Journal of Economic Theory **101**, 252–280 (2001)
doi:10.1006/jeth.2000.2723, available online at http://www.idealibrary.com on IDEAL®

Learning While Searching for the Best Alternative

Klaus Adam

European University Institute, Via dei Roccettini 9, 50016 San Domenico di Fiesole, Florence, Italy

Received June 24, 1999; revised July 24, 2000;
published online March 29, 2001

This paper delivers the solution to an optimal search problem where the searcher faces more than one search alternative and is learning about the attractiveness of the respective alternatives during the search process. The optimal sampling strategy is characterized by simple reservation prices that determine which of the search alternatives to sample and when to stop searching. The reservation price criterion is optimal for a large class of learning rules, including Bayesian, nonparametric, and ad-hoc learning rules. The considered search problem contains as special cases many earlier contributions to the search literature and thereby unifies and generalizes two directions of research; search with learning from a single search alternative and search without learning from several search alternatives. *Journal of Economic Literature* Classification Numbers: D81, D83. © 2001 Elsevier Science

1. INTRODUCTION

Economic problems involving search due to uncertainty about the location of objects are copious and hence have received a considerable amount of attention. After the igniting article of Stigler [19] economists themselves have been searching, namely for sampling strategies that are optimal in different situations involving uncertainty (Lippman and McCall [11], McKenna [12]). This paper follows this tradition and determines the optimal search strategy for a class of search problems that is characterized by two main features, learning during the search process and distinguishable search alternatives.

To be explicit, consider the following job search example falling into the class of problems I consider. A job searching unemployed worker faces a number of job offering firms where each firm might either be willing to hire this worker and offer some wage or reject the worker's application. The fundamental uncertainty in the worker's search process consists of the fact that the worker does not know which firms are willing to hire at which wage and which ones would reject the application. Thus, the worker has to search for a good offer by applying to firms, observing the outcomes, and deciding whether to accept an offer or whether to continue searching.

0022-0531/01 $35.00

Learning is introduced by allowing for the natural possibility that the searcher is not only uncertain about which firm offers which wage but also uncertain about the prevailing wage offer distribution. The searcher, possessing priors about the offer distribution, can use a search outcome, i.e., a job offer from a particular firm, to learn about the wage offer distribution and update his priors.

It is equally natural to suppose that the searcher faces different types of firms and has different priors about the vacancies offered by firms of different type. One might think of a migration model where the worker has to decide whether to apply to domestic firms or to firms located abroad. Alternatively, one might think of workers facing firms operating within different industrial sectors or, even more explicitly, of newly graduated Ph.D. students facing different types of employers such as universities, international organizations, and private businesses.

In abstract terms, a search problem involving learning adds to the uncertainty about the location of objects further uncertainty about the objects' values. The presence of distinguishable search alternatives captures the fact that search opportunities typically differ from each other and that search involves a thorough choice among the available alternatives. Random sampling of offers, as commonly assumed, is then suboptimal.

If the search is sequential with full recall of previous offers, then I find that the optimal search strategy for the class of search problems involving learning and distinguishable search alternatives is characterized by a simple reservation price for each search alternative. The reservation price of an alternative is simply a real number that is assigned to the alternative and the higher this number, the more attractive it is to search the corresponding alternative. The reservation prices for all alternatives together determine which of the search alternatives to sample, and when to stop searching. The optimal strategy is very simple and prescribes always searching in the alternative with the highest reservation price and stopping searching as soon as the best offer exceeds the reservation prices of all available alternatives.

The optimal search strategy is a generalization of the one found by Weitzman [22] for the case without learning. The main difference is that we allow the reservation prices to change during the search process as new information arrives through new search outcomes. In this way, it is optimal for the searcher to stay reactive to the search outcomes and, for example, direct his search towards another search alternative if the outcomes of the previously searched alternative have been disappointing.

The optimality of the search strategy holds for a large class of learning rules for which, roughly speaking, the reservation prices keep decreasing as additional search outcomes are observed. Learning rules with this property include Bayesian learning as well as non-parametric and ad-hoc learning.

In addition to answering the question on how to search optimally in a situation involving learning and distinguishable search alternatives, the result of this paper should be of twofold interest to economists.

First, the answer to the normative question allows for positive modeling of economic behavior within the neoclassical maximization paradigm. There are many situations of economic interest that involve both of the above features and where the findings of this paper are applicable. Adam [1], for example, uses the present results to study the effects of consumers' ability to direct search efforts in an equilibrium search model. Besides the search for low prices the present framework also allows the modeling of firms' research for new products or technologies. Examples include oil companies searching for new oil fields to exploit or pharmaceutical companies' research for medical drugs. Such firms are confronted with different potential oil fields and different promising research approaches, respectively, and learn about the attractiveness of their alternatives during the (re)search process. Finally, the results can be applied to model investment decisions if investment is interpreted as the search for good investment projects.

Second, the result contains several earlier contributions to the search literature as special cases and thereby contributes to the unification and generalization of the search theoretical framework. Although learning and distinguishable search alternatives have already been considered in the literature only one of these features was present at a time (Rothschild [16], Rosenfield and Shapiro [15], Morgan [13], Talmain [20], Chou and Talmain [4], Bikchandani and Sharma [2] considered learning but assumed indistinguishable search alternatives; Salop [17], Weitzman [22], Vishwanath [21] studied distinguishable search alternatives but abstracted from learning) and many of the search problems studied in earlier contributions are contained in the class of problems considered in this paper.[1]

It is worth noting that removing learning or distinguishable search alternatives considerably reduces the complexity and realism of the search problems studied. On one hand, assuming indistinguishable search alternatives removes the choice decision from the search problem. All search alternatives are (at least believed to be) the same and the search problem then reduces to the question of when is the optimal time to stop searching. On the other hand, abstracting from learning implies that the value of a search outcome (e.g., of a job offer) consists solely in its payoff (i.e., the wage), since search outcomes do not convey any valuable information (e.g., about the wage offer distribution). As a result the optimal search strategy

[1] Exceptions are: Vishwanath [21] dealing with non-sequential search; Morgan [13] dealing mainly with the existence of reservation price functions; and Rothschild [16] who does not allow for the recall of previous offers.

has to condition only on the best of all observed offers (i.e., the best wage offered so far) and not on the whole sequence of observed offers.

Finally, note that the problem considered in this paper differs from simple armed bandit problems but that it is related to bandit superprocesses.

First, consider the difference with respect to the simple bandit problem. In such a decision problem the player receives a reward every time the arm of some bandit is pulled and nothing otherwise. In contrast to this, in the considered search problem a number of arms are pulled without actually receiving a reward. Only when the searcher decides not to pull any further arms (i.e., to stop searching) does he receive the best of all previously observed rewards.

Next, consider bandit superprocesses which are a generalization of simple bandit processes allowing for multiple arms per bandit. Adding a second "stopping arm" to each standard bandit (as in Glazebrook [9]) allows for the possibility that the payoff is obtained at the end of search when the stopping arm is pulled. Glazebrook shows that if the value of the stopping option is non-decreasing in the number of searches, then the optimal policy is characterized by some simple selection rule for the arms and the indices given by Gittins and Jones [6] for single-armed bandits, where the single armed bandits are constructed by applying the selection rule for the arms to the bandit superprocess. However, Glazebrook does not evaluate the derived indices and the monotonicity conditions that would allow for a straightforward explicit calculation fail to hold in the present case (e.g., Propositions 4.2 and 4.5 in Gittins [5]).[2] Thus, the contribution of this paper could also be considered as delivering an explicit expression for these indices in the absence of such monotonicity.[3]

The next section sets up the search problem I consider and explains how other search problems with identical search alternatives or without learning are special cases of the one considered here. Section 3 describes as a benchmark the optimal search strategy when the searcher knows the payoff distributions and is not learning. Section 4 contains the main part of the paper. I delineate the class of admitted learning rules and present the optimal search strategy for the case with learning. I also explain why the sampling rule of the benchmark problem generalizes to the case with learning. In Section 5 I ask whether one can also hope for optimality of the search rule with more general learning rules than the ones I consider. Except for a very special case the answer is found to be negative. A conclusion summarizes the findings. The appendix contains the proofs.

[2] Note that although we have decreasing reservation prices with our learning rules there is always a positive probability that the search outcome is above the reservation price.

[3] For similar exercises see Glazebrook [7] and [8].

2. THE MODEL

A search problem is characterized by a searcher facing a (possibly infinite) number of search opportunities. Each search opportunity can be thought of as a box that contains an uncertain reward. The searcher has the possibility to open any box at a cost and find out what reward is contained in the box. I want to allow the boxes to differ from each other, not only with respect to the actual reward they contain but also with respect to the probability with which they contain (or are believed to contain) certain rewards. One can think of this as different boxes having different colors on the outside, while equal boxes are of equal color. Each color then represents a search alternative and the searcher, being able to observe these colors, has to choose among them in every search step.

More formally, let the boxes be indexed by the natural numbers and let the set $J = \{1, 2, ...\}$ contain all the available boxes. Each box $j \in J$ has some color $i \in \{1, 2, ..., I\}$, i.e. there are I different colors or search alternatives. The color of a box is observable at no cost. To simplify the language a box of color i will sometimes be referred to as an i-box.

There are M^i boxes of color i where M^i can be finite or infinite. Boxes of the same color are identical and are characterized by the triple $\{c^i, t^i, d^i(\cdot)\}$ where c^i is the cost of opening an i-box, t^i is the time span that passes from opening the box until its reward is observed and the function d^i: $R \mapsto [0, 1]$ describes the probability distribution of the rewards from opening the box. The parameters c^i and t^i are known to the searcher while $d^i(\cdot)$ is unknown. The functions $d^i(\cdot)$ can have support on $\mathbf{R}$ and the random variables described by them are assumed to have finite mean if $M^h < \infty$ for all $h = 1, 2, ..., I$ and to have finite variance in all other cases.

For a given point in time I denote by r^i the number of already opened i-boxes. x^i_n is the outcome from opening the nth box of color i. The vector $X^i_{r^i} = (x^i_1, x^i_2, ..., x^i_{r^i})$ contains the observed outcomes of opening i-boxes.

The searcher sequentially samples boxes and can open a closed box of color i by paying the amount c^i. He has to wait a time span t^i and then receives an offer drawn from $d^i(\cdot)$.[4] Recall of previously drawn offers is allowed. If the search stops the searcher gets y, which is the maximum of the so far drawn offers and some outside opportunity x^o the searcher possesses independently from the search outcomes:

$$y = \max\{x^o, x^1_1, x^1_2, ..., x^1_{r^1}, ..., x^I_1, x^I_2, ..., x^I_{r^I}\}.$$

[4] That search costs c^i have to be paid some time t^i before the search result is observed is not restrictive. Problems where c^i is payed at the time when the search results are observed fit into the problem by appropriately discounting search costs.

The searcher maximizes discounted expected payoffs minus costs with a discount rate $0 \leqslant r < \infty$.

Uncertainty has two sources. First, offers from boxes are drawn from some probability distribution. Second, there is uncertainty about the prevailing distribution from which offers are drawn. Uncertainty about $d^i(\cdot)$ may be represented by beliefs in the form of a probability distribution $p^i(\theta)$ over some parameter θ that indexes the set of possible true probability distributions $d^i(\cdot \mid \theta)$ for boxes of color i, where the true distribution function $d^i(\cdot)$ is equal to $d^i(\cdot \mid \theta^i)$ for some specific value θ^i of the parameter. Beliefs $p^i(\theta)$ about boxes of color i are updated using the observed search outcomes $X^i_{r^i}$ from i-boxes. Updated beliefs are denoted by $p^i(\theta \mid X^i_{r^i})$. Given these beliefs one can calculate an expected true probability distribution $f^i(x \mid X^i_{r^i})$ for the boxes of each color by integrating out the uncertainty about the parameter θ:

$$f^i(x \mid X^i_{r^i}) = E[d^i(x \mid \theta) \mid X^i_{r^i}] = \int_{\Theta} d^i(x \mid \theta)\, p^i(\theta \mid X^i_{r^i})\, d\theta.$$

For expository reasons, $f^i(\cdot \mid \cdot)$ has been derived from the Bayesian learning mechanism above. Since I do not want to confine myself to rational learning, I also allow $f^i(\cdot \mid \cdot)$ to be directly specified by some non-rational ad-hoc learning rule.[5] In both cases, rational and non-rational learning, the functions $f^i(\cdot \mid \cdot)$ determine a joint prior probability for any sequence $(x^i_1, x^i_2, ..., x^i_n)$ of search outcomes with

$$\Pr(x^i_1, x^i_2, ..., x^i_n) = f^i(x^i_1) \cdot f^i(x^i_2 \mid x^i_1) \cdot \cdots \cdot f^i(x^i_n \mid x^i_1, x^i_2, ..., x^i_{n-1}). \quad (1)$$

In the rest of the paper I will use cumulative density functions (c.d.f.) with $F^i(\cdot \mid X^i_{r^i})$ denoting the c.d.f. of $f^i(\cdot \mid X^i_{r_i})$. Given the probability distribution (1) defined by the learning rule, the searcher maximizes the discounted expected payoff minus costs

$$\max_S E[e^{-r\tau_s} y_{\tau_s} - C_s], \quad (2)$$

where τ_s is the stopping time under sampling rule S, y_{τ_s} is the offer accepted in τ_s, and $E[C_s]$ is the expected discounted sampling costs under S. Clearly, if learning is non-rational then (2) differs from expected utility maximization because the search is only optimal given the employed learning rule. If learning is Bayesian, then (2) is identical to expected utility maximization.

I want to make two comments with regard to the above setup. First, it is a quite restrictive but crucial assumption that the functions F^i depend only on observations of i-boxes, i.e. outcomes of boxes of color $h \neq i$ do not

[5] The random variables described by $f^i(\cdot \mid \cdot)$ are assumed to have finite mean if $M^h < \infty$ for all $h = 1, 2, ..., I$ and to have finite variance otherwise.

reveal information about the parameter θ^i of i-boxes. For a Bayesian learner this is an implicit assumption on having prior one on the parameters $(\theta^1, ..., \theta^I)$ being chosen independently.

Second, the setup comprises as special cases models without learning and several search alternatives, and models with learning but identical boxes. In the cases when there is only one box of each color no learning will take place and the model reduces to the one studied by Weitzman [22].[6] In cases when all the boxes have the same color, the model reduces to the search problems considered (amongst other problems) in Rosenfield and Shapiro [15], Talmain [20], Bikchandani and Sharma [2], Chou and Talmain [4].

3. BENCHMARK: OPTIMAL STRATEGY WITHOUT LEARNING

This section presents the optimal sampling rule when there is only one box of each color and hence no learning takes place.[7] Such a problem is equivalent to a search problem with full information when the searcher's expected payoff distributions equal the true payoff distributions. The results presented here will serve as a helpful reference point for our later considerations and the main result is due to Weitzman [22].

For expository reasons consider the following simple but instructing example.

EXAMPLE 3.1. Suppose that there are only 2 boxes, a red one and a green one. Table I describes the payoff distributions $d^i(\cdot)$ for each box. For simplicity I will refer to the zero outcome as a "failure" and to the strictly positive outcome as a "success". With search costs for opening a box equal to 20, no discounting and the value of the outside option equal to zero, the expected payoffs from opening a *single* box are shown in Table II.

Since the red box has a higher expected value than the green one, it might seem better to sample the red box first. If the result of doing so is a failure, it still pays to sample the green box because it has a positive expected payoff. If the result of the red box is a success, then sampling the green box yields a negative expected gain. The expected payoff of this sampling order is therefore readily calculated to be

$$-20 + 0.9 \cdot 70 + 0.1(-20 + 0.15 \cdot 200) = 44.$$

[6] The searcher might still learn about the box of a particular color by opening it, yet at the time learning takes place there are no other boxes of that color left.

[7] Remember that we ruled out learning across boxes of different color.

TABLE I

Red	Payoff	0	70
	With probability	0.1	0.9
Green	Payoff	0	200
	With probability	0.85	0.15

Yet, sampling the green box first and then in case of a failure the red box is the optimal sampling order. Its expected value is

$$-20 + 0.15 \cdot 200 + 0.85(-20 + 0.9 \cdot 70) = 46.55.$$

A simple intuition exists as to why the expectation criterion does not work in deciding upon which box to open first: It ignores the option value of the possibility to continue searching if there is a low search outcome. This option value is relatively small in the case of a failure of the red box, namely $0.1 \cdot (-20 + 0.15 \cdot 200) = 1$ (the probability of a failure of the red box times the expected value of opening the green box), but relatively high in the case of a failure of the green box, namely $0.85 \cdot (-20 + 0.9 \cdot 70) = 36.55$. Adding the first option value to the expected value of the red box gives 44, which is the value of the non-optimal sampling order. Adding the second option value to the expected value of the green box gives 46.55, the value of the optimal sampling order. Thus, although the immediate payoff from sampling the green box is lower than the immediate payoff from sampling the red box, the higher option value of continued search more than compensates for this.

It turns out that it is not necessary to calculate the option values of continuing to search to determine the right sampling order. There is a simple way of calculating an index for every search alternative that is only based on the payoff distribution of the respective alternative. This is important to know because the option value of continued search can be a fairly complicated object, especially if one has many boxes of many different colors and, as in the next section, learning during the search process. The

TABLE II

	Expected payoff
Red	43
Green	10

index has already been suggested by Lippman and McCall [11]. In the following I will describe how it is calculated and give some intuition on why it works.

Suppose the best offer from previous searches is y, then the expected gain over y from opening an i-box and stopping the search with what is best is given by

$$Q^i(y) = \left(\beta^i \int_{-\infty}^{y} y\, dF^i(x) + \beta^i \int_{y}^{\infty} x\, dF^i(x) - c_i\right) - y$$

$$= \beta^i \int_{y}^{\infty} (x-y)\, dF^i(x) - (1-\beta^i)\, y - c_i,$$

where $\beta^i = e^{-rt^i} \leqslant 1$ is the discount factor.

Define as the reservation price R^i of an i-box that value of the best offer y at which the searcher would be indifferent between the following two actions:

1. Stopping the search with y, and
2. Sampling an i-box and stopping with what is the best offer then

That is

$$Q^i(R^i) \equiv 0.$$

Note that R^i can be calculated using the payoff distribution of i-boxes only, ignoring any value derived from continuing to search.

The values R^i are the indices characterizing the optimal search strategy. It is based on these indices as follows:

Step 1. Calculate the reservation prices for each box.

Step 2. If there is no closed box with a reservation price higher than the current best offer y then stop searching and accept y, otherwise continue with Step 3.

Step 3. Open the box with the highest reservation price and go back to step 2.

A simple check of the reservation prices of the two boxes in our previous example reveals that $R^{red} = 47.8 < 66.7 = R^{green}$.[8] The rule therefore confirms the optimality of sampling the green box first.

A simple intuition exists as to why the above sampling rule is the optimal one. Consider the following alternative interpretation of the

[8] In the case of example 3.1 the reservation price formula boils down to $R^i = x_h^i - \frac{c^i}{p_h^i}$ where x_h^i is the value of the high payoff and p_h^i is the probability of obtaining it.

reservation prices. It is well known that the optimal strategy for a search problem with an infinite number of *i*-boxes (and no other alternatives) is a reservation price strategy. The optimal reservation price for such a problem is the same as the one calculated above. Moreover, the reservation price is the value of a secure payoff that makes the searcher indifferent between accepting a secure payoff and having the opportunity to sample *i*-boxes. $R^i > R^h$ can then be understood as the return from sampling *i*-boxes being higher than the return from sampling *h*-boxes. Search opportunities with higher reservation prices should therefore be sampled first.

4. OPTIMAL STRATEGY WITH LEARNING

This section contains the main results of the paper. I begin by presenting the reservation prices and discussing their properties. Then I delineate the class of admitted learning rules and present the optimal sampling strategy for a learning searcher. The optimal strategy is found to be a generalization of the benchmark strategy. In the last subsection I explain why this is the case.

4.1. *The Reservation Prices*

As in the case of known distributions, one can define Q^i as the expected gain of opening one more *i*-box and stopping search thereafter over stopping immediately. With learning the expected distribution of the search outcomes of *i*-boxes, $F^i(\cdot \mid X^i_{r^i})$, now depends on the information contained in the previously observed search outcomes $X^i_{r^i}$. Therefore, the expected gain Q^i is now a function of the available information

$$\begin{aligned} Q^i(X^i_{r^i}, y) &= \beta^i \int_{-\infty}^{y} y \, dF^i(x^i_{r^i+1} \mid X^i_{r^i}) \\ &\quad + \beta^i \int_{y}^{\infty} x^i_{r^i+1} \, dF^i(x^i_{r^i+1} \mid X^i_{r^i}) - c_i - y \\ &= \beta^i \int_{y}^{\infty} (x^i_{r^i+1} - y) \, dF^i(x^i_{r^i+1} \mid X^i_{r^i}) - (1-\beta^i) \, y - c_i, \end{aligned}$$

where $\beta^i = e^{-rt^i} \leqslant 1$ is the discount factor.

Analogously to the full information case, one can define the reservation price of boxes from alternative *i*.[9]

[9] Existence and uniqueness is guaranteed by the conditions of Lemma 7.1 in the appendix.

DEFINITION. The reservation price $R^i(X^i_{r^i})$ for boxes from alternative i is the value of y that solves $Q^i(X^i_{r^i}, y) = 0$.

Again, the reservation price of i-boxes is that value of the best offer y which makes the searcher indifferent between stopping, and searching once more and stopping thereafter.

Note that the reservation prices R^i are now also a function of the current information $X^i_{r^i}$. Reservation prices may therefore change over time as new information becomes available. Yet, how they might change in the future does not enter into the calculation of the reservation prices. Therefore, for given beliefs and given expected distribution function $F(\cdot \mid X^i_{r^i})$, the reservation prices are independent from the searcher's learning rule.

The $R^i(X^i_{r^i})$ again have an alternative interpretation as the reservation price of an optimally behaving (non-learning) searcher facing an infinite number of boxes with payoff distribution $F^i(\cdot \mid X^i_{r^i})$.

4.2. *Learning Rules*

We saw in the previous section that the reservation prices depend only on current beliefs and are independent from the potential future evolution of these beliefs, i.e. from the learning rule. If we want to characterize the optimal search strategy based on this momentary picture of beliefs, we have to restrict the admitted learning rules so that this picture is sufficiently informative about the future.

We can express the necessary requirements on the learning rules in terms of an assumption on the evolution of reservation prices as learning proceeds. All learning rules with falling reservation prices are admitted. Formally,

Assumption A1. Let $X^i_{r^i+1} = (X^i_{r^i}, x^i_{r^i+1})$, then

$$R^i(X^i_{r^i+1}) \leqslant R^i(X^i_{r^i}) \qquad \text{or} \qquad R^i(X^i_{r^i+1}) \leqslant x^i_{r^i+1} \quad \forall i, X^i_{r^i}, x^i_{r^i+1}.$$

Assumption A1 requires that after observing an additional search outcome of an i-box, the new reservation price $R^i(X^i_{r^i+1})$ is either smaller than the old reservation price $R^i(X^i_{r^i})$ or smaller than new the offer $x^i_{r^i+1}$.[10]

This can be interpreted as follows: Either the searcher receives a low search outcome and lowers in response to that the beliefs about the attractiveness of the sampled search alternative, which in turn leads to a lower reservation price; or the searcher receives a good outcome indicating that

[10] Since $R^i(X^i_{r^i}) < y$ implies $Q^i(X^i_{r^i}, y) < 0$, A1 insures that the one period gains $Q^i(X^i_{r^i}, y)$ stay negative once they have become negative at some point of time. A1 therefore implies the sufficient condition used in Rosenfield and Shapiro ([15], Theorem1) to establish the optimality of a myopic stopping rule.

the search alternative is more attractive than previously believed and increases the reservation price. In the latter case, it is important that the increase in the reservation price is moderate enough to ensure that the second of the above inequalities holds.

What is ruled out are so-called strong positive learning effects. These are search outcomes revealing a lot of good news about the attractiveness of a search alternative. In fact, so much good news that if the searcher was given the value of such a search outcome as the outside option, he would terminate the search but as one told him that this outside option was drawn from the search alternative, he would want to continue searching.

In the following I give examples of learning rules that fulfill A1 and that have been used in the search literature dealing with identical boxes.[11] The optimal sampling strategy I derive holds for any of the following learning rules. The searcher might even apply different learning schemes to different search alternatives.

1. Let the offer distribution be multinomial with N possible outcomes $x_1, x_2, ..., x_N$ and the probability of observing outcome x_i be equal to θ_i. If learning is Bayesian and the searcher has Dirichlet priors about the vector θ, i.e.

$$p(\theta \mid \alpha_1, \alpha_2, , ..., \alpha_N) \propto \theta_1^{\alpha_1 - 1} \theta_2^{\alpha_2 - 1} \cdots \theta_N^{\alpha_N - 1} \qquad \text{with} \quad \alpha_i > 0,$$

then reservation prices are decreasing (e.g. Talmain [20]). The generalization of the multinomial Dirichlet case to a Dirichlet process with an infinite number of possible outcomes also implies declining reservation prices (see Bikchandani and Sharma [2]).

2. A class of ad-hoc learning rules (generalizing the learning rule of the previous point) where the posterior distribution is a convex combination of the prior and the empirical distribution with the weight on the empirical distribution non-decreasing with additional observations:

$$F^i(x \mid X^i_{r^i}) = (1 - a_{r^i})\, F(x) + a_{r^i} H(x \mid X^i_{r^i}).$$

With $a_{r^i+1} \geqslant a_{r^i}$, $F(x)$ being the prior distribution before search started, and $H(\cdot \mid X^i_{r^i})$ being the empirical distribution based on the observations $X^i_{r^i}$ (Bikchandani and Sharma [2]).

3. A non-parametric learning procedure used in Chou and Talmain ([4]) that makes no assumptions on the underlying class of probability distributions and constructs $F(\cdot \mid X^i_{r^i})$ according to the maximum entropy

[11] In some of the references increasing reservation prices can be found because the search problem has been posed in terms of searching for the lowest price instead of searching for the highest reward, as is done here.

principle. Suppose the searcher knows that the outcomes are distributed between some interval $[a, b]$. The conditional probability of some outcome x, having observed $x^1 \leqslant x^2 \leqslant \cdots \leqslant x^{r^i}$ (not necessarily in this order) is obtained by assigning each interval $[a, x^1]$, $[x^1, x^2]$, ..., $[x^{r^i}, b]$ a uniformly distributed probability mass $1/(r^i+1)$ (and a point mass if $x^i = x^{i+1}$).

4. Let the offer distribution be exponential with unknown origin θ:

$$f(x \mid \theta) = ae^{a(x-\theta)} \qquad \text{for} \quad x \leqslant \theta.$$

Learning is Bayesian and the priors are such that the logarithm of the prior distribution $\log(p(\theta))$ is concave (see Rosenfield and Shapiro [15]).

4.3. *Results*

The following theorem states the optimal sampling strategy for the search problem with learning and contains the main result of this paper. Its proof is deferred to the appendix. The optimal rule is just the benchmark rule applied to repeatedly updated reservation prices.

THEOREM 4.1. *Given* A1 *holds, the following strategy is optimal*:

Step 1. *With the available information calculate the reservation prices for each alternative and continue with Step* 2.

Step 2. *If there is no closed box with a reservation price higher than the current best offer y, then stop searching and accept y, otherwise continue with Step* 3.

Step 3. *Search the alternative* (*or one of the alternatives*) *with the maximum reservation price and go back to Step* 1.

The theorem tells us that the reservation prices are sufficient to determine the optimal sampling strategy although they are based solely on current beliefs. Returning to section 3's interpretation of the reservation prices as rates of return, the rule tells us to sample the alternatives with the *currently* highest returns.

The optimality of this focus on current beliefs might be surprising. In a learning context information is valuable, since it enables the searcher to make better search decisions in the future. In general, it might therefore be worth giving up payoffs in the short term to obtain information that allows superior future decisions.

In the considered search problem there is no trade-off between the information gain and the payoff gain and focusing on the payoff gain alone is sufficient to obtain optimality. The reason is to be found in the restrictions on the learning rules that I imposed. Here learning exhibits enough monotonicity to prevent the searcher from optimally going through a "payoff-valley" to potentially reach a higher "payoff-mountain" later on.

TABLE III

	Payoff	0	1
Red	With probability	0.5	0.5
Green	With probability	0.7	0.3

Obviously, the possibility of strong learning could give an incentive to go through the "payoff-valley" and therefore it had to be ruled out.

To obtain some intuition on why the remaining learning processes do not give similar incentives consider the following example: Imagine having two search alternatives with equal reservation prices, a blue one and an orange one. Suppose also that there is only one blue box left, while there are still many orange boxes. Sampling the blue box does not reveal information about any other search opportunity, while sampling an orange box reveals information about the remaining orange boxes. Nevertheless, theorem 4.5 indicates that first opening the orange box is not superior to first opening the blue box. Considering the rate of return interpretation of reservation prices, it is clear why: When first sampling an orange box, the remaining orange boxes will have a lower rate of return (due to A1).[12] Yet, at this point of the search process this knowledge is irrelevant because the best alternative to stopping searching is opening the blue box. The knowledge about the reservation prices of the remaining orange boxes only becomes relevant after *both* boxes, the blue and an orange one, have been opened. As a result, the early information gain achieved by first opening an orange box does not matter for the optimal sampling order.

The optimal sampling procedure above is only slightly different from the benchmark sampling rule. The informed searcher has to calculate reservation prices only once, while a learning searcher has to permanently adapt them due to new information. Step 3 of the rule therefore points back to Step 1. For the rest, the rule remains unchanged. This slight change, however, substantially alters optimal search behavior, as the following example illustrates.

EXAMPLE 4.1. Suppose that there are only two search alternatives, a red one and a green one, but many boxes of each alternative. Boxes have only two kinds of outcomes: "success", identified with a payoff equal to 1, or "failure", identified with a payoff equal to zero. The true probabilities for success and failure for the respective alternatives are indicated in Table III. Assume a discount factor equal to 1, sampling costs for both boxes equal to 0.1 and the value of the outside option equal to 0.

[12] I abstract here from the possibility that the search stops to make the argument as simple as possible.

(a) Optimal sampling strategy under full information. Knowing the true probabilities of outcomes, the reservation prices are $R^{red} = 0.8$ and $R^{green} = 0.\overline{6}$. Hence, an informed searcher prefers to open red boxes and stops with the first success. Suppose that the searcher encounters a sequence of failures. Optimally, his strategy is to continue opening red boxes until they have all been opened and only then to switch to the opening of green boxes. Green boxes are opened until a success is encountered or all of them have been searched. Notice the following feature of the optimal strategy: Since the ranking of alternatives is constant during the search process, the searcher does not switch sampling from one alternative to another, unless there are no boxes of that alternative left.

(b) Optimal sampling strategy with learning. Now consider a searcher that is uncertain about the true underlying probability distribution and is learning by taking a convex combination of his prior distribution and the empirical distribution function (This is the second learning rule in Section 4.2):

$$F^i(x \mid X^i_{r^i}) = (1 - a_{r^i})\, F(x) + a_{r^i} H(x \mid X^i_{r^i}).$$

Let the weight on the empirical distribution be $a_{r^i} = r^i/(1 + r^i)$ and the searcher's priors $F(x)$ be unbiased in the sense that they are equal to the true underlying probability function as shown in Table 3. At the beginning of search the reservation prices are therefore equal to the ones of an informed searcher, but as the searcher makes additional observations they are adjusted downwards. The ranking of alternatives is therefore changing during the search process. The searcher might well search green boxes before all the red boxes have been opened. Negative results from searching red boxes "bid" down their reservation price and make the searcher believe that green boxes are more interesting. The same reasoning applied to green boxes might cause a switch back to sampling red boxes again. In further contrast to the full information case, sampling might even stop with a failure and not all the boxes searched because beliefs having worsened so much that the outside option looks more profitable than continued search. The previous effects can be seen in Table IV for the above learning rule and a sequence of failures. The table reads as follows. The first column indicates the search stage, the second the number of so far made observations of red and green boxes (i.e. the number of observed failures of each), the following two columns show the current reservation prices. The last column gives the optimal search strategy according to Theorem 4.1. The searcher reduces the reservation prices and switches between sampling red and green boxes in response to failures until finally the reservation prices of both boxes are so low that the outside option appears more attractive than continued search.

TABLE IV

t	(r^{red}, r^{green})	R^{red}	R^{green}	Optimal strategy
0	(0, 0)	0.8	0.66	search a red box
1	(1, 0)	0.6	0.66	search a green box
2	(1, 1)	0.6	0.33	search a red box
3	(2, 1)	0.4	0.33	search a red box
4	(3, 1)	0.2	0.33	search a green box
5	(3, 2)	0.2	0.0	search a red box
6	(4, 2)	0.0	0.0	stop searching and take the outside option

4.4. *An Equivalent Search Problem without Learning*

This section explains why the benchmark rule translates to the case with learning. I show that for each search problem P with learning one can construct an equivalent search problem P^e without learning. P^e is equivalent to P in the sense that for any search rule S in P there exists a corresponding search rule S^e in P^e, such that S yields in P the same expected payoff as S^e in P^e. After showing that the optimal rule in P^e is the benchmark rule, it is easy to see that the corresponding rule in P exists and is the generalization of the benchmark rule stated in Theorem 4.1.

Consider a search problem with learning P. P is described by the number I of alternatives, the numbers M^i of boxes of each alternative, the prior beliefs and the learning rule. A sampling rule S for P is a mapping from the set of available information $(X^1_{r^1}, X^2_{r^2}, ..., X^n_{r^n})$ to the set of integers $\{0, 1, ..., I\}$, where $S=0$ indicates stop searching and $S=i$ for $i \geqslant 1$ indicates continue searching with an i-box.

At the beginning of the search the M^i boxes of alternative i have an expected distribution function $F^i(\cdot)$ and an associated reservation price R^i_0. After opening an i-box, the expected distribution of outcomes for the remaining M^i-1 boxes from i becomes $F^i(\cdot \mid x^i_1)$ and the associated reservation price is $R^i(x^i_1)$. If another box of this alternative is opened, then the remaining M^i-2 boxes have expected distribution $F^i(\cdot \mid x^i_1, x^i_2)$ and the associated reservation price is $R^i(x^i_1, x^i_2)$, etc.

Alternatively, one could interpret the previous observation as follows: The searcher has only one box with reservation price R^i_0 (after sampling one i-box the reservation price of the remaining i-boxes changes), one box with reservation price $R^i(x^i_1)$, another one with $R^i(x^i_1, x^i_2)$, and so on. Based on this alternative interpretation I construct the equivalent search problem P^e. P^e has the same number of boxes as P. The first M^1 boxes have reservation prices

$$R^1_0, R^1(w^1_1), R^1(w^1_1, w^1_2), ..., R^1(w^1_1, w^1_2, ..., w^1_{M^1-1})$$

respectively, where the $R^1(\cdot)$ are the reservation price functions of the 1-boxes in P. Similarly, let the next M^2 boxes in P^e have reservation prices

$$R_0^2,\, R^2(w_1^2),\, R^2(w_1^2, w_2^2),\, ...,\, R^2(w_1^2, w_2^2, ..., w_{M^1-1}^2).$$

Continue to assign reservation prices to boxes in the above manner until each box in P^e has one reservation price. For the moment, take the values w_j^i as given. The box with the highest reservation price among the first M^1 boxes in P^e that has not been opened yet will be called "the best 1-box", the best unopened box of the next M^2 boxes "the best 2-box", and so on.

From Assumption A1 we know that (for any values of the w_j^i) the reservation prices of the boxes can be ordered as[13]

$$R_0^i \geqslant R^i(w_1^i) \geqslant R^i(w_1^i, w_2^i) \geqslant \cdots \geqslant R^i(w_1^i, w_2^i, ..., w_{M^1-1}^i).$$

If the w_j^i were given, P^e would be the benchmark search problem. Yet, to make P^e equivalent to P one must choose stochastic w_j^i, with w_1^i drawn from $F^i(\cdot)$, w_2^i from $F^i(\cdot|w_1^i)$, w_3^i from $F(\cdot\,|\,w_1^i, w_2^i)$, and so on. The likelihood of encountering a box with a certain reservation price in P^e is then the same as in P. This allows to construct a sampling rule S^e for P^e that achieves an (objective) expected payoff that is equal to the (subjective) expected payoff of S in P: S^e is the same as S but is evaluated at $W_{r^1}^1, W_{r^2}^2, ..., W_{r^n}^n$ and specifies sample "the best i-box", whenever S would specify sample some i-box, i.e. $S^e = S(W_{r^1}^1, W_{r^2}^2, ..., W_{r^n}^n)$ with $S^e = 0$ indicating stop searching and $S^e = i$ with $i \geqslant 1$ indicating sample "the best i-box".

Even though the w_j^i are stochastic, the optimal search rule in P^e is still the benchmark rule because it is optimal for each given sequence of the w_j^i and because the realizations of the w_j^i are independent from the sampling decisions. The benchmark rule for P^e states sample the box with the highest reservation price amongst "the best i-boxes". The optimal rule in P is the rule corresponding to this rule and states sample an i-box with the maximum reservation price, which is precisely the optimal sampling rule from Theorem 4.1.

5. LIMITATIONS AND EXTENSIONS

In this Section I discuss whether Theorem 4.1 holds when the assumption on the independence of boxes from different alternatives is relaxed.

[13] I ignore the potential increase in the reservation price admitted by A1 because it leads to the termination of the search.

TABLE V

Red	Payoff	0	0.5	1
	With probability	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
Green	Payoff	0	0.6	0.95
	With probability	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Unfortunately, the optimality of the search rule turns out to be sensitive to such a relaxation.

The following example shows that even with decreasing reservation prices Theorem 4.1 generally does not hold.

EXAMPLE 5.1. There are only two alternatives, red and green, and only one box of each alternative. The searcher's expected payoff distributions are shown in Table V. The searcher's outside option is $y = 0.5$. Without discounting and search costs equal to 0.1 for both alternatives, the reservation prices are $R^{red} = 0.7$ and $R^{green} = 0.65$.

Let learning be such that when sampling the red box the new reservation price for the green box drops below 0.5 when outcomes 0 or 0.5 have been observed and that it is anything smaller than 1 otherwise. Interpret this as low outcomes of the red box revealing that low outcomes of the green box are more likely. As a result, search optimally stops after sampling the red box, irrespective of the outcome. Sampling the red box first, as suggested by Theorem 4.1, yields $\frac{2}{3}0.5 + \frac{1}{3}1 = \frac{2}{3}$. However, the alternative strategy of opening the green box first and stopping search thereafter yields $\frac{1}{3}0.5 + \frac{1}{3}0.6 + \frac{1}{3}0.95 = 0.68\overline{3} > \frac{2}{3}$.

To see why the sampling rule might be sub-optimal in this more general setting consider the equivalent search problem P^e without learning from Section 4.4. For the benchmark rule to be optimal in P^e (and its corresponding rule in P) it was crucial that the searcher could not influence the sequence of reservations prices through the sampling decisions. In the above example this does not hold because the reservation price of the green box depends on whether the searcher sampled the red box.

For special cases the sampling rule might generalize to dependent alternatives. Theorem 4.1 requires only that the reservation prices R^i of i-boxes do not depend on observations x^j of j-boxes ($j \neq i$). This is possible even though the x^j affect the distribution F^i. When x^j leaves unchanged the values of F^i above the current best offer y, then the reservation price R^i of i-boxes with $R^i > y$ will be unaffected by x^j. Since boxes with reservation prices below the current best offer do not affect the value of the search, this special case of dependent boxes is covered by Theorem 4.1.

6. CONCLUSIONS

This paper constructed the optimal sampling strategy for a search problem where the searcher faces different search alternatives and is learning about these alternatives during the search process. I thereby unified and generalized two kinds of earlier contributions: search problems with learning but identical search opportunities and search problems with distinguishable search alternatives but without learning.

The optimal sampling rule is characterized by a simple reservation price criterion. The rule implies that search opportunities with higher reservation prices should be sampled before ones with lower reservation prices. In contrast to the full information case, the ordering of different search alternatives in terms of reservation prices keeps changing during the search process. Learning therefore makes a substantial difference to the optimal sampling order. At the same time the sampling rule retains its simple structure and learning can be accounted for without complicating the analysis.

The independence of different search alternatives has been found to be crucial for the optimality of the sampling rule and finding conditions on the learning process that allow for an extension of the results to the case of dependent search alternatives is left for future research.

APPENDIX

LEMMA 7.1. *If either $\beta_i < 1$ or $c_i > 0$, then a unique reservation price exists.*

Proof of Lemma 7.1. The function $Q^i(X^i_{r^i}, \cdot)$ is continuous, differentiable and decreasing.

$$\frac{d}{dy} Q^i(X^i_{r^i}, y) = \frac{d}{dy}\left[\beta_i \int_y^\infty (x^i - y)\, dF^i(x^i \mid X^i_{r^i})\right] - (1-\beta_i) \tag{3}$$

$$= \beta_i \int_y^\infty -1\, dF^i(x^i \mid X^i_{r^i}) \tag{4}$$

$$- [(x^i - y)\, dF^i(x^i \mid X^i_{r^i})]_{x^i = y} - (1-\beta_i) \tag{5}$$

$$= -\beta_i (1 - F^i(y \mid X^i_{r^i})) - (1-\beta_i) \tag{6}$$

$$\leqslant 0 \tag{7}$$

Since

$$Q^i(X^i_{r^i}, -\infty) = \infty \tag{8}$$

$$Q^i(X^i_{r^i}, +\infty) = \begin{Bmatrix} -\infty & \text{if } \beta_i < 1 \\ -c^i & \text{if } \beta_i = 1 \end{Bmatrix} \tag{9}$$

a solution exists. If $\beta_i < 1$, then $\frac{d}{dy} Q^i < 0$ and the solution is also unique. If $\beta_i = 1$, then $\frac{d}{dy} Q^i < 0$ only if $F^i(y \mid X^i_{r^i}) < 1$. With $c^i > 0$ this is guaranteed at the reservation price: $F^i(R^i(X^i_{r^i}) \mid X^i_{r^i}) = 1$ implies $Q^i(X^i_{r^i}, R^i(X^i_{r^i})) < 0$ which
contradicts the definition of the reservation price. ∎

Proof of Theorem 4.1.[14] begin by proving the optimality of the stopping rule (i.e., Step 2 of the theorem). If there is some i-box with $R^i > y$ ($R^i < y$), then the one period gain $Q^i > 0$ ($Q^i < 0$). Therefore, as long as there is some closed i-box with $R^i > y$ stopping cannot be optimal, since opening an i-box and stopping then gives already a higher payoff. If all closed boxes have a reservation value below y, then A1 insures that reservation prices will also be below the best offer in all future search steps. Gains from continued search will always be negative and stopping is therefore optimal.

Suppose that S is a sampling rule where stopping is optimal as derived above. In addition, suppose that S specifies at some search stage to sample a k-box with reservation price R^k and in case that the stopping rule prescribes continuation in the next search step an l-box with $R^l > R^k$. I will show that S cannot be optimal. To do so I will construct an alternative sampling rule S' and show that S' has higher expected valued than S. S' is like S but interchanges the sampling order such that the box with the higher reservation price R^l is sampled first and the one with the lower reservation price R^k thereafter.[15]

Before constructing S' and proving the claim I have to introduce some notation. At the search stage where S specifies to sample a k-box, let the previous observations of search outcomes be $\{X^i_{r^i}\}_{i=1}^{I}$ and the current best offer $y = \max\{X^1_{r^1}, X^2_{r^2}, ..., X^I_{r^I}\}$. Define

$$R^j = \max_{i \mid r^i < M^i} R^i(X^i_{r^i})$$

$$R^{h(x^j_{r^j+1})} = \max\{R^j(X^j_{r^j+1}) \mid r^j + 1 < M^j, \max_{i \mid i \neq j \wedge r^i < M^i} \{R^i(X^i_{r^i})\}\}.$$

[14] The structure of the first part of the proof augments Weitzman's [22] proof by the stochastic elements that account for the learning process.

[15] Notice that the sampling order of S' is feasible. $k = l$ is not possible, since reservation prices of k-boxes must decrease when sampling (optimally) proceeds.

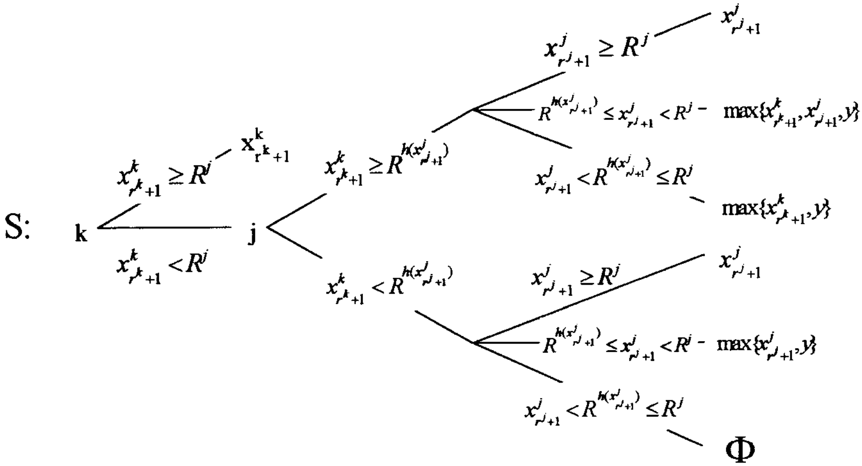

FIG. 1. Strategy *S*.

j-boxes have currently the highest reservation price of all closed boxes and h-boxes are the ones that have the highest reservation price after one j-box has been sampled and the search outcome $x^j_{r^j+1}$ been observed.[16] h may depend on $x^j_{r^j+1}$ because the decrease of the reservation price of j-boxes depends on $x^j_{r^j+1}$.

By assumption we know that

$$R^k(X^k_{r^k}) < R^l(X^l_{r^l}) \leqslant R^j$$

We should distinguish two cases: $l \neq j$ and $l = j$. The first case is the easier one: The highest reservation price R^j remains unaffected by the sampling of a k- and an l-box. The optimal stopping criterion is therefore the same in both search stages: Stop if the current best offer is larger than R^j and continue otherwise. In the second case the best reservation price drops to $R^{h(x^j_{r^j})}$ after sampling the l-box ($l = j$), changing the optimal stopping condition. I will only consider this more complicated case.[17]

Recall that the rule S specifies to sample first a k-box and in case of continuation a j-box with the stopping decision being optimal as derived above. Figure 1 gives a graphical representation of the strategy for the first two search steps. Depending on the search outcome several cases can be distinguished that are represented by branches. The values written at the end of these branches represent the payoffs for the respective cases. If

[16] $r^i < M^i$ is a condition insuring that there is still an unopened i-box.

[17] The results for the first case can be obtained by replacing j by l and $R^{h(x^j_{r^j})}$ by R^j in the following.

search outcomes fall into the case represented by the lowest branch, then search continues. Φ denotes the value of continued search with rule S for this case.

The proposed alternative strategy S' differs from S for the first two search steps but is identical to S for later search steps: S' specifies to first sample a j-box (instead of a k-box). If the new best offer $\max\{y, x^j_{r^j+1}\} \geqslant R^{h(x^j_{r^j+1})}$, then S' specifies to stop search. Otherwise it prescribes to sample a k-box and to continue as prescribed by the rule S. Figure 3 represents the sampling rule S' graphically. Again, Φ denotes the value of continuing search with rule S' after the first two search steps.[18]

The following notation will prove useful to calculate the expected payoffs of S and S':

$$\Pi^k = \Pr(x^k_{r^k+1} \geqslant R^j)$$

$$\Pi^j = \Pr(x^j_{r^j+1} \geqslant R^j)$$

$$\lambda_k = \Pr(R^j > x^k_{r^k+1} \geqslant R^{h(x^j_{r^j+1})})$$

$$\lambda_j = \Pr(R^j > x^j_{r^j+1} \geqslant R^{h(x^j_{r^j+1})})$$

$$\mu^k = \Pr(R^{h(x^j_{r^j+1})} > x^k_{r^k+1} \geqslant R^k)$$

$$w^k = E[x^k_{r^k+1} \mid x^k_{r^k+1} \geqslant R^j]$$

$$w^j = E[x^j_{r^j+1} \mid x^j_{r^j+1} \geqslant R^j]$$

$$\tilde{v}^k = E[\max\{x^k_{r^k+1}, y\} \mid R^j > x^k_{r^k+1} \geqslant R^{h(x^j_{r^j+1})}]$$

$$\tilde{v}^j = E[\max\{x^j_{r^j+1}, y\} \mid R^j > x^j_{r^j+1} \geqslant R^{h(x^j_{r^j+1})}]$$

$$v^k = E[x^k_{r^k+1} \mid R^j > x^k_{r^k+1} \geqslant R^{h(x^j_{r^j+1})}]$$

$$u^k = E[x^k_{r^k+1} \mid R^{h(x^j_{r^j+1})} > x^k_{r^k+1} \geqslant R^k]$$

$$d = E[\max\{x^j_{r^j+1,}, x^k_{r^k+1}, y\} \mid R^j > x^k_{r^k+1} \geqslant R^{h(x^j_{r^j+1})}, \\ R^j > x^j_{r^j+1} \geqslant R^{h(x^j_{r^j+1})}]$$

$$\Phi = E[\Psi(\bar{S}\backslash\{j, k\}, \max\{x^j_{r^j+1}, x^k_{r^k+1}, y\}, S) \mid R^{h(x^j_{r^j+1})} > x^k_{r^k+1}, \\ R^{h(x^j_{r^j+1})} > x^j_{r^j+1}, R^j \geqslant R^{h(x^j_{r^j+1})}].$$

All probabilities and expectation operators are conditional on the information $\{X^i_{r^i}\}^I_{i=1}$. The function $\Psi(\bar{S}\backslash\{j, k\}, \max\{x^j_{r^j+1}, x^k_{r^k+1}, y\}, S)$ represents the value of continued search when the set of closed boxes is $\bar{S}$

[18] This value is the same as with rule S because S' equals S for all steps after the second.

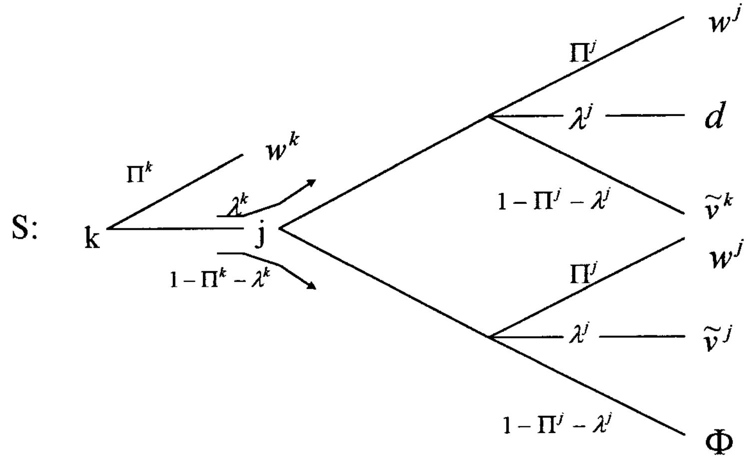

FIG. 2. Strategy *S*.

less one *j*- and one *k*-box, the best offer is $\max\{x^j_{r^j+1}, x^k_{r^k+1}, y\}$ and the sampling rule *S*. Figure 2 describes the probabilities and the expected payoffs of strategy *S* for the cases distinguished in Fig. 1 using the above notation. Similarly, Fig. 4 treats for strategy S' and the cases of Fig. 3. Looking at these figures reveals that the expected payoffs of the strategies *S* and S' are

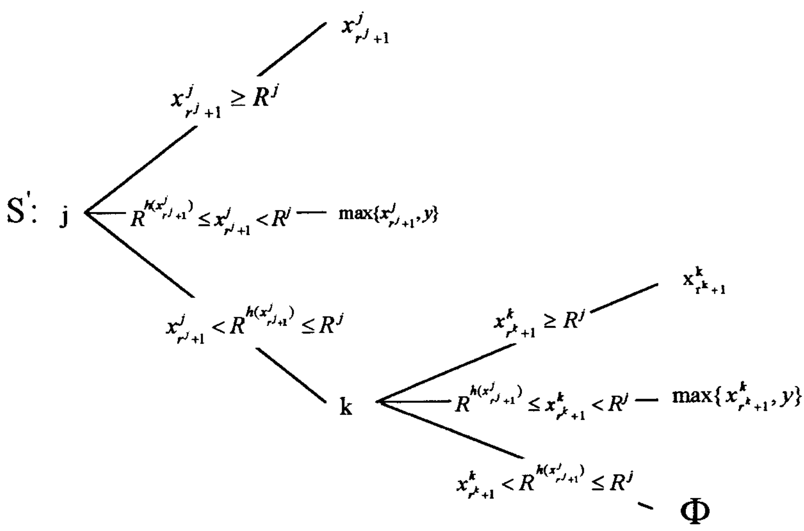

FIG. 3. Strategy S'.

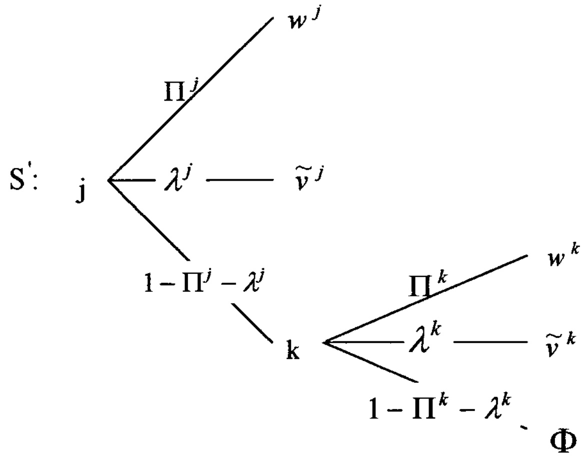

FIG. 4. Strategy S'.

$$S = -c^k + \beta^k \Pi^k w^k + \lambda^k \beta^k (-c^j + \Pi^j \beta^j w^j + \lambda^j \beta^j d \tag{10}$$

$$+ (1 - \Pi^j - \lambda^j)\, \beta^j \tilde{v}^k) + (1 - \Pi^k - \lambda^k)\, \beta^k (-c^j + \Pi^j \beta^j w^j \tag{11}$$

$$+ \lambda^j \beta^j \tilde{v}^j + (1 - \Pi^j - \lambda^j)\, \beta^j \Phi) \tag{12}$$

$$S' = -c^j + \Pi^j \beta^j w^j + \lambda^j \beta^j w^j + (1 - \Pi^j - \lambda^j)\, \beta^j (-c^k + \Pi^k \beta^k w^k \tag{13}$$

$$+ \lambda^k \beta^k w^k + (1 - \Pi^k - \lambda^k)\, \beta^k \Phi) \tag{14}$$

The payoff difference between S' and S is

$$\begin{aligned} S' - S = {} & (c^k - \Pi^k \beta^k w^k)(1 - (1 - \Pi^j - \lambda^j)\, \beta^j) \\ & - (c^j - \Pi^j \beta^j w^j)(1 - (1 - \Pi^k)\, \beta^k) \\ & + \lambda^j \beta^j (-\lambda^k \beta^k d + (1 - \beta^k (1 - \Pi^k - \lambda^k))\, \tilde{v}^j) \\ & + (1 - (1 - \Pi^k - \lambda^k)\, \beta^k)\, \lambda^j \beta^j. \end{aligned} \tag{15}$$

From the definition of the reservation prices we have

$$c^j = \Pi^j \beta^j (w^j - R^j) - (1 - \beta^j)\, R^j \tag{16}$$

$$\begin{aligned} c^k = {} & \beta^k (\Pi^k (w^k - R^k) + \lambda^k (v^k - R^k) \\ & + \mu^k (u^k - R^k)) - (1 - \beta^k)\, R^k \end{aligned} \tag{17}$$

Substituting (16) and (17) into (15) gives:

$$\begin{aligned} S' - S = {} & (R^j - R^k)(1 - \beta^j(1 - \Pi^j))(1 - \beta^k(1 - \Pi^k)) \\ & + (v^k - R^k)(\lambda^k \beta^k (1 - \beta^j(1 - \Pi^j))) \\ & + (\tilde{v}^j - R^k)(\lambda^j \beta^j (1 - \beta^k(1 - \Pi^k))) \\ & + (u^k - R^k)(\mu^k \beta^k (1 - \beta^j(1 - \Pi^j - \lambda^j))) \\ & + (\tilde{v}^j + v^k - R^k - d)\, \lambda^k \beta^k \lambda^j \beta^j. \end{aligned} \tag{18}$$

Furthermore,

$$\begin{aligned} d = {} & E[\max\{\max\{x^j_{r^j+1}, y\} - R^{h(x^j_{r^j+1})}, x^k_{r^k+1} - R^{h(x^j_{r^j+1})}\} \\ & + R^{h(x^j_{r^j+1})} \mid R^j > x^k_{r^k+1} \geqslant R^{h(x^j_{r^j+1})}, R^j > x^j_{r^j+1} \geqslant R^{h(x^j_{r^j+1})}] \\ \leqslant {} & E[\max\{x^j_{r^j+1}, y\} - R^{h(x^j_{r^j+1})} + x^k_{r^k+1} - R^{h(x^j_{r^j+1})} \\ & + R^{h(x^j_{r^j+1})} \mid R^j > x^k_{r^k+1} \geqslant R^{h(x^j_{r^j+1})}, R^j > x^j_{r^j+1} \geqslant R^{h(x^j_{r^j+1})}] \\ = {} & \tilde{v}^j + v^k \\ & - E[R^{h(x^j_{r^j+1})} \mid R^j > x^k_{r^k+1} \geqslant R^{h(x^j_{r^j+1})}, R^j > x^j_{r^j+1} \geqslant R^{h(x^j_{r^j+1})}] \\ \leqslant {} & \tilde{v}^j + v^k - R^k. \end{aligned}$$

The last inequality is due to the fact that by definition $R^{h(x^j_{r^j+1})} \geqslant R^k$ for any realization of $x^j_{r^j+1}$. Therefore, any term in (18) is greater or equal zero with the first term being strictly greater than zero. This proves the sub-optimality of any strategy of the form S. Optimal strategies must sample boxes in the order of decreasing reservation prices. However, this does not mean that at each search stage the box with the highest reservation price has to be sampled as the theorem prescribes. I will turn attention to this point in the following.

Suppose T is a sampling rule that stops according to the optimal stopping rule and samples boxes in order of decreasing reservation prices. However, suppose that T specifies at some search stage not to sample the box with the currently highest reservation price. I will show that T cannot be optimal by proving that there exists a strategy T' that has a higher expected value.

Suppose again that available observations are $\{X^i_{r^i}\}^I_{i=1}$ and that j-boxes have the highest reservation price equal to R^j. T specifies to sample a k-box with $R^k < R^j$. Thereafter (in case of continued search), T prescribes to sample $l, m, n, \ldots$-boxes with $R^k \geqslant R^l \geqslant R^m \geqslant R^n \geqslant \cdots$.[19] Since a

[19] m might depend on the outcome $x^k_{r^k+1}$, similarly the types $n, l, \ldots$ might depend on previous observations. For notational simplicity, we will ignore this dependence.

sampling of a j-box is incompatible with the assumption of sampling in order of decreasing reservation prices, j-boxes will never be sampled. The optimal stopping rule then implies that search stops only if $y \geqslant R^j$. To calculate the expected value of search rule T define for $\alpha = j, k, l, m, \ldots$:

$$\Pi^\alpha = \Pr(x^\alpha_{r^a + \#\alpha} \geqslant R^j)$$

$$w^\alpha = E[x^\alpha_{r^\alpha + \#\alpha} | x^\alpha_{r^a+1} \geqslant R^j],$$

where $\#\alpha$ is the number of alternative α-boxes in the sequence $k, l, m, \ldots, \alpha$.[20] Π^α is the probability that search stops when sampling box α. w^α is the expected value of x^α given that search stops.[21]

The expected value of T is easily calculated to be

$$\begin{aligned} T = {} & [-c^k + \beta^k \Pi^k w^k] \\ & + \beta^k (1 - \Pi^k)[-c^l + \beta^l \Pi^l w^l] \\ & + \beta^k (1 - \Pi^k)\, \beta^l (1 - \Pi^l)[-c^m + \beta^m \Pi^m w^m] \\ & + \beta^k (1 - \Pi^k)\, \beta^l (1 - \Pi^l)\, \beta^m (1 - \Pi^m)[-c^n + \beta^n \Pi^n w^n] \\ & + \cdots. \end{aligned} \tag{19}$$

Now consider the following alternative strategy T'. T' uses the same stopping rule as T: Stop if $y \geqslant R^j$ and continue otherwise. However, T' samples first a j-box and then (in case of continuation) $k, l, m, n, \ldots$ boxes. The expected value of T' is

$$T' = [-c^j + \beta^j \Pi^j w^j] + \beta^j (1 - \Pi^j)\, T.$$

Remembering from the definition of the reservation price that

$$c^\alpha = \beta^\alpha \Pi^\alpha (w^\alpha - R^j) + \beta^\alpha \lambda^\alpha (v^\alpha - R^\alpha) - (1 - \beta^\alpha)\, R^\alpha, \tag{20}$$

where

$$\lambda^\alpha = \Pr(R^j > x^\alpha_{r^a + 1 + \#\alpha} \geqslant R^\alpha)$$

$$v^\alpha = E[x^\alpha_{r^a + \#\alpha} \mid R^j > x^\alpha_{r^a + 1 + \#\alpha} \geqslant R^\alpha].$$

[20] Remember that each alternative $k, l, m, \ldots$ is a number from the set $\{1, 2, \ldots, I\}$. For example, if $\alpha = n$ and $k, l, m, n = 1, 4, 3, 4$ then $\#\alpha = 2$, i.e. it is the second box of alternative 4.

[21] Probabilities and expectations are again conditional on the available information $\{X^i_{r^i}\}^I_{i=1}$.

One can calculate the payoff difference between T' and T to be

$$\begin{aligned} T' - T &= [-c^j + \beta^j \Pi^j w^j] + [\beta^j(1-\Pi^j) - 1]\, T \\ &= \beta^j \Pi^j R^j + (1-\beta^j)\, R^j + [\beta^j(1-\Pi^j) - 1]\, T \\ &= [\beta^j(\Pi^j - 1) + 1](R^j - T). \end{aligned} \tag{21}$$

The first bracket in the last line of (21) is strictly positive.[22] It remains to show that $R^j > T$. Substituting (20) into (19) and recognizing that $\beta^\alpha \lambda^\alpha (v^\alpha - R^\alpha) \geqslant 0$ we obtain

$$\begin{aligned} T &= [\beta^k \Pi^k R^k + (1-\beta^k)\, R^k - \beta^k \lambda^k (v^k - R^k)] \\ &\quad + \beta^k (1-\Pi^k)[\beta^l \Pi^l R^l + (1-\beta^l)\, R^l - \beta^l \lambda^l (v^l - R^l)] \\ &\quad + \beta^k (1-\Pi^k)\, \beta^l (1-\Pi^l)[\beta^m \Pi^m R^m + (1-\beta^m)\, R^m \\ &\quad - \beta^m \lambda^m (v^m - R^m)] \\ &\quad + \cdots \\ &\leqslant [\beta^k \Pi^k R^k + (1-\beta^k)\, R^k] \\ &\quad + \beta^k (1-\Pi^k)[\beta^l \Pi^l R^l + (1-\beta^l)\, R^l] \\ &\quad + \beta^k (1-\Pi^k)\, \beta^l (1-\Pi^l)[\beta^m \Pi^m R^m + (1-\beta^m)\, R^m] \\ &\quad + \cdots \\ &= [1 - \beta^k (1-\Pi^k)]\, R^k + \beta^k (1-\Pi^k)[1 - \beta^l (1-\Pi^l)]\, R^l \\ &\quad + \beta^k (1-\Pi^k)\, \beta^l (1-\Pi^l)[1 - \beta^m (1-\Pi^m)]\, R^m. \end{aligned}$$

Defining

$$s^\alpha = \beta^\alpha (1-\Pi^\alpha) \geqslant 0$$

we can write

$$\begin{aligned} T &= [1-s^k]\, R^k + s^k [1-s^l]\, R^l + s^k s^l [1-s^m]\, R^m + \cdots \\ &= R^k + s^k [\underbrace{(R^l - R^k)}_{\leqslant 0} + s^l [\underbrace{(R^m - R^l)}_{\leqslant 0} + s^m [\cdots \\ &\leqslant R^k \\ &< R^j. \end{aligned}$$

Thus (21) is strictly positive and strategies of the form T cannot be optimal.

[22] From the definition of the reservation price we have $\Pi^j > 0$.

The only strategy that is not of the form S or T and that has not been proven to be suboptimal is the sampling strategy of Theorem 4.1. It uses the optimal stopping rule, samples boxes in the order of decreasing reservation prices and always chooses the box with the highest reservation price. Since an optimal strategy exists (either due to the finiteness of expectations in the case of a finite number of search opportunities or due to the assumption of finite variance in the case of infinitely many search opportunities, see DeGroot [3, Chap. 12 and 13]), this establishes the optimality of the proposed rule. ∎

ACKNOWLEDGMENTS

This paper draws upon chapter 1 of my Ph.D. thesis at the European University Institute, Florence, Italy. I thank Pierpaolo Battigalli, Matthias Brückner, and Ramon Marimon for insightful discussions and helpful comments. Financial support from the Deutscher Akademischer Austauschdienst is gratefully acknowledged. The usual disclaimer applies.

REFERENCES

1. K. Adam, Competitive pricing in markets with search and information frictions, Economics Department Working Paper, European University Institute, Florence, Italy, 2000. [http://www.iue.it]
2. S. Bikchandani and S. Sharma, Optimal search with learning, *J. Econ. Dynamics and Control* **20** (1996), 333–359.
3. M. DeGroot, "Optimal Statistical Decisions," McGraw–Hill, New York, 1970.
4. Chien fu Chou and G. Talmain, Nonparametric search, *J. Econ. Dynamics and Control* **17** (1993), 771–784.
5. J. C. Gittins, "Multi-armed Bandit Allocation Indices," Wiley, New York, 1989.
6. J. C. Gittins and D. M. Jones, A dynamic allocation index for the sequential design of experiments, *in* "Progress in Statistics" (J. Gani, Ed.), North-Holland, Amsterdam, 1972.
7. K. D. Glazebrook, A profitability index for alternative research projects, *OMEGA*, *Inter. J. Management Sci.* **4** (1976), 79–83.
8. K. D. Glazebrook, Some ranking formulae for alternative research projects, *OMEGA*, *Inter. J. Management Sci.* **6** (1978), 193–194.
9. K. D. Glazebrook, Stoppable families of alternative bandit processes, *J. Appl. Probab.* **16** (1979), 843–854.
10. M. G. Kohn and S. Shavell, The theory of search, *J. Econ. Theory* **9** (1974), 93–123.
11. S. A. Lippman and J. J. McCall, The economics of job search: A survey, *Econ. Inquiry* **14** (1976), 155–189 and 347–367.
12. C. J. McKenna, Theories of individual search behavior, *in* "Surveys in the Economics of Uncertainty" (J. D. Hey and P. J. Lambert, Eds.), Blackwell, New York, 1987.
13. P. B. Morgan, Distributions of the duration and value of job search with learning, *Econometrica* **53** (1985), 1199–1232.
14. J. F. Reinganum, Nash equilibrium search for the best alternative, *J. Econ. Theory* **30** (1983), 139–152.

15. D. B. Rosenfield and R. D. Shapiro, Optimal adaptive price search, *J. Econ. Theory* **25** (1981), 1–20.
16. M. Rothschild, Searching for the lowest price when the distribution of prices is unknown, *J Polit. Econ.* **82** (1974), 689–711.
17. S. C. Salop, Systematic job search and unemployment, *Rev. Econ. Studies* **40** (1973), 191–201.
18. R. Slonim, Learning in a search-for-the-best-alternative experiment, *J. Econ. Behav. Organization* **25** (1994), 141–165.
19. J. G. Stigler, The economics of information, *J. Polit. Economy* **65** (1961), 213–225.
20. G. Talmain, Search from an unknown distribution: An explicit solution, *J. Econ. Theory* **57** (1992), 141–157.
21. T. Vishwanath, Parallel search for the best alternative, *Econ. Theory* **2** (1992).
22. M. L. Weitzman, Optimal search for the best alternative, *Econometrica* **47** (1979), 641–654.