of our subjects were consistent in the policies that they used. None of the policies used could account for more than 30% of the decisions. In the classic TAB problem and the OAB problem we find that our results contradict earlier studies. We also found that our heuristics outperformed the subjects in some of the studies and in a couple of the studies the subjects outperformed the heuristics.

CHAPTER 1: DECISION MAKING UNDER UNCERTAINTY: THE BANDIT PROBLEM

INTRODUCTION

Very often one is faced with the task of choosing between two goods or services, the characteristics of which may be uncertain. It could be something that is as simple as choosing an overnight carrier for the delivery of some important papers or more complex as choosing between two treatments that are available for a certain disease. How would we make choices between these goods and services or treatments? A problem like this, with an individual decision maker (DM) making choices under uncertainty, can be classified as a bandit problem. The structure of a typical bandit problem has a single DM, who at discrete or continuous time intervals over an indefinite or finite horizon may choose one action from a set of two or more possible actions (i.e., the arms of the bandit), where this set is constant over time. (The term bandit refers to slot machines in a casino). Each arm, if employed in a period, generates a reward to the DM according to some time invariant distribution. The DM may begin with some prior belief of the arm's true type. It is common knowledge that the arms are independent of each other; hence employing the arm for at least a single trial can only generate partial knowledge of the arm's type. Finally, the DM is assumed to be interested only in maximizing the sum of her expected rewards.

For the purpose of this thesis, we will be dealing only with discrete time (as most of the literature does), and assume that the number of arms always equals two. Some

theoretical research has been done on multiple arms which will not be pursued in this thesis, except to say that extending the problem to a multiple armed problem does not appear to have posed much problem to the theoreticians (e.g. Rodman, 1978; Whittle, 1980). Another variation that is seen in the theoretical literature that will not be dealt with in this thesis is the notion of dependent arms. In all our experiments, we will consider only independent arms i.e., the outcome of one arm is not dependent on the outcome of the other arm. While theoretically it may be interesting to analyze dependent arms, we decided to keep this investigation simple and look at independent arms. One justification for this is that there are plenty of examples in the real world where one has to make choices between two goods or services that are independent of each other. In three of the experiments we deal with variations of the two-armed bandit problem (TAB) (where the characteristics of both arms are unknown). In the fourth and last experiment we deal with a subset of this class of problems, i.e., the one-armed bandit problem (OAB). In this variant, the characteristics of one of the arms are known, but not of the other. The problem here then would be to decide how to choose between the known and unknown arms.

To better elucidate the problem, let us pursue the medical example a little further. Two independent treatments are available for a certain disease, patients arrive at a clinic one at a time, and one of the treatments must be used on each. These treatments could be surgery, drugs, or even some form of "alternative medicine." For simplicity, assume that the response to the treatments is either positive (success/ patient is cured) or

negative (failure/ patient is not cured). Information about the effectiveness of the treatments accrues as they are used. The overall objective is to cure as many patients as possible. An inevitable but obviously beneficial result of any trial is the acquisition of information concerning the treatments. This accumulating information allows for better treatment of patients appearing later in the sequence of trials than those appearing early. This seemingly innocent but important problem is surprisingly difficult to solve. This is an example of a TAB problem. This problem is more complicated than testing two drugs, where traditionally one could run two samples, one under each treatment, and then determine the best treatment by some statistical comparison. The TAB problem arises in situations where one cannot afford the time or cost of taking large samples before acting and making such comparisons.

The Two-Armed Bandit Problem

Bellman (1956) credits W. R. Thompson, a prominent biostatistician, with being the first to formulate the TAB problem (1933), but Fredrik Mosteller (1955) introduce the "catchy" appellation in connection with experiments in the psychology of learning. The TAB problem pervades many fields of current research interest. It arises in experimental methodology, medicine, control theory, artificial intelligence, learning theory, psychology, strategic management, etc. (Witten, 1976).

One important extension of the TAB to the real world could involve strategic partnering. This is of particular interest to managers who are in the process of finding a strategic partner for their business. Perhaps the most significant manifestation of this rise

in inter-organizational cooperation in recent years has been the dramatic increase in joint ventures, joint R&D agreements, technology exchange, direct investment, licensing, and a host of other such arrangements (Gulati, 1995). Many of the large auto manufacturers seem to find strategic partnering a necessary fact of life to function in the competitive world of today. Consider a manufacturer who might be looking for suppliers with whom to develop a partnership. Few suppliers will be familiar to him due to their earlier interactions with him. But if he is now looking to find a new supplier, how does he decide between any two he might find suitable. If, after narrowing the list of suppliers down and using criteria that are important to him, he finds that there is a newcomer who has found his way to the final list, how does he then decide between a supplier with whom he has done business and a newcomer, especially if the newcomer looks equally good on paper? This process of choosing could be likened to an OAB problem or to a TAB problem if we have to choose between two newcomers. Each time we place an order with them, we gather more information about them, thereby helping us in our final and most important decision, i.e., who will become our strategic partner. We need to keep in mind that many of the large corporations have been unsuccessful in the marketplace due to bad decisions. A popular story that is often repeated in Strategic Management classes is the case of Wang Computers that was the first to introduce word processing software, but decided to pass up the opportunity to work with Apple computers and strike out alone. That unfortunate decision along with other similar crucial ones eventually led to its downfall.

Similarly, models of searching and learning have become quite popular in the field of microeconomics. For example, in labor economics one often hears of a worker who periodically receives information about her current job's true characteristics (and hence the wages she can expect), and has the opportunity to remain with her current employer or switch to a new job where again information about future wages might accrue (Viscusi, 1979). This is particularly true today when the labor market is extremely tight and one is tempted to believe stories floating around about the demand for a worker with a particular type of skill for example, programming in Java.

The preceding examples provide a variety of scenarios in the real world where solutions to the one-armed or two-armed bandit problem would become useful. One would assume that the choices are made strategically, considering not only the immediate utility (recovery) derived from each option but also how the information gained from each choice may affect the subsequent choices. To generalize the problem, the TAB can be stated as a conflict between taking those actions that yield probabilistically immediate rewards and those (such as acquiring information) whose benefits manifest themselves only later. In other words, it may be wise to sacrifice some potential early payoff for the prospect of information that will allow for more informed choices later. This aspect prompted Whittle (1982) to claim that a bandit problem 'embodies in essential form a conflict evident in all human action'. The 'information versus immediate payoff' makes the general problem challenging but interesting.

SOLUTIONS

One approach used in the literature is the minimax approach in which nature is regarded as an opponent in a zero-sum, two-person game. Nature chooses the probabilities of the two arms p_1 and p_2 in the unit square, according to some *a priori* restriction. The DM's goal is to minimize the expected difference between what is achieved and what could be achieved were p_1 and p_2 known, in other words try to minimize regret. Nature's goal is to maximize this expected difference.

However, the vast majority of the bandit literature takes the Bayesian approach for solving the problem, where the utility of the strategy is averaged over p_1 and p_2 with respect to some measure which represents information that is present about the various processes separate from the current experiment. However, many regard this measure as subjective. In this setting, the objective is to maximize the expected number of successes in the first *n* trials. The two arms are independent with prior knowledge of their parameters assumed to be described by beta distributions. We are going to use this approach for the purposes of this paper.

Bush and Mosteller noted that the "optimal decision rule ... for playing the TAB for 'n' trials seems not to be known (1955)." Since then, the optimal policy has been obtained numerically through the dynamic programming technique of backward induction. This technique is particularly useful in solving multistage decision making problems. Formally, the TAB problems can be characterized in the following manner. Decisions are to be made at each of a finite number (N) of discrete time stages. The value

of N is assumed to be known. At the beginning of stage k, k=1,2,..., N, the DM obtains information I_k about the present state of a given system and executes a decision D_k. The system then changes through a certain transformation, and the DM is faced with a new state at stage k+1. If the transformation from stage k to stage k+1 depends upon I_k and D_k only, the process is considered to be deterministic. However, if it involves uncertainty U_k about some external event, the process is considered to be stochastic or adaptive, according to the nature of U_k. A unique feature of the adaptive process is the use of information derived from earlier observations to reduce uncertainty about the process. The TAB problem falls in the adaptive class since we assume that some learning takes place.

"When making a decision, the DM attempts to maximize some well-specified criterion function, depending upon I_k, D_k, and U_k. A set of decisions, one for each of the N stages, is called a policy. The policy that maximizes the criterion function is considered to be optimal. In dynamic tasks, to behave optimally the DM has to consider the effect of each of his decisions on the future states of the system and, consequently, on his future behavior. Hence, the optimal policy has the following independence of path property. After any number of decisions, say k, the effect of the remaining N-k stages of the decision process depends only upon the present state k and the subsequent decisions (Bellman, 1961). The independence of path property necessitates the use of a backward induction technique, starting from the end of the process and moving toward the beginning.

The most important aspect of dynamic programming in experimental decision-making research is to provide a normative baseline. However, before deriving the optimal policy we need to characterize some of the elements in the problem. As stated earlier, the probabilities of success for each arm are p_1 and p_2. If lever 1 is operated and a success is obtained, the DM gets a reward of x_1 units. Similarly, if lever 2 is operated and a success is obtained, the DM gains a reward of x_2 units. No reward is given in the case of failure. We also assume that the number of stages, N, is known to the DM when he begins the process, and his task is to maximize his earnings over the N stages.

If p_1 and p_2 were known to the DM, and $x_1=x_2=x$, the normative solution for the DM would be to choose either lever 1 or 2 depending on which one had a higher probability. However, p_1 and p_2 are not known exactly, only their distribution functions can be assumed known. The optimal policy for this case can be obtained as follows.

The Bayesian revision of probabilities: The independence between the two arms allows the consideration of each one separately. Consider then an OAB problem (no decisions are involved) where at each pull a success occurs with probability p and a failure with probability (1-p). Consider first the single-stage problem. At the outset, the knowledge about p is assumed to be represented by the prior probability density function $f(p)$, $0 \leq p \leq 1$. It is assumed that the DM possesses the information that out of n trials (where n the number of trials completed at any given time, and is a subset of N), m successes have been scored ($0 \leq m \leq n$). His task is to calculate first the posterior

probability density function f(p|m,n) and later the expected value of p, i.e., $\int pf(p|m,n)dp$. Bayes' rule yields

$$f(p \mid m,n)=\frac{pr(m,n \mid p)f(p)}{\int_0^1 pr(m,n \mid p)f(p)dp}. \tag{1}$$

The integral in the denominator does not depend on p, so it becomes a normalizing factor for the posterior probability distribution. The term pr(m,n|p) is the probability of obtaining m success in n trials, given a value of p. A familiar result from elementary theory for Bernoulli trials is:

$$pr(m,n|p)=\binom{n}{m}p^m(1-p)^{n-m}. \tag{2}$$

Hence,

$$f(p|m,n)=\frac{p^m(1-p)^{n-m}f(p)}{\int_0^1 p^m(1-p)^{n-m}f(p)dp}. \tag{3}$$

Multiplying[1] both sides by p and integrating over the region $0 \le p \le 1$ yields the posterior expected value of p, denoted by $\hat{p}$.

[1] Equation 3 is the density of a beta distribution with parameters m+1 and n-m+1. For graphic displays of beta distribution see Mosteller and Tukey (1968).

$$\hat{p} = \frac{\int_0^1 p^{m+1}(1-p)^{n-m} f(p)dp}{\int_0^1 p^{m}(1-p)^{n-m} f(p)dp}. \qquad (4)$$

For the particular case where the prior distribution of p is rectangular, i.e., f(p) =1 for any p, $0 \leq p \leq 1$, we obtain (using a standard of integrals)

$$\hat{p} = \frac{m+1}{n+2}. \qquad (5)$$

The Bayesian estimate of p acts "as if" there were already two observations on hand, a success and a failure. Therefore, when m = n = 0, $\hat{p} = 1/2$ as expected.

Consider next a TAB problem with independent probabilities of success p_1 and p_2, assume a common rectangular prior distribution for each arm, and let the fractions of success on arms 1 and 2 be m_1 / n_1 and m_2 / n_2, respectively. The expected values of the posterior probabilities after n_1+n_2 trials are, respectively,

$$\hat{p}_1 = \frac{m_1+1}{n_1+2}, \text{ and } \hat{p}_2 = \frac{m_2+1}{n_2+2} \qquad (6)$$

The justification for assuming rectangular probability densities is that they represent the minimum possible a priori information about the values of p_1 and p_2.

This concludes the evaluation role of the Bayesian DM. It may be added here that the maximum likelihood estimates of p_1 and p_2 are m_1 / n_1 and m_2 / n_2, respectively. The latter values approach the Bayesian estimates when m_1, m_2, n_1, and n_2 are all large.

It should be emphasized that the probability evaluation and decision rules of the DM are independent. Thus, the decision rule may be maintained while the probabilities may be revised in a non-Bayesian fashion. For example, the DM may update the information but not extract from the observation all the information it conveys. Edwards (1968) refers to such a DM as a "conservative Bayesian." Other models of revision of subjective probability may be incorporated in the dynamic programming solution.

The Dynamic Programming Solution

With a Bayesian approach, the statistics m_1, n_1, n_2, and m_2 are sufficient for the computation of the probability densities, and hence may be used to define a state. Figure 1 graphically describes the change from state k to state k+1, k=1, 2, ..., n. The DM obtains information about state k, updates it, and then makes one of two decisions i.e., whether to pull lever 1 or lever 2. The system then moves on to state k+1 through a Bernoulli process with the success parameters p_1 and p_2, according to whether the decision was "Arm 1" or "Arm 2", respectively. Thus, when the horizon is finite, backward induction can be used to determine optimal strategies. One first finds the maximal expected payoff at the very last stage of every possible (n-1) history (sequence of selections and results), optimal or otherwise. Proceeding to the penultimate stage, one maximizes the expected payoff from every possible (n-2) history. Continuing backwards, while remembering the optimal arms at each partial history, yields all optimal strategies.

In a nutshell, the dynamic programming approach transforms the n stage maximization problem involving n decisions into a backward succession of n single stage maximization problems. Functions are built up recursively, one by one, until the entire process is expressed in terms of functional equations.

Let the optimal expected value of the task from the state (m_1,n_1,m_2,n_2) be denoted by $E_k(m_1,n_1,m_2,n_2)$. Then for stage N,

$$E_N(m_1,n_1,m_2,n_2) = \max\{H_N(1), H_N(2)\} ,$$

(7)

where

$$H_N(1) = p_1x \quad \text{and } H_N(2) = p_2x. \tag{8}$$

For stage k, $1 \leq k \leq N-1$,

$$E_k (m_1,n_1,m_2,n_2) = \max\{H_k(1),H_k(2)\} , \tag{9}$$

where

$$H_k(1) = p_1[x+E_{k+1} (m_1+1,n_1+1,m_2,n_2)] + (1-p_1)[E_{k+1}(m_1,n_1+1,m_2,n_2)],$$

(10)

and

$$H_k(2) = p_2[x+E_{k+1} (m_1,n_1,m_2+1,n_2+1)] + (1-p_2)[E_{k+1}(m_1,n_1,m_2,n_2+1)] \tag{11}$$

In other words, the expression $H_k(1)$ means that if lever 1 is pulled, a reward x can be expected with probability p_1, and the system then moves into the next state (m_1+1,n_1+1,m_2,n_2). The quantity E_{k+1} is assumed to have been computed in the previous stage" (Horowitz, 1973).

The optimal policy supports the "stay on a winner" rule, i.e., if on stage n the DM pulls an arm that yields a success, it is optimal for him to pull that arm on stage n+1, regardless of his decisions on the first n-1 stages. However, this rule does not hold when p_1 and p_2 are mutually dependent (Bradt, Johnson and Karlin, 1956). When $p_1 = p_2$ and $n_1>n_2$, the optimal policy is to pick the lever on which there is less information.

A computer program for the numerical implementation of the dynamic programming technique is listed in Appendix 1 - 1. The program assumes the Bayesian method of revision of probabilities and rectangular prior distributions, i.e., p_1 and p_2 are computed according to (6).

Since the number of states increases with N^4 (where N is the total number of trials), one faces problems of the computer memory being overrun. Hence, we took advantage of the symmetry of the two arms and reduced the computation. Only states where $n_1 \geq n_2$ were considered. An example of the program output for N=5 is also included in Appendix 1 - 2. For each state, the values of $\hat{p}_1$, $\hat{p}_2$, $H_k(1)$, $H_k(2)$, $H_k(2) - H_k(1)$, and the optimal decision are specified. The states for which $\hat{p}_1 = \hat{p}_2$ and $n_1 > n_2$ are denoted by "*". The optimal policy, then, is to decide upon a lever with the lower probability of success, a state denoted by "**". The latter situation represents instances

where the TAB problem does not have as its solution the "common sense" policy of deciding upon the lever with the higher posterior probability of success. An exception occurs in the last stage N, when no information gained can be used, and therefore $\hat{p}_1 > \hat{p}_2$ always implies $H_N(1) > H_N(2)$. The program was tested successfully through comparison with Horowitz's (1973) calculations for N=5.

The optimal solutions of the TAB problem and its variations have been investigated by many researchers including Robbins (1952,1953), Berry (1972), and more recently by Glazebrook and Owen (1991). It was the statistician Herbert Robbins who recognized the significance of the TAB problem as a general question of economical design of statistical experiments. For a TAB problem with both arms unknown and no prior distribution, Robbins suggested a selection strategy that depends on the history only through the last selection and the result of that selection; namely, the "stay on a winner, switch on a loser" rule. Robbin's objective was to maximize the long run proportion of successes. He shows how this rule uniformly dominates random selection. The most important contributions of Bradt, Johnson and Karlin (1956) are for the case in which one arm is known. They prove that if the known arm is optimal at any stage then it is also optimal thereafter. They also give an example in which p_1 and p_2 are dependent and for which the unique optimal procedure switches on the winner and stays on the loser!

Zelen (1969) proposes another simplistic procedure for the treatment of N patients. He uses the Robbin's "stay with a winner" rule for the first n patients in the trial and then uses the apparently superior treatment for the remaining N-n patients. Zelen

finds that $n \cong N/3$ yields nearly the maximal expected proportion of successes for the entire sequence. Berry (1972) considers the case of two independent Bernoulli arms where the discount sequence is n-horizon uniform or, in other words, a success at any stage is worth 1. The "stay with a winner" rule is proved in this setting. Jones examines two sub optimal rules which he names the "one-step-ahead" and "stay with the winner". In the former a decision is made by comparing the expected returns of the next trial for each arm. It takes into account the entire history of the earlier trials but does not take into account the effect of a decision on future decisions. The latter rule depends only on the results of the last trial. He shows how the efficiencies of both the suboptimal rules decrease with an increase in N (number of trials) and that the former is always more efficient than the latter (Jones, 1975). A good source for a survey of the theoretical work is Berry and Friedst (1985), and Berry (1985).

Previous Experimental Work

While many people have conducted in-depth theoretical work on this problem, there appears to be just a handful of empirical studies (using human subjects) that have been conducted (Brand, Sakoda and Woods (1957), Cane (1962), Horowitz (1973), Meyer and Shi (1995)). Brunswick (1939) was among the first to consider the effect of varying the probability of reward in a two-choice learning task. However, he used animal subjects, and hence the study is not of much interest to us. Brand et al. designed an experiment to provide a more satisfactory test of Brunswick's formulation of the influence of probability of the reward on two-choice learning using human subjects. Two

conditions, probability ratio of reward and probability differences of reward on the two choices were used. The results show that performance measures varied positively with probability difference of reward, but showed no consistent variation with probability ratio of the reward. In another study, the same authors tried to test how subjects would vary their choices between the alternatives if they believed that the occurrence was patterned according to a schedule. Hence, they varied the instruction sets given to the subjects. One set of subjects was told that the rewards were random and another set that the rewards followed some fixed pattern. They did not find any significant difference between the two sets of subjects except in one condition. Questions may be raised about misinformation that was presented to the subjects especially because the subjects were told that $p_1 + p_2$ always equal one and that was not necessarily the case. There appeared to be no justification for providing misinformation. Cane (1962) used humans and rats in her two-armed bandit experiments. She refutes the results of other learning experiments that are used to test the assumptions upon which linear mathematical models of learning rest and the assumptions are found to be invalid.

In his dissertation, Horowitz compares optimal strategies (obtained by dynamic programming) with three sub optimal strategies. His main objective is to analyze the way subjects behave in a two-armed bandit setting. His three sub-optimal strategies are degradations of the optimal model. They include the random model (complete loss of memory), one stage memory (loss of memory of all but the last stage), and the one stage horizon policy (the DM is myopic limiting his concern to the present stage only and not to

the future stages). He attempts to improve on earlier work by not restricting time within or between trials. He also displayed the number of decisions and successes on each arm, thus encouraging the processing of accumulated information. His most notable finding was that individuals tended to systematically oversample less promising options when placed in environments with low base rates of success (even though there was a real chance for monetary loss since subjects were penalized for every failure). The sensitivity analysis of the three sub-optimal policies shows that ignoring the past is more costly than ignoring the effect of the present decision on future behavior of the system. The random policy was outperformed by the others under every condition tested. This study can be criticized for offering a bonus for the best performance among those participating, because of possible subject contamination, since there is a difference between one who is trying to achieve the largest number of successes and one who is trying to maximize the expected number of successes.

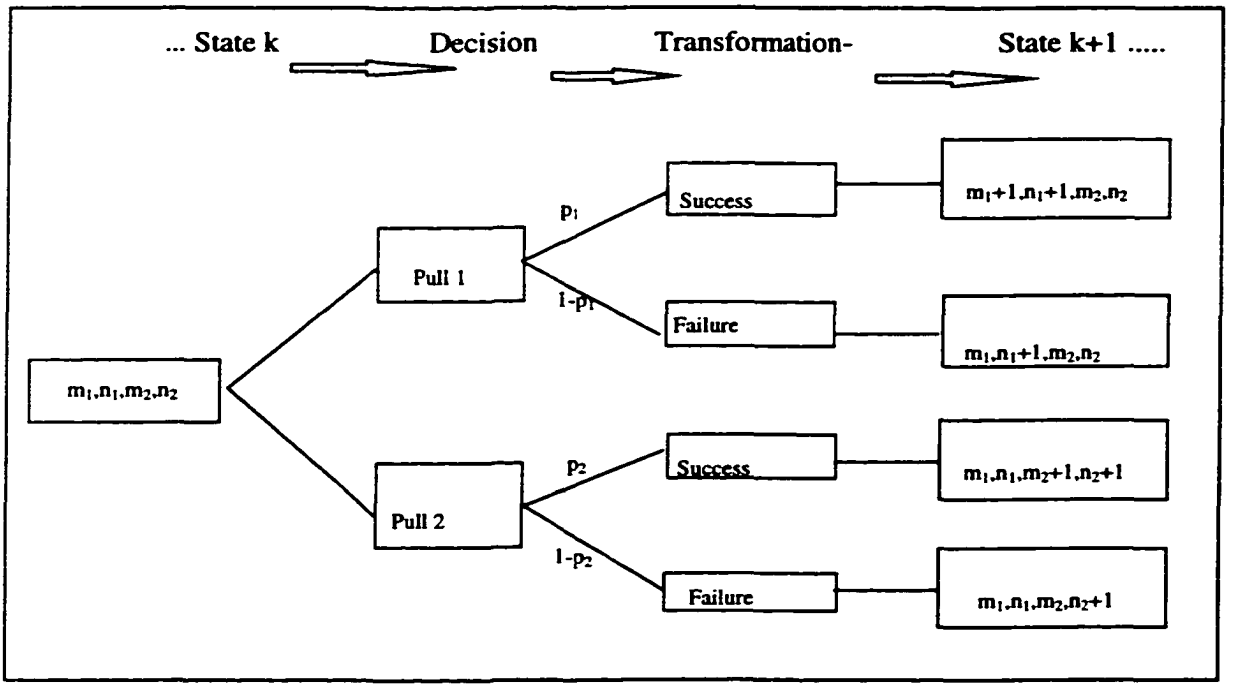
Meyer and Shi (1995) explored the process by which individuals learn from feedback when making recurrent choices among ambiguous alternatives. Their subjects were asked to play an OAB problem (choosing between two airlines, one with known on-time departure probability, and the other an airline with unknown probability). Subjects make repeated choices and are asked to maximize on-time departures. Similar to Horowitz, their data reveal a tendency to overexperiment with unpromising options and a tendency to increasingly switch between the two choices as the average base rate of success decreases.

The present study tries to go beyond all of these studies. We try to extend the TAB problem to cover various possibilities that, though more realistic, were not investigated in previous studies. We do so by looking at three variations of the TAB problem. The first experiment is the classical TAB experiment with two stationary unknown arms. The second experiment is a variation of the OAB problem where one arm is known, and the other arm is unknown but increasing in the probability of success. In the third experiment, both arms have increasing probabilities of success. In the last experiment, subjects face one arm with increasing probabilities and the other arm with decreasing probabilities. These variations are definitely more realistic since one is often faced with similar situations in the real world and very often we embark on an information gathering expedition before we actually make a choice. Examples are addressed in the respective chapters.

The chapters that follow are ordered according to the difficulty of the tasks. In the first TAB, we compare the subjects' decisions to the optimal policy and to two other sub-optimal polices. In the next chapter, we study the OAB problem. Here, we compare the subjects' decisions to two heuristics that have been taken from the literature to see how the subjects fare against them. In the fourth chapter, we study the TAB problem with the increasing arms, and in the fifth chapter we study a TAB problem with increasing and decreasing arms. In each of these cases, we compare subjects' decisions to several heuristics to see how our subjects fare against them and if the heuristics can surpass the performance of the subjects. The main focus of this thesis is to uncover behavioral

regularities in a class of very difficult tasks with non-stationary probabilities of success. There are innumerable studies done by Kahneman and Tversky (1982) and others that show that individuals do not approach decision problems in a normative manner particularly not in a Bayesian fashion. Hence, we wish to formulate and experimentally test several simple heuristics as potential models describing human behavior.

Figure 1 - 1: Characteristics of the dynamic programming elements in the TAB problem



COMPUTER PROGRAM IN 'C' FOR THE NUMERIC IMPLEMENTATION OF THE DYNAMIC PROGRAM

```
#include <stdio.h>
#include <stdlib.h>

#define P_max(a,b)   ((a) > (b) ? (a) : (b))
#define P_min(a,b)   ((a) < (b) ? (a) : (b))

typedef struct ListElt  {
 int      m1,m2,i,k;
 float  p1,p2,h1,h2;
}ListElt;

typedef struct ListNode {
 ListElt e;
 struct ListNode *next;
}ListNode;

typedef struct ListType {
 int count;
 ListNode *head;
}ListType;

static ListType *L;
int k=3,Flag=1;
FILE *Out;


void PrintFile()
{
```

```
 ListNode *t;
 ListElt  e;

 Out = fopen("output","w");
fprintf(Out,"K  M1/N1 M2/N2  P-1    P-2    DELTA    H-1    H-2    DIFF.
DECISION\n");
 for(t=L->head;t!=NULL;t=t->next) {
 e = t->e;
 fprintf(Out,"%1d  %2d %1d  %2d %1d %7.4f %7.4f %7.4f %7.4f %7.4f %7.4f",
  e.k,e.m1-1,e.k-e.i,e.m2,e.i-1,e.p1,e.p2,e.p1-e.p2,e.h1,e.h2,e.h1-e.h2);
 if((e.h1-e.h2) < 0.0)
   fprintf(Out," IIst arm");
 if((e.h1-e.h2) > 0.0)
   fprintf(Out," Ind arm");
 if((e.h1-e.h2) == 0.0)
   fprintf(Out," Both arms");
 fprintf(Out,"\n");
 }
 fclose(Out);
 Out=NULL;
}

void Sort()
{
 ListNode *t;
 ListElt temp;
 short flag=1;

 while(flag==1)        {
```

```
 flag = 0;
 for(t=L->head;t->next!=NULL;t=t->next)  {
  if((t->e.k<t->next->e.k)||
        ((t->e.k==t->next->e.k)&&(t->e.m1<t->next->e.m1))||
        ((t->e.k==t->next->e.k)&&(t->e.m1==t->next->e.m1)&&
        (t->e.i<t->next->e.i))){
        temp = t->next->e;
        t->next->e = t->e;
        t->e = temp;
        flag = 1;
        }
  }
 }
}

void Insert(e)
ListElt e;
{
 ListNode *temp;
 ListElt T;

 for(temp=L->head;temp!=NULL;temp=temp->next)        {
        T=temp->e;
        if((T.k==e.k)&&(T.m1==e.m1)&&(T.m2==e.m2)&&(T.i==e.i)) return;
        }
 temp=(ListNode *)malloc(sizeof(ListNode));
 temp->e = e;
 temp->next=L->head;
 L->head=temp;
```

```
 L->count++;
}
double EXP(MM1,INNN,MM2,K)
int MM1,INNN,MM2,K;
{
 float H_I,H_II,P_I,P_II,x,y,n;
 double temp;
 ListElt R;

 x = MM1; y = MM2; n = INNN;
 if(K==(k+1))
    return (0.0);
 P_I = (x)/(K-n+2);
 P_II = (y)/(n+1);
 H_I = P_I*(1+ EXP(MM1+1,INNN,MM2,K+1)) + (I-
P_I)*(EXP(MM1,INNN,MM2,K+1));
 H_II = P_II*(1+ EXP(MM1,INNN+1,MM2+1,K+1))+(1-
P_II)*(EXP(MM1,INNN+1,MM2,K+1));
 temp = P_max(H_I,H_II);
 R.m1=MM1;R.m2=MM2;R.k=K;R.i=INNN;
 R.p1=P_I;R.p2=P_II;R.h1=H_I;R.h2=H_II;
 if((Flag==0)||((P_I > P_II)&&(H_II > H_I))||
   ((P_I < P_II)&&(H_II < H_I)))    {
        Insert(R);
 }
 return temp;
}
main(argv,argc)
int argc;
```

```
char *argv[];
{
int     c;

L=(ListType *)malloc(sizeof(ListType));
L->head=NULL;L->count = 0;
printf("Input the number of stages:");
scanf("%d",&k);
if(k<1) exit;
printf("\n");
printf("Do you need only critical state info?(0 or 1) :");
scanf("%d",&c);
if(c<0) exit;
if(c==1) Flag=1; else Flag=0;
printf("\n");
EXP(1,1,1,1);
if (L->head!=NULL) {
        Sort();
        PrintFile();
        }
}
```

SAMPLE OF THE OUTPUT GENERATED BY THE DYNAMIC PROGRAM FOR N=5 CASE

K	M1	/N1	M2	/N2	P-1	P-2	DELTA	H-1	H-2	DIFF.	DECISION
5	4	4	0	0	0.8333	0.5	0.3333	0.8333	0.5	0.3333	Ist arm
5	3	3	0	1	0.8	0.3333	0.4667	0.8	0.3333	0.4667	Ist arm
5	3	3	1	1	0.8	0.6667	0.1333	0.8	0.6667	0.1333	Ist arm
5	3	4	0	0	0.6667	0.5	0.1667	0.6667	0.5	0.1667	Ist arm
5	2	2	0	2	0.75	0.25	0.5	0.75	0.25	0.5	Ist arm
5	2	2	1	2	0.75	0.5	0.25	0.75	0.5	0.25	Ist arm
5	2	2	2	2	0.75	0.75	0	0.75	0.75	0	Both arms
5	2	3	0	1	0.6	0.3333	0.2667	0.6	0.3333	0.2667	Ist arm
5	2	3	1	1	0.6	0.6667	-0.067	0.6	0.6667	-0.067	IInd arm
5	2	4	0	0	0.5	0.5	0	0.5	0.5	0	Both arms
5	1	1	0	3	0.6667	0.2	0.4667	0.6667	0.2	0.4667	Ist arm
5	1	1	1	3	0.6667	0.4	0.2667	0.6667	0.4	0.2667	Ist arm
5	1	1	2	3	0.6667	0.6	0.0667	0.6667	0.6	0.0667	Ist arm
5	1	1	3	3	0.6667	0.8	-0.133	0.6667	0.8	-0.133	IInd arm
5	1	2	0	2	0.5	0.25	0.25	0.5	0.25	0.25	Ist arm
5	1	2	1	2	0.5	0.5	0	0.5	0.5	0	Both arms
5	1	2	2	2	0.5	0.75	-0.25	0.5	0.75	-0.25	IInd arm
5	1	3	0	1	0.4	0.3333	0.0667	0.4	0.3333	0.0667	Ist arm
5	1	3	1	1	0.4	0.6667	-0.267	0.4	0.6667	-0.267	IInd arm
5	1	4	0	0	0.3333	0.5	-0.167	0.3333	0.5	-0.167	IInd arm
5	0	0	0	4	0.5	0.1667	0.3333	0.5	0.1667	0.3333	Ist arm
5	0	0	1	4	0.5	0.3333	0.1667	0.5	0.3333	0.1667	Ist arm
5	0	0	2	4	0.5	0.5	0	0.5	0.5	0	Both arms
5	0	0	3	4	0.5	0.6667	-0.167	0.5	0.6667	-0.167	IInd arm
5	0	0	4	4	0.5	0.8333	-0.333	0.5	0.8333	-0.333	IInd arm
5	0	1	0	3	0.3333	0.2	0.1333	0.3333	0.2	0.1333	Ist arm
5	0	1	1	3	0.3333	0.4	-0.067	0.3333	0.4	-0.067	IInd arm
5	0	1	2	3	0.3333	0.6	-0.267	0.3333	0.6	-0.267	IInd arm
5	0	1	3	3	0.3333	0.8	-0.467	0.3333	0.8	-0.467	IInd arm
5	0	2	0	2	0.25	0.25	0	0.25	0.25	0	Both arms
5	0	2	1	2	0.25	0.5	-0.25	0.25	0.5	-0.25	IInd arm
5	0	2	2	2	0.25	0.75	-0.5	0.25	0.75	-0.5	IInd arm
5	0	3	0	1	0.2	0.3333	-0.133	0.2	0.3333	-0.133	IInd arm
5	0	3	1	1	0.2	0.6667	-0.467	0.2	0.6667	-0.467	IInd arm
5	0	4	0	0	0.1667	0.5	-0.333	0.1667	0.5	-0.333	IInd arm
4	3	3	0	0	0.8	0.5	0.3	1.6	1.3	0.3	Ist arm
4	2	2	0	1	0.75	0.3333	0.4167	1.5	1.0833	0.4167	Ist arm
4	2	2	1	1	0.75	0.6667	0.0833	1.5167	1.4167	0.1	Ist arm
4	2	3	0	0	0.6	0.5	0.1	1.2	1.1333	0.0667	Ist arm
4	1	1	0	2	0.6667	0.25	0.4167	1.3333	0.9167	0.4167	Ist arm

4	1	1	1	2	0.6667	0.5	0.1667	1.333	0.9167	0.4167	Ist arm
4	1	1	2	2	0.6667	0.75	-0.083	1.4167	1.5167	-0.1	IInd arm
4	1	2	0	1	0.5	0.3333	0.1667	1	0.8333	0.1667	Ist arm
4	1	2	1	1	0.5	0.6667	-0.167	1.1667	1.3333	-0.167	IInd arm
4	1	3	0	0	0.4	0.5	-0.1	0.9	1.0333	-0.133	IInd arm
4	0	0	0	3	0.5	0.2	0.3	1	0.7	0.3	Ist arm
4	0	0	1	3	0.5	0.4	0.1	1.0333	0.9	0.1333	Ist arm
4	0	0	2	3	0.5	0.6	-0.1	1.1333	1.2	-0.067	IInd arm
4	0	0	3	3	0.5	0.8	-0.3	1.3	1.6	-0.3	IInd arm
4	0	1	0	2	0.3333	0.25	0.0833	0.6667	0.6	0.0667	Ist arm
4	0	1	1	2	0.3333	0.5	-0.167	0.8333	1	-0.167	IInd arm
4	0	1	2	2	0.3333	0.75	-0.417	1.0833	1.5	-0.417	IInd arm
4	0	2	0	1	0.25	0.3333	-0.083	0.6	0.6667	-0.067	IInd arm
4	0	2	1	1	0.25	0.6667	-0.417	0.9167	1.3333	-0.417	IInd arm
4	0	3	0	0	0.2	0.5	-0.3	0.7	1	-0.3	IInd arm
3	2	2	0	0	0.75	0.5	0.25	2.25	2.0083	0.2417	Ist arm
3	1	1	0	1	0.6667	0.3333	0.3333	2	1.6667	0.3333	Ist arm
3	1	1	1	1	0.6667	0.6667	0	2.1222	2.1222	0	Both arms
3	1	2	0	0	0.5	0.5	0	1.6167	1.6667	-0.05	IInd arm *
3	0	0	0	2	0.5	0.25	0.25	1.5	1.2583	0.2417	Ist arm
3	0	0	1	2	0.5	0.5	0	1.6667	1.6167	0.05	Ist arm
3	0	0	2	2	0.5	0.75	-0.25	2.0083	2.25	-0.242	IInd arm
3	0	1	0	1	0.3333	0.3333	0	1.1111	1.1111	0	Both arms
3	0	1	1	1	0.3333	0.6667	-0.333	1.6667	2	-0.333	IInd arm
3	0	2	0	0	0.25	0.5	-0.25	1.2583	1.5	-0.242	IInd arm
2	1	1	0	0	0.6667	0.5	0.1667	2.7222	2.5611	0.1611	Ist arm
2	0	0	0	1	0.5	0.3333	0.1667	2.0556	1.8889	0.1667	Ist arm
2	0	0	1	1	0.5	0.6667	-0.167	2.5611	2.7222	-0.161	IInd arm
2	0	1	0	0	0.3333	0.5	-0.167	1.8889	2.0556	-0.167	IInd arm
1	0	0	0	0	0.5	0.5	0	2.8889	2.8889	0	Both arms

K = stage
m_1= number of successes on arm 1
n_1= number of trials on arm 1
m_2= number of successes on arm 2
n_1= number of trials on arm 2
p_1 and p_2=probability of success on arm 1 and 2
H-1 and H-2= expected values of the task from state K conditional upon decision 1 or 2 respectively and upon optimality of remaining decisions.

CHAPTER 2: STANDARD TWO-ARMED BANDIT PROBLEM

INTRODUCTION

This chapter examines the standard (or baseline) case of the two-armed bandit (TAB) problem. This is the case where the two arms are independent of each other and the probability of success on each of them is fixed (stationary) and unknown. In particular, for each arm the probability of success on trial n+1 is independent of the DM's success or failure on the previous trial n.

To motivate this case, consider the following example. In the jet setting corporate world of today where time is money, you have been given the task of picking an official airline for the company. You have narrowed the field to two private airlines. They are both aggressively trying to court you, and they have extremely competitive rates, but you know nothing about their punctuality record. (For the sake of simplicity, assume that there are only two possible outcomes: either you arrive on time or you do not.) You decide that there is only one way you can find out about this. You will try each airline out multiple times, and you will award the contract based on their performance for you. This is your only option since there are no sources for you to get punctuality records etc. (Assume that you do not have too much time to make a decision, so sampling the two carriers for an indefinite number of trials is also out of question.) We will assume here that there is no change in the airlines performance (at least during the period of the contract) and that you incur no cost in switching from one carrier to another. How you

determine which is the best private airline for you when sampling is restricted is the essence of the standard TAB problem.

Since choosing between two startup airlines or overnight package carriers may evoke unnecessary associations for the subjects in a lab setting, we decided to model the situation by choosing between two urns, with a mixture of white and black balls in each urn.

METHOD

Subjects

Twenty subjects participated in the experiment. The subjects consisted of undergraduate students, graduate students, and a few university employees who responded to advertisements in the school newspaper and postings on bulletin boards around campus. The advertisements asked for volunteers to participate in an individual decision making experiment, and informed them that their payment would be contingent on their performance in a computer-controlled experiment.

Procedure

Subjects arrived at the Behavioral Decision Lab (BDL) in groups of six or seven. After signing in at the reception desk, each subject was randomly assigned to one of several identically furnished rooms. Each room was equipped with a personal computer. Subjects were asked to enter their name and social security number on the screen before they could begin the experiment. The instructions were available on-line. A hard copy of the instructions was also available (see Appendix 2 - 1). Paper and pencil were available

for the subjects to make any notes during the course of the experiment. The instructions explained in detail the task that the subjects had to perform and presented extensive examples. The subjects were informed that their primary purpose was to maximize their earnings (by selecting white balls), given the uncertainty of the task. Subjects were asked to summon the experimenter if they had any questions either during or after reading the instructions. Before the start of the actual trials, subjects went through a practice round of 10 trials to familiarize them with the task. At the end of the 10 practice rounds, they were told that the real trials were about to begin and that they would be paid for correct decisions.

Subjects were asked to imagine two opaque urns (jars) in front of them, each with a mixture of white and black balls in an unknown proportion. They were to reach into the urn and pick a ball, observe its color, and replace it. Each white ball represented a success and each black ball represented a failure. Since the experiment was computerized (for ease of administration), the subjects obviously had no urns into which they could reach. Instead, they were asked to pick between two urns on-line, and the computer would randomly select a ball from the urn. (This would be the equivalent of choosing between two urns and then reaching in and selecting a ball randomly with replacement.) If the ball picked was white, the subject earned a "franc" (a fictitious unit of money); if it was black, he/she earned nothing. (There was no penalty for picking a black ball.)

Subjects were informed that each urn contained several white and black balls, that sampling was with replacement, and that the percentage of white balls in each urn (which

was not known) could take on any value between 0 and 100. If the subjects understood and believed the instructions, then their prior distributions for both arms could be described by the uniform distribution. Of course, we do not know how the subjects interpreted the instructions. If the subjects misinterpreted the instructions, the assumption of a uniform prior may be inappropriate.

Each subject faced a series of eight tasks, (see Table 2 - 1). They were obtained by obtained by crossing three factors: The difference in probabilities (Δ=0.05 or Δ=0.15), the values of p_l (p_l = 0.8 or p_l = 0.4), and number of trials (N=10 or N=40). Once the subjects made a choice between the urns and entered the number in the computer, the computer randomly picked a ball from the urn. Subjects were paid in francs depending on the color of the ball that was picked. Francs were converted to US dollars at the rate of \$1 = 5 francs for the 10 trial tasks, and at the rate of \$1 = 10 francs for the longer 40 trial tasks. The conversion rates were chosen to ensure that the subjects would pay as much attention to the shorter tasks as to the longer tasks (in the pre-tests few subjects reported a tendency to ignore the shorter tasks and pay more attention to the longer tasks). At any stage during the experiment, the subject could observe the results of the previous trials within the same task by using the **F1** key. The computer provided the subjects with information that included the trial number, the urn that they had chosen, and the outcome of the random drawing by the computer. At the end of each task, subjects were informed that they were about to start a new task, and that they would sample from a new set of urns (see Appendix 2 - 1 for sample screens).

After completing the eight tasks, the subjects were paid their cumulative earnings, debriefed, and dismissed from the lab. The experiment lasted approximately 50 minutes. Mean earnings per subject were $ 12.74.

RESULTS

Descriptive Analysis of Subjects' Policies

Clearly, the tasks we set for the subjects were not easy. The results are expected to depend on the difficulty of the task. When the difference (Δ) between the probability of success in each of the urns is lower, it is much more difficult to differentiate between the urns. It should be easier to differentiate between the urns better when the difference is higher. Additionally, it should be easier to do it when the length of the task is 40 trials as opposed to the 10 trials.

As mentioned earlier, there were eight tasks altogether, four lasting for ten trials and four for 40 trials. The tasks were ordered according to their length. Subjects were assigned numbers depending on their order of showing up for the experiment. The subject number determined whether he/she saw the 10-trial tasks or the 40-trial tasks first. Even numbered subjects faced the 10-trial tasks first, and odd numbered subjects faced the 40-trial tasks first. Each task followed a similar procedure. Subjects were asked to choose between two opaque urns. They observed the different combinations of the parameters shown in Table 2 - 1 depending on whether they were even or odd numbered subjects. Table 2 - 2 shows the order of the conditions as presented to the subjects. Table 2 - 3 summarizes the frequencies of the choices and the number of successes and failures

achieved by the subjects. For the N=10 problems (upper part of Table 2 - 3), the subjects completed 800 trials (20 subjects participated in 4 conditions each of which repeated for 10 trials). They achieved successes (picked a white ball) in 52.75% of the trials (422/800). Similarly for the N=40 problems (lower part of Table 2 -3), the subjects completed 3200 trials and achieved success in 56.16% of the trials, (1797/3200). When N=10 and N=40 they chose Urn 1 approximately 50% of the time, thus exhibiting no response bias. A sequence of t-tests ($p < .05$) showed that there were no significant differences between their frequency of choices in either the 10 trial or 40 trial tasks. There were also no significant differences for the order of presentation. In other words, there were no significant differences between the choices of the even numbered and the odd numbered subjects. Henceforth, the results of the even numbered and odd numbered subjects will be combined.

The next analysis that we conducted looked at the number of times subjects made the right choice (picked the urn with the higher probability) in the last 10% and 20% of the trials for each of the 8 tasks. We are looking for the number of times subjects made the correct choice on

1. the last trial when N=10 (Table 2 - 4),
2. the last two trials when N=10 (Table 2 - 5),
3. the last 4 trials when N=40 (Table 2 - 7), and
4. the last eight trials when N=40 (Table 2 - 8).

We also recorded the number of times the subjects made the right choice in all the trials of each task.

1. all trials when N=10 (Table 2 - 6)

2. all trials when n = 40 (Table 2 - 9).

The idea was to compare the proportion of correct choices in the entire task with the proportion of correct choices in the last 20% and last 10% to find out whether there was any improvement in performance.

Considering first Table 2 - 4, we notice that on the last trial the subjects had higher percentages of failures when the probabilities were low. They fared better when the probabilities were high. The percentage of success increased from 35% and 30% in the low probability cases to 65% and 85% in the high probability cases, which is a statistically significant difference. When we consider the last two trials (Table 2 - 5), we find basically the same pattern. When the probabilities are high, subjects achieved a higher rate of success (77.5%). When the probabilities are low, the rate of success is significantly lower (45% and 25%). It is puzzling as to why subjects performed worse than chance in the low probability cases. Perhaps the task probability combinations and the short duration of each task (N=10) made it difficult for the subjects.

One would expect that the difference between p_1 and p_2 should have a consistent effect on the number of times the correct urn (urn with the higher probability of success) was chosen. This was not the case. Within each task, we find that this hypothesis did not

hold. However, when we compare between the two high probability cases and the two low probability cases we find that the hypothesis does hold except in Table 2 – 4 between the high probability cases. A sequence of t-tests($d.f._{199}$, $\alpha_{.05}$) shows that the choices of the urns did differ from chance (0.5) in all the conditions except for condition (0.4, 0.25). That can be confirmed by looking at Table 2 – 6. It appears that when the probabilities are high, there may be less motivation to pick the urn with the higher probability since the subjects perceive the success rates to be high anyway on both the urns. This is definitely an interesting finding even though the results were not always significant. In retrospect, choosing $\Delta = 0.05$ (where $\Delta = |p_1 - p_2|$) was not the best decision. It made the task unduly difficult for the subjects especially in the N=10 case.

Table 2 - 6 summarizes the results of all the subjects in the various conditions when N=10. When comparing Table 2 - 6 to Tables 2 - 4 and 2 - 5, we find that the overall patterns are similar, i.e., in the low probability cases the success rates are low whereas the success rates in the higher probability cases are considerably higher. We also find that when we compare between the low probability tasks and the high probability tasks, urn 1 was chosen more often when $\Delta = 0.15$ as opposed to 0.05.

The overall success rates for all the subjects in the entire sequence of trials in the low probability conditions are 33% and 23% (Table 2 - 6). In the higher probability conditions, success rates are 76% and 78%, which are significantly higher. There does not appear to be a significant learning going on over the trials, since the results are mixed for the conditions. In three of the conditions ((0.4, 0.25), (0.4, 0.35), and (0.8, 0.75)), there

was a moderate, though not always significant, increase in the number of times the subjects achieved a success in the last trial as compared to the overall trails. But in the other condition (0.8, 0.65), there was a significant drop in success rate in the last trial. We also checked to see if there were any significant difference between the $\Delta = 0.15$ conditions and between the $\Delta = 0.05$ conditions. The t-tests show that there is no significant difference whether the highest probability is 0.8 or 0.4.

The results are much clearer when N=40. Table 2 - 7 and Table 2 - 8 show the number of times that the subjects chose each of the urns in the last 10% of the trials (4), and the last 20% of the trials (8). Table 2 - 9 summarizes the results across all the various conditions for N=40. Subjects always chose the urn with the higher probability more often, even in the low probability cases. When we compare the results of Table 2 - 9 (all trials) to the results in Table 2 - 7 (last 10%) and Table 2 - 8 (last 20%), we find that the success rates have not increased in the last 10% of the trials compared to all the trials except in condition (0.4, 0.25) (21.25% in Table 2 - 7 to 32.625% in Table 2 - 9). In the high probability conditions they are almost identical (77.5% and 72.5% in Table 2 - 7 and 77.5% and 75.375% in Table 2 - 9). However, there is a decrease in the proportion of success in the condition (0.4, 0.35) (46.25% in Table 2 - 7 to 39.5% in Table 2 - 9).

We used the t-test ($d.f._{799}$, $\alpha_{0.01}$) of proportion to test if the urn choices of subjects differed from chance (0.5), and found that this was significantly different in all the four conditions.

We also checked to see if there was a consistent impact of Δ on the choices of the low probability cases in all the three tables. In the high probability cases, we find that while the effect is not as pronounced when all trials are considered, it is significant when we look at only the last 10% and the last 20% of the trials. In other words, when $\Delta = 0.15$ one should find them choosing the higher urn more often that when $\Delta = 0.05$. This hypothesis does not hold when we compare between the urns.

Overall, subjects performed much better in the 40-trial than the 10-trial conditions. As expected, in the longer tasks they had time to distinguish between the urns and were able to choose the higher urn more often. Subjects also did better (more successes) when p_1 and p_2 were high than low. This might suggest that they expected high probabilities of success from the beginning and if their expectations were met, they might satisfice, if not they continue, switching and experimenting more with the other arm.

Generic Policies

Next, we tried to assess the decision rules tat the subjects were using. We examined individual decisions closely to determine the policies that they might be using. At first, we tried to classify the subjects according to four generic policies that we had identified. However, as we studied the actions of the subjects it appeared that they were following other policies besides the ones we had identified. In the end, we found that most of the subjects could be classified in terms of seven different policies. We first describe the seven policies and then classify the subjects.

In the first policy ("stay on same arm") that we identified, subjects stayed with the same urn regardless of the trial outcome. There might have been a few times (less than 10%) that they appeared to "peak" at the other urn, but for the majority of the trials (90% or more) they stayed on the same urn.

In the second policy ("stay on winner, switch on loser"), the subjects stayed on the same urn following a success, and immediately switched to the other urn following a failure.

In the third policy ("fixed sampling"), the subjects sampled between the two urns for 'n' trials (this 'n' was not fixed for the subjects and when N=40 it varied between 3 - 15 trials), and then switched to the urn that they preferred and stayed on that urn till the end of the task.

In policy four ("random"), the subjects did not show any consistency in their switching behavior. They switched between urns on both wins and losses. They also did not stay on the same arm for a fixed number of times.

In the fifth policy identified ("fixed switching"), the subjects switched between the urns regardless of the result, after two or three trials. This policy is clearly distinguishable from the fourth policy even though the subjects may appear to be similar. Under policy four, the subjects switched from one urn to another after an arbitrary number of trials, regardless of the results.

In the sixth policy ("stay on a loser, switch on a winner"), the subjects appeared to be switching urns after a win, and staying on the same urn after a loss.

Finally, in the seventh policy ("switch after a few losses") subjects did not switch urns after the first loss that they encountered. Rather, they waited for a few losses before switching. Unlike the second policy, they did not switch on the first instance of a loss, but rather waited for a few losses (restricted between 2-4). It seemed as if they were willing to give the losing urn another chance before switching.

The decisions of the subjects were all tabulated and then scrutinized in an attempt to understand what they were doing. We tried to see if there was a pattern emerging while they made their decisions. In some cases, there was a clear and consistent pattern. For example, some subjects always shifted their choice to the other urn after encountering a failure i.e., they appeared to be following the second policy that we identified and they would do this regardless of the condition or task they faced. Some subjects were not as sure about what they wanted to do, so they would keep sampling the two urns with no particular pattern. After a few trials (this number of trials would vary between subjects and between tasks too, usually between 3-15 trials), they would pick one urn and stay with that urn till the end of that task (third policy). There were a few subjects who followed multiple policies. This was more apparent in the 40-trial tasks. However, for ease of computation we classified subjects under the policy that they followed for the majority of the trials. So while a subject may appear to be switching on a loser for the first few trials, he may adopt the policy of staying on a particular urn after, say, trial 12. Then, he would be classified under the latter policy for that particular task. If, on the other hand, for the first 20 trials a subject followed the "stay on a winner" policy and then picked an urn and stayed on it regardless of the outcomes, he was classified as having

followed both policies. Again, this generic policy classification is by mere observation of the data. We are trying to recreate what the subjects were doing by studying the data closely. For example, subject 18 appeared to be almost consistent in his decision policy no matter what the task. He used the fifth policy almost consistently in all the tasks i.e. he switched after 2 or 3 trials. He did not follow this policy only on task (0.4, 0.25; 40). It must be mentioned that none of the other subjects were as consistent in their decision policies.

In 21% of all the tasks (Figure 2 - 1) the subjects stayed on the same urn through most of the trials. When the probabilities were high, regardless of the number of trials, this was the most often chosen strategy. Subjects 14, 15, 16, 19 and 20 consistently chose this policy when the probabilities of a success in the urns were high. Subjects chose it only four times when the probabilities of winning were low. However, none of the above mentioned subjects chose it for the low probabilities.

Approximately 20% of the subjects appeared to be trying to sample the urns to decide which urn to pick, and then pick one urn and stay on it no matter what the results were. In other words, they were following the third policy. The trial number that they chose to shift to a particular urn and stay there appears to be quite arbitrary. Some subjects switched as early as the fourth or fifth trial in the 40 trial task and then stayed on the urn of choice for the rest of the trials. Yet, others waited till the 15th trial before deciding to stay on one urn. In the 10-trial tasks, it appears that subjects switched somewhere between the third and the fifth trial to stay on the same urn. Subject 10

seemed to be partial to this rule using it in seven out the eight tasks that he faced. It was used more in the 40-trial tasks.

About 19% of the subjects appeared to be following the fourth policy. In other words, we could not detect any of the policies in their behavior. They appeared to be shifting between the urns both on successes and failures. In the N=40 cases it was used more often when the probabilities were lower.

Policy 7 appears to be the next most commonly used policy. In 14% of the tasks, the subjects seemed to be switching after two or three losses. It seems like they picked an urn and were not willing to change unless they faced a series of two or three losses, and only then they felt that they needed to change or risk losing more money (or, rather, not winning much money). This was mostly used when the probabilities of winning were low and just once when the probability of winning was high. This was consistent across both the 10 and 40-trial tasks.

In 11% of the cases, the subjects switched after 2 or 3 trials no matter what the results, i.e., they used policy number 5. They were not willing to stay on the same urn for more than two or three trials at a time. This policy was used rarely in the low probability tasks, and more often in the higher probability tasks. As mentioned earlier, Subject 18 was partial to this policy using it in seven out of eight tasks.

In about of 10% of the cases, the subjects followed Policy 2. They tended to stay on a winner and switched the choice of urn when they encountered a failure. (This was a surprising finding since this would appear to be the most logical policy to follow.)

Finally, in about 5% of the tasks the subjects appeared to be switching following a success and stayed on the same urn following a failure. This behavior may be attributed to the gambler's fallacy. Subjects might have believed that after a loss they were bound to get a win, and after a win they felt that they had exhausted their "good luck" so it was better to move to the other urn. This happened regardless of the fact that they were told at the beginning of the experiment that the number of white and black balls in the urn were sampled independently.

What were the subjects doing when they followed these various policies? Were they following it consciously or blindly? Since they were motivated by moderate sums of money, we have every reason to believe that they were thinking about their choices carefully before making them.

Conversations with the subjects after they had been paid and debriefed revealed that many of them noticed "patterns in the appearances of the white and black balls" and hence they tailored their choice of urns to the "patterns" that they saw. Some subjects attempted to guess what ball would be picked next. It warrants mentioning that most subjects appeared to be switching policies between tasks with few exceptions. Subject 18 and subject 10 were the only ones who were consistent. Subject 18 followed policy 5 consistently for seven of the eight tasks. Subject 10 earned $13.90 and Subject 18 earned $13.40 for their participation, which were both higher than the average earnings for the entire group. Thus, a major finding is considerable difference between and even within subjects. This evidence refutes any model that prescribes the same policy or rule for all subjects. We do not have personal data (e.g. measure of attitude towards risk, answers to

personality inventories) that would help us in independently classify the subjects or account for the differences between them that we have uncovered. Another clear conclusion is that no one model can account for the results of all subjects. Attempts to come up with a single model when the tasks are so difficult are futile.

Decision Policies

As explained in the earlier chapter, we want to compare to each other three alternative models or decision policies to study behavior in the TAB problem.

1. The optimal policy with a restricted horizon

2. The One Stage Memory (OSM) model

3. The One Stage Horizon model (OSH).

We will explain each of these models and the reasons for choosing them in the following paragraphs. These three decision models are used as a benchmark to account for the results of all the subjects.

Optimal Policy

Using dynamic programming, which is a computational technique that prescribes decisions for a class of multistage decision making problems, it is possible to calculate the optimal policy for the baseline TAB problem. The TAB problem can easily be characterized as a series of decisions that are made at each of finite number of discrete time stages. The optimal model assumes revision using the Bayesian approach. The strategy requires only the sufficient statistics, namely, the number of successes and

failures on the two arms. Thus, it assumes perfect memory (past) and perfect knowledge about the horizon (future). Since it would be unreasonable to believe that subjects could behave in this optimal manner (and because the computers could not handle a problem this large in a moderate amount of time), in the N=40 case we avoided comparing the optimal policy to the decisions made by the subjects. Rather, we used a restricted optimal policy, which assumes perfect memory, revision of probabilities in a Bayesian manner and a limited horizon. Earlier studies have found that the horizon is at most equal to three stages (Rapoport, 1966). We looked for the case when the horizon ranged from 1 to 10 only for the tasks where N=10. Notice that the 10-stage case is the actual optimal policy when T=10. Table 2 - 10 to Table 2-13 show the number of matches with the optimal policy each subject had.

Consider each of the tables individually. Table 2 - 10 shows the percentage of matches with the optimal policy for condition (0.4 and 0.25; 10). On average, we find that 67.5% of the decisions appear to have matched the optimal policy. There were no changes when we increased the horizon to 10 stages. There is a wide variation in the matches made by the various subjects. For example, consider subject 10. He matched the optimal policy on every trial. Recall that he tried each urn for 'n' trials and then stayed on the same urn for the rest of the trials. However, on the other hand, subject 8 matched the optimal policy only in 40% of the trials. He switched every time he encountered a loss. Subjects 2, 3, 12 and 18 matched the optimal policy 60% of the time. Only four of the subjects matched it less than 60% of the time.

In Table 2 - 11, when the probabilities were 0.4 and 0.35, the average percentage of matches was marginally higher in the one-stage and two-stage horizon analysis. It decreases by 3% as the horizon increases beyond two stages. Subject 10 matched the optimal policy 100% in the first two stages, but as the number of stages increased the percentage decreased to 80%. Subjects 14 did extremely well this time by matching the optimal policy 100% of the time. He used the "stay on a winner, switch on a loser" policy. Subject 9 still had a lower than average success in the matches with the optimal policy. He switched on both losses and wins. Subject 1 was consistent with 80% match.

When the probabilities were increased to 0.8 and 0.65 (Table 2 - 12), subject 10 increased his matches to 100%. Subject 1 was consistent with 80% matches, and subject 9 started improving the matches to 60% in the one stage horizon and 50% matches when the horizon was increased. The number of subjects who increased their matches to 100% increased from one in the low probability cases to seven in this task. All the subjects who matched the optimal policy 100% of the time stayed on the same urn on all trials. However, subject 6 appears to have done rather poorly with just 30% of matches with the optimal policy. Subject 6 switched on both wins and losses in this particular task. Other subjects (5 and 8) who used this policy also fared poorly in comparison to the average for the task. The overall average number of matches increased to 77% in the one-stage horizon and 76 % in the longer horizon cases.

When the probabilities of success in the urns are increased to (0.8, 0.75) (Table 2 - 13), there is an overall increase in the number of matches to 82.5% in the one- and two-

stage horizon cases and 82% in all the others stage horizon case. Subjects 12, 13, 14, 16, 17 and 19 matched the optimal policy 100% of the time. They all stayed on the same urn throughout the entire task. Subject 1 improved to 90% matches using the same strategy. Subject 9 matched only 50% of the trials. But subject 10 who was doing extremely well in matching the optimal policy in the lower probability cases performed worse than average, and matching only 70% of the choices. Subject 5 who switched on both wins and losses fared the worst, matching the optimal policy only 30% of the times.

Considering all four tables together, we find that the average number of matches increases as the probabilities of winning in the urns increase, and that there is a slight increase in the matches when the horizon is less than three. As the horizon increases there is a marginal decrease in the number of matches to the optimal policy. This lends support to our hypothesis that it would be difficult to plan beyond one or two stages and corroborates findings by Rapoport (1966) in earlier studies. However, a closer look shows that the changes in the numbers were caused by just 2 of the subjects. All the other subjects remained the same for all the stages showing that there was really no difference to them or that in their case (the choices that they made and the results that were realized) the optimal policy was not very sensitive to changes in the horizon.

The policies that had the highest number of matches to the optimal policy in the higher probability tasks were to stay on the same urn. When the probabilities were lower, we find that the "stay on a winner, switch on a loser" rule seems to be having the maximum number of matches with the optimal policy. Another policy that also had

matches in the 80% or above range was the switch after 2 or 3 losses. None of the other policies fared as well when compared with the optimal policy.

Approximately 60% of the subjects seemed to match the optimal policy quite well when the probabilities increased. Subject 5 matched the optimal solution decreasingly as the probabilities of success were increased. Subject 1 who had done consistently well in the first three tasks suddenly appeared to do worse than average when the probabilities of winning were the highest. The improvement in the performance with the increase in probabilities is encouraging and easy to explain.

To check the sensitivity of the optimal policy to the length of horizon, we conducted a simulation for 1000 runs. We simulated the exact conditions the subjects faced (i.e., the probabilities of success in each of the urns), and had the computer pick the urns on each trial instead of the subject playing it. In other words, we generated data for 1000 artificial subjects each facing 10 trials. We ran each of the tasks separately. We collected the data and ran them through the same analysis as the actual subjects data that we collected, and then compared the decisions of the urns picked using the optimal policy to see if the simulated subjects seem to be as insensitive to the horizon as our actual subjects had been. Our results show that the real subjects did <u>better</u> than the simulated subjects. The average number of matches with the optimal policy in each of the conditions does not appear to vary drastically from condition to condition, the range of the average going from .57 to .596 (See Table 2 - 14). However, there appears to be another difference between the two. Whereas the actual subjects did marginally better

when the horizons were 1 and 2, the simulated subjects did better in all conditions when the horizon was 6 except in the (0.4, 0.35) condition when the one- stage horizon did the best. It appears that the simulations did not corroborate our findings.

The next two policies that we examined constitute degradations of the optimal policy. Recall that the optimal model assumes perfect memory (past) as well as perfect knowledge of the horizon (future). Since it is suspected that slight degradations of the optimal policy would not result in significant changes in the results, we chose to proceed with two degradations. One policy assumes that the subject has complete loss of memory with the exception of the information gathered in the last trial (the urn he chose and the ball that was picked). The other policy assumes that the subject behaves like there is just one more stage to go before the game terminates.

One Stage Memory Policy (OSM)

This policy assumes that the DM focuses only on the last stage only and ignores the previous history of the game. With this restriction, the optimal policy is the "stay on a winner, switch on a loser" rule. This simplistic nevertheless easy to implement policy does not require the DM to use either dynamic programming or Bayesian revision of probabilities. This policy outperforms a random policy particularly when $|p_1 - p_2|$ is large (Horowitz, 1973). Looking at each of the tasks, we find that in general there is an increase in the number of matches as the probabilities increase with two exceptions. In the 10-trial tasks we find that there was a drop in the proportion of matches when the probabilities increased to (0.4, 0.35). Similarly in the 40-trial tasks, we find that there was a small

insignificant drop in the number of matches when the probabilities increased to (0.8, 0.75) (Column 2 of Table 2 - 15). In the higher probability tasks (10 trials) we find that subject 13 matched the OSM policy 100% of the time. Subjects 15 and 19 matched it 100% in the (0.8, 0.65; 10) tasks, and subject 16 matched it 100% of the time (0.8, 0.75; 10) tasks. In the 40-trial tasks, we find that subject 3 alone matched the OSM policy 100% in the (0.8, 0.65) task. On average, we find that subjects match the OSM policy at least 50% of the time in all the tasks.

One-Stage Horizon Policy (OSH)

This policy assumes that the subject behaves as if there is only one more stage to go before the game terminates. This is a more complicated policy than the OSM policy. The optimal policy in this case is: Choose arm 1 whenever $\hat{p}_1 > \hat{p}_2$, and be indifferent when $\hat{p}_1 = \hat{p}_2$. where $\hat{p}_1$ and $\hat{p}_2$ are the expected values of the posterior probabilities that can be easily calculated as noted in Chapter 1. This is easy to do when the subject keeps tabs on the number of losses and number of trials. That may not be as easy when the subject participates in an experiment or when she may be worried that she may not complete the task in the allotted time (no time pressure was placed on the subjects, and any pressure that they felt would have been solely of their own creation). This policy fared better than the OSM policy in all the conditions. The number of matches increased with the increase in probabilities except in two cases. In condition (0.4, 0.35; 10), the number of matches fell to 51.67% from 59.44%, and again fell when the probabilities increased to (0.8, 0.75; 40) from 75.9% to 71.54%. This same pattern held even for the OSM policy.

It must be noted that there are a sizable number of decisions where both the OSH and the OSM policy had the same results, which is evident from the fourth column in Table 2 - 15.

Discussion

The three policies, namely, the optimal policy, the OSM policy, and the OSH policy have succeeded in capturing some of the DM's decisions. The optimal policy requires dynamic programming to compute and has done well in capturing the decisions of the subjects in the 10-trial tasks. There was a marked improvement in performance as the probabilities of success in the urns increased. How we rate this performance depends on the yardstick we use to measure success. We must keep in mind that the optimal policy requires complicated calculations beyond the ability of the subjects, and assumes perfect memory and perfect knowledge of the horizon. Subjects seem to have done well, given these tough requirements. It appears that the optimal policy managed to capture at least 67% of the subjects' decisions for the low probability task and up to 82% of the subjects' decisions when the probabilities of success in the urns increased. With the simulation, it appears that the number of matches ranged from 57.5% to 58.5%, which is not considered a significant increase.

The OSM policy also seems to have done quite well in capturing the decisions of the subjects. In all the conditions taken together, the OSM policy was matched 62.61% of the time and ranges from approximately 49% - 72% for each of the tasks individually. We would have expected this number to be higher, since this was an extremely easy policy to

follow. The OSH policy performed slightly better than the OSM policy averaging 66.63% over all the tasks and ranging from approximately 51% to 76%. This is surprising since this was the more difficult policy to follow.

It is also important to mention that while subjects were using different rules for the different tasks, the optimal policy seem to explain the majority of their decisions. This contradiction can probably be explained by the fact that many policies could potentially yield the same decision. Even while comparing the OSH and OSM policies in Table 2-15, we find that a good proportion of the decisions matched both policies.

Our study tends to contradict Horowitz's study. One of his main findings was that subjects tended to oversample the less promising options when the base rates of success were low. In our study we find that, the subjects tend to oversample the less promising options when N=10 and when the difference (Δ) is .05. When Δ increases to 0.15 this effect vanishes. Thus the base rates of success does not seem to have an impact on the oversampling of less promising options. Subjects did that in our study in the low and high conditions. Further when N=40 we find that this effect disappears regardless of Δ. One probable explanation for this is that when N=10 there isn't enough time for subjects to gather sufficient information and act on it.

What can we infer from all this? Against all expectations, the subjects seem to be following the optimal policy quite closely at least for the N=10 case as compared to the OSM and OSH policy. This is remarkable given the complexity of the task. Why they do not perform at least as well using one of the simpler heuristics is surprising. Whereas we

would have expected the OSM policy to outperform the other two policies, it actuality it performed the worst.

Table 2 - 1: Combinations of the parameters in the baseline experiment

$\Delta \longrightarrow$	.05		.15	
p $\longrightarrow$	.8	.4	.8	.4
N_1 $\longrightarrow$ (10)	(.8,.75)	(.4,.35)	(.8,.65)	(.4,.25)
N_2 $\longrightarrow$ (40)	(.8,.75)	(.4,.35)	(.8,.65)	(.4,.25)

where Δ = difference between p_1 and p_2
p = probability of each arm
N = Number of replications/trials in each task

Table 2 - 2: Order of conditions for baseline experiment

Odd Numbered subjects

N	p_1	p_2
10	.65	.8
	.4	.25
	.35	.4
	.75	.8
40	.8	.75
	.25	.4
	.8	.65
	.4	.35

Even Numbered Subjects

N	p_1	p_2
40	.75	.8
	.4	.25
	.65	.8
	.35	.4
10	.8	.65
	.25	.4
	.4	.35
	.8	.75

Table 2 - 3: Summary of Decisions by Result

N=10			
	Result		
Urn	**Failure**	**Success**	
1	192	193	385
2	186	229	415
	378	422	800

N=40			
	Result		
Urn	**Failure**	**Success**	
1	724	885	1609
2	679	912	1591
	1403	1797	3200

Table 2 - 4: Number of times subjects choosing the correct urn in the last trial when N=10

Condition	Decision	Result		Total
		Failure	Success	
(0.4, 0.25)	Urn 1	5	4	9 (45%)
	Urn 2	8	3	11 (55%)
	Total	13 (65%)	7 (35%)	20
(0.4, 0.35)	Urn 1	5	2	7 (35%)
	Urn 2	9	4	13 (65%)
	Total	14 (70%)	6 (30%)	20
(0.8, 0.65)	Urn 1	3	7	10 (50%)
	Urn 2	4	6	10 (50%)
	Total	7 (35%)	13 (65%)	20
(0.8, 0.75)	Urn 1	2	9	11 (55%)
	Urn 2	1	8	9 (45%)
	Total	3 (15%)	17 (85%)	20

Table 2 - 5: Number of times subjects choosing the correct urn in last two trials when N=10

Condition	Decision	Result		Total
		Failure	Success	
(0.4, 0.25)	Urn 1	9	11	20 (50%)
	Urn 2	13	7	20 (50%)
	Total	22 (55%)	18 (45%)	40
(0.4, 0.35)	Urn 1	13	2	15 (37.5%)
	Urn 2	17	8	25 (62.5%)
	Total	30 (75%)	10 (25%)	40
(0.8, 0.65)	Urn 1	5	19	24 (60%)
	Urn 2	4	12	16 (40%)
	Total	9 (22.5%)	31 (77.5%)	40
(0.8, 0.75)	Urn 1	6	16	22 (55%)
	Urn 2	3	15	18 (45%)
	Total	9 (22.5%)	31 (77.5%)	40

Table 2 - 6: Number of times subjects choosing the correct urn when N=10.

Condition	Decision	Result		Total
		Failure	Success	
(0.4, 0.25)	Urn 1	60	42	102 (51%)
	Urn 2	74	24	98 (49%)
	Total	134 (67%)	66 (33%)	200
(0.4, 0.35)	Urn 1	73	9	82 (41%)
	Urn 2	81	37	118 (59%)
	Total	154 (77%)	46 (23%)	200
(0.8, 0.65)	Urn 1	22	82	104 (52%)
	Urn 2	26	70	96 (48%)
	Total	48 (24%)	152 (76%)	200
(0.8, 0.75)	Urn 1	19	71	90 (45%)
	Urn 2	25	85	110 (55%)
	Total	44 (22%)	156 (78%)	200

Table 2 - 7: Number of times subjects choosing the correct urn in the last 4 trials when N=40

Condition	Decision	Result		Total
		Failure	Success	
(0.4, 0.25)	Urn 1	35	14	49 (61.25%)
	Urn 2	28	3	31 (38.75%)
	Total	63 (78.75%)	17 (21.25%)	80
(0.4, 0.35)	Urn 1	24	28	52 (65%)
	Urn 2	19	9	28 (35%)
	Total	43 (53.75%)	37 (46.25%)	80
(0.8, 0.65)	Urn 1	12	46	58 (72.5%)
	Urn 2	6	16	22 (27.5%)
	Total	18 (22.5%)	62 (77.5%)	80
(0.8, 0.75)	Urn 1	16	29	45 (56.25%)
	Urn 2	6	29	35 (43.75%)
	Total	22 (27.5%)	58 (72.5%)	80

Table 2 - 8: Number of times subjects choosing the correct urn in the last 8 trials when N=40

Condition	Decision	Result		Total
		Failure	Success	
(0.4, 0.25)	Urn 1	71	32	103 (64.375%)
	Urn 2	45	12	57 (35.625%)
	Total	116 (72.5%)	44 (27.5%)	160
(0.4, 0.35)	Urn 1	56	54	110 (68.75%)
	Urn 2	37	13	50 (32.25%)
	Total	93 (58.125%)	67 (41.875%)	160
(0.8, 0.65)	Urn 1	17	95	112 (70%)
	Urn 2	14	34	48 (30%)
	Total	31 (19.375%)	129 (80.625%)	160
(0.8, 0.75)	Urn 1	28	58	86 (53.75%)
	Urn 2	18	56	74 (46.25%)
	Total	46 (28.75%)	114 (71.25%)	160

Table 2 - 9: Number of times subjects choosing the correct urn when N=40.

Condition	Decision	Result		Total
		Failure	Success	
(0.4, 0.25)	Urn 1	296	181	477 (59.625%)
	Urn 2	243	80	323 (40.375%)
	Total	539 (67.375%)	261 (32.625%)	800
(0.4, 0.35)	Urn 1	269	218	487 (60.875%)
	Urn 2	215	98	313 (39.125%)
	Total	484 (60.5%)	316 (39.5%)	800
(0.8, 0.65)	Urn 1	80	412	492 (61.5%)
	Urn 2	100	208	308 (38.5%)
	Total	180 (22.5%)	620 (77.5%)	800
(0.8, 0.75)	Urn 1	106	355	461 (57.625%)
	Urn 2	91	248	339 (42.375%)
	Total	197 (24.625%)	603 (75.375%)	800

Table 2 - 10: Percentage of Matches with Optimal Policy for Condition (0.4, 0.25; 10)

Subject #	1 stage	2 stage	3 stage	4 stage	5 stage	6 stage	7 stage	8 stage	9 stage	10 stage
1	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
2	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
3	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
4	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
5	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
6	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
8	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
9	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
10	1	1	1	1	1	1	1	1	1	1
11	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
12	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
13	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
14	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
15	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
16	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
17	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
18	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
19	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
20	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
Average	**0.675**	**0.675**	**0.675**	**0.675**	**0.675**	**0.675**	**0.675**	**0.675**	**0.675**	**0.675**

Table 2 - 11: Percentage of Matches with Optimal Policy for Condition (0.4, 0.35:10)

Subject #	1 stage	2 stage	3 stage	4 stage	5 stage	6 stage	7 stage	8 stage	9 stage	10 stage
1	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
2	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
3	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
4	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
5	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
6	0.6	0.6	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
7	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
8	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
9	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
10	1	1	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
11	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
12	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
13	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
14	1	1	1	1	1	1	1	1	1	1
15	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
16	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
17	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
18	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
19	0.9	0.9	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
20	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Average	**0.73**	**0.73**	**0.705**	**0.705**	**0.705**	**0.705**	**0.705**	**0.705**	**0.705**	**0.705**

Table 2 - 12: Percentage of Matches with Optimal Policy for Condition (0.8, 0.65;10)

Subject #	1stage	2stage	3stage	4stage	5stage	6stage	7stage	8stage	9stage	10stage
1	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
2	0.8	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
3	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
4	1	1	1	1	1	1	1	1	1	1
5	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
6	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
7	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
8	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
9	0.6	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
10	1	1	1	1	1	1	1	1	1	1
11	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
12	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
13	1	1	1	1	1	1	1	1	1	1
14	1	1	1	1	1	1	1	1	1	1
15	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
16	1	1	1	1	1	1	1	1	1	1
17	1	1	1	1	1	1	1	1	1	1
18	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
19	1	1	1	1	1	1	1	1	1	1
20	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
Average	**0.77**	**0.765**	**0.76**	**0.76**	**0.76**	**0.76**	**0.76**	**0.76**	**0.76**	**0.76**

Table 2 - 13: Percentage of Matches with Optimal Policy for Condition (0.8, 0.75;10)

Subject #	1stage	2stage	3stage	4stage	5stage	6stage	7stage	8stage	9stage	10stage
1	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
2	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
3	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
4	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
5	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
6	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
8	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
9	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
10	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
11	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
12	1	1	1	1	1	1	1	1	1	1
13	1	1	1	1	1	1	1	1	1	1
14	1	1	1	1	1	1	1	1	1	1
15	1	1	1	1	1	1	1	1	1	1
16	1	1	1	1	1	1	1	1	1	1
17	1	1	1	1	1	1	1	1	1	1
18	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
19	1	1	1	1	1	1	1	1	1	1
20	0.7	0.7	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
Average	**0.825**	**0.825**	**0.82**	**0.82**	**0.82**	**0.82**	**0.82**	**0.82**	**0.82**	**0.82**

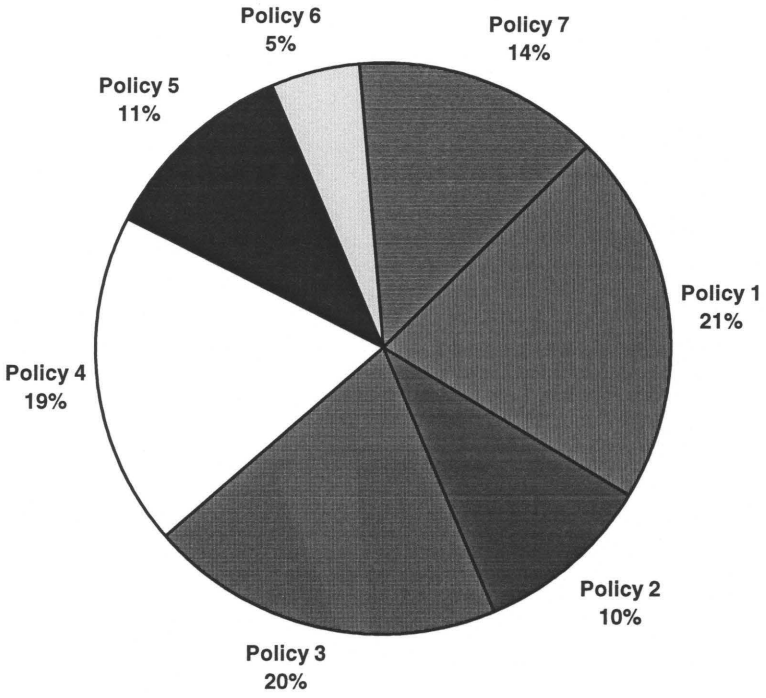
Table 2 - 14: Summary of Percentages of matches with Optimal Policy for Simulation.

COND	1 stage	2 stage	3 stage	4 stage	5 stage	6 stage	7 stage	8 stage	9 stage	10 stage
(0.4, 0.25)	.578	.578	.578	.578	.567	.586	.566	.578	.574	.574
(0.4, 0.35)	.591	.59	.59	.59	.584	.575	.575	.591	.587	.587
(0.8, 0.65)	.579	.584	.584	.584	.574	.588	.57	.583	.579	.579
(0.8, 0.75)	.584	.588	.589	.588	.58	.596	.578	.589	.584	.584

Table 2 - 15: Comparison of the percentage of matches with the OSM and the OSH policies

CONDITION	OSM	OSH	BOTH
(0.4, 0.25) (10)	57.78%	59.44%	38.33%
(0.4, 0.35) (10)	49.44%	51.67%	32.22%
(0.8, 0.65) (10)	68.89%	68.89%	52.22%
(0.8, 0.75) (10)	70%	77.78%	57.22%
(0.4, 0.25) (40)	53.85%	59.87%	32.44%
(0.4, 0.35) (40)	59.10%	67.95%	41.79%
(0.8, 0.65) (40)	71.79%	75.90%	48.63%
(0.8, 0.75) (40)	70.00%	71.54%	54.23%

Figure 2 - 1: Distribution of decisions between the generic strategies in TAB 1



Where:
Policy 1 - Stay on the same urn

Policy 2 - Stay on a winner, Switch on a loser

Policy 3 - Sample for n-trials, then pick an urn and stay on it

Policy 4 - Switch on wins and losses (catch all policy)

Policy 5 - Switch after 2 - 3 trials

Policy 6 - Gamblers Fallacy

Policy 7 - Switch after 2 -3 losses

INSTRUCTIONS TO SUBJECTS FOR STANDARD TAB

You are about to participate in a computer controlled individual decision making experiment in which your payoff is dependent on your performance. The money you will earn will paid to you at the end of the experiment. A research foundation has contributed the money to finance this study. The purpose of this experiment is to study how individuals behave when making a sequence of choices between two alternatives each of which yield a success with an unknown probability. A sequential test of two new medicines on a series of patients, is one example of this process. Only one of two drugs can be administered to each patient and each treatment is either a success or a failure. Clearly one wishes to maximize the number of success but only by trying each of the drugs can the physician learn which is more effective.

To take a simpler task, imagine that there are two urns or opaque jars in front of you marked 'I' and 'II.' Each urn has some black and some white balls in it, but you cannot see how many. The total number of balls may vary from one urn to another. You do not know this number, nor do you know the proportions of the black and white balls in each urn.

Your task is very simple. You will make a series of 'T' drawings (trials) the urns. 'T' will be known to you and in the situations you will face today it will either be 10 or 40. At the beginning of each trial, you have to decide which urn you are going to draw a

ball from. You then reach into the selected urn, pick up a ball blindly, record the color, and return the ball to the same urn. If you get a white ball, you will be paid 1 unit of money which we call a "franc." If you get a black ball, you get nothing. Once you have noted the color of the ball, you replace it, shake the urn and move on to the next trial. Clearly as you make more draws from a particular urn, you learn more about the proportion of white and black balls in that urn. Of course, you have to trade this value off against the value of drawing from the urn you currently think has the highest fraction of white balls. The essence of the task is to gather information about the urns and, at the same time, make decisions that increase your payoff.

Your objective is to maximize the money that you earn over 'T' trials by trying to choose as many white balls as you can. Rather than having you draw from the two urns, which is a tedious process, we have computerized the entire process. Your only action thus will be to choose one of the two urns. The computer will then randomly pick a ball from the urn you have chosen, will record your payoff for the trial, and will inform you of your accumulated payoff from the beginning of the task.

Today you will perform 8 of these tasks (each task is a series of 'T' decisions) in the complete experiment. At the beginning of each task, you will be told the value of 'T,' which will either be 10 or 40 and you will be reminded that the computer has selected two new urns, each with a different mix of white and black balls, which will remain fixed over the upcoming 'T' decisions. **The percentage of white balls in each of the two urns can be any value between 0 and 100 with all values in this range being equally likely.**

Please note that the likelihood (or probability) of the computer drawing a white ball will correspond exactly to the percentage of white balls in the selected urn. This likelihood will not change from trial to trial, just as it would not change in the urn if you replaced the ball you have selected after each trial.

Once the task begins you will see the following information on the prior to each decision you make:

History of previous play:			
trial #	urn 1	urn 2	cumulated payoff
You are on trial number (1.......T)			[out of T for this task]
Which urn would you like to sample from? (I or 2)			

For example, suppose you sampled from urn 1 on the first trial and you picked a black ball, and from urn 2 the next time and picked a white ball. Your screen would then look similar to this:

History of previous play			
	outcome		
trial #	urn 1	urn 2	cumulated payoff
1	Black	----	0
2	-----	White	1
You are on trial number (1.......T)			[out of T for this task]
Which urn would you like to sample from? (1 or 2)			

Once you enter either '1' or '2,' the computer will randomly draw a ball from that urn and present you with the following screen.

On trial number (1.....T)
Urn chosen on this trial (I or II)
Ball picked up on this trial (Black or White)
Your earnings for this trial = 0 or 1 franc.
[Press any key to continue]
[Press F1 to see history]

Following the presentation of the outcome of the trial, the computer updates your record and returns you to the first screen for the next trial. This process will be repeated for 'T' trials. At any time during the task you will be able to check the history for trials that are no longer visible (the history screen will hold the results for about 20 trials) by using the **F1** key that is located at the top left hand side of your keyboard.

As mentioned, you will participate in a series of 8 tasks. The number of trials will differ from one task to another. The two urns will also change between the tasks. In other words, the proportion of the white and black balls will change between the tasks.

Between each task you will see this screen:

You are about to begin a new sequence of 'T' trials.
You will sample from a new pair of urns.
The urns are still labeled I and II but they are not the same ones used in the previous. task.

Please keep in mind that the computer has not been programmed to play against you. Rather, the number of white balls and black balls in each urn has been predetermined and will not change from trial to trial within the next T choices. Your task is to draw as many white balls as you can over the next T trials.

Your concern is to make as much money as you can, given these conditions. You are being paid in a fictitious currency called a "franc". "Francs" will be converted into US dollars at the end of the experiment at the rate $1= 5 francs for the 10 trial tasks and at the rate of $1 = 10 francs for the 40 trial task. You will be paid before leaving.

If you have any questions about the instructions, please inform the experimenter immediately. You may refer to the written copy of the instructions placed on your desk at any time. Please take whatever time necessary to make your decisions in this experiment.

CHAPTER 3: ONE-ARMED BANDIT WITH NON-STATIONARY PROBABILITIES

INTRODUCTION

In the traditional one-armed bandit problem (OAB), one arm has a known probability of success and the other arm has an unknown probability. Sampling is with replacement for both arms. Some experimental work has been done on OAB in the past (Meyer and Shi, 1995). As stated in the first chapter, the central finding of their work was that when making decisions over time subjects may not always choose the option with the highest expected payoff. Rather, they might prefer experimenting with the riskier arm more than is prescribed by the optimal model.

When Meyer and Shi investigated this problem, they were concerned with the static task environment that the subjects faced in earlier experiments that studied choices under ambiguity, and hence decided to add the dynamic component to the environment. We decided to go one step further with our investigation. Instead of dealing with the traditional OAB problem in a dynamic environment, we violated the assumption of stationarity. In our version, there is one arm with known and fixed probability (just like the traditional OAB) and a second arm that has a probability of success that increases following each failure. In particular, in the fixed arm case sampling is with replacement, whereas in the unknown arm sampling case it is with replacement following a success and without replacement following a failure.

To motivate this experiment, consider the following example. Suppose you were part of the national defense strategic team of a small country trying to determine the effectiveness of a new surface to air missile system. All reports on the new system sound promising, however, you are trying to compare it to its alternative (status quo) with a well-established record. These new systems are not fool-proof, and it is your job to make the decision by testing the effectiveness of this new missile system. During the testing period, adjustments can be made on the new system to improve it every time it fails to hit the target. The status quo has been around for a while and its accuracy level is well known; hence, no adjustments are made on this system.

This task tries to capture the essence of the complexities of the one-armed bandit (OAB) problem. As the name suggests, in the OAB the probability of getting a success on one of the arms is stationary and known. On the other arm, the probability of success is unknown. However, it is known that the probability of success increases with each failure. On all conditions, the fixed probability of success in the known arm (p_1) is higher than the initial probability of success in the unknown arm (p_2). (This fact is unknown to the subjects however.)

Each subject faced three conditions labeled I, II, and III (see Table 3 - 1) repeated twice. In each condition, the subjects were presented with three variations of probabilities for urn 2. In other words, subjects were presented with three variations of each condition. The variations were achieved by changing the rate of change in the probability of success for urn 2. We achieved this by merely changing the total number of balls in the urn but maintaining the same initial proportions. For example, in condition I, subjects knew that

the probability of success in urn 1 was .3 and stationary, but on urn 2 they had an unknown and non-stationary probability of success. This was determined to be 0.2 for the first trial in condition I. But, the rate of change was different in each variation, since in the first variation (1a) there were a total of 40 balls in the urn, in the second variation (1b) there were 20 balls in the urn, and in the third variation (1c) there were only 10 balls in the urn. Thus, in the third variation, if the subject were to draw a black ball from urn 2 in the first trial, his chances of getting a white ball from urn 2 in the next trial would increase from .2 (2/10) to .22 (2/9). If the subject had been in the first variation and had drawn a black ball from urn 2 on the first trial, his chances of getting a white ball on the next trial improved from .2 (8/40) to .2051(8/39) in the second trial. Note that after sampling 8 black balls from urn 2 in condition 1c, the probability of success increased from 0.2 to 1.0 and remained there for the remaining trials (if any). To achieve certain success in sampling urn 2 in condition 2b, for example, all 14 black balls were to be sampled (in a total of 30 trials). Thus, the rate of change was faster in the third variation compared to the second variation, and in the second variation compared to the first. Subjects had to complete 18 tasks totally, i.e., they faced each combination of urn 1 and urn 2 two times. See Table 3 - 2 for the order of combinations presented to the subjects.

In condition I, the probability of success in urn 1 was stationary at 0.3. In condition II, the probability of success in urn 1 was .5, and in condition III, it was 0.7. Thus, we have three levels, which we can call low, medium, and high levels of probabilities of success in urn 1. In urn 2 on the other hand, the probabilities are unknown and non-stationary. In every combination, urn 2 started with a lower probability of

success than urn 1. However, depending on the total number of balls in urn 2 and on the color of the balls that were drawn from the urn, the probabilities could increase rather rapidly. The white balls were replaced back in the urn, but the black balls were removed from the urn. Thus, sampling of the white balls only was with replacement. The white balls were replaced back into urn 2 but the black balls were removed from it.

METHOD

Subjects

Advertisements were placed in the school newspaper and postings were made on bulletin boards around campus asking for volunteers. Subjects were told that their payments would be contingent on their performance. Twenty subjects participated in the experiment. The subjects consisted of undergraduate students, graduate students, and a few university employees. None of them had participated in the previous (Chapter 2) experiment.

Procedure

The experiment was conducted in the same manner as the previous experiment (see Chapter 2 for details of the procedure). Subjects received the instructions online and on paper. They were informed that their primary task was to maximize their earnings. They were allowed to summon the experimenter if they had questions. As in the two-armed bandit (TAB) experiment, subjects faced 10 practice trials before the actual trials began. The history of their decisions on each task could be observed by using the F1 key. At the end of each task, the subjects were informed that they were about to start a new

task, and that they would sample from a new set of urns (see Appendix 3 - 1 for the actual instructions and sample screens). Each task lasted for 30 trials.

In Table 3 - 2 we describe the order of conditions as presented to the subject. There were two replications of the experiment. We tried to present the subjects with a low, medium and high condition for urn 1 and for urn 2. We also varied the presentation order of the rates of change. For example, in the first task, the subjects saw a low condition for urn 1, and it was paired with the slowest rate of change for urn 2. In the second task, the subjects were presented with the medium condition for urn 1 and the medium rate of change for urn 2, and so on. Subjects were never told that the tasks would be repeated.

Each time a subject achieved a success she was paid a fictitious unit of currency called a "franc". Francs were converted to real currency at the end of the experiment at the rate of 20 francs = 1 U.S. dollar. After completing the eighteen tasks, the subjects were paid their cumulative earnings, debriefed, and dismissed from the lab. The experiment lasted approximately an hour. The mean earnings per subject were $13.85.

Results

Generic Policies

Unless certain assumptions are made about the subjects' belief in the rate of change, and initial probability of success on urn 2 in trial 1, no optimal policy can be derived. Therefore, we try to focus on describing the policies actually used -- to the extent that they can be identified.

The first policy that we identified is constant switching behavior. Subjects prefer to switch constantly between urns rather than to choose an urn and stay with it. There was no defined pattern to their switching. Sometimes they would switch on every trial, and at other time sometimes they would switch after 2-3 trials. It didn't appear that their decision was based on the result of the previous trial.

Subjects who followed the second policy, sampled both urns for n trials and then picked an urn to stay on. While this "n" might have differed between subjects and conditions, we found that in general most subjects switched between both urns for 10-15 trials (33%-50%) before they decided to pick an urn and stay on it. In the third policy, the subject preferred to stay on one urn for the entire length of the task.

We originally came up with these three decision policies. These three polices have been used by subjects in the previous experiment (Chapter 2). When we tried to classify the decisions of the subjects, we found that many of their decisions could not be captured by these three policies. None of the policies used by the subjects in chapter 2 applied to the subjects in this experiment. Therefore, we identified new policies that the subjects appeared to be following.

The fourth policy is one where subjects stayed on one urn for most of the trials. Occasionally, they would try the other urn. Some subjects would try the other urn sequentially for a few. We did not attempt to restrict the pattern of sampling the other urn, but we did restrict the number of times they sampled from the other urn to a total of 6 or 7 trials. (Anything more than that classify them under a different policy.) Subject 7

used this policy frequently. For example, in condition 1a he started by choosing urn 2 and stayed on it for four trials. Then he switched to urn 1 for 2 trials and switched back to urn 2 for 3 trials. He tried urn 1 again for one trial but switched back to urn 2 for four trials. He sampled from urn 1 for 3 trials and then stayed on urn 2 for 8 trials. He sampled from urn 1 on the next trial and stayed on urn 2 for the remaining trials. Thus, he sampled urn 1 for seven trials and urn 2 for twenty three trials.

Subjects who used the fifth policy tried both urns, but unlike those using the first policy, they would stay on an urn for at least 6-7 trials at a time. They did not exhibit the frequent switching behavior seen in the first policy. Most subjects made about 3-5 switches during the duration of a task. Take subject 3 for example. In the first replication of condition III she stayed on urn 1 for 15 trials, switched to urn 2 for 9 trials and then switched back to urn 1 for the rest of the trials.

In sixth and last policy, subjects stayed on one urn for "n" trials and then switched to the other urn and stayed there until the end of the task. Only a single switch was made through the entire task. This policy is called the one switch policy. These policies are summarized in Table 3 - 3.

Three policies stood out as the most popular, namely, policy 1, policy 3, and policy 4. That is, the most frequently used policies were the constant switching policy, staying on the same urn policy, and staying on one urn for most of the trials policy. Together they accounted for the decisions of approximately 75% of the subjects (Figure 3 - 1).

Table 3 - 4 lists the individual subject policy decisions, and Table 3 - 5 shows the number of times a policy was used in each of the tasks. We find that policy 1 (constant switching) was used in 24.72% of the tasks. It was used most often in the first two conditions that the subjects faced, namely, task 1a and task 2b. As they faced more tasks, subjects appeared to be prefer using different strategies. We find that some subjects used this policy in the first few tasks that they saw and then hardly ever used it in the latter tasks (subjects 1 and 2). Notice that it is never used in the second replication by these two subjects. Subjects 11, 12 and 15 used it in eleven of the tasks that they faced. Other subjects (9, 13 and 16) never used this policy.

Policy 2 (switch for n trials, then pick an urn to stay on) was used on just 7.22% of the tasks. It was used most often by subjects 1 and 19 (4 times), and maybe once or twice by other subjects. Seven of the subjects never used this policy even once.

Policy 3 (stay on the same urn) was the second most often used policy (25.28%). However, it was more often used in the second replication (55 times) than the first (36 times). If we consider their choice of urns, urn 1 was the favored urn 46 times, and urn 2 the favored urn 45 times. It appears that the two urns were chosen almost equally. Subjects favored this policy when the probability of success in urn 1 was high, and they chose urn 1 more often (32/46). When the probability of success on the known urn was low or medium, subjects who chose this policy tended to favor urn 2 (40/45).

Policy 4 was the most used policy (26.94% of all trials). Under this policy, subjects stay on one of the arms most of the time and occasionally try the other urn. As

stated earlier, trying the other urn was restricted to no more than 6-7 times. It was used 45 times in the first replication and 52 times in second replication. As in the previous policy, subjects who chose this policy chose urn 1 more often (58 vs. 39) and used it most frequently when the probability of success on the known urn was high (38/58). Urn 2 was chosen 35 times when the probabilities were low and medium. This appeared to be the intuitive thing to do. The subjects appear to be playing it safe by choosing the urn with the high known probability, but taking a risk when the known probabilities are low or medium.

Policy 5 (Switching after 6-7 trials) was chosen in just 9.44 % of the tasks. Under this policy, subjects stayed on one of the urns for around six or seven trials. There wasn't the constant switching behavior of the first policy. It was used 22 times in the first replication but dropped to just 12 times in the second replication. However, there did not appear to be a clear pattern of choice in this case. The subjects did not appear to be using it more often in one task or another. Subjects 3 and 13 were most partial to this policy using it eight times and six times, respectively.

Policy 6 (One switch policy) was used the least (6.39%). However, even this policy was used 23 times so it could not be ignored. Subjects who used this policy chose one urn for a certain number of trials and then switched to the other urn and stayed there. In most cases, they would switch when the success rates were low, but sometimes they did it even when they were doing well. No clear pattern of choices emerged for this policy either. In the first replication, it was never used when the probability on urn 1 was low, but used a few (4) times in the low probability cases in the second replication.

Descriptive Analysis of the data

Next, we examined the choices between the urns that the subjects actually made. Table 3 - 6 shows that urn 1 was chosen more often than urn 2 (5937 vs. 4863). The most number of successes was achieved in urn 1 (3361 i.e. approximately 31% of all decisions made). To study this in more detail, consider at Table 3 - 7. This table displays the results of the decisions made by the subjects. In tasks 1a, 1b and 1c, the subjects faced the low probability of success (0.3) in urn 1 (fixed, known and unchanging). Subjects tended to choose urn 2 (unknown and changing probability of success) significantly more often and achieved a 44% success rate for that urn compared to the 30% success rate for urn 1. In the medium probability (0.5) case (tasks 2a, 2b and 2c), we find that urn 1 was chosen marginally more often than urn 2 (2006 vs. 1594). The success rate on urn 1 was a little over 49% and the success rate achieved on urn 2 was 48%. In the high probability tasks (0.7) (tasks 3a, 3b and 3c) subjects overwhelmingly chose urn 1 over urn 2 (2932 vs. 668) and achieved the specified success rate (70.80%). Although subjects chose the known urn more often, those who did choose urn 2 did achieve an overall success rate of around 55.39%. It may be of interest to note that there was some similarity on behavior with the subjects in the previous experiment (chapter 2). When the subjects faced the high probability conditions in the standard TAB, their tendencies were to stay on the same arm (can be considered as known since they had some experience trying it) and not test the other arm. When the subjects faced the low probability conditions, the tendency to switch was much higher.

We also study how the subjects changed their behavior between the two replications. Table 3 - 8 shows the decisions that the subjects made by replication. We find that the same pattern of behavior holds for both replications. In the low probability tasks, the subjects chose urn 2 more often. It appears that urn 2 was chosen more in replication 2 than in replication 1. However, there was no way for the subjects to know that this was a repeated task. In the medium condition, urn 2 was chosen more often in both the replications except in the task 2c (replication 2). Subjects had a high rate of success in urn 2 (about 67%) so that would definitely have looked better than the assured 50% success rate and that could have motivated subjects to stay with the unsure choice. In the high probability tasks, subjects were partial to the known urn. They seemed content with the success rate and stayed on that urn most (at least 75%) of the time. However, they were not completely averse to trying the unknown urn. A logistic regression does not show a significant effect for replication at $\alpha = 0.05$.

FURTHER ANALYSES

Next we attempted to test the success rate of alternative simple heuristics that the subject might have used to solve this difficult non-stationary task. We simulated three simple heuristics to see how their performance would compare against the policies used by the subjects. Each simulation was repeated 100 times. We chose three very simple policies to simulate. The first heuristic used the simple rule of "stay on a winner and switch on a loser" for urn 1, and "stay on a loser and switch on a winner" for urn 2 for the first n trials. At the end of n trials, the two probabilities were compared and the urn

with the higher probability was chosen for the rest of the trials. In other words, at the end of *n* trials the program computes the "estimated probability" of success in urn 2 and compares it to the known probability of success in urn 1. The estimated probability is based on the current history for that task. If they had encountered 3 wins in 6 tries on urn 2, then the estimated probability of success would be 0.5.

The second heuristic simulated the simple rule "try urn 2 for *n* trials, then compare the estimated probability of success with the known probability of success in urn 1 and choose the urn with the higher probability for the rest of the trials." For both heuristics, *n* was equal to 10 ± 2. Therefore, n varied from 8 to 12. Additionally, for both heuristics, if the known probability on urn 1 and the estimated probability on urn 2 were equal at the end of *n* trials, the heuristic chose to go with urn 2 for the remaining trials because of the increasing probability of success.

The third heuristic was the simplest of them all, pick the unknown urn for all trials. The idea behind this was that since the probabilities of success were increasing, there might be a chance it would outperform being conservative and staying with the known urn for all trials.

The rational for these heuristics comes from the literature. Zelen (1969) proved that a good heuristic to approximate the optimal policy for the TAB was the "Stay on a winner, switch on a loser" for the first n trials and then pick the one with the higher probability of success for the rest of the trials. His calculations showed that the best results are achieved when $n \cong N/3$. We wanted to check if a variation of that heuristic

would work well in our OAB experiment. We wanted to compare the average earnings of the subjects with the potential earnings that our policy could have generated.

Tables 3 – 9 and 3 – 10 list the outcomes from the first simulation. The numbers in the cells represent the average number of successes for the 100 runs of the simulation. We did find that with each change in task there was a slight increase in the average level of success as expected. In Table 3 - 9, we find that there was no significant change in the success level when we varied 'n' from 8 to12. Therefore, we averaged all the levels of *n* together (Table 3 - 10) for the 30 trials and each of the tasks individually. If we compare the results of having used this heuristic to just having chosen urn 1 (where the expected rate of return was 0.3 in condition I, 0.5 in condition II, and 0.7 in condition III, the simulation it does worse), the heuristic does not perform as well having a slightly lower probability of success in most of the tasks except in task 1c and task 2c. Yet, if the subjects had followed this heuristic, they would have averaged approximately $22.05. That is much higher than the average earnings ($13.85) of the subjects in this experiment. Therefore, in that sense, this heuristic outperformed the decision rules that the subjects were using.

The second heuristic that we used was "try urn 2 for n trials and then compare the actual probability of urn 1 with the estimated probability in urn 2 and stay on the urn with the higher probability." In this case, too, we first collapsed the trials and checked to see if there was any significant difference in the variations in *n* (Table 3 - 11). Whereas n = 8 yielded the best results in 5 out of the 9 tasks, the increase in success rate was marginal. Consequently, we collapsed the various cases of *n*. Table 3 - 12 lists the success rates for

all the trials across all tasks. This heuristic, too, when compared with the fixed probabilities on urn 1 does not do too well. It performed better in tasks 1b, 1c and 2c. In all the other tasks, it performed worse. However, it seems to have outperformed the first simulations in the low and medium conditions. The performance on the high condition was marginally lower. Overall, if subjects had employed this heuristic to make their choice they would have earned on average $22.85. This is considerably higher than the average earnings by the actual subjects and the potential earnings if the first heuristic had been used.

The third simulation that we used was "stay on the unknown (increasing) urn for all trials". The average results are listed in Table 3 - 13. As expected, this heuristic performs the best. When the rate of change is slow ('a' conditions), we find that, on the average, the heuristic performs worse (.28 vs. .3, .40 vs. .5 and .50 vs. .7) than staying on the known arm for Conditions I, II and III. When the rate of change is medium ('b' conditions), we find that when the known probabilities are 0.3 and 0.5, the heuristic produces higher returns on average, but when the probability on the known arm is 0.7, it does not perform as well (.48 vs. .3, .55 vs. .5, and .63 vs. .7). When the rate of change is the fast ('c' conditions), the heuristic outperforms the expected returns on the known arm in all three conditions (.73 vs. .3, .76 vs. .5 and .80 vs. .7). Overall, this heuristic performed exceedingly well. If the subjects had used this heuristic, they could have earned $25.85.

Discussion

The subjects were willing to take a risk when the probability of success on the known arm was low. However, as the probability on the known arm increased to medium or high, the subjects preferred not to venture into the unknown. One explanation for this behavior could be that subjects are exhibiting some sort of satisficing behavior (Simon, 1957). It appears that that when the subjects do 'well', according to some preconceived threshold, they are satisfied with the certain alternative (arm). When that preconceived threshold is not met, they keep searching (switching arms) in order to increase their payoffs. Another heuristic that comes to mind that may explain this behavior is 'anchoring and adjustment' (Slovic and Lichtenstein, 1971). Subjects could be anchoring on the known probabilities and then using that to adjust their expectations. In other words, they know that they can receive a certain payoff with certainty in each condition. If the anchor is low, they may prefer the uncertain but increasing arm, but the anchor is set at a higher figure (.5 and .7) they may prefer to stay with the known arm.

Meyer and Shi (1995) show that "subjects do tend to under-experiment with promising options and over-experiment with unpromising options, and a tendency to increasingly switch between the two arms as the average base rate of success decreases." (We must keep in mind that their experiment did differ from ours in that they held the unknown arm fixed and unchanging where as in our experiment we had an increasing rate of success with each failure.) Our results are just the opposite of theirs: our subjects do choose the promising option more often than the unpromising one. We also found that the switching behavior was not necessarily restricted to the unpromising options. On the

other hand, we found that some subjects had a tendency to switch no matter what the task. Yet, others never ever used this policy. Thus, in our case, it was more a product of subject preferences rather than a task. One other observation was that this switching behavior appeared more often in the early tasks but seemed to disappear as the subjects got more experienced. Now the question remains: are these differences in results a product of the experimental design or can they be something else?

Of the three simulations that we conducted, we found that the third simulation performed the best, on average in terms of earnings. This simulation required subjects to stay on the increasing arm for all trials. Subjects could have earned $12 more on average if they had used this policy. Both the other heuristics also outperformed the subjects in terms of earnings by at least 8 dollars.

In conclusion, we can say our subjects acted intuitively oversampling the promising (increasing) urn when the known probabilities were low but holding back increasingly when the known probabilities changed from low to medium or high.

Table 3 - 1: Distribution of Balls in the urns for One Armed Bandit Experiment

Condition	Color	Urn 1	Urn 2		
I	White	3 **(1)**	8 **(1a)**	4 **(1b)**	2 **(1c)**
	Black	7	32	16	8
II	White	5 **(2)**	12 **(2a)**	6 **(2b)**	3 **(2c)**
	Black	5	28	14	7
III	White	7 **(3)**	16 **(3a)**	8 **(3b)**	4 **(3c)**
	Black	3	24	12	6

Table 3 - 2: Order of conditions for One Armed Bandit Experiment.

Replication	Urn 1	Urn 2	Replication	Urn 1	Urn 2
1	1	1a	2	1	1c
1	2	2b	2	2	2c
1	3	3c	2	3	3c
1	1	1b	2	1	1a
1	2	2c	2	2	2a
1	3	3a	2	3	3a
1	1	1c	2	1	1b
1	2	2a	2	2	2b
1	3	3b	2	3	3b

Table 3 - 3: Summary of the six policies subjects used

Policy Number	Policy Name	Description of Policy
Policy 1	Constant Switching	Constant switching between urns
Policy 2	Switch for 'n' trials, and pick an urn to stay on	Switch between urns for 'n' trials and then stay on one urn for the rest. (For most subjects 'n' varied between 10 -15)
Policy 3	Stay on same urn	Stay on the same urn for all trials
Policy 4	Stay on same urn for most trials	Stay on the same urn for most trials with no more than 7 tries at sampling the other urn.
Policy 5	Switching after 6-7 trials	Switching between urns after 6-7 trials
Policy 6	One switch policy	Stay on an urn for 'n' trials and then make one switch and stay on the other urn for the rest of the trials.

Table 3 - 4: Classification of subject decisions in the order that they saw the tasks.

Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Cond 1a	1	1	1	3	1	4	4	4	3	1	1	1	5	2	1	4	1	1	1	1
Cond 2b	1	1	2	4	4	1	1	1	3	5	4	1	5	1	1	4	1	1	1	2
Cond 3c	2	1	5	3	3	5	4	5	3	6	5	4	6	6	4	3	3	4	4	5
Cond 1b	1	1	5	4	4	3	4	5	3	3	1	1	3	2	1	4	5	4	1	1
Cond 2c	2	1	5	3	5	3	4	4	3	4	1	1	6	1	1	3	1	1	1	1
Cond 3a	1	2	5	3	3	6	4	2	6	4	1	4	5	3	4	3	4	1	4	1
Cond 1c	1	5	2	4	3	3	4	4	3	3	1	1	3	3	1	3	2	2	4	5
Cond 2a	1	4	5	1	3	3	1	1	3	4	1	6	5	6	1	3	5	4	5	4
Cond 3b	2	4	4	3	3	3	4	4	6	4	4	1	5	4	4	3	4	4	3	1
Cond 1c	4	6	4	2	3	3	3	4	3	3	1	1	3	3	1	3	4	3	2	1
Cond 2c	2	4	4	5	3	3	4	2	6	3	1	1	3	2	2	3	4	4	2	4
Cond 3c	4	4	3	3	3	3	3	5	3	4	4	1	6	4	2	3	3	3	3	1
Cond 1a	4	4	5	4	3	5	5	4	3	1	4	1	4	6	1	3	1	3	6	1
Cond 2a	4	5	5	3	6	4	4	1	3	2	1	1	4	5	1	3	4	1	1	1
Cond 3a	4	4	3	3	3	4	4	5	6	6	4	4	5	3	4	3	4	6	3	1
Cond 1b	4	2	5	3	3	6	3	4	3	2	1	4	3	3	1	3	1	1	2	1
Cond 2b	3	4	1	4	3	6	1	4	6	3	1	4	4	1	1	3	4	1	2	4
Cond 3b	3	4	1	3	3	3	4	4	3	2	4	4	6	3	4	3	4	6	4	5

Note: The top half is the first replication and the bottom half is the second replication.

Table 3 - 5: Number of times a policy was used in each of the conditions.

	Policy 1	Policy 2	Policy 3	Policy 4	Policy 5	Policy 6
Condition 1a	12	1	2	4	1	0
Condition 1b	7	1	4	5	3	0
Condition 1c	4	3	7	4	2	0
Condition 2a	6	0	4	4	4	2
Condition 2b	11	2	1	4	2	0
Condition 2c	9	1	4	3	2	1
Condition 3a	4	2	4	6	2	2
Condition 3b	2	1	5	10	1	1
Condition 3c	1	1	5	5	5	3
Condition 1a	5	0	4	6	3	2
Condition 1b	5	3	7	3	1	1
Condition 1c	4	2	9	4	0	1
Condition 2a	7	1	3	5	3	1
Condition 2b	6	1	4	7	0	2
Condition 2c	2	5	5	6	1	1
Condition 3a	1	0	6	8	2	3
Condition 3b	1	1	7	8	1	2
Condition 3c	2	1	10	5	1	1

Table 3 - 6: Summary of subject decisions and the outcomes

DECISION	RESULT		
	FAILURE	SUCCESS	TOTAL
URN 1	2576	3361	5937 (54.97%)
URN 2	2561	2302	4863 (45.02%)
TOTAL	5137 (47.56%)	5663 (52.43%)	10800

Table 3 - 7: Results of the decisions made by subjects (by condition).

Condition	Decision	Result		Total
		Failure	Success	
1a	Urn 1	280	118	398 (33.16%)
	Urn 2	598	204	802 (66.83%)
	Total	878	322 (26.83%)	1200
1b	Urn 1	261	123	384 (32%)
	Urn 2	519	297	816 (62%)
	Total	780	420(35%)	1200
1c	Urn 1	161	56	217 (18.08%)
	Urn 2	317	666	983 (81.92%)
	Total	478	722 (60.17%)	1200
2a	Urn 1	348	314	662 (55.17%)
	Urn 2	344	194	538 (44.83%)
	Total	692	508 (42.33%)	1200
2b	Urn 1	382	387	769 (64.08%)
	Urn 2	260	171	431(35.92%)
	Total	642	558 (57.25%)	1200
2c	Urn 1	288	287	575 (47.97%)
	Urn 2	225	400	625 (52.08%)
	Total	513	687 (66.92%)	1200
3a	Urn 1	297	725	1022 (85.16%)
	Urn 2	100	78	178 (14.83%)
	Total	397	803 (66.92%)	1200
3b	Urn 1	293	664	957 (79.75%)
	Urn 2	110	133	243 (20.25%)
	Total	403	797 (66.42%)	1200
3c	Urn 1	266	687	953 (79.41%)
	Urn 2	88	159	247 (20.58%)
	Total	354	846 (70.5%)	1200

Table 3 - 8: Decisions made by subjects in each replication

REPLICATION 1				
Condition	Decision	Result		Total
		Failure	Success	
1a	Urn 1	150	67	217
	Urn 2	279	104	383
	Total	429	171	600
1b	Urn 1	134	64	198
	Urn 2	259	143	402
	Total	393	207	600
1c	Urn 1	93	26	119
	Urn 2	159	322	481
	Total	252	348	600
2a	Urn 1	177	137	314
	Urn 2	180	106	286
	Total	357	243	600
2b	Urn 1	175	172	347
	Urn 2	154	99	253
	Total	329	271	600
2c	Urn 1	180	181	361
	Urn 2	99	140	239
	Total	279	321	600
3a	Urn 1	143	363	506
	Urn 2	53	41	94
	Total	196	404	600
3b	Urn 1	163	345	508
	Urn 2	51	41	92
	Total	214	386	600
3c	Urn 1	127	328	455
	Urn 2	50	95	145
	Total	177	423	600

REPLICATION 2				
Condition	Decision	Result		Total
		Failure	Success	
1a	Urn 1	130	51	181
	Urn 2	319	100	419
	Total	449	151	600
1b	Urn 1	127	59	186
	Urn 2	260	154	414
	Total	387	213	600
1c	Urn 1	68	30	98
	Urn 2	158	344	502
	Total	226	374	600
2a	Urn 1	171	177	348
	Urn 2	164	88	252
	Total	335	265	600
2b	Urn 1	207	215	422
	Urn 2	106	72	178
	Total	313	287	600
2c	Urn 1	108	106	214
	Urn 2	126	260	386
	Total	234	366	600
3a	Urn 1	154	362	516
	Urn 2	47	37	84
	Total	201	399	600
3b	Urn 1	130	319	449
	Urn 2	59	92	151
	Total	189	411	600
3c	Urn 1	139	359	498
	Urn 2	38	64	102
	Total	177	423	600

Table 3 - 9: Aggregate results of Simulation 1

	Result 1a	Result 1b	Result 1c	Result 2a	Result 2b	Result 2c	Result 3a	Result 3b	Result 3c
'n' = 8	26.30	33.43	44.13	45.03	46.30	54.77	63.57	65.47	66.57
'n' = 9	27.87	32.40	44.63	42.90	46.13	52.33	64.73	64.97	67.27
'n' = 10	26.67	29.8	37.67	43.57	45.60	53.27	64.17	64.56	66.87
'n' = 11	26.75	33.43	42.35	45.03	46.57	54.77	3.49	65.47	66.58
'n' = 12	26.57	31.80	46.97	43.67	46.40	53.47	62.5	63.53	66.43
Ave.	26.74	32.17	43.51	44.04	46.14	53.72	63.70	64.79	66.74

Simulation 1: Stay on winner and switch on loser in urn 1 for first n trials; Stay on loser switch on winner in urn 2 for first n trials; then compare known probability of urn 1 with estimated probability of urn 2, and stay on the urn with the higher probability for rest of trials

Table 3 - 10: Summary of Simulation1 across all conditions.

Trial No	Result 1a	Result 1b	Result 1c	Result 2a	Result 2b	Result 2c	Result 3a	Result 3b	Result 3c
1	23.8	28.6	28.2	40	40.4	38.6	55.4	58.6	56.8
2	23.6	21.6	23	39	39.6	38	57	55.6	58.2
3	20.4	24.4	21.6	34.8	39.6	37	57.8	58.8	55.2
4	25	22	26.4	35	38.6	44	58.8	62.8	57
5	24.4	24.2	26.4	34.6	38.4	44.6	60.2	61.2	58.8
6	20	22.8	28	34.8	40.4	46.8	59.4	59.4	61.8
7	21.4	27.2	35	40	37.4	44	58.2	62.6	64.6
8	20.4	22.6	31	34.8	36.6	46	59.2	52.6	65.8
9	25	28	34.6	38.8	46.6	43.6	56.6	63	64.4
10	23.8	29	35.4	42.2	43.2	45.8	60.6	63.2	64.4
11	25.8	28.6	43	37.6	45.2	47.4	61.8	65.8	65.8
12	28.8	26	34.8	47.4	44.2	50.8	67.4	66.2	66.6
13	25.2	31.8	35.8	46	44.2	51.6	66.8	66.4	67.2
14	26.8	28	36.2	46	47.2	43.8	65.4	64.8	65
15	30.2	29.4	38	45.2	43.8	49.8	65	67	65.2
16	26.8	33.2	44	40.6	46.8	54.2	65	68.8	67.6
17	27.6	30.6	47.4	47.8	46.8	59.6	66.2	63.2	72.4
18	27.2	34.6	51.2	44.8	47	59.6	66.4	65.4	64.6
19	29.2	32.2	50.2	45.4	46.2	60.2	64.6	68.4	65.4
20	28.8	32.6	54.4	50.4	52	59	67.6	67.4	71
21	28	37.8	52.2	49.6	49.6	65	67.6	62.8	72.2
22	30.6	39	55.8	50.8	51	60	63.6	67.2	65.4
23	26.8	36	59.2	43.6	52.8	62.4	70.4	66.8	69.2
24	31	38.2	55.6	50.4	51.4	61.6	66.4	69	73.8
25	29.4	40.8	57.4	50.2	49	61.6	64.6	64.8	73.4
26	27.4	41.6	60.8	50.2	48.2	65.6	66.6	74.2	77
27	31	41.2	59.6	52.8	54.8	64.6	69.6	67.6	74.4
28	31	41.4	58.2	47	51.8	66.4	66	71.4	71.8
29	33	42.8	60	55.6	53.4	70.6	68	65.4	74.4
30	29.8	49	61.8	45.8	58.2	69.4	69	73.4	72.8
Average	26.74	32.17	43.51	44.04	46.15	53.72	63.71	64.79	66.74

Table 3 - 11: Aggregate of Simulation 2

	Result 1a	Result 1b	Result 1c	Result 2a	Result 2b	Result 2c	Result 3a	Result 3b	Result 3c
'n' = 8	28.10	34.33	49.97	43.73	49.83	60.27	61.73	63.27	67.43
'n' = 9	26.83	33.13	50.40	42.77	45.27	54.97	60.23	63.33	65.17
'n'= 10	27.50	32.07	50.10	42.83	46.23	62.47	60.73	61.87	66.23
'n'= 11	26.93	33.70	60.77	43.37	46.63	55.97	60.47	61.87	64.47
'n'= 12	27.23	35.53	73.33	42.77	47.37	66.93	58.73	62.83	65.73
Average	27.32	33.75	56.91	43.09	47.07	60.12	60.38	62.63	65.81

Simulation 2: Try urn 2 for first 'n' trials. Then compare the known probability of urn 1 with the estimated probability of urn 2 and stay on the urn with the higher probability for the rest of the urns.

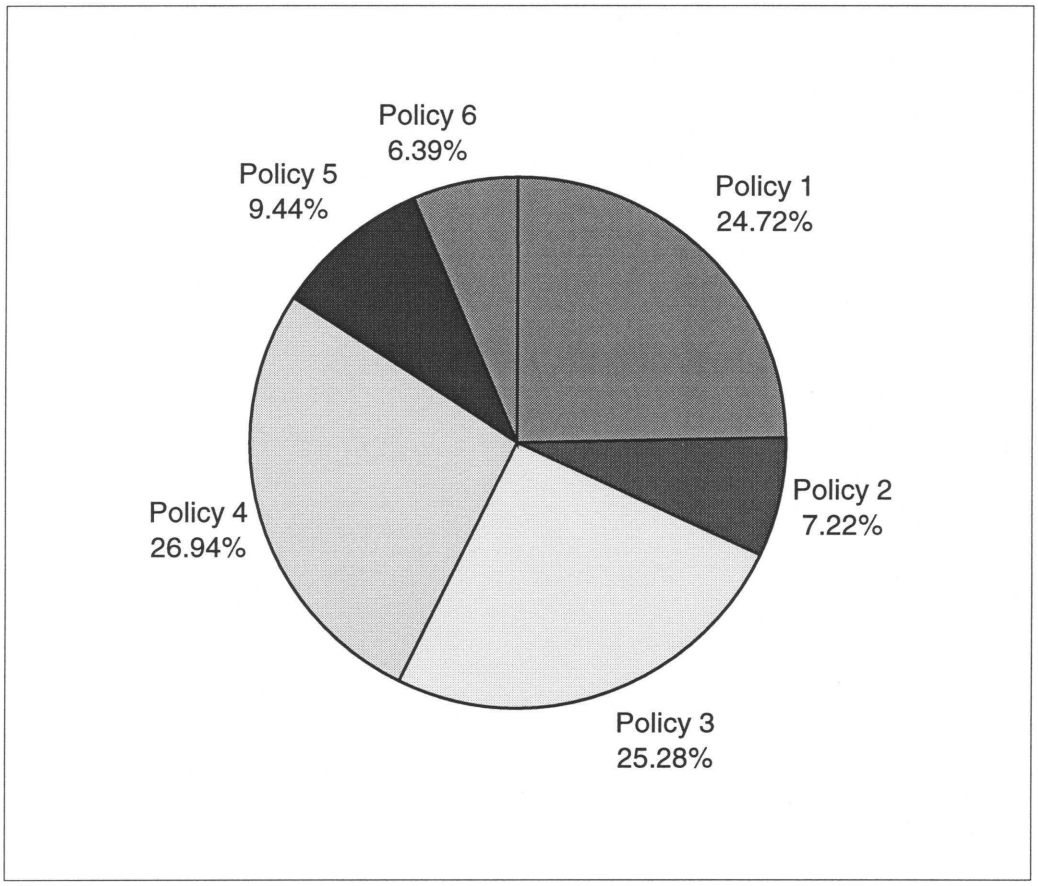
Table 3 - 12: Summary of Simulation 2 across all conditions

Trial No	Result 1a	Result 1b	Result 1c	Result 2a	Result 2b	Result 2c	Result 3a	Result 3b	Result 3c
1	20.4	22.4	20.2	28.6	31	29	44	40.4	40.6
2	20.2	19.6	23	27	33.2	31.8	42.2	43.4	42.8
3	21.4	19.6	24	28.8	34.2	32.6	41.4	41	47.4
4	19.8	25.4	30	30.4	33.4	44.2	41	43.2	49.4
5	20.6	24.6	31	32.4	34.2	46.8	45	43.8	52.6
6	18	21.8	29.4	34.8	38.6	42.8	41.2	47.2	52
7	20.8	26.8	38	36	36.8	47.6	43	47.6	59
8	23.2	25.4	38	39	37.2	51.8	44.4	54.8	67.4
9	24.6	27.2	46.6	38.8	43.8	62.4	51	56.2	68.2
10	28.6	32	50	39.2	44.6	57.4	56.6	61.2	73.2
11	26.4	32.8	49.8	43.2	45.4	56.6	61.8	65.2	72.2
12	27.4	34	51.8	48.2	46.2	58.6	61.4	66.2	69.4
13	32.2	32.4	56.6	44.4	46.6	60.4	67.4	69	69.2
14	24.6	29.8	63.2	49.2	48.6	61.4	69	69.4	70.6
15	25.6	32.6	62.8	54.4	46.2	59.6	68.2	72.6	68.8
16	29.6	32	67.2	47	49.4	65.4	70.6	67.8	71.2
17	32.2	32.2	66.6	47.4	54.4	66.6	69	71.4	72.4
18	29.4	36.2	69.4	47.6	56.2	68.4	66.8	69.6	72.4
19	33.2	37.8	71.2	50.6	52.2	70	68.8	71.4	70.4
20	28.4	37.4	73.8	44.4	51.2	72	65.4	70.6	69.2
21	32.8	39.8	72	46.2	53.6	69.2	67.6	69	74.4
22	32.8	37.6	74.4	49.8	49.2	69.4	70.4	68.2	73.6
23	31.4	39.6	75.2	49.4	50.8	70.8	68	72.2	73.4
24	30.2	42	72.8	48	56.8	71.2	67	71	68.8
25	27.2	38.8	76.2	48	60.2	71.4	69.4	69	68.6
26	30.8	46	74.4	43.8	52.4	72.8	70.4	68.8	68
27	31.8	45.6	74.2	47.6	56	74.2	73.2	73.6	71.8
28	31.2	44	74	51.4	59	72	68.8	71.6	73
29	30.8	48.2	75.2	52.2	56.2	74.4	71.2	69.6	73.2
30	34	49	76.4	45	54.4	72.8	67.2	74	71
Average	27.32	33.75	56.91	43.09	47.07	60.12	60.38	62.63	65.81

Table 3 - 13: Summary of Simulation 3 across all conditions.

Trial No.	Result 1a	Result 1b	Result 1c	Result 2a	Result 2b	Result 2c	Result 3a	Result 3b	Result 3c
1	15	24	26	26	38	32	42	45	40
2	18	17	16	29	42	37	45	39	40
3	18	25	19	38	37	31	43	38	47
4	22	25	40	37	27	37	35	39	61
5	16	23	29	26	33	40	40	40	49
6	17	26	34	41	42	47	52	45	57
7	23	28	37	32	30	59	42	50	60
8	22	22	41	34	47	48	36	52	58
9	28	20	40	35	29	51	47	54	69
10	30	28	55	44	35	62	44	60	67
11	25	40	64	38	40	75	40	59	80
12	31	32	77	35	48	78	49	55	78
13	30	36	76	37	53	85	42	60	80
14	23	40	85	35	43	84	49	67	85
15	31	36	92	38	49	83	51	61	90
16	23	49	90	35	55	88	52	72	87
17	27	43	91	48	54	94	52	61	88
18	30	47	95	34	58	95	50	65	96
19	30	48	99	47	56	94	57	63	95
20	28	58	98	45	64	98	52	71	92
21	31	68	99	43	69	96	54	74	96
22	27	69	98	47	69	99	53	81	96
23	32	68	100	43	69	98	56	77	99
24	32	70	99	48	74	98	62	79	99
25	41	75	100	41	88	96	58	78	97
26	40	85	100	41	86	96	59	85	100
27	41	83	100	53	85	100	60	80	100
28	39	87	100	51	88	100	63	89	99
29	41	86	100	45	83	99	62	81	99
30	51	91	100	55	88	100	64	92	99
	28.73	48.30	73.33	40.03	55.97	76.67	50.37	63.73	80.10

Figure 3 - 1: Distribution of decisions between strategies in the OAB



Policy 1: Random Switching Behavior
Policy 2: Switch for 'k' trials, and then stay on one urn for the rest of the trials
Policy 3: Stay on the same urn for all trials
Policy 4: Stay on one urn for most trials (except 5 or 6 trials)
Policy 5: Switch after 6 or 7 trials.
Policy 6: Stay on an urn for 'n' trials, then switch and stay on the other for the rest of the trials.

INSTRUCTIONS TO SUBJECTS FOR THE ONE ARMED BANDIT

You are about to participate in a computer-controlled decision making experiment in which your payoff is contingent on your performance. The money you will earn will be paid to you at the end of the experiment. A research foundation interested in clinical decision-making has contributed the money to finance the study.

The major purpose of this experiment is to study how individuals behave when asked to make a sequence of choices between two alternatives, one of which yields a success with known probability, and the other yields a success with an unknown probability. The task that we have in mind is the sequential testing of two new medicines, (whose properties have not been established), on a series of patients. At any time only one drug can be administered to a patient, and the treatment is scored as either a success or a failure. Only when the result of the treatment is known, is the next patient administered one of the two drugs. Clearly, one would like to maximize the number of successes; however, only by trying each of the drugs can the physician learn which is more effective.

It would be simpler to describe the task that you are about to perform in terms of two urns, labeled 1 and 2. Suppose that each urn has some black and some white balls in it. The total number of balls may vary from one urn to another. You are not informed of the total number of balls in each box, or the proportion of black and white balls in one of the urns.

Your task is simple. You will make a series of drawings from the urns, one on each trial. The total number of drawings (denoted by T), will be known to you when the task starts. At the beginning of each trial, your task is to decide which urn to draw a ball from. Choose an urn, draw a ball blindly, and record its color (either black or white).

Your payment on each trial depends on the ball you draw.

If you draw a white ball, you will be paid one unit of money, called "franc". (Francs will be converted into US dollars at the end of the experiment). Please replace the white ball back into the urn from which you drew it.

If you draw a black ball, you will receive no payment. If the ball is withdrawn from urn 1, please replace the black ball into the urn; however, if the ball is withdrawn from urn 2, please remove the ball back from the urn.

This completes the trial. To repeat, once you have noted the color of the ball, you replace the ball if it is white, back into the urn from which you withdrew it. However, if it is black, you replace it if it was drawn from urn 1, but remove it if it was drawn from urn 2. You get paid one franc for a white ball and none for a black ball. Hence, each time you draw a black ball from urn 2, you increase your chances of getting a white ball on the next trial. The essence of this decision making task is to gather information about the unknown urn and at the same time make decisions that increase your payoff.

Your objective in the task is to draw as many white balls as possible, thereby increasing your earnings.

Example

Supposing the two urns are composed as follows and that the number of balls is known:

URN	# WHITE BALLS	# BLACK BALLS
1	8	4
2	10	5

Suppose that you choose urn 1 on trial 1. If you draw a white ball (and receive a payment of 1 franc), then on the next trial the composition of the urns will be the same. If you draw a black ball from the other urn (and receive a payment of zero), then on the next trial the composition of the two urns will be:

URN	# WHITE BALLS	# BLACK BALLS
1	8	4
2	10	4

Having drawn a black ball from urn 2, the chances of drawing a white ball have increased from 10/15 to 10/14 i.e. from 67% to 71%.

Continuing the same example, supposing that your drawings are as follows:

TRIAL	URN	COLOR	PAYOFF	# W IN URN 1	# B IN URN 1	# W IN URN 2	# B IN URN 2
1.	1	WHITE	1	8	4	10	5
2.	1	WHITE	1	8	4	10	5
3.	1	BLACK	0	8	4	10	5
4.	2	WHITE	1	8	4	10	5
5.	2	BLACK	0	8	4	10	4
6.	2	BLACK	0	8	4	10	3

Thus, after 6 trials, you have earned 3 francs. The chances of drawing a white ball from urn 1has not changed but the chances of drawing a white ball from urn 2 has improved to 10/13 (77%).

Drawing from two urns and keeping track of the number of balls drawn is a tedious task. Therefore, to simplify the task, we have computerized the entire process. Your only decision will be to choose an urn on each trial. The computer will then randomly draw a ball from the urn you have chosen, exactly as in the example above, and will inform you of your payoff for the trial and your accumulated payoff from the beginning of the experiment.

You will participate in 18 different tasks. At the beginning of each task, you will be informed of the total number of trials in the task. You will also be reminded that the new task has two different urns with possibly different compositions of white and black balls.

You may wonder how the (unknown) proportion of white balls in each urn has been determined. The only information that we can give you is that the percentage of white balls in each of the two urns can assume any value between 0 and 100 with all values in this range being equally likely.

Once the task begins the computer will display the following information on the screen before you make your decision:

History of previous play

Trial # urn 1 urn 2 cumulated payoff

You are on trial number (1...T)

The probability of winning on urn 1 is _____.

Which urn would you like to sample from? (1 or 2)

For example, suppose you sampled from urn 1 on the first trial and you picked a black ball, and from urn 2 the next time and picked a white ball. Your screen would then look similar to this:

History of previous play

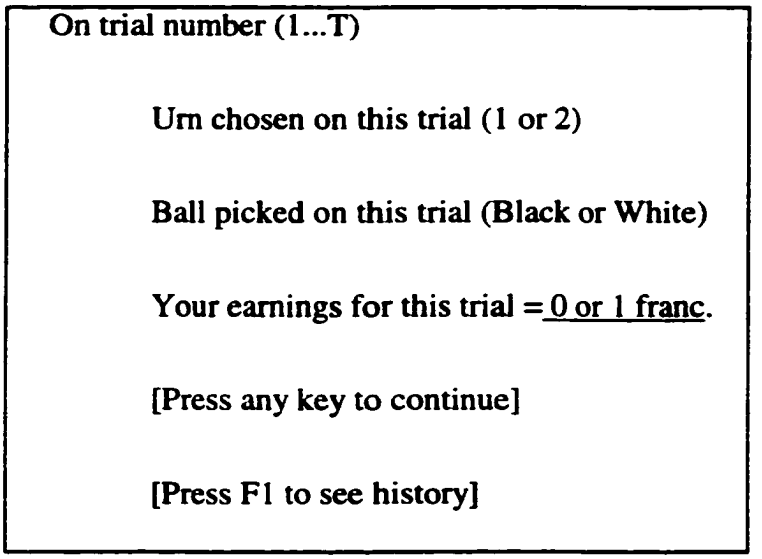
	Outcome		
Trial #	urn 1	urn 2	cumulated payoff
1	Black	----	0
2	-----	White	1

You are on trial number (1...T)

The probability of winning on urn 1 is _____.

Which urn would you like to sample from? (1 or 2)

Once you enter either '1' or '2', the computer will randomly draw a ball from that urn and present you with the following screen.



Following the presentation of the outcome of the trial, the computer will update your record and return you to the first screen for the next trial. This process will be repeated for 'T' trials. At any time during the task you will be able to check the history for trials that are no longer visible (the history screen will hold the results for about 20 trials) by using the F1 key that is located at the top left hand side of your keyboard.

As mentioned, you will participate in a series of 18 tasks. The number of trials will differ from one task to another. The two urns will also change between the tasks. In other words, the proportion of the white and black balls originally in each of the urns will change between the tasks.

After each task, you will see this screen:

You are about to begin a new sequence of '30' trials.

You will sample from a new pair of urns.

The urns are still labeled 1 and 2 but they are not the same ones used in the previous task.

Please keep in mind that the computer has not been programmed to play against you. Rather, the number of white balls and black balls in each urn has been predetermined and will change from trial to trial (depending on the ball that is picked) within the next 30 choices. Your task is to draw as many white balls as you can over the next 30 trials.

Your concern is to make as much money as you can, given these conditions. You are being paid in a fictitious currency called a "franc". "Francs" will be converted into US dollars at the end of the experiment at the rate of $1 = 20 francs. You will be paid before leaving.

If you have any questions about the instructions, please inform the experimenter immediately. You may refer to the written copy of the instructions placed on your desk at any time. Please take whatever time necessary to make your decisions in this experiment.

CHAPTER 4: TWO-ARMED BANDIT WITH TWO INCREASING PROBABILITIES

INTRODUCTION

In the last two chapters, we dealt with the standard TAB and a variation of the OAB problems. In this chapter, we investigate a variant of the TAB problem. In the classic TAB, the arms are stationary (i.e., the probabilities of success do not change with each trial). In the present variation, we introduce increasing probabilities of success with the encounter of each failure. If the subject encounters a success there is no change in the probability of success, however, every time he encounters a failure, the probability of success increases in the next trial.

An analogy for the third experiment can be drawn from the medical example described in the introduction. Patients are seen one by one at a treatment center, and the physician has to decide which treatment to use on the patients. However, with each patient who does not recover, the treatment is fine-tuned and improved. This can be done because now the doctors/surgeons have incrementally increased their knowledge. It could be that the doses of the medication are slightly altered or they may actually perfect the surgical procedure a little more. (We are making an assumption here that the impetus to improve does not exist after a success or it could be that they get complacent after a success.) Now, we are dealing with a case where the probabilities of success are non-stationary and increasing.

Each subject faced four conditions (see Table 4 - 1). Each condition was repeated four times in a round robin fashion. Each round consisted of all four conditions and each subject faced four rounds. Thus, subjects faced a total of 16 tasks. In addition, in the even numbered rounds we also switched around the labeling of the urns (i.e., what was called urn 1 in round 1 was now called urn 2). This was done so that the subjects would not realize that we were repeating the same tasks and also to discourage subjects from picking an urn and staying with it till the end. Table 4 -2 lists the labeling of the urns as presented to the subjects. With the original labeling, urn 2 always had the faster rate of change since there were fewer balls in urn 2, except for condition 4 where both urns were identical.

In condition 1, both urns have an identical probability of success (0.4), but they increase at different rates. Urn 2 had exactly half the number of black and white balls (as compared to urn 1) in it, hence the rate of change in urn 2 was higher than in urn1. In condition 2, the two urns have different starting probabilities of success and their rate of change is different. The urn that had the lower percentage of white balls (3/15 =0.2) also had fewer balls; hence, the rate of increase is much faster than the other urn, which starts at a higher level (16/40=0.4) but also has a larger number of balls. In condition 3, the two urns start with very low probabilities of success (0.05 in urn 1 and 0.1 in urn 2) but increase rapidly at different rates. The justification for this is two treatments that are totally new with no history whatsoever, hence no one really has any idea what the success rates are. However, with the passage of time the treatments are improved. In condition 4, both urns have equal probabilities of success (0.6), and identical rates of increase. This

condition was introduced as a control condition to see how subjects would behave when the two urns are identical, and if they could pick it up over the course of thirty trials that they played.

METHOD

Subjects

Subjects were recruited in the usual manner (see chapters 2 and 3) through advertisements/bulletin board postings around campus. Subjects were told that their payments would be contingent on their performance. Any subject who had taken part in prior bandit experiments were not allowed to participate.

Procedure

Refer to Chapter 2 for details. All the experiments were conducted in separate rooms at the Behavioral Decision Lab. All were computerized and all instructions were available online. A hard copy of the instructions was also available (see Appendix 4 - 1). Subjects were allowed to make any notes they wanted over the course of the experiment using a blank piece of paper and pencil. The instructions explained the task at hand in detail and gave extensive examples. The subjects were informed that their primary purpose was to maximize their earnings (by selecting white balls), given the uncertainty of the tasks. At the end of a practice round with 10 trials, they were told that the real trials were about to begin and that they would be paid for correct decisions.

Subjects were asked to imagine two urns in front of them and were told that their task would be to pick a ball from an urn. Every time they picked a white ball (success),

they would figuratively replace the ball in the urn. Each time they picked a black ball, it was removed from the urn. The idea was that each time they encountered a failure, they removed the black ball from the urn thereby increasing their chances of getting a white ball (success) in the future. In other words, p_1 and p_2 increased as a failure was encountered for the corresponding urns. The total number of trials for each task was known and fixed at 30.

The subjects were informed that each urn contained several white and black balls, but the actual composition of the urns would be unknown. They were also told that sampling was with partial replacement in both urns (i.e. only the white balls would be replaced, the black balls were removed). Thus, the proportion of white balls in each urn was unknown and could vary from trial to trial.

At every stage in the experiment, the subjects were informed of the trial number. They could also observe the outcome of the earlier trials within the same task by pressing the F1 key. At the end of each task, the subjects were informed that they were about to start a new task, and that they would sample from a new set of urns (see Appendix 4 -1 for sample screens).

Each time the subject achieved a success he was paid a "franc", a fictitious unit of currency. Francs were converted to dollars at the rate of 15 francs = 1 U.S. dollar. After completing the sixteen tasks, the subjects were paid their cumulative earnings, debriefed, and dismissed from the lab. The experiment lasted approximately an hour, and the mean earnings per subject were $ 16.20.

RESULTS

Generic Policies

We attempted to identify and explain what the subjects were doing when they faced the various conditions. All of the policies that the subjects used in this experiment were almost identical to the ones used in the OAB experiment, with minor variations. This is not surprising since the OAB experiment had one increasing arm and one stationary arm. (For a summary of the policies see Table 4 - 3.) The only difference between the policies in the two experiments was the duration for which they carried out some of the variable policies. For example, in policy 4 of the OAB (stay on the same urn for most trials) subjects sampled the other urn only 5-6 times, but in the present experiment some tried it up to 7 times. In the case of the OAB and policy 5, subjects switched after 6-7 trials whereas in this experiment subjects switched after 5-6 trials.

The first policy that we identified was the constant switching between the urns. Subjects were not consistent in their pattern of switching behavior; in other words, they were exhibiting inconsistent behavior. Sometimes they would switch after a win and at other times they would switch after a loss. Sometimes they would switch after a few trials and at other times they would switch after 1 trial.

In the second policy, we found that the subjects would switch for "n" trials and then pick an urn to stay on it. In most cases (at least 70% of the time), they chose the urn in which they had achieved the higher success rate. However, occasionally, they chose the urn with the lower probability. We fixed 'n' be $\leq$ 15 (i.e., no more than 50% of the

trials), since otherwise the subjects would have been classified under another policy. However, most (approximately 75%) subjects picked an urn to stay on after 8-10 trials. A couple of subjects switched for 15 trials before deciding on which urn to stay.

In the third policy, subjects chose an urn and stayed on it for the entire duration of the task. We call this the stay on an urn policy.

Beside these three main policies, other policies were also followed. In the fourth policy that we identified, subjects stayed on one urn for most of the trials. We restricted the number of times they could try the other urn to a maximum of 5-6 times (i.e. no more than 20% of the trials). If they tried it more than that, we classified them under another policy.

In the fifth policy, subjects exhibited switching behavior again. Unlike the first policy (where they repeatedly switched urns in an inconsistent manner), in this policy subjects stayed on an urn for approximately 5-7 trials at a time, hence there were fewer switches.

The last policy is one where subjects stayed on an urn for 'n" trials, then switched to the other urn and stayed there. It appears that for some reason they were displeased with the returns on the current urn and decided to switch and stay with the new urn. In this policy we set $n \leq 10$. This policy will be called the one switch policy.

We attempted to categorize the decisions of the subjects to categorize them under the various policies identified above. The first thing we did was to determine what each subject did for every task that he/she faced. We then classified their actions for each task

into a policy. Some policies were straightforward. For example, if a subject stayed on urn 1 for every trial in the task we could easily classify him as using policy 3. If, on the other hand, she repeatedly switched between urns, we could say that she used policy 1. On occasion, things were not as clear. The subjects would appear to be switching repeatedly for, say, 10 trials and then they would stay on each urn for say 6-7 trials at a time. In that case, we had to make the decision to go with policy 5, since that was the policy the subjects used for the majority of the trials.

A graphical representation of the choices of the subjects can be seen in Figure 4 - 1. Policy 1 was used most often. Subjects repeatedly switched between urns in 24.38 % of the tasks. Subjects 9 and 12 used this policy in 14 of their 16 tasks. Subject 2 used it 10 times, and subject 8 used it 9 times. All the others used it less than 6 times. Three subjects (4, 11 and 20) never used this policy.

Policy 2 (switching for 'k' trials, and then picking one urn and staying on it for the rest of the trials) was the second most popular policy (23.75%). It was never used more than 9 times by any subject but it was used by all subjects at least once.

Policy 3 (staying on the same urn) was used in 15.94% of the tasks. At least 6 subjects never used this policy even once. Only three subjects used it more than 5 times. This is not surprising given that subjects may be hesitant to just stay on an urn without ever trying the other urn unless they were especially lucky (many white balls) in the initial trials.

Policy 4 (staying on an urn for most of the trials) was used in 15.63% of the tasks. Only subjects 15 and 19 used it more than 5 times. Subjects 2, 9 and 12 never used it.

Policy 5 was used in 12.81% of the tasks. This was the switching policy where subjects stayed on an urn at least 5-7 trials before switching. Subjects 10, 16 and 19 used it more than 5 times. Seven of the subjects never used it even once.

Last, but not least, we have policy 6 that was used in 7.5% of the cases. Seven subjects never used the policy and none of the subjects used the policy more than 5 times.

While it may be difficult to understand why the subjects chose one policy over another, we can look at their earnings to see if that justified their behavior. Subjects 5 and 6 had the highest earnings. Looking at their policy usage (See Tables 4 - 4 and 4 - 5), one finds that they used policy 3 and 4 most often. Subjects who used policy 1 very frequently (12 and 9) were among the lowest earners. While we cannot generalize, it seems that subjects who used policy 3 more often were among the higher earners, while the lowest 3 earners never used this policy even once (12, 9 and 19). Subject 19 did not use policy 1 very frequently, but his most used policy was policy 5 (also based on switching between urns). Similarly, the fourth and fifth lowest earners (subjects 10 and 16) were among the highest users of policy 5.

Descriptive Analyses of the subjects' decisions

Table 4 - 6 represents the choices made by the subjects collectively. Urn 1 was chosen in approximately 54% of the trials and urn 2 was chosen in 46% of the trials. Overall, subjects achieved success in a little over 50% of the trials. The breakdown of

success rate is almost identical for urns 1 and 2. Subjects achieved a 51% success rate in urn 1 as opposed to a success rate of a little less than 50% in urn 2. Thus, while there is a significant difference ($\alpha < .05$) between the number of times urn 1 and 2 were chosen, there is no significant difference between the success and failure rates in each of the urns. Subjects failed to choose urn 2 (original labeling) which had the faster rate of change more often.

In Table 4 - 7, we break down the decisions of subjects by the various conditions that they faced. We find that in conditions 1 and 4 there is a significant difference ($\alpha <$ .05) between the number of times urn 1 and urn 2 are chosen (59.71% vs. 40.29% and 54.5% vs. 45.5%) In both cases, urn 1 was chosen with higher frequency. In condition 2 the split between urn 1 and urn 2 are identical. Urn 1 was chosen slightly more often than urn 2 in condition 3 but the difference is not significant (51.5% vs. 48.5%). Thus, in this experiment it is clear that the subjects preferred to choose urn 1 to urn 2. The only logical explanation for the subjects preference for urn 1 could be the natural tendency of the subjects to start with urn 1. The way the tasks were presented to them, there was no other factor that would help them to distinguish between the two urns in the first trial.

To see if the failures and successes played a part in the choice of urns, we categorized the decisions of the subjects by task in Table 4 - 7. In condition 1 there does not appear to be a significant difference between the failures and successes (1214 vs. 1186). However, there is a significant difference between the two in the other three conditions. In condition 2 (1366 vs. 1034) and 3 (1420 vs. 980) there were more failures

than successes, but in condition 4 there were significantly more successes (1632 vs. 768). This is not suprising since there were a large proportion of white balls in both the urns (60%) in condition 4. However, it is difficult to explain why they had so many failures in condition 3, since this condition had the lowest number of total balls. Urn 2 was not chosen as often as urn 1 but the failure rate in both the urns was approximately the same. However, it must be noted that the starting success rate was very low in this condition.

While we hoped that subjects would not always start a task with urn 1, our analysis shows that in trial 1 subjects chose urn 1 (227) more than twice as often as urn 2 (93). However, in the second trial, we find that there is no significant difference between the number of times urn 1 (168) and urn 2 (152) were chosen. Thus, any bias in favor of urn 1 in the first trial was short lived. We can only surmise that the high percentage of failures in urn 1 (62.5%) was the reason for the switch. However, the failure rate for the subjects who chose urn 2 was even higher at 76.3%.

We take this analysis further by analyzing the decisions in each of the replications. In Table 4 - 8, we list the results of the decisions made by the subjects by condition and replication. In condition 1, we find that the subjects behave in the same manner in all replications. In all four replications, urn 1 was chosen more often but the effect is more pronounced in replication 2 and 3 with it being chosen approximately 65% of the time as opposed to replications 1 and 4, where it was chosen only about 55% of the time. However, the success rate in each of the urns varied quite a bit from replication to replication. The success rate in urn 1 was higher in replication 2 and 3, but it was higher in urn 2 in the other two replications.

In condition 2, we find that urn 1 was chosen more often in replication 1 and 3 and urn 2 was chosen more often in replication 2 and 4. There was a similar effect for the success rate, i.e., for the odd-numbered replications, urn 1 had a higher success rate (48.35% and 45.93%) and in the even numbered replications urn 2 had a higher success rate (49.32% and 48.11%). It must be noted here that the starting rate of success in urn 1 was .4 (16/40 balls) whereas the starting rate of success in urn 2 was .2 (3/15 balls). In this condition, it appears that the urn that was chosen more often achieved a higher percentage of successes. This did not always hold true in condition 1.

In condition 3, we find that in replication 1 and 3, urn 2 was chosen more often and in replication 2 and 4, urn 1 was chosen more often. The urn that was chosen more often registered a higher number of successes too. For example, urn 1 was chosen in approximately 73% of the trials in replication 4 and it achieved a success rate of approximately 60%. Urn 1 started with a probability of success equal to .05 (1/20 balls) and urn 2 started with a probability of .1 (1/10 balls).

Finally, in condition 4, we find that in replication 1 and 4 the two urns were chosen approximately an equal number of times. However, in replication 2 and 3, urn 1 was chosen more often. The success rates registered in each of the urns are very similar and that is to be expected since both urns started out with exactly the same composition of balls and the starting probabilities of success were quite high at 0.6.

Overall, we don't find too many consistencies between conditions and replications. In conditions 2 and 3 we find that even though the two urns started out with

different probabilities of success, the urn that was chosen more often achieved a higher success rate. In condition 1, we find that even though urn 1 was always chosen more often it did not always achieve a higher success rate. In condition 4, we find that the urns started out with equal probabilities, and even though they were not chosen with equal probabilities, they achieved almost similar success rates in both the urns.

FURTHER ANALYSES

Given these observations, we decided to simulate three policies and compare how subjects' decisions fared against them. We wanted to compare the actual earnings of the subjects with the potential earnings that they could have realized if they had used one of these heuristics. Each simulation was run 100 times for 30 trials. As in the experiment subjects faced, each condition was repeated 4 times. The first two heuristics were variations of the "stay on a loser, switch on a winner" policy. The rationale behind this heuristic is simple and intuitive. If subjects are to maximize their earnings from the task, they need to exhaust the black balls in the urns as quickly as possible. The only way they can exhaust the urns would be to stay and sample the same urn when they encountered a black ball. In the first variation we decided to use this policy for the first n trials and then pick the urn with the higher estimated probability and stay on it. As in the earlier simulations, n was equal to 10 ± 2. If, at the end of n trials, both urns had equal estimated probabilities, we ran the heuristic for another trial and the two probabilities would be compared to pick the urn with the higher probability. This condition was evoked in approximately 1% of the simulations. The second variation was to use this policy for the

entire duration of the task. In other words we would use the stay on a loser, switch on a winner heuristic for all trials.

The third policy simulated was " switch after 'n' wins, otherwise stay". Notice that this policy is a variation of the second policy. Instead of switching on the first win, we would wait for *n* wins before switching. We hoped this policy would provide a trade off between the number of winners and increase in the probability of success. We decided to use 2 and 3 as values for *n*. If the results looked promising, we planned to try other values for *n*.

Tables 4 - 9 and 4 - 10 list the outcomes of the first simulation. Table 4 - 9 lists the aggregate results for each condition for the five values of '*n*' used (8-12). We find that the heuristic performed best in condition 4 and worst in condition 3. This could be due to the respective high and low initial probabilities of success. There does not appear to be any significant increase in the success level as we varied 'n' between 8 and 12. However, when we look at the total number of successes across all conditions, 'n' = 8 performed the best even though the difference is not much. Therefore, we averaged all the levels of 'n' together in Table 4 - 10 for the 30 trials and each of the conditions individually. In each of the conditions, we observed the expected increase in the percentage of successes as the trial numbers increase. However, as expected, the increase was most remarkable in condition 3 because of the limited number of balls in both the urns. This heuristic performed the best in condition 4 where the two urns were identical. Using our simulation the subjects would have averaged $14 as opposed to the $16.20 that the subjects actually averaged.

Since n=8 performed the best overall, we decided to check and see if the results would hold if we reduced the value to 6 and 7. We find that the performance increases when n=7 but remains almost identical to n=8 when n=6.

Simulation 2 was another variation of the "stay on a loser, switch on a winner" policy. However, in this simulation, subjects used the same policy throughout the task. The results of this policy can be seen in Table 4 -11. The table lists the percentage of successes achieved in the simulation. We find that this simulation does worse that simulation 1. The difference is most remarkable in condition 3. It averaged only a 17% success rate. The poor performance of this simulation can probably be explained by not staying on one urn but switching all the way until the end of each task. This does not allow us to exhaust the black balls in any one urn and that would have been most possible in this condition since one of the urns had only 10 balls. Subjects would have averaged only $11.50 if they had used this heuristic.

The results of the third simulation are captured in Tables 4 -12 and 4 -13. We felt that by increasing the number of wins that the subject stayed on an urn would improve their chances of success. We ran the simulation for n =2 and 3. We found that that there was a slight improvement over simulation 2 (just 2 successes), but it did not better the results of simulation 1. Since the improvement in results over simulation 2 was incremental, that it was felt that it was not worth pursuing this simulation any further. If this simulation had been followed, subjects would have averaged around 173 wins and that could have earned them approximately $11.50. The success rates are almost identical

to simulation 2. It appears that there is no significant difference in the overall success rate on average whether you switch after 1, 2, or 3 wins.

Discussion

Analyses of the subjects' decisions show that this was also a difficult task for them to master. The average success rate in each of the tasks ranged from 40% - 68%. Condition 3 had the lowest success rate and condition 4 had the highest success rate.

In condition 1, subjects chose urn 1 more frequently. Both urns 1 and 2 in this condition started with the same probability but urn 2 would have increased at a faster rate, due to the fewer number of balls in it.

In condition 2, we find that subjects chose urn 1 and urn 2 the same number of times. The proportion of success and failures achieved were also identical. They achieved fewer successes in this condition than failures. In this condition, urn 1 started with a higher probability of success but urn 2 had a faster rate of increase.

In condition 3, subjects again could not differentiate between urns 1 and 2. While the starting proportion of success was very low in both cases, urn 2 had a very fast rate of increase. Subjects also achieved the lowest rate of success in this condition just 40%. However, when we looked at the data for the individual replications we find that in two of the replications (1 and 3) urn 2 was chosen more often and in the other two, urn 1 was chosen more often. Also as in condition 2, the urn that was chosen more often, registered a higher success rate.

In condition 4, the two urns started out with an identical number of black and white balls in them. Urn 1 was chose more often (probably due to the proclivity of subjects to start with urn 1). Since subjects' decisions were being reinforced with a high success rate, they may have decided to stay on it. The same argument could hold for urn 2 too, since the percentage of successes in the two urns were almost identical in all replications. The overall rate of success achieved in this condition is not surprisingly the highest averaging approximately 68%.

The subjects used almost identical policies in this case of the TAB with increasing probabilities as they did in the OAB case where one arm was fixed and known and the other arm was increasing. This is a major finding since the subjects were naïve and care had been taken not to make sure that none of the subjects participated in more than one experiment. We do need to keep in mind that the conditions were not identical and that the TAB tasks were tougher since they had two increasing arms.

This study is unique because no one has attempted to do anything similar in the literature. Therefore, comparisons are not feasible. Subjects tend to exhibit high switching behavior. This could be because they were perplexed by the task or did not understand it. It could also be that the task was difficult for them. The two switching policies (i.e., Policies 1 and 2) accounted for nearly 50% of the trials. However, it must be noted that the highest earners, namely, subjects 6 and 5, used policy 3 or "stay on an urn" for the majority of the tasks. The subject who used policy 1 most often (14 times) was among the lowest earners (ranked 18th). Subject 9 who ranked last also used this policy 14 times.

However, when we tried to simulate a structured variation of the switching policies we found that the actual subjects outperformed the simulations. Simulation 2 did not perform as well compared to the real subjects or simulation 1. Simulation 3 performed a little better than simulation 2 on average but it was by a miniscule amount (2 successes). So another finding is that there is not much difference in the earnings on average whether the subjects switched after 1, 2, or 3 wins.

The general conclusion that one draws from this experiment is that it does not pay to switch. It is better to pick an urn and stay on it for the entire duration of the task. However, our first heuristic performed the best when switches were made for 7 trials and then the estimated probabilities were compared and the urn with the higher probability was picked to stay on till the end of the task.

Table 4 - 1: Distribution of colored balls for TAB 2

Condition	Color of Ball	Number of Balls in Urn 1	Number of Balls in Urn 2
1	White	16	8
	Black	24	12
2	White	16	3
	Black	24	12
3	White	1	1
	Black	19	9
4	White	18	18
	Black	12	12

Table 4 - 2: Order of conditions that subjects saw (TAB 2).

	Order of Conditions				Labeling on Urns
Round 1	1	2	3	4	Original label
Round 2	2	3	4	1	Reversed label
Round 3	3	4	1	2	Original label
Round 4	4	1	2	3	Reversed label

Table 4 - 3: Summary of the policies subjects used in TAB 2.

Policy Number	Policy Name	Description of Policy
Policy 1	Constant Switching	Repeated switching between urns
Policy 2	Switch for 'n' trials then pick an urn to stay on	Switch between urns for 'n' trials and then stay on one urn for the rest. 'n' ≤ 15.
Policy 3	Stay on same urn	Stay on the same urn for all trials
Policy 4	Stay on same urn for most trials	Stay on the same urn for most trials with no more than 5-6 tries at sampling the other urn.
Policy 5	Switch after 5-6 trials	Switching between the urns after 5-6 trials on one urn
Policy 6	One switch policy	Stay on an urn for 'n' trials and then make one switch and stay on the other urn for the rest of the trials

Table 4 - 4: Policy choice of subjects for each of the tasks that they saw (TAB 2).

	Sub 1	Sub 2	Sub 3	Sub 4	Sub 5	Sub 6	Sub 7	Sub 8	Sub 9	Sub 10	Sub 11	Sub 12	Sub 13	Sub 14	Sub 15	Sub 16	Sub 17	Sub 18	Sub 19	Sub 20
Cond 1a	1	1	1	2	1	4	5	1	1	5	3	1	1	2	2	5	1	1	5	4
Cond 1b	2	1	2	2	3	3	4	1	1	5	3	1	2	2	2	1	5	1	1	2
Cond 1c	3	1	2	4	3	3	3	2	1	4	3	1	4	1	4	5	5	2	1	6
Cond 1d	2	1	1	2	3	3	6	6	1	5	3	1	4	5	4	4	3	1	5	4
Cond 2a	2	2	4	2	1	5	1	1	1	1	4	1	2	5	2	5	1	1	4	4
Cond 2b	2	1	2	5	3	6	1	1	2	1	3	1	2	5	2	5	3	1	5	6
Cond 2c	4	1	5	2	4	6	2	4	1	2	6	1	1	1	4	5	5	4	4	2
Cond 2d	4	1	1	6	6	4	6	4	1	5	3	1	2	5	4	6	5	4	5	6
Cond 3a	2	2	2	4	4	2	6	1	2	2	6	1	2	4	2	4	1	2	5	2
Cond 3b	5	2	2	2	3	1	5	1	1	1	6	2	2	1	2	2	5	4	4	2
Cond 3c	2	1	2	2	4	4	2	1	1	2	2	1	2	1	1	5	2	5	4	3
Cond 3d	2	2	2	2	2	3	4	1	1	2	2	2	2	2	2	5	4	2	2	6
Cond 4a	3	2	2	3	3	3	2	1	1	5	3	1	4	3	4	3	1	5	5	5
Cond 4b	6	1	2	3	3	3	3	4	1	6	3	1	2	6	4	3	3	3	5	2
Cond 4c	2	3	3	3	3	3	3	4	1	3	3	1	6	4	4	5	3	3	4	5
Cond 4d	6	1	4	3	3	3	6	5	1	5	3	1	4	4	6	4	3	3	4	6

Note: A suffix of *a* denotes replication 1, a suffix of *b* denotes replication 2, a suffix *c* denotes replication 3, and a suffix *d* denotes replication 4.

Table 4 - 5: Number of times a policy was used in each of the conditions (TAB 2).

	Policy 1	Policy 2	Policy 3	Policy 4	Policy 5	Policy 6
Condition 1a	10	3	1	2	4	0
Condition 1b	7	7	3	1	2	0
Condition 1c	5	3	5	4	2	1
Condition 1d	5	2	4	4	3	2
Condition 2a	8	5	0	4	3	0
Condition 2b	6	5	3	0	4	2
Condition 2c	5	4	0	6	3	2
Condition 2d	4	1	1	5	4	5
Condition 3a	3	10	0	4	1	2
Condition 3b	5	8	1	2	3	1
Condition 3c	6	8	1	3	2	0
Condition 3d	2	13	1	2	1	1
Condition 4a	4	3	7	2	4	0
Condition 4b	3	3	8	2	1	3
Condition 4c	2	1	10	4	2	1
Condition 4d	3	0	6	5	2	4

Table 4 - 6: Summary of subject decisions and the outcomes (TAB2)

DECISION	RESULT		
	FAILURE	SUCCESS	TOTAL
URN 1	2536	2641 (51.01%)	5177 (53.93%)
URN 2	2232	2191(49.53%)	4423 (46.07%)
TOTAL	4768 (49.67%)	4832 (50.33%)	9600

Table 4 - 7: Results of the decisions made by subjects (by condition) TAB 2.

Cond	Decision	Result		Total
		Failure	Success	
1	Urn 1	714	719 (50.17%)	1433 (59.71%)
	Urn 2	500	467 (48.29%)	967 (40.29%)
	Total	1214	1186 (49.42%)	2400
2	Urn 1	683	517 (43.08%)	1200 (50%)
	Urn 2	683	517 (43.08%)	1200 (50%)
	Total	1366	1034 (43.01%)	2400
3	Urn 1	722	514 (41.58%)	1236 (51.5%)
	Urn 2	698	466 (40.03%)	1164 (48.5%)
	Total	1420	980 (40.83%)	2400
4	Urn 1	417	891 (68.12%)	1308 (54.5%)
	Urn 2	351	741 (67.86%)	1092 (45.5%)
	Total	768	1632 (68%)	2400

Table 4 - 8: Results of the decisions made by subjects (by condition x replication) TAB 2

Replication	Condition	Decision	Result		Total
			Failure	Success	
1	1	Urn 1	184	150 (44.91%)	334 (55.67%)
		Urn2	128	138 (51.88%)	266 (44.33%)
		Total	312	288 (48.00%)	600
2		Urn 1	170	215 (55.84%)	386 (64.17%)
		Urn2	129	86 (40%)	215 (35.83%)
			299	301 (50.17%)	600
3		Urn 1	207	188 (47.59%)	395 (65.83%)
		Urn2	111	94 (45.85%)	205 (34.17%)
			318	282 (47%)	600
4		Urn 1	153	166 (52.04%)	319 (53.17%)
		Urn2	132	149 (53.02%)	281 (46.83%)
			285	315 (52.5%)	600
1	2	Urn 1	204	191 (48.35%)	395 (65.83%)
		Urn2	146	59 (28.78%)	205 (34.17%)
			350	250 (41.67%)	600
2		Urn 1	142	89 (38.53%0	231 (38.5%)
		Urn2	187	182 (49.32%)	369 (61.5%)
			329	271 (45.17%)	600
3		Urn 1	186	158 (45.93%)	344 (57.33%)
		Urn2	158	98 (38.28%)	256 (43.67%)
			344	256 (42.67%)	600
4		Urn 1	151	79 (34.35%)	230 (38.33%)
		Urn2	192	178 (48.11%)	370 (61.67%)
			343	257 (42.83%)	600

Results of the decisions made by subjects (by condition x replication) TAB 2

Replication	Condition	Decision	Result		Total
			Failure	Success	
1	3	Urn 1	189	19 (9.13%)	208 (34.67%)
		Urn2	171	221 (56.38%)	392 (65.33%)
			360	240 (40.00%)	600
2		Urn 1	171	198 (53.66%)	369 (61.5%)
		Urn2	206	25 (10.82%)	231 (38.5%)
			377	223 (37.17%)	600
3		Urn 1	187	36 (16.14%)	223 (37.17%)
		Urn2	168	209 (55.44%)	377 (62.83%)
			355	245 (40.83%)	600
4		Urn 1	175	261 (59.86%)	436 (72.67%)
		Urn2	153	11 (6.71%)	164 (27.33%)
			328	272 (45.33%)	600
1	4	Urn 1	98	202 (67.33%)	300 (50%)
		Urn2	96	204 (68%)	300 (50%)
			194	406 (67.66%)	600
2		Urn 1	107	233 (68.53%)	340 (56.67%)
		Urn2	78	182 (70%)	260 (46.33%)
			185	415 (69.17%)	600
3		Urn 1	111	252 (69.42%)	363 (60.5%)
		Urn2	84	153 (64.56%)	237 (39.5%)
			195	405 (67.5%)	600
4		Urn 1	101	204 (66.89%)	305 (50.83%)
		Urn2	93	202 (68.47%)	295 (49.17%)
			194	406 (67.67%)	600

Table 4 - 9: Aggregate results of Simulation 1 (TAB 2)

	Condition 1	Condition 2	Condition 3	Condition 4	TOTAL
'n' = 6	52.18	47.66	46.13	67.51	213.48
'n' = 7	52.73	49.33	50.50	68.10	220.66
'n' = 8	52.18	47.66	46.03	67.54	213.42
'n' = 9	48.57	45.17	43.18	68.18	205.09
'n' = 10	50.39	44.35	44.98	67.63	207.34
'n' = 11	50.29	43.83	40.44	67.39	201.95
'n' = 12	49.14	42.78	38.90	66.18	197.01
Average	50.12	44.76	42.71	67.39	

Simulation 1; Stay on a loser, Switch on a winner for the first 'n' trials. At the end of 'n' trials compare the estimated probability of success on each urn and stay on the urn with the higher probability for the rest of the trials. (All numbers are the average number of success across all 100 simulations).

Table 4 - 10: Summary of Simulation 1 across all conditions (TAB2)

Trial #	Condition 1	Condition 2	Condition 3	Condition 4
1	39.85	29.1	7.25	59.5
2	41.65	29.25	8.1	60.85
3	42.4	27.45	8.6	58.6
4	41.6	30.15	9.85	62.45
5	43.2	30.5	10.2	62
6	41.6	30.75	12	63.35
7	45.05	31.6	10.8	63.2
8	44.05	32.05	11.55	61.85
9	42.75	34.85	13.05	64.2
10	43.95	37.1	15.75	63.15
11	46.6	37.6	17.3	61.65
12	44.55	40.35	21.95	64.85
13	42.35	38.8	23.1	61.65
14	45.85	40.7	27.2	62.7
15	47.75	43.1	31.3	65.65
16	48.9	43.4	34.75	67.15
17	49.9	46.2	41.05	65.35
18	51.8	46.2	45.7	67.95
19	52	46.9	52.15	69.1
20	54.85	49.85	57.5	70
21	55	52.25	64.25	71.6
22	55.8	55.45	68.9	71.85
23	57.85	55.3	74.9	73.2
24	58.05	56.6	79.8	73.35
25	48	58.85	82.95	73.35
26	61.85	58.9	85.45	74.95
27	62.2	61.35	88.6	75.3
28	63.1	63.85	90.7	76.8
29	65.55	66.15	91.9	78.7
30	65.4	68.1	94.55	77.25
Average	50.115	44.76	42.71	67.385

Table 4 - 11: Summary of Simulation 2 over all conditions (TAB 2)

TRIAL #	CONDITION 1	CONDITION 2	CONDITION 3	CONDITION 4
1	39	30.5	6.5	58.5
2	43	32.75	8.75	59.25
3	39.25	27.25	7.75	60.5
4	42.75	31.5	9.75	61
5	48	31	9.25	57.75
6	45.75	29	11.5	62.25
7	40.5	32.75	10.75	64.25
8	43.5	32.5	11.25	67.25
9	42.75	37.5	14.5	63.5
10	44.25	32.5	14.75	60.5
11	47.5	36	12.25	68.75
12	48	42.75	10.25	60
13	44.5	37	11.5	66
14	49	37.75	14	62.75
15	49.5	41.5	13.5	68.25
16	50	38.25	13.5	65.25
17	48.5	43.75	18	67.75
18	49.75	40.5	18	63.75
19	47.75	42.75	19.75	70.75
20	51.75	48.25	17	65.75
21	55	42.5	16.5	66.75
22	51	47.5	19.5	67.5
23	52.75	50.5	18.5	73.25
24	51.25	47.25	22.25	68
25	53.75	45.75	25.5	71.25
26	51.75	51.25	24.25	71.5
27	49.5	48.5	32	72.75
28	51.75	54.75	31	70.25
29	53.25	54	38	74
30	56.25	54.5	39.5	65.75
Average	48.04	40.74	17.32	65.83

Simulation 2: Stay on a loser and switch on a winner for all trials. These numbers represent the average number of successes in each of the conditions for the individual trials over all the simulations. These numbers represent the average of the number of successes over all the 100 simulations for each of the individual conditions.

Table 4 - 12: Aggregate results of Simulation 3 (TAB2)

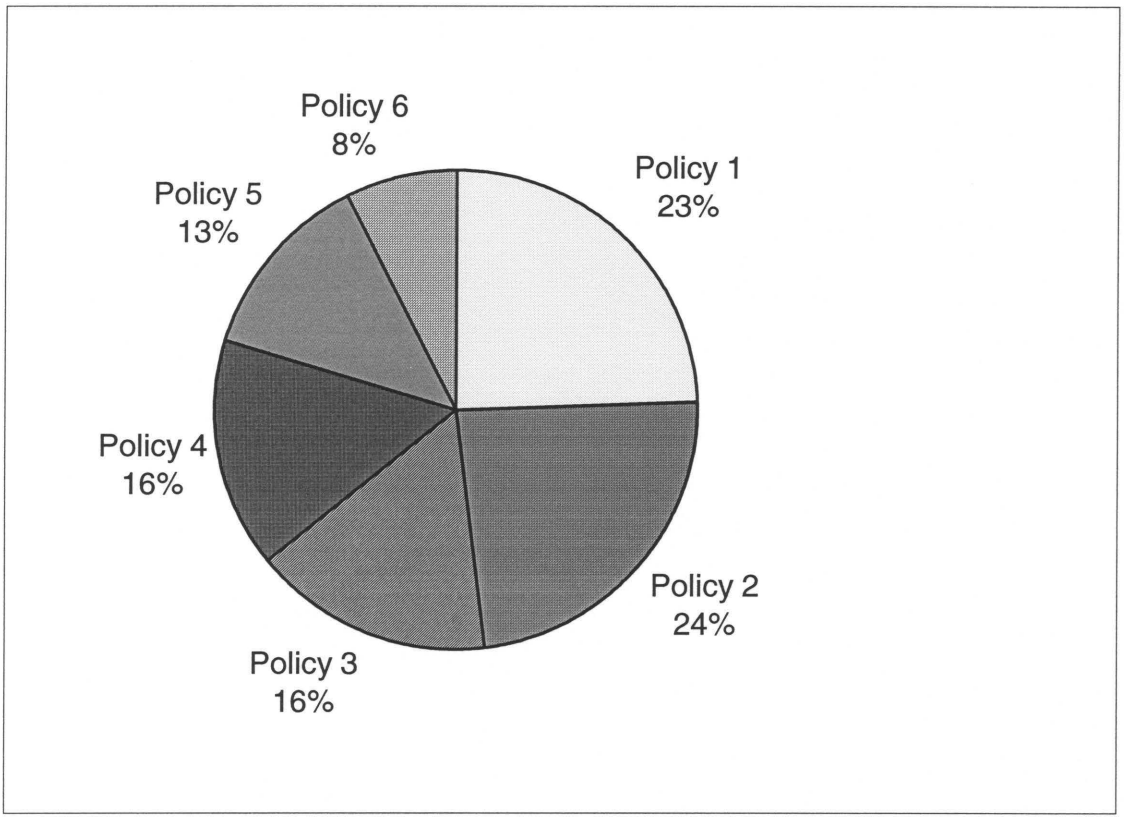
	Condition 1	Condition 2	Condition 3	Condition 4	TOTAL
'n' = 2	47.8	41.05	17.53	66.39	172.77
'n' = 3	48.59	41.51	17.80	65.69	173.59
Average	48.195	41.28	17.665	66.04	

Simulation 3; Stay on a loser, Switch after n wins. (All numbers are the average number of success across all 100 simulations).

Table 4 - 13: Summary of Simulation 3 over all conditions (TAB 2)

TRIAL #	CONDITION 1	CONDITION 2	CONDITION 3	CONDITION 4
1	42.75	32.88	8.25	58.00
2	41.63	29.50	8.38	63.63
3	41.63	32.63	9.50	62.38
4	42.75	30.75	10.00	61.88
5	42.50	30.00	10.63	62.88
6	42.50	31.25	12.38	62.38
7	42.38	31.00	13.38	63.13
8	44.38	32.88	15.63	63.13
9	44.13	34.38	17.63	62.75
10	46.63	33.00	23.38	61.38
11	44.63	37.75	25.88	63.88
12	44.75	35.63	21.25	63.38
13	46.00	39.25	9.13	64.00
14	47.75	37.88	8.25	63.88
15	47.88	38.13	11.75	64.63
16	49.13	40.38	13.38	67.25
17	51.75	42.88	15.25	66.88
18	48.50	44.50	13.63	66.13
19	49.25	45.63	16.75	64.75
20	53.13	46.88	23.38	70.13
21	50.88	49.00	24.88	70.00
22	51.13	49.38	23.00	70.13
23	53.63	48.63	16.25	69.00
24	51.88	49.00	18.88	71.75
25	52.75	49.50	19.75	69.13
26	50.25	51.25	20.88	70.25
27	53.25	50.75	21.38	72.13
28	54.88	53.25	29.25	72.00
29	57.13	52.25	33.13	69.00
30	56.13	58.25	35.05	71.50
Average	48.20	41.28	17.67	66.04

Figure 4 - 1: Distribution of subjects' choices among various policies (TAB2)



Policy 1: Constant Switching
Policy 2: Switch for 'n' trials, then pick an urn to stay on
Policy 3: Stay on same urn
Policy 4: Stay on same urn for most trials
Policy 5: Switch after 5-6 trials
Policy 6: One switch policy

INSTRUCTIONS TO SUBJECTS FOR TAB - 2

You are about to participate in a computer-controlled decision making experiment in which your payoff is contingent on your performance. The money you will earn will be paid to you at the end of the experiment. A research foundation interested in clinical decision-making has contributed the money to finance the study.

The major purpose of this experiment is to study how individuals behave when asked to make a sequence of choices between two alternatives, each of which yields a success with a probability that is initially unknown. The task that we have in mind is the sequential testing of two new medicines, whose properties have not been established, on a series of patients. At any time only one drug can be administered to a patient, and the treatment is scored as either a success or a failure. Only when the result of the treatment is known, another patient is administered one of the two drugs. Clearly, one would like to maximize the number of successes; however, only by trying each of the drugs can the physician learn which is more effective.

It would be simpler to describe the task that you are about to perform in terms of two urns, labeled 1 and 2. Suppose that each urn has some black and some white balls in it. The total number of balls may vary from one urn to another. You are not informed of the total number of balls in each box, or the proportions of black and white balls in each urn.

Your task is simple. You will make a series of drawings from the urns, one on each trial. The total number of drawings (denoted by T), will be known to you when the task starts. At the beginning of each trial, your task is to decide which urn to draw a ball from. Choose an urn, draw a ball blindly, and record its color (either black or white).

Your payment on each trial depends on the ball you draw. If you draw a white ball, you will be paid one unit of money, called "franc". (Francs will be converted into US dollars at the end of the experiment.) Please replace the white ball in the urn from which it was drawn.

If you draw a black ball, you will receive no payment. Please remove the black ball from the urn from which it was drawn.

This completes the trial. To repeat, once you have noted the color of the ball, you replace it, if it is white, and remove it, if it is black. You get paid one franc for a white ball and zero for a black ball. Clearly, as you draw more balls from a particular urn, you learn more about its composition. Moreover, because the total number of balls in each urn is fixed, you increase your chances of drawing a white ball on the next trial each time you draw a black ball. The essence of this decision making task is to gather information about the urns and at the same time make decisions that increase your payoff.

Your objective in the task is to draw as many white balls as possible, thereby increasing your earnings.

Example

Supposing the two urns are composed as follows and that the number of balls is known:

URN	# WHITE BALLS	# BLACK BALLS
1	8	4
2	4	6

Supposing that you choose Urn 1. If you draw a white ball (and receive a payment of 1 franc), then on the next trial the composition of the urns will be the same. If you draw a black ball (and receive a payment of zero), then on the next trial the composition of the two urns will be:

URN	# WHITE BALLS	# BLACK BALLS
1	8	3
2	4	6

Having drawn a black ball from Urn 1, the chances of drawing a white ball have increased from 8/12 to 8/11.

Continuing the same example, supposing that your drawings are as follows:

TRIAL	URN	COLOR	PAYOFF	# W IN URN 1	# B IN URN 1	# W IN URN 2	# B IN URN 2
1.	1	WHITE	1	8	4	4	6
2.	1	WHITE	1	8	4	4	6
3.	1	BLACK	0	8	3	4	6
4.	2	WHITE	1	8	3	4	6
5.	2	BLACK	0	8	3	4	5
6.	1	BLACK	1	8	2	4	5

Thus, after 6 trials, you have earned 4 francs. The chances of drawing a white ball from Urn 1 are now 8/10 (80%) and of drawing a white ball from urn 2, 4/9 (44.4%).

Drawing from two urns and keeping track of the number of balls drawn is a tedious task. Therefore, to simplify the task, we have computerized the entire process. Your only decision will be to choose an urn on each trial. The computer will then randomly draw a ball from the urn you have chosen, exactly as in the example above, and will inform you of your payoff for the trial and your accumulated payoff from the beginning of the experiment.

You will participate in 16 different tasks. At the beginning of each task you will be informed that the new task has two different urns with possibly different compositions of white and black balls.

You may wonder how the (unknown) proportion of white balls in each urn has been determined. The only information that we can give you is that the percentage of white balls in each of the two urns can assume any value between 0 and 100 with all values in this range being equally likely.

Once the task begins the computer will display the following information on the screen before you make your decision:

History of previous play:

trial # urn 1 urn 2 cumulated payoff

You are on trial number (1.....T)

Which urn would you like to sample from? (1 or 2)

For example, suppose you sampled from urn 1 on the first trial and you picked a black ball, and from urn 2 the next time and picked a white ball. Your screen would then look similar to this:

History of previous play

outcome

trial #	urn 1	urn 2	cumulated payoff
1	Black	----	0
2	-----	White	1

You are on trial number (1.....T)

Which urn would you like to sample from? (1 or 2)

Once you enter either '1' or '2,' the computer will randomly draw a ball from that urn and present you with the following screen.

On trial number (1.....T)

Urn chosen on this trial (I or II)

Ball picked up on this trial (Black or White)

Your earnings for this trial = 0 or 1 franc.

[Press any key to continue]

[Press F1 to see history]

Following the presentation of the outcome of the trial, the computer updates your record and returns you to the first screen for the next trial. This process will be repeated for 'T' trials. At any time during the task you will be able to check the history for trials that are no longer visible (the history screen will hold the results for about 20 trials) by using the F1 key that is located at the top left hand side of your keyboard.

As mentioned, you will participate in a series of 16 tasks. The two urns will change between the tasks. In other words, the proportion of the white and black balls will change between the tasks.

Between each task you will see this screen:

You are about to begin a new sequence of trials.

You will sample from a new pair of urns.

The urns are still labeled I and II but they are not the same ones used in the previous. task.Please keep in mind that the computer has not been programmed to play against you. Rather, the number of white balls and black balls in each urn has been predetermined and will change from trial to trial (depending on the ball that is picked) within the next 30 choices. Your task is to draw as many white balls as you can over the next 30 trials.

Your concern is to make as much money as you can, given these conditions. You are being paid in a fictitious currency called a "franc". "Francs" will be converted into US dollars at the end of the experiment at the rate of $1 = 10 francs. You will be paid before leaving.

If you have any questions about the instructions, please inform the experimenter immediately. You may refer to the written copy of the instructions placed on your desk at any time. Please take whatever time necessary to make your decisions in this experiment.

CHAPTER 5: TWO-ARMED BANDIT PROBLEM WITH INCREASING AND DECREASING PROBABILITIES

INTRODUCTION

In the earlier chapters, we studied the classic TAB problem, the OAB problem, and the TAB with two increasing arms. The next problem that we would like to address is the TAB with one increasing and one decreasing arm. The analogy for the third TAB experiment that we conducted can be drawn from the manufacturing industry. In a factory at any given time, one can find old machinery working side by side new machinery. Sometimes both produce the same widgets. Suppose there are two competing brands of machinery available for the manufacture of widgets. However, the two brands behave differently. The "old faithful" that has been working fairly well all these years has been showing signs of wear and tear. It has been reliable but you now realize that each run is taking its toll on the machine. The reliability decreases with each use. In other words, you realize that the error rate is slowly increasing on the old machine. On the other hand, you have a brand new machine that has just been added to your floor. This machine has the latest technology installed in it. It can learn from its mistakes. It can be fine-tuned each time it makes a mistake (we assume that fine-tuning needs to be done only when mistakes are made) so that the productivity can be increased with each run. How would one schedule jobs on the two machines? That is the essence of this TAB problem.

We are trying to study how subjects would behave if they were to encounter one arm with increasing probabilities and the other arm with decreasing probabilities. Hence,

in this experiment we are again dealing with non-stationary probabilities. Subjects are not aware of the probabilities at the start of the experiment. They are aware that the probability in arm 1 increases with each failure encountered, and that the probability of failure increases in arm 2 with each success encountered.

Each subject faced four conditions (see Table 5 - 1). Each condition was repeated four times in a round robin manner. Each round consisted of all four conditions and each subject faced four rounds. Thus, subjects faced 16 tasks in all.

In condition 1, both urns began with identical probabilities of success. In addition, they also had the same number of white and black balls in the urn. In condition 2, both urns began with different probabilities of success. In fact, urn 1 had exactly half the probability of success that urn 2 had. In condition 3 both urns began with the same probability of success, however, urn 1 had twice the number of balls that urn 2 had. In condition 4, both urns began with a different number of total balls and different initial probabilities of success. Urn 1 had an equal or lower initial rate of success than urn 2. In all the conditions, the rate of increase in the success of urn 1 was lower than the rate of decrease of failure in urn 2. (However, the subjects were not privy to this fact.)

Method

Subjects

Subjects were recruited from around campus in the same manner as the earlier experiments. They were told that their payments would be contingent on performance.

Twenty subjects participated in the experiment. Care was taken to make sure that the subjects had not participated in any of the earlier bandit experiments.

Procedure

(See Chapter 2 for details.) Subjects arrived at the Behavioral Decision Lab (BDL) to participate in the experiment. All instructions were furnished online as well as on paper. The subjects were informed that their primary purpose was to maximize their earnings (by selecting white balls), given the uncertainty of the tasks. At the end of the 10 practice rounds, subjects were told that the real trials were about to begin and that they would be paid for correct decisions.

Subjects were asked to imagine two urns in front of them. They were told that their task would be to pick a ball from an urn. Sampling of white balls in urn 1 was with replacement while black balls were removed after sampling. On the other hand, in urn 2 they faced the opposite scenario. Sampling of black balls was with replacement and sampling of white balls was without replacement. Each time the subjects encountered a failure on urn 1, they increased their chances of getting a white ball (success) in the future. Alternatively each time they encountered a success in urn 2, the white ball was removed thereby decreasing their chances of getting a success in the future. The percentage of white balls in each urn was unknown and could vary from trial to trial. The total number of trials for each task was known and fixed at 30.

As in the earlier experiments, subjects had to choose which urn they wanted, and the computer would pick a ball for them. If it was a white ball, they were paid a franc,

and it (the ball) was returned to the urn if picked from urn 1, but removed from the urn if picked from urn 2. If a black ball was picked from urn 1, it was removed from the urn, but if it was picked from urn 2, it was replaced in the urn. At every stage in the experiment, subjects were informed which trial they were on. They could also observe the results of the previous trials in the same task by using the F1 key. At the end of each task, subjects were informed that they were about to start a new task, and that they would sample from a new set of urns (see Appendix 5 -1 for sample screens).

Each time a subject achieved a success he was paid one "franc". Francs were converted to dollars at the rate of 20 francs = 1 U.S. dollar. After completing the sixteen tasks, the subjects were paid their cumulative earnings, debriefed, and dismissed from the lab. The experiment lasted approximately an hour and the mean earnings per subject were $ 10.08.

RESULTS

Generic Policies

This was not an easy experiment to identify policies for, because subjects were exhibiting switching behavior and it was difficult to distinguish between the various switching patterns. After a careful examination, we realized that there were discernable patterns to their switching behavior. Table 5 - 2 lists all the policies that were used by the subjects. We identified 5 policies that subjects appeared to be using.

Instead of clubbing all the switching policies under one policy, we attempted to differentiate between the switching policies. As mentioned earlier we found that we could

categorize the switching behavior into three groups. We decided to distinguish between switching after 1 - 3 trials, 4 - 6 trials and after more than 7 trials.

In the first policy that we identified, subjects exhibited frequent switching behavior that we defined as switching after less than 3 trials. Most subjects switched after both wins and losses on both the urns. This is similar to the constant switching policies seen in the previous chapters.

In the second policy, subjects switched after 4 - 6 trials. We called this the moderate switching policy. In the third policy subjects switched after 7 trials. We called this the infrequent switching policy. We found that many subjects were willing to wait for a while to see what would unfold before they decided to switch. In both these policies, as in the first policy, we do not differentiate between wins and losses.

In policy 4, subjects stayed on the same urn through the entire task. In the fifth and last policy that we identified, subjects chose to stay on one urn mostly, but occasionally would test the other urn. We restricted this "testing" or checking out to less than 8 or 9 trials. This policy should not be confused with policy 3 where subjects switched after more than 7 trials. In policy 5, subjects would switch to the other urn sometimes after 1 or 2 trials, and at other times after 5 or 6 trials. Subjects never strayed to the other urn for more than 9 trials in a given task, but there was no fixed pattern in their attempts to "peek" at the other urn. This policy is similar to "stay on an urn most of the time" that we saw in the chapters 3 and 4.

Figure 5 -1 portrays the distribution of the choices between the various policies. Collectively, in 59 tasks (18.44%), subjects chose policy 1. That is, they switched frequently between the urns after less than 3 trials. In 19.38% of the tasks, subjects switched after 4-6 trials, i.e., they followed policy 2. Among the switching polices (the way we categorized it for the experiment), policy 3 was the most used. It was used in approximately 25% of the tasks. This meant that subjects were willing to wait and see what would happen before switching urns. Together, these switching policies were used in over 60% of the tasks. All subjects used at least one of the switching policies, except for subject 14. Subjects 3, 4, 5, 7, and 10 chose not to use policy 1 even once.

Subjects who followed policy 4 decided to stay on the same urn for the entire task. One subject (14) used this policy consistently (he chose to stay on Urn 1), no matter what the task. All subjects who chose this policy stayed on urn 1. This was the least used policy (10.69%). Eleven of the subjects did not use this policy even once.

Policy 5 was most frequently used (26.88% or 86 tasks). In this policy, subjects chose the same urn in most trials except for 8-9 trials. While this would not have been surprising if urn 1 had been chosen, there were a few subjects who chose to stay on urn 2 for most of the trials even though they had been informed that it had a decreasing rate of success.

Looking at Tables 5 -3 and 5 - 4, we can easily see that not all policies were used by all the subjects. Subjects who used policy 1 most often were among the lower earners (subjects 2, 8 and 17). As mentioned before, subject 14 was the highest earner he used

policy 4 exclusively choosing urn 1 in all the trials. Subjects who used policies 3 and 5 also did moderately well in their earnings. The top four performers were those who used strategies that limited the number of switches, i.e., they used policies 3, 4, or 5.

Descriptive Analysis of subjects decisions

Tables 5 - 5 and 5 - 6 represent the choices made by the subjects collectively and in each of the tasks individually. Urn 1 was chosen in approximately 60.5% of the trials and urn 2 was chosen in 39.5% of the trials. Subjects achieved success in approximately 41% of the trials overall. There is a significant difference in the number of successes and failures. The subjects achieved almost 42% success in urn 1 as opposed to a success rate of a little less than 41% in urn 2. Thus, while there was a significant difference between the number of times urn 1 and 2 were chosen, there was no significant difference between the success and failure rates between the two urns. This means that overall, subjects were able to keep the failure rate from increasing very high in urn 2.

If we break it down by condition (Table 5 - 6), we find that in all conditions, urn 1 was chosen significantly more often than urn 2. This is not surprising, since urn 2 had a decreasing rate of success as opposed to urn 1 that had an increasing rate of success. In every condition, we find that there were a higher number of failures than successes (1383 vs. 1017, 1313 vs. 1087, 1415 vs. 985, 1510 vs. 890). In conditions 1 and 3 we find that the success rate in urn 1 was higher than the success rate in urn 2. However, in conditions 2 and 4 we find that the success rate in urn 2 was higher. Recall that in condition 1, the two urns were identical, and in condition 3, the initial percentage of success were the

same in both urns, however, the number of balls in urn 2 was half those in urn 1. In condition 2, the initial rate of success in urn 1 (0.3) is half the initial success rate in urn 2 (0.6), and similarly the initial rate of success in urn 1 (0.2) in condition 4 is a little less than half that in urn 2 (0.5). Therefore, that explains why the success rate in urn 2 was higher than the success rate in urn 1. Subjects were able to sample from urn 2 without letting it affect their earnings potential too badly. It appears that subjects were able to identify the urn with the higher initial probability and were able to exploit it although that was the urn with the decreasing probabilities.

When we look at each condition individually, as in Table 5 - 7, we observe that the patterns we observe in Table 5 - 6, hold consistently across replications in Table 5 -7 too. That is, in all replications of all conditions, subjects chose urn 1 more often and the difference between the two were significant in all replications. However, when we study the success rate in the individual urns we find that in conditions 1 and 3, urn 1 had a higher success rate, but in conditions 2 and 4, urn 2 had a higher success rate (in all replications). This is consistent with the overall results in Table 5 - 6 too. Thus, another important finding is that even though the tasks were presented in a round robin manner, subjects behaviors were consistent across the various replications.

FURTHER ANALYSIS

As in the case of TAB with increasing probabilities on both arms, we came up with three heuristics that we thought would yield good results (increase the number of success). We ran the heuristics 100 times and compared their performance to the

subjects' performance. We are trying to find a systematic way to increase the number of successes. Two of the heuristics are based on the basic idea of getting a few wins from urn 2, then switching to the increasing urn, and staying on it until the end of the task. The third heuristic follows the simple policy of sampling only from the increasing urn exclusively and ignoring the decreasing urn.

The first heuristic that we simulated had two variables n and x in it. The heuristic was "Alternate for n trials, then compare the estimated probabilities in both urns and stay on the higher urn. If urn 2 was picked as the higher urn, stay on it until x losses are encountered. Then switch to urn 1 and stay on it till the end of the task.." We set

$n = 10 \pm 2$ and $1 \leq x \leq 5$.

Each of these conditions was run 100 times. We ended up with 25 simulations that we needed to analyze. This heuristic allowed us to sample both urns for a fixed number of trials and then, given the estimated probabilities, we would pick the higher urn. This heuristic is a variation of one that was used with the TAB with two increasing arms.

The second simulation was a simpler one: "Stay on urn 2 for x losses and then switch to urn 1 and stay there till the end of the task." As in the earlier simulation, $1 \leq x \leq 5$.

The third simulation was the simplest of the three: "stay on the increasing arm for all trials."

The results of the simulations were suprising. The heuristics did not perform as well we expected them to. In fact, the subjects seemed to have outperformed the heuristic on average. Initially, we ran 25 simulations (5 combinations of *n* and *x*) but looking at the results, we found that as *n* increased the results deteriorated (See Table 5 - 8). In other words, the longer we sample the urns, the number of successes decreases. This was an interesting finding, hence we thought it might be interesting to see what happened if we reduced *n* to include 6 and 7. That is, now *n* ranged between 6 and 12. We found that as we reduced '*n*', the number of successes increased marginally from 180.93 to 181.85. Additionally we found that as *x* increased from 1 to 5, the number of successes dropped lower. So we decided to explore further and decided to vary *n* to include 1 - 5 too. We did not conduct all variations of x but restricted it to just 1, since it was obvious that the success would decrease as x increased.

Table 5 -9 includes the results from varying *n* between 1 and 5. Notice how the number of successes increases as the value of *n* drops. This result validated what we observed from the subjects' results. The highest scoring subject was the one who stayed exclusively on urn 1. The simulation also seems to suggest that it does not help to even sample urn 2. The simulation performs well compared to the subjects. On average, subjects received 198 successes and our simulation 1_1 achieved 187 successes on average. If the two outliers in the subject pool are removed, the average drops to 182 successes.

Simulation 2 yielded results similar to Simulation 1 (see Table 5 - 10). The longer one stayed on urn 2 the lower the performance. Therefore, this simulation also supports

the fact that it did not really make sense to try urn 2 for very long. The highest earnings were realized when the losses were restricted to 1. What this suggests is that as soon as the subjects encountered a single failure on urn 2 they were supposed to switch to urn 1 and stay there till the end of the task. As in the earlier simulation, we started by examining only cases where losses ranged between 1 and 5, but later we decided to extend it to 10 losses just to see how much decline would be realized. On average, we found that there would be a decrease of approximately 30 if we extended the number of losses to 10.

The results of simulation 3 (Table 5 - 11) were not surprising. When the policy required that only urn 1 be sampled, the number of successes increased to nearly 195. This was an improvement over the other two heuristics but the subjects did better averaging around 204 success. Our subjects managed to capture wins from urn 2 and then capture wins from urn 1. Although their success rate compared to the other experiments was lower, this was creditable since it was a more difficult experiment than the others. They had to deal with both increasing and decreasing arms.

Discussion

This experiment seemed to be the toughest for the subjects. However, subjects did well in achieving moderate level of success given the difficulty of the task. Examining subjects behavior, we find that they switched urns frequently. Subjects switched between the urns on approximately 60% of the tasks. But the most frequently used policy was one where they stayed on one urn for most of the trials. What was suprising about this policy

was that a few subjects chose to stay on urn 2 for most of the trials, though a majority of them chose to stay on urn 1. The subjects could have maximized their earnings by reducing the number of switches, and we find that our highest earner facing these tasks did that exclusively. The results were quite consistent across replications. When the initial probabilities were equal, urn 1 was chosen more frequently. When the initial probabilities were very different, even though urn 1 was sampled more often the probability of success in urn 2 was much higher across all replications. The subjects were also very consistent in their choices, because although the tasks were presented to them in a round robin manner their choices held the same pattern.

The simulations also concur with the conclusion reached by the descriptive analysis of the subjects' decisions. The longer one stayed on urn 2 the lower the earnings. Additionally, using the results of both simulations 1 and 2, we find that it did not make sense to sample urn2 at all. Even if one were to sample it, the simulation suggests that in order to maximize earnings one had to switch to urn 1 after one or two losses. Waiting any longer than that decreased the earnings capacity. That was validated by simulation 3 that used the policy to sample only urn 1, and it would have achieved the highest earnings among the three. All three heuristics performed moderately well. None of them was able to outperform the subjects but they came quite close. So while the tasks were extremely difficult, subjects managed to do moderately well and we couldn't come up with a heuristic that could outperform their behavior.

Table 5 - 1: Distribution of colored balls for TAB 3

Condition	Color of Ball	Number of Balls in Urn 1	Number of Balls in Urn 2
1	White	16	16
	Black	24	24
2	White	9	18
	Black	21	12
3	White	16	8
	Black	24	12
4	White	4	15
	Black	16	15

Table 5 - 2: Summary of the Policies subjects used in TAB 3

Policy Number	Policy Name	Description of Policy
Policy 1	Frequent switching	Switch urns after 1-3 trials
Policy 2	Moderate switching	Switch after 4-6 trials
Policy 3	Infrequent switching	Switch after > 7 trials
Policy 4	Stay on the same urn	Stay on the same urn for all trials
Policy 5	Stay on the same urn for most trials	Stay on the same urn for all trials except 8 or 9 trials

Table 5 - 3: Policy choice of subjects for each of the tasks that they saw (TAB 3).

	Sub 1	Sub 2	Sub 3	Sub 4	Sub 5	Sub 6	Sub 7	Sub 8	Sub 9	Sub 10	Sub 11	Sub 12	Sub 13	Sub 14	Sub 15	Sub 16	Sub 17	Sub 18	Sub 19	Sub 20
Cond 1a	2	2	2	5	3	2	5	5	1	5	1	5	1	4	1	3	1	2	5	5
Cond 1b	2	1	3	4	5	3	5	1	5	4	2	3	5	4	5	3	5	1	3	2
Cond 1c	2	1	3	5	3	1	3	3	5	4	3	2	3	4	5	5	1	3	2	3
Cond 1d	3	1	2	3	5	2	2	1	1	5	2	5	3	4	5	5	1	5	4	1
Cond 2a	2	2	5	5	3	5	3	1	5	3	2	1	2	4	1	2	1	3	5	1
Cond 2b	2	1	5	5	2	2	3	2	1	4	1	5	2	4	3	5	1	2	3	5
Cond 2c	2	1	3	3	3	2	4	3	5	5	2	3	1	4	3	3	5	5	4	4
Cond 2d	1	1	2	3	5	3	3	5	1	3	2	5	3	4	3	4	1	3	5	3
Cond 3a	2	2	3	2	2	2	5	3	5	5	2	5	2	4	3	5	2	5	5	3
Cond 3b	2	1	3	2	5	3	2	5	5	3	1	4	1	4	3	4	2	5	1	4
Cond 3c	2	1	5	5	5	2	3	5	5	5	1	3	1	4	5	5	1	5	1	5
Cond 3d	5	1	3	5	5	4	3	1	3	3	3	1	1	4	5	5	1	5	4	3
Cond 4a	2	5	2	2	2	2	3	1	5	4	3	1	2	4	5	3	1	2	1	2
Cond 4b	2	2	5	2	3	4	3	1	3	5	5	1	1	4	3	5	3	2	1	3
Cond 4c	5	1	3	3	3	3	3	2	5	3	2	5	5	4	2	1	5	3	5	3
Cond 4d	2	1	5	3	2	3	4	1	5	5	3	5	1	4	3	3	1	3	5	4

Note: A suffix of *a* denotes replication 1, a suffix of *b* denotes replication 2, a suffix *c* denotes replication 3, and a suffix *d* denotes replication 4.

Table 5 - 4: Number of times a policy was used in each of the conditions.

	Policy 1	Policy 2	Policy 3	Policy 4	Policy 5
Condition 1a	5	5	2	1	7
Condition 1b	3	3	5	3	6
Condition 1c	3	3	8	2	4
Condition 1d	5	4	3	2	6
Condition 2a	5	5	4	1	5
Condition 2b	4	6	3	2	5
Condition 2c	2	3	7	4	4
Condition 2d	4	2	8	2	4
Condition 3a	0	8	4	1	7
Condition 3b	4	4	4	4	4
Condition 3c	5	2	2	1	10
Condition 3d	5	0	6	3	6
Condition 4a	4	8	3	2	3
Condition 4b	4	4	6	2	4
Condition 4c	2	3	8	1	6
Condition 4d	4	2	6	3	5

Table 5 - 5: Summary of subject decisions and their outcomes

DECISION	RESULT		
	FAILURE	SUCCESS	TOTAL
URN 1	3376	2435 (41.90%)	5811 (60.53%)
URN 2	2245	1544 (40.75%)	3789 (39.47%)
TOTAL	5621	3979 (41.45%)	9600

Table 5 - 6: Results of the decisions made by subjects (by condition).

Condition	Decision	Result		Total
		Failure	Success	
1	Urn 1	842	701	1543
	Urn 2	541	316	857
	Total	1383	1017	2400
2	Urn 1	792	514	1306
	Urn 2	521	573	1094
	Total	1313	1087	2400
3	Urn 1	824	738	1562
	Urn 2	591	247	838
	Total	1415	985	2400
4	Urn 1	918	482	1400
	Urn 2	592	408 (40.8%)	1000
	Total	1510	890	2400

Table 5 - 7: Results of the decisions made by subjects (by condition x replication) TAB 3

Replication	Condition	Decision	Result		Total
			Failure	Success	
1	1	Urn 1	203	153 (42.98%)	356 (59.33%)
		Urn2	153	91 (37.30%)	244 (40.67%)
		Total	356	244 (40.67%0	600
2		Urn 1	227	179 (44.09%)	406 (67.67%)
		Urn2	126	68 (35.05%)	194 (32.33%)
			353	247 (41.17%)	600
3		Urn 1	206	175 (45.93%)	381 (63.50%)
		Urn2	141	78 (35.62%)	219 (36.5%)
			347	253 (42.17%)	600
4		Urn 1	206	194 (48.5%)	400 (66.67%)
		Urn2	121	79 (39.50%)	200 (33.33%)
			327	273 (45.50%)	600
1	2	Urn 1	192	129 (40.19%)	321 (53.50%)
		Urn2	129	150 (53.76%)	279 (46.50%)
			321	279 (46.50%)	600
2		Urn 1	196	120 (37.97%)	316 (52.67%)
		Urn2	138	146 (51.41%)	284 (47.33%)
			334	266 (44.33%)	600
3		Urn 1	200	125 (38.46%)	325 (54.17%)
		Urn2	131	144 (52.36%)	275 (45.83%)
			331	269 (44.83%)	600
4		Urn 1	204	140 (40.70%)	344 (57.33%)
		Urn2	123	133 (51.95%)	256 (42.67%)
			327	273 (45.50%)	600

Results of the decisions made by subjects (by condition x replication) TAB 2 - (contd)

Replication	Condition	Decision	Result		Total
			Failure	Success	
1	3	Urn 1	197	185 (48.43%)	382 (63.67%)
		Urn2	152	66 (30.28%)	218 (36.33%)
		Total	349	251 (41.83%)	600
2		Urn 1	212	161 (43.16%)	373 (62.17%)
		Urn2	164	63 (27.75%)	227 (37.83%)
			376	224 (37.33%0	600
3		Urn 1	223	195 (46.65%)	418 (69.67%)
		Urn2	122	60 (32.97%)	182 (30.33%)
			345	255	600
4		Urn 1	192	197 (50.64%)	389 (64.83%)
		Urn2	153	58 (27.49%)	211 (35.17%)
			345	255 (42.5%)	600
1	4	Urn 1	219	112 (33.84%)	331 (55.17%)
		Urn2	153	116 (43.12%)	269 (44.83%)
			372	228 (38.00%)	600
2		Urn 1	229	117 (33.82%)	346 (57.67%)
		Urn2	156	98 (38.58%)	254 (42.33%)
			385	215 (35.83%)	600
3		Urn 1	241	139 (36.58%)	380 (63.33%)
		Urn2	129	91 (41.36%)	220 (36.67%0
			370	230 (38.33%)	600
4		Urn 1	229	114 (33.24%)	343 (57.17%)
		Urn2	154	103 (40.08%)	257 (42.83%)
			383	217 (36.17%)	600

Table 5 - 8: Results of Simulation 1 (TAB3)

	6_1	6_2	6_3	6_4	6_5	7_1	7_2	7_3	7_4	7_5
Cond1	47.49	46.01	45.85	45.49	44.93	47.09	46.27	45.85	45.33	45.23
Cond2	44.68	44.45	44.26	43.68	44.17	44.49	44.4	43.79	43.85	43.78
Cond3	46.83	46.33	45.79	45.02	44.41	47.36	46.44	46.02	45.18	44.92
Cond4	42.85	40.94	39.28	38.01	37.67	41.76	40.71	39.25	38.17	38.03
Total	**181.85**	**177.73**	**175.18**	**172.2**	**171.18**	**180.7**	**177.82**	**174.91**	**172.53**	**171.96**

	8_1	8_2	8_3	8_4	8_5	9_1	9_2	9_3	9_4	9_5
Cond1	46.99	45.92	45.7	45.27	43.97	46.61	45.84	45.44	44.27	44.59
Cond2	45.2	44.69	44.63	43.93	44.24	44.68	44.35	43.75	44.1	44.94
Cond3	46.96	45.12	44.82	44.39	43.51	45.9	45.87	44.79	45.36	44.1
Cond4	41.78	39.97	38.92	37.56	37.43	41.41	39.59	38.5	38.35	37.33
Total	**180.93**	**175.7**	**174.07**	**171.15**	**169.15**	**178.6**	**175.65**	**172.48**	**172.08**	**170.96**

	10_1	10_2	10_3	10_4	10_5	11_1	11_2	11_3	11_4	11_5
Cond1	46.16	45.28	44.58	44.18	43.89	46.26	45.32	44.85	44.258	43.46
Cond2	44.51	44.71	44.51	43.725	43.42	44.48	43.99	43.84	45	44
Cond3	45.36	44.43	44.37	43.59	43.05	45.28	44.58	44.01	44.15	43.46
Cond4	41.1	39.18	37.93	37.18	36.88	40.76	38.73	37.93	37.59	36.43
Total	**177.13**	**173.6**	**171.39**	**168.67**	**167.24**	**176.78**	**172.62**	**170.63**	**170.99**	**167.35**

	12_1	12_2	12_3	12_4	12_5
Cond1	45.13	44.575	44.08	44	43.21
Cond2	44.48	44.075	44.59	44.58	44.58
Cond3	44.28	44.23	43.58	43.5	42.79
Cond4	39.19	38.25	38.26	36.94	36.62
Total	**173.08**	**171.13**	**170.51**	**169.02**	**167.2**

Simulation 1: Alternate for *n* trials, the compare and stay on the higher urn. If urn 2 is chosen, then stay on it for *x* losses and switch back to urn 1 after that to stay on it till the end of the task. Each of the numbers in the cells are percentages of success over the entire simulation. The column headings use the notation *n_x*. For example 12_1 implies that it alternated for 12 trials, then compared the estimated probabilities and if urn 2 , stayed on it for 1 loss before switching to urn 1 and staying there till the end of the task.

Table 5 - 9: More results from Simulation 1 (TAB3)

	1_1	2_1	3_1	4_1	5_1
Cond1	48.99	48.41	47.98	47.78	47.73
Cond2	44.51	44.72	44.8	44.88	45.11
Cond3	49.14	48.1	48.39	47.82	47.88
Cond4	44.71	43.88	44.08	42.87	43.32
TOTAL	**187.35**	**185.11**	**185.25**	**183.35**	**184.04**

Table 5 - 10: Results from Simulation 2 (TAB3)

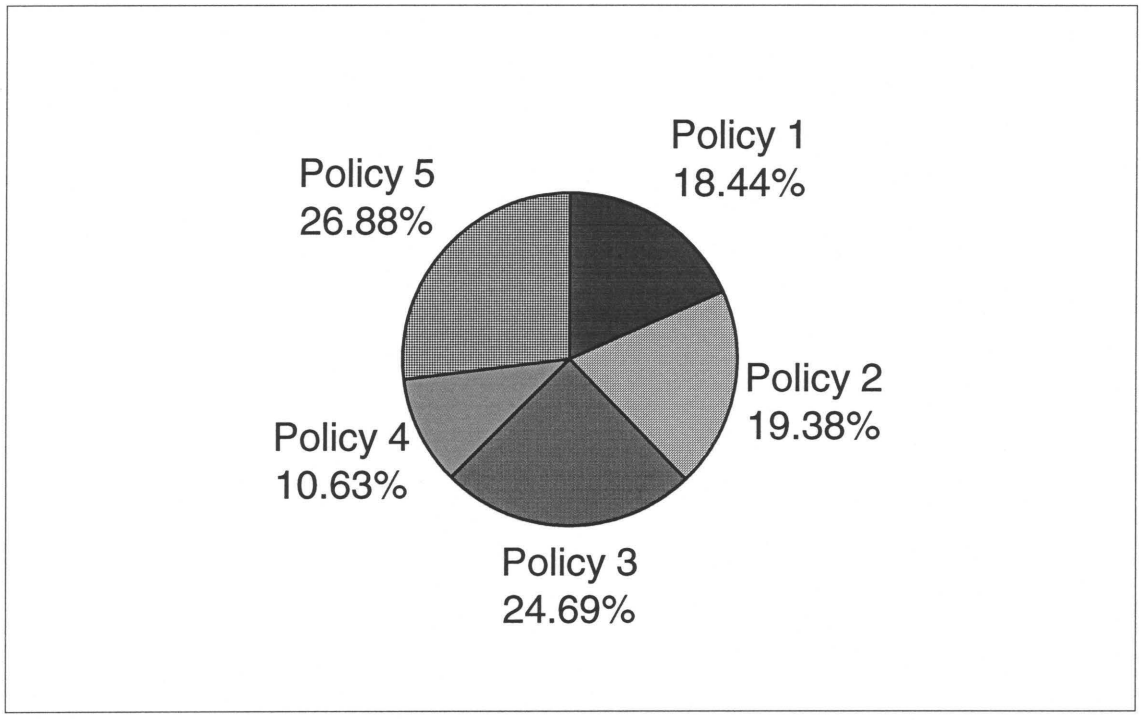
	COND1	COND2	COND3	COND4	TOTAL
losses=1	49.65	44.39	49.25	45.61	188.9
losses=2	49.48	43.19	48.45	43.64	184.76
losses=3	46.39	44.21	47	41.6	179.2
losses=4	45.73	43.93	45.11	39.59	174.36
losses=5	45.19	43.49	44.13	38.63	171.44
losses=6	43.57	44.63	43.03	37.48	168.71
losses=7	42.54	44.21	41.1	37.1	164.95
losses=8	41.91	43.73	40.71	36.47	162.82
losses=9	40.93	43.77	39.52	36.33	160.55
losses=10	40.01	43.82	38.25	35.86	157.94

Simulation 2: Stay an Urn 2 for x losses, then switch to urn 1 and stay there till the end of the task. $1 \leq x \leq 10$. Numbers in the cells represent averages over 100 repetitions of the simulation.

Table 5 - 11: Results of Simulation 3 (TAB3)

Trial	CONDITION 1	CONDITION 2	CONDITION 3	CONDITION 4
1	38.00	25.00	37.25	19.25
2	43.75	29.75	41.00	21.25
3	44.50	33.00	42.00	21.75
4	41.00	29.25	42.50	21.75
5	45.50	31.00	42.25	25.50
6	41.50	32.00	44.00	26.75
7	44.75	36.25	41.75	22.50
8	47.75	33.00	42.50	26.00
9	50.25	37.00	46.75	28.00
10	46.25	35.75	45.00	32.00
11	46.50	39.00	44.75	33.75
12	48.00	38.50	43.75	36.50
13	46.75	41.25	46.25	35.25
14	48.25	42.25	53.50	39.75
15	48.00	44.25	49.50	45.00
16	49.25	40.75	50.00	46.75
17	51.50	47.25	52.75	45.50
18	56.25	48.00	52.00	43.25
19	50.50	44.25	50.75	53.75
20	55.25	44.50	52.25	52.75
21	55.50	49.00	55.00	62.75
22	55.75	48.75	56.50	67.25
23	62.25	56.50	55.50	67.75
24	55.50	55.75	58.00	74.50
25	61.75	57.00	59.00	73.25
26	59.75	59.50	59.75	80.00
27	59.75	60.00	63.00	78.75
28	63.00	65.50	61.00	87.00
29	62.25	67.25	64.50	88.50
30	59.00	65.50	61.75	88.75
Ave	**51.27**	**44.56**	**50.48**	**48.18**

Figure 5 - 1: Distribution of subjects decisions between various policies (TAB 3)



Policy 1: Frequent Switching
Policy 2: Moderate Switching
Policy 3: Infrequent Switching
Policy 4: Stay on the same urn
Policy 5: Stay on the same urn for most trials

INSTRUCTIONS TO SUBJECTS FOR TAB - 3

You are about to participate in a computer-controlled decision making experiment in which your payoff is contingent on your performance. The money you will earn will be paid to you at the end of the experiment. A research foundation interested in clinical decision-making has contributed the money to finance the study.

The major purpose of this experiment is to study how individuals behave when asked to make a sequence of choices between two alternatives, each of which yields a success with a probability which is initially unknown. The task that we have in mind is the sequential testing of two new medicines, (whose properties have not been established), on a series of patients. At any time only one drug can be administered to a patient, and the treatment is scored as either a success or a failure. Only when the result of the treatment is known, is the next patient administered one of the two drugs. Clearly, one would like to maximize the number of successes; however, only by trying each of the drugs can the physician learn which is more effective.

It would be simpler to describe the task that you are about to perform in terms of two urns, labeled 1 and 2. Suppose that each urn has some black and some white balls in it. The total number of balls may vary from one urn to another. You are not informed of the total number of balls in each box, or the proportions of black and white balls in each urn.

Your task is simple. You will make a series of drawings from the urns, one on each trial. The total number of drawings (denoted by T), will be known to you when the task starts. At the beginning of each trial, your task is to decide which urn to draw a ball from. Choose an urn, draw a ball blindly, and record its color (either black or white).

Your payment on each trial depends on the ball you draw.

If you draw a white ball, you will be paid one unit of money, called "franc". (Francs will be converted into US dollars at the end of the experiment). If the white ball is drawn from urn 1, please replace the white ball in the urn. If the white ball is drawn from urn 2, however, remove it from the urn (i.e. it is not replaced back into the urn).

If you draw a black ball, you will receive no payment. If the ball is withdrawn from urn 1, please remove the black ball from the urn; however, if the ball is withdrawn from urn 2, please replace the ball back into the urn.

This completes the trial. To repeat, once you have noted the color of the ball, you replace it, if it is a white ball from urn 1 or a black ball from urn 2, and remove it, if it is black ball from urn 1 or a white ball from urn 2. You get paid one franc for a white ball and zero for a black ball. Hence, each time you draw a black ball from urn 1, you increase your chances of getting a white ball on the next trial, and every time you draw a white ball from urn 2, you decrease your chances of getting a white ball in the next trial. The essence of this decision making task is to gather information about the urns and at the same time make decisions that increase your payoff.

Your objective in the task is to draw as many white balls as possible, thereby increasing your earnings.

Example

Supposing the two urns are composed as follows and that the number of balls is known:

URN	# WHITE BALLS	# BLACK BALLS
1	8	4
2	10	5

Suppose you choose Urn 1 in the first trial. If you draw a white ball (and receive a payment of 1 franc), then on the next trial the composition of the urns will be the same. If you draw a black ball from the same urn (and receive a payment of zero), then on the next trial the composition of the two urns will be:

URN	# WHITE BALLS	# BLACK BALLS
1	8	3
2	10	5

Having drawn a black ball from Urn 1, the chances of drawing a white ball have increased from 8/12 to 8/11 i.e. from 67% to 73%.

Continuing the same example, supposing that your drawings are as follows:

TRIAL	URN	COLOR	PAYOFF	# W IN URN 1	# B IN URN 1	# W IN URN 2	# B IN URN 2
1.	1	WHITE	1	8	4	10	5
2.	1	WHITE	1	8	4	10	5
3.	1	BLACK	0	8	3	10	5
4.	2	WHITE	1	8	3	9	5
5.	2	BLACK	0	8	3	9	5
6.	1	BLACK	0	8	2	9	5

Thus, after 6 trials, you have earned 3 francs. The chances of drawing a white ball from Urn 1 are now 8/10 (80%) and of drawing a white ball from urn 2, 9/14 (64%).

Drawing from two urns and keeping track of the number of balls drawn is a tedious task. Therefore, to simplify the task, we have computerized the entire process. Your only decision will be to choose an urn on each trial. The computer will then randomly draw a ball from the urn you have chosen, exactly as in the example above, and will inform you of your payoff for the trial and your accumulated payoff from the beginning of the experiment.

You will participate in **16** different tasks. At the beginning of each task you will be informed of the total number of trials in the task. You will also be reminded that the new task has two different urns with possibly different compositions of white and black balls.

You may wonder how the (unknown) proportion of white balls in each urn has been determined. The only information that we can give you is that the percentage of white balls in each of the two urns can assume any value between 0 and 100 with all values in this range being equally likely.

Once the task begins the computer will display the following information on the screen before you make your decision:

History of previous play:

trial # urn 1 urn 2 cumulated payoff

You are on trial number (1.....T)

Which urn would you like to sample from? (1 or 2)

For example, suppose you sampled from urn 1 on the first trial and you picked a black ball, and from urn 2 the next time and picked a white ball. Your screen would then look similar to this:

History of previous play

	outcome		
trial #	urn 1	urn 2	cumulated payoff
1	Black	----	0
2	-----	White	1

You are on trial number (1.....T)

Which urn would you like to sample from? (1 or 2)

Once you enter either '1' or '2,' the computer will randomly draw a ball from that urn and present you with the following screen.

On trial number (1.....T)

Urn chosen on this trial (I or II)

Ball picked up on this trial (Black or White)

Your earnings for this trial =_____' franc.

[Press any key to continue] [Press F1 to see history]

Following the presentation of the outcome of the trial, the computer will update your record and return you to the first screen for the next trial. This process will be repeated for 'T' trials. At any time during the task you will be able to check the history for trials that are no longer visible (the history screen will hold the results for about 20 trials) by using the **F1** key that is located at the top left hand side of your keyboard.

As mentioned, you will participate in a series of 8 tasks. The number of trials will differ from one task to another. The two urns will also change between the tasks. In other words, the proportion of the white and black balls originally in each of the urns will change between the tasks.

After each task you will see this screen:

The urns are still labeled I and II but they are not the same ones used in the previous. task.

Please keep in mind that the computer has not been programmed to play against you. Rather, the number of white balls and black balls in each urn has been predetermined and will change from trial to trial (depending on the ball that is picked) within the next 30 choices. Your task is to draw as many white balls as you can over the next 30 trials.

Your concern is to make as much money as you can, given these conditions. You are being paid in a fictitious currency called a "franc". "Francs" will be converted into

US dollars at the end of the experiment at the rate of $1 = 20 francs. You will be paid before leaving.

If you have any questions about the instructions, please inform the experimenter immediately. You may refer to the written copy of the instructions placed on your desk at any time. **Please take whatever time necessary to make your decisions in this experiment**.

CHAPTER 6: CONCLUSION

DISCUSSION

"Decision theory is a complex, somewhat ill-defined, body of knowledge developed by mathematicians, statisticians, economists, and psychologists attempting to prescribe how decisions should be made and to describe systematically what variables affect decisions." (Rapoport and Wallsten, 1972). These authors suggest that whereas the distinction between individual and group decision-making is obvious, classification of decision tasks and theories of individual decision making is not, and the various classifications that have been proposed are not always clear. They distinguish between single-stage and multi-stage games depending on the number of decisions the DM is required to make. They go further to distinguish between two classes of multistage decision tasks, sequential and dynamic. In sequential decision tasks, stage-to-stage changes in the state of the system do not depend on the DM's previous decisions. Noncontingent probability learning, revision of opinion, and optimal stopping tasks fall in this category. In a dynamic decision task, stage to stage changes in the state of the system are affected by the DM's previous decisions as well as by the states of the system at the preceding stages. The Reader Control Problem (Rapoport, 1966a, 1966b, 1967a), multistage inventory control tasks (Rapoport, 1967b; Pierskalla, 1969; Rapoport and Calder, 1972), multistage betting games (Rapoport, 1970; Rapoport and Jones, 1970), and the TAB fall under this class of problems. Most of these dynamic tasks are very complex either in the presentation of the problem or in the number of choices available to the DM.

The TAB in comparison is much simpler with respect to the way it is presented to the DM and the range of decisions on each trial (simple binary choice). Thus, it is an excellent choice for the study of behavior in dynamic tasks.

Solutions to the TAB problem have been around for at least seventy years. This area of research began with the Thompson article in the Biometrika (1933), where he posed the problem, and was then extended by various other statisticians, mathematicians, and economists like Robbins, Berry, Gittins, Jones, and Rosthschild to name a few. Much of the earlier research focussed on characterizing the optimal solution. For the most part, problems of this nature have been approached via dynamic programming formulations. Various extensions of the original work have been developed incorporating precedence constraints, arrivals, and a variety of more general reward/decision structures. Research has also looked at various alternatives to the dynamic programming approach. While there has been a preoccupation with the theoretical solutions, very little effort has been directed to examining the performance of naïve subjects when faced with similar tasks.

The bandit problem in our experiments was presented to the subjects in a simplified manner with the use of urns and colored balls. The purpose of this thesis was twofold: to present the subjects with a series of more realistic and consequently progressively more difficult tasks, and to account for their behavior. Our subjects faced dynamic tasks that allowed for learning. Instead of just pulling the imaginary arms of a slot machine, they were also presented with realistic cover stories to help them understand the task better and put it in perspective for them. Therefore, whether the task

was presented to them as the testing of two medications or deciding between which of two missile systems to use it was definitely more pertinent.

We focussed on both individual and aggregate behavior. Given that the problems we are tackling are so ill defined, it precluded us from developing the normative solutions. Hence, we took a descriptive approach since there was very little, if any, theoretical work done for the specific versions of the bandit problem that we studied. Faced with the dilemma, we chose to sacrifice mathematical tractability for realism. These particular problems were chosen because of their applicability to real life decision choices and to slowly start adding to the literature in this area.

Our research focused on the classic TAB problem and three variations of it. In all the experiments, subjects were told that their main task was to maximize their earnings over the course of the experiments. In the classic TAB problem, subjects had to choose between two urns. Sampling was with replacement. For the subsequent experiments, we gradually increased the difficulty of the task. In the second experiment, subjects had to choose between a known urn (fixed and known probability of success) and an unknown but increasing urn (probability of success on the arm increases following a failure in that urn). In the third experiment, subjects faced two unknown but increasing urns (probability of success in an urn increases after a failure in that urn). In the fourth and last experiment, subjects faced one increasing urn (probability of success increases after a failure in that urn) and one decreasing urn (probability of failure increases after a success in that urn).

Our approach was two pronged. Subjects participated in the experiments, and we did a descriptive analysis of their decisions. Experimental economics tends to gloss over individual differences in their analyses. Economists seem to prefer to capture aggregate behavior while our approach is more psychological in nature, giving particular attention to individual decision-makers. In the final analysis, it is individual behavior rather than aggregate behavior that we want to understand. For the classic TAB experiment, we compared the decisions of our subjects to the optimal policy and to two degradations of the optimal policy. The optimal policy had been calculated by Bellman (1957) using dynamic programming which is a computational technique that prescribes decisions for a class of multistage decision making problems. This provided the normative baseline. For the other two heuristics, assumptions about the past (memory) and the future (horizon) were degraded. These two heuristics had been studied by Horowitz (1973) in his dissertation. He justified their use stating

" ... it was suspected that relative to the TAB optimal policy slight degradations would result in negligible losses, we chose to concentrate on two extreme degradations: (1) the loss of all memory with the exception of the information included in the last stage, and (2) the limitation of the horizon on the next stage only."

Since we had no theoretical results to compare our subjects decisions to for the other three experiments, we ran simulations of three heuristics for each of the experiments. Then we compared the subjects actual earnings to the potential earnings that could have been realized had the heuristics been used.

Subjects faced a series of tasks in each experiment. None of the subjects was allowed to participate in more than one experiment. In each of the experiments, subjects faced at least 3 different tasks and each task was repeated 2 - 4 times. The tasks were differentiated from each other by the number of white and black balls placed in each of the urns. Subjects were informed when a new task was about to start. They were also told very clearly that they would be choosing from a new set of urns with a different combination of balls.

Several results from each of the experiments warrant further mention. We identified various policies that subjects were using while making decisions. It should be noted that not one of the policies (in any of the experiments) could account for even 50% of the subjects' decisions. At best, a few of the policies could account for a little over 25% of the decisions in some of the experiments. This tells us that subjects were extremely heterogeneous in their decision-making. Subjects were not consistent in their use of policies either. There were only a few subjects who used the same policy consistently in all the tasks that they faced. However, they were an exception. Most subjects switched between policies but did not use them consistently when the tasks were repeated.

In the classic TAB problem, subjects faced tasks of two lengths. In the N=10 tasks we found that when the probabilities of success are low, subjects performed worse than chance. However, when the probabilities were higher, they achieved close to 77% success. We also find that overall subjects tended to oversample from Urn 2, which had the lower rate of success. In his study, Horowitz found that subjects tended to oversample

the less promising urns when the probabilities are low. Our studies do not support that finding. However, we found that there appeared to be a tendency among our subjects to oversample from the less promising urn when the difference between p_1 and p_2 (Δ=. 15) was high. This effect vanished when the Δ was low (Δ=. 05). We also compared the subjects decisions to the optimal policy and to two degradations of the optimal policy. We found that the subjects matched the optimal policy 67% of the time in the low probability tasks and 82% of the time in the high probability tasks. We checked the sensitivity of the optimal policy by running simulations. We found that the real subjects actually outperformed the simulation. The simulation matched the optimal policy only 58% of the time on average. Another difference between the subjects and the simulations is that while subjects did marginally better when the horizon was 1 and 2, corroborating studies done by Rapoport (1966), the simulation performed better when the horizon was equal to 6.stages. The other two heuristics tested were the OSM and OSH policies. Subjects matched the policy at least 49% of the time in each of the tasks, and over all the tasks matched it approximately 61% of the time. The OSH policy fared better than the OSM policy matching 66% of the subjects decisions. This suggests to us that subjects tend to be myopic. They prefer to consider the future than the past. It could be that subjects figure what has happened has happened, and nothing can be changed but there is something that can be anticipated in the future. Even though they are considering a horizon, this policy suggests looking one stage only into the future. In general, we found that as the probabilities increase the matches increase except for condition (0.4, 0.35)

where there was a slight drop in the number of matches with both the OSM and OSH policies. At least 30% of the decisions in every task matched both the OSM and OSH policies.

When N=40 the results were much clearer. Subjects chose the urn with the higher probability of success in all tasks. This clearly contradicts Horowitz's study. Our subjects do not oversample less promising options when probabilities are low. Thus, unlike the N=10 case, there was no effect for Δ either. Subjects matched the OSM and OSH policies on average 66% of the time. As in the N=10 case, the OSH matched more of the subjects' decisions than the OSM policies.

In the OAB experiment, our subjects were forced to choose between a known option and an unknown one. We found that subjects were willing to take risks when the known probabilities were low but not as willing when the known probabilities increased to medium and high. Studies have shown that individuals are often "ambiguity averse" when faced with a choice between two ambiguous lotteries with positive expected values, preferring options with more certain to less certain odds (e. g., Curley and Yates, 1989, Einhorn and Hogarth, 1985, Heath and Tversky, 1990). Our subjects appear to go a step further when the known probabilities are medium or high but not so when the known probabilities are low. We concluded that subjects were either exhibiting some sort of satisficing behavior or using the "anchoring and adjustment" heuristic. This also contradicts the findings by Meyer and Shi who say, "subjects tend to underexperiment with promising options and overexperiment with unpromising options." Three heuristics were considered, and the one that performed the best was the "stay on the increasing arm"

for all trials. One of the other heuristics was the "stay on a winner for the first N/3 trials and then stay on the better arm for the remaining N-n trials." This was adapted from Zelen's (1969) paper. The other heuristic was "try urn 2 for n trials and compare the estimated probability with known probability on urn 1 and stay on the higher urn. Both these heuristics also outperformed the subjects on average.

In TAB2 with two increasing arms, we find a slight tendency to chose urn 1 more often than urn 2 in each of the tasks except condition 2 where there was an even split between the two urns. In all conditions, there were fewer balls in urn 2, and hence the rate of increase would have been higher in urn 2 than urn 1. However, the initial probability of success was higher or equal in urn 1. It wasn't very clear if subjects could distinguish between the varying rates of increase in this experiment since in at least conditions 2 and 3 there was no significant difference between the number of times the urns were chosen. We compared the subjects' decisions to three simulations. All three simulations were variations of the "stay of a loser, switch on a winner policy". The main idea behind these policies was to try to exhaust the black balls in the urns as soon as possible so that we could have long runs of success. Subjects out-performed all three simulations. We named the first policy, the "stay on a loser, but switch on a winner for the first 'n' trials." At the end of 'n' trials, we compare the estimated probability on each urn and stay on the urn with the higher probability. We found that on average n=7 stages had the highest earnings. We find that the reason it did better than the other two simulations is because of condition 3. In this condition, we find that the other two simulations did dismally, earning on average just 17 francs. The reason is that the latter two heuristics did not exhaust the

black balls quickly enough since there the policies encourage switching between the urns. The first simulation stayed on one urn after 'n' trials thereby eliminating black balls from the urn over the length of the task.

In TAB3 with one increasing arm and one decreasing arm, we find that subjects switched frequently between the urns. Hence, subjects did not earn very much in this experiment. Subjects who switched less frequently earned more than those who switched more often. The three simulations that we conducted concurred with these results. The simulations suggest that it does not make sense to sample urn 2 (decreasing urn) for too long (no more than 2 losses). However, the subjects with a judicious mix of policies outperformed all three simulations that we used.

The last two studies were more difficult than the former two. We had no other studies to compare our results to. We also do not have theoretical results to benchmark our results against. Subjects outperformed all the heuristics that we used. It taught us never to underestimate subjects. When both arms are increasing, our simulations show that it is important to test both urns and compare the estimated probabilities before staying on one urn. It did not pay to switch back and forth between the urns frequently. Although we introduced a decreasing arm in the last experiment, we find that the same conclusion holds. That is, if one switched back and forth between the urns often one reduced his/her payoff. However, in the last experiment it didn't pay to stay on urn 2 for very long. The simulations seem to suggest that one shouldn't stay on urn 2 for more than a single loss.

Overall, we have mixed results. In the classic TAB problem and the OAB problem we find our results contradict earlier studies. Nevertheless, our results are just as interesting. Other studies corroborate similar findings. However, what we would like to see is if these results are robust. Will these same results hold if the probabilities were changed or the length of the tasks increased or some other new factor introduced?

LIMITATIONS

The present study can be further improved in several ways. First, one wonders if this task was too difficult for the subjects. Would that account for some of the subjects' poor performance? Second, we did not include manipulation checks to see if the subjects understood the tasks. Whereas we made the instructions as clear as possible with detailed examples, did they truly understand the difference between sampling with and without replacement? We did have them go through practice rounds but the screens they saw did not explain the task any more than it would have in the real trials. In retrospect, we should have given them a questionnaire to check whether they understood what they were doing.

Third, in the classic TAB problem the values of Δ were too small, given the stochastic nature of the task. Additionally, crossing Δ with N might not have been a very good idea either. In particular the values of Δ for $N = 10$ should have been larger so that subjects could clearly differentiate between the two urns since there weren't sufficient trials to gather information and then act on it.

Fourth, the question arises whether we presented too many different tasks to the subjects. We tried to vary the tasks to prevent boredom and monotony, but we could have introduced a high level of difficulty unconsciously. This might have contributed to some confusion within the subjects and the lack of clear results.

FURTHER RESEARCH

This discussion immediately leads us to the question of further research. All the limitations stated above can be easily addressed in future studies. Nevertheless, other interesting problems can be studied in this class of problems. Studies can be further extended to include multiple arms. As mentioned in the first chapter, much theoretical work has been done on multiple arms. It can easily be justified since the very crux of the bandit problem is making choices when faced with ambiguous situations. More often that not, we are faced not with just two but multiple scenarios that we must chose from. Another extension could be introducing a penalty for making a wrong decision or maybe introduce the notion of switching costs. Most decisions in real life have a penalty associated with it. For example, one can't switch from one job to another without incurring a cost. That can be easily captured in an experimental situation. A third interesting variation would be to study the model of a multi-person two-armed bandit; where the players gather information in two ways, namely, by experimentation and by observing other players' actions. While experimentation will be a direct way to learn, observing other players' actions would be an indirect but important way to gather information as those actions by others may reflect their private information. Finally, a fourth variation might be the case where when subjects decided they want to switch urns,

the switch is delayed and takes place only after a prespecified number of trials. This switchover period may be known or unknown to the subjects and could be fixed or changing.

This thesis is, hopefully, only the beginning of a whole series of studies that can be conducted in this arena. Several suggestions for further research have been mentioned above. The idea is to contribute to the experimental literature in this area since it is severely lacking. Interest in the bandit literature is growing as more uses are found for it. We have already discussed its applications in the area of medical research, job hunting, pricing, and learning. By researching multiple arm bandits or switching costs, we would be rendering the problems more likely to resemble real life.

REFERENCES

Banks, J and Sundaram, R. (1992). A Class of Bandit Problems Yielding Myopic Optimal Strategies, *Journal of Applied Probability*, v29, n3, 625-632.

Banks, J. and Sundaram, R. (1992). Denumerable-Armed Bandits, Econometrica, v60, n5, 1071-1096.

Banks, J. S., and Sundaram, R. K. (1994). Switching costs and the Gittens Index, *Econometrica*, v62, n3, 687-694.

Banks, J., Olson, M., and Porter, D. (1997). An Experimental Analysis of the Bandit Problem, *Economic Theory*, 10, 55-77.

Bellman, R. (1956). A problem in the Sequential Design of Experiments. *Sankhya* 16, 221-229.

Bellman, R. (1961). *Adaptive Control Processes*. Princeton: Princeton University Press.

Bergmann, D. and Valimaki, J. (1996). Learning and Strategic Pricing, *Econometrica*, v64, n5, 1125-1149.

Berry, D. A. (1972). A Bernoulli Two Armed Bandit, *The Annals of Mathematical Statistics*, v.43, No. 3, 871-897.

Berry, D. A. (1978). Modified Two Armed Bandit Strategies for Certain Clinical Trials, *Journal of the American Statistical Association*, v.73, 339-345.

Berry, D. A. (1985). One and two-armed bandit problems. *Encyclopedia of Statistical Sciences*, vol. VI (eds. by S. Kotz and N. L. Johnson), Wiley, New York.

Berry, D. A., and Fristedt, B. (1984*). Bandit problems: sequential allocation of experiments*. London, Chapman and Hall.

Bolton, P, and Harris, C. (1999). Strategic Experimentation, *Econometrica*, v67, n2, 349-374.

Bradt, R. N., Johnson, S. M. and Karlin, S.(1956). On Sequential Designs for Maximizing the Sum of n Observations, *The Annals of Mathematical Statistics*, v 27, 1060-1074.

Brand, H., Woods, P. J., and Sakoda, J. M. (1956). Anticipation of Reward as a Function of Partial Reinforcement, *Journal of Experimental Psychology*, 52, 18-22.

Bush, R. R. and Mosteller, F. (1955). *Stochastic Models for Learning*. New York: Wiley.

Cane, V. R. (1962). Learning and Inference. *Journal of the Royal Statistical Society*, 125, 183-209.

Colton, T. (1963). A Model for Selecting One of Two Medical Treatments. *American Statistical Association Journal*, June, 388-400.

Cover, T. (1968). A Note on the Two-Armed Bandit Problem with Finite Memory, Information and Control, 12, 371-377.

Curley, S. P., and Yates, F. J. (1989). An empirical evaluation of descriptive models of ambiguity reaction in choice situations, *Journal of Mathematical Psychology*, v.33 (4), 397-427.

De Groot, M. H. (1970). *Optimal Statistical Decisions*, New York: McGraw Hill.

Einhorn, H. J., and Hogarth, R. M. (1985). Ambiguity and uncertainty in probabilistic inference, *Psychological Review*, v. 92 (4), 433-461.

Fushimi, M. (1973). An improved version of a Sobel-Weis play-the-winner procedure for selecting the better of two binomial populations. *Biometrika*, 60, 3, 517-523.

Glazebrook, K. D. (1979). Stoppable Families of Alternative Bandit Processes, *Journal of Applied Probability*, 16, 843-854.

Glazebrook, K. D. and Garbe, R. (1996). Reflections on a New Approach to Gittins Indexation, *Journal of the Operational Research Society*, v47, n10, 1301-1309.

Glazebrook, K. D. And Owen, R. W. (1991). New results for Generalized Bandit Problem, *International Journal of System Science*, vol. 22, no. 3, 479-494.

Gulati, R. (1995) Does Familiarity Breed Trust? The Implications of Repeated Ties for Contractual Choice in Alliances, *Academy of Management Journal*, 38: 85-112.

Haveman, H. (1993). Follow the leader: Mimetic Isomorphism and Entry into New Markets, *Administrative Science Quarterly*, v. 38, n4, 593-627.

Heath, C., & Tversky, A. (1990). Preference and belief: Ambiguity and competence in choice under uncertainty. In Borcherding, K., Oleg, I., et al (Eds.), *Contemporary issues in decision-making* (pp 93-123). Amsterdam, Netherlands: North-Holland.

Hoel, D. G. (1972). An Inverse Stopping Rule for play-the-Winner Sampling. *Journal of American Statistical Association*, v.67, n.337, 148-151.

Horowitz, Abraham V. (1973). Experimental Study of the Two Armed Bandit Problem, Unpublished Ph.D. Dissertation, The University of North Carolina, Chapel Hill, NC.

Jones, P. W. (1975). The Two-Armed Bandit, *Biometrika*, 62, 523-524.

Jones, P. W. (1976). Some Results for the Two Armed Bandit Problem, *Math. Operationsforsch. u. Statist.* vol. 3. 471-475.

Jones, P. W. (1978). On the Two-Armed bandit with one Probability Known, *Metrika*, v 25, 235-239.

Kahneman, D., Slovic, P, and Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*, Cambridge, Cambridge University Press.

Kolonko, M and Benzing, H. (1985). The Sequential Design of Bernoulli Experiments including Switching Costs, *Operations Research*, v33, n2, 412-426.

Meyer, R. and Shi, Y. (1995). Sequential Choice under Ambiguity: Intuitive Solutions to the Armed Bandit Problem, *Management Science*, v. 41, no. 5, 817-834.

Mitchell, W. (1989) Whether and When? Probability and timing of Incumbents' Entry into Emerging Industrial Subfields, *Administrative Science Quarterly*, 34, 208-230.

Parkhe, A. (1993). Strategic Alliance Structuring: A Game Theoretic and Transaction Cost Examination of Interfirm Cooperation, *Academy of Management Journal*, v. 36, no. 4, 794-829.

Rapoport, A. (1967). A Study of a Multi-Stage Decision Making Task with an Unknown Duration. *Human Factors*, Feb, 54-61.

Rapoport, A., and Wallsten, T. S. (1972). Individual Decision Behavior, *Annual Review of Psychology*, 23, 131-176.

Robbins, H. E. (1952). Some Aspects of the Sequential Design of Experiments, *Bulletin of the American Mathematical Society*, 55, 527-535.

Rodman, L (1978). On the many-armed bandit problem. *Annals of Probability*. 6: 491-498.

Rothschild, M. (1974). A Two Armed Bandit Theory of Market Pricing, *Journal of Ecnomic Theory*, v9, n2, 185-202.

Schmalensee, R. (1975). Alternative Models of Bandit Selection, *Journal of Economic Theory*, v10, n3, 333-342.

Smith, V. C, and Pyke, R. (1965). The Robbins-Isbell Two Armed Bandit Problem with Finite Memory, *The Annals of Mathematical Statistics*, v36, n5, 1375-1386.

Thompson, W. R. (1933). On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samplers. *Biometrika*, 25, 285-294.

Viscusi, W. K. (1979). Job hazards and the worker quit rates: An analysis of adaptive worker behavior, *Internat. Econom.* Rev. 20: 29-58.

Wahrenberger, D. L., Antle, C.E., and Klimko, L. A. (1977) Bayesian Rules for the Two-Armed Bandit Problem, Biometrika, 64, 172-174.

Whittle, P. (1980). MultiArmed Bandits and the Gittins Index, *The Journal of the Royal Statistical Society, Series B (Metholodogical)*, v42, n2, 143-149.

Whittle, P. (1981). Arm-Acquiring Bandits, *The Annals of Probability*, vol. 9, 284-292.

Yakowitz, S. J. (1969). *Mathematics of Adaptive Control Processes*. New York: American Elsevier Publicity Company.