*Econometrica*, Vol. 47, No. 3 (May, 1979)

# OPTIMAL SEARCH FOR THE BEST ALTERNATIVE

BY MARTIN L. WEITZMAN[1]

This paper completely characterizes the solution to the problem of searching for the best outcome from alternative sources with different properties. The optimal strategy is an elementary reservation price rule, where the reservation prices are easy to calculate and have an intuitive economic interpretation.

## INTRODUCTION

A BROAD CLASS of economic search problems can be cast in the following form. There are a number of different opportunities or sources, each yielding an unknown reward. The uncertainty about the reward from a source can be eliminated, at a fee, by searching or sampling. Each source has its own independent probability distribution for the reward, search cost, and search time. Sources are sampled sequentially, in whatever order is desired. When it has been decided to stop searching, only one opportunity is accepted, the maximum sampled reward. Under this formulation, what sequential search strategy maximizes expected present discounted value?

A powerful solution concept applies to the above model. Each source is assigned a reservation price—an invariant critical number analogous to an internal rate of return. The reservation price of a source is easily computed, depends *only* on the features of that source, and has an intuitive economic interpretation.

The selection rule is to search next that unsampled source with highest reservation price. The stopping rule is to terminate search whenever the maximum sampled reward is above the reservation price of every unsampled source. This simple characterization of an optimal policy is the basic result of the present paper. Fundamental properties are derived and interpreted.

## AN EXAMPLE

The following example conveys the flavor of the basic problem analyzed in this paper.

Suppose the research department of a certain large organization has been assigned the task of finding a new and cheaper way to produce some commodity. Two substitute technologies are being considered, the benefits of which are uncertain and would not be known until development work is completed. Because they produce the same commodity, no more than one technology would actually be used even if both were developed. It is estimated that a production process based on the so-called alpha technology might yield a total savings of 100 with

[1] For their helpful comments, I would like to thank K. Burdett, M. Manove, M. Obstfeld, and M. Rothschild.

probability .5 and of 55 with probability .5. The alternative omega process with probability .2 might deliver a possible savings of 240 but it would not offer any improvement at all with probability .8. Research and development, which must be done to remove the uncertainty, costs 15 for the alpha process and takes one period, whereas it costs 20 for omega and takes two periods. The interest rate is 10 per cent per period. Table I summarizes the relevant information.

TABLE I

| Project | $\alpha$ | | $\omega$ | |
|---|---|---|---|---|
| Cost | 15 | | 20 | |
| Duration | 1 | | 2 | |
| Reward | 100 | 55 | 240 | 0 |
| Probability | .5 | .5 | .2 | .8 |

The problem is to find a sequential search strategy which maximizes expected present discounted value.

It is easy to show that developing only alpha or only omega is better than not researching either project. The expected value of researching alpha is

$$-15+\left(\frac{1}{1.1}\right)[.5(100)+.5(55)]=55.5,$$

whereas for omega it is

$$-20+\left(\frac{1}{1.1}\right)^2[.2(240)+.8(0)]=19.7.$$

Thus, at least one project should be developed.

The next logical question is which alternative should be researched first?

By any of the standard economic criteria, alpha dominates omega. Alpha has lower research cost, shorter development lag, higher expected reward, greater minimum reward, less variance.

Most economists or engineers might guess that alpha should be developed first. They would probably be reacting to the fact that the expected value of alpha is so much higher than omega.

However, there is a crucial difference between the value of a project and the order in which it should be researched. Alpha is worth more in the sense that the expected value of an optimal program without it is lower than without omega. Nevertheless, and somewhat paradoxically, it turns out that the optimal sequential strategy is to develop omega first.

This can be shown as follows. Suppose alpha is developed first. If the payoff turns out to be 55, it would then be worthwhile to develop omega, because the expected value of that strategy would be

$$-20+\left(\frac{1}{1.1}\right)^2[.2(240)+.8(55)]=56$$

which is greater than the value at that point of not developing omega, 55.

However, since

$$-20+\left(\frac{1}{1.1}\right)^2[.2(240)+.8(100)]=85.8,$$

it would not be economical to develop omega if alpha had a 100 payoff.

So the expected value of an optimal policy beginning with developing alpha is

$$-15+\left(\frac{1}{1.1}\right)\left[.5(100)+.5\left(-20+\left(\frac{1}{1.1}\right)^2[.2(240)+.8(55)]\right)\right]=55.9.$$

A similar calculation shows the expected value of an optimal policy which starts by developing omega is

$$-20+\left(\frac{1}{1.1}\right)^2\left[.2(240)+.8\left(-15+\left(\frac{1}{1.1}\right)[.5(100)+.5(55)]\right)\right]=56.3$$

Thus, the optimal policy for this example has the counter-intuitive property that omega is researched first.

The remainder of the paper is devoted to placing this kind of problem in a more general context, deriving a simple decision rule, and explaining its properties.

## PANDORA'S PROBLEM[2]

There are $n$ closed boxes at the beginning of our scenario. Box $i$, $1\leq i\leq n$, contains a potential reward of $x_i$ with probability distribution function $F_i(x_i)$, independent of the other rewards. It costs $c_i$ to open box $i$ and learn its contents, which become known only after a time lag of $t_i$. Instantaneous learning is the special case $t_i=0$.

An initial amount $x_0$ is available, representing a fallback reward that could always be collected if no sampling were undertaken or if every sampled reward happened to be less than $x_0$. In many applications it is natural to set $x_0=0$. All costs and benefits are converted to present values by the discount rate $r$.

At each stage Pandora must decide whether or not to open a box. If she chooses to stop searching, Pandora collects at that time the maximum reward she has thus far uncovered. Should Pandora wish to continue sampling, she must select the next box to be opened, pay at that time the fee for opening it, and wait for the outcome. Then will come the next decision stage. Note a characteristic asymmetry: the *sum* of search costs is paid *during* search, whereas the *maximum* reward is collected *after* search has been terminated.

Pandora worships maximized expected present discounted value. She needs to know what she should do to be consistent with this fundamental conviction. Pandora wants a sequential decision rule that will tell her at each stage whether or not to continue searching, and if so, which box to open next.

Pandora's problem can be formally posed in a dynamic programming format. Let the collection of $n$ boxes, denoted $I$, be partitioned into any set $S$ of sampled

[2] With apologies to connoisseurs of Greek mythology.

boxes and its complement $\bar{S}$ of closed boxes. That is,

$$S \cup \bar{S} = I, \qquad S \cap \bar{S} = \varnothing,$$

where

$$I = \{1, 2, \ldots, n\}.$$

The variable $y$ will represent the maximum sampled reward (from the opened boxes *and* the initial fallback reward)

$$(1) \qquad y = \max_{i \in S \cup \{0\}} x_i.$$

It is intuitively obvious (and easily verified) that all relevant information about the previously opened boxes is summarized by $y$; knowing the individual values of $x_i$ for $i \in S \cup \{0\}$ is superfluous to making a correct decision because all probability distributions are independent.

The state of the system at any time is given by the statistic $(\bar{S}, y)$. Define $\Psi(\bar{S}, y)$ as the expected present discounted value of following an optimal policy from this time on when the set of closed boxes is $\bar{S}$ and the maximum sampled reward is $y$.

For each subset $\bar{S}$ of $I$ and every $y$, the state valuation functions $\Psi$ must satisfy the fundamental recursive relation

$$(2) \qquad \Psi(\bar{S}, y) = \max\left\{y, \max_{i \in \bar{S}}\left\{-c_i + \beta_i\left[\Psi(\bar{S} - \{i\}, y)\int_{-\infty}^{y} dF_i(x_i) + \int_{y}^{\infty} \Psi(\bar{S} - \{i\}, x_i)\, dF_i(x_i)\right]\right\}\right\}$$

where[3]

$$(3) \qquad \Psi(\varnothing, x) = x,$$

$$(4) \qquad \beta_i = e^{-rt_i}.$$

Equation (2) is just the principle of optimality for dynamic programming. At stage $(\bar{S}, y)$ Pandora could terminate search, collecting reward $y$. Or, she might open box $i$, for each $i \in \bar{S}$, which results in expected discounted net gain

$$-c_i + \beta_i\left[\Psi(\bar{S} - \{i\}, y)\int_{-\infty}^{y} dF_i(x_i) + \int_{y}^{\infty} \Psi(\bar{S} - \{i\}, x_i)\, dF_i(x_i)\right].$$

The value of an optimal policy at $(\bar{S}, y)$ is the maximum of these alternatives.

In principle, the state valuation functions $\{\Psi(\bar{S}, y)\}$ could be recursively built up by systematic induction on the number of closed boxes. Using (2), (3), state valuation functions could be constructed first for all sets consisting of one closed box, then for all sets of two boxes, of three, four, etc. The actual computation is likely to be a combinatoric task of unwieldy proportions unless the number of boxes is very small.

[3] For some applications it may be appropriate to interpret $1 - \beta_i$ as the probability that investigating source $i$ results in a catastrophic accident, nullifying rewards and terminating search. Note that nothing formally prevents $c_i$ from being negative; in such situations $-c_i$ would be interpreted as a component of gain additive across sources.

At any stage $(\bar{S}, y)$, Pandora's optimal decision is that policy which maximizes the right hand side of (2). If two or more policies tie, it makes no difference how the tie is broken. Note that although an optimal strategy is implicitly contained in equation (2), the form of that strategy is little more than a complete enumeration of what to do in all possible situations.

The economic search literature has dealt extensively with the situation where, in effect, all boxes are identical. For this special case the issue of choosing *which* box to open does not arise. The essential question is when to stop. The answer is: search continues until a reward greater than some "reservation price" is discovered. The reservation price for sampling with recall is that hypothetical cutoff value of the maximum reward which would make it just equal to the expected net gain of opening exactly one more box.[4]

The contribution of the present paper is to show that with alternative search opportunities the optimal policy is a straightforward analogue of the above idea. Each (different) box is assigned a (different) reservation price, calculated by the same formula as before, which now serves as a basis for the optimal stopping *and* selection rule. The reservation price of a box determines its ordinal ranking, prescribing when it should be opened relative to the other boxes. Thus, all the advantages of a simple rate of return criterion apply in a search context.

## SOME INTERPRETATIONS

The formulation presented in this paper is general enough to cover, at a high level of abstraction, economic search models from a variety of settings.

Take for example the standard job search model with wage offers retained. The current framework allows the situation where the job searcher may choose to sample from various firms having different characteristics. The lump sum reward is most appropriately interpreted as the discounted present value of all future wages. Search costs, which presumably include a psychic component, are net of any side compensation (unemployment benefits or wages from a currently held job). The possibility of reaccepting current work while searching on-the-job is accommodated by making the fallback reward $x_0$ equal to the present discounted value of the current wage. Other modifications are also possible.

Searching for the lowest price on some commodity available from different stores is also an example amenable to the analysis developed in this paper. Let the good have some intrinsic utility measured in dollar terms. The reward available from a store is the difference between the utility of the commodity and its price. Search costs should include the opportunity loss of forgoing the item in question while search continues, as well as the more orthodox cost of visiting a store to obtain a price quotation. The option of not buying the good at all can be represented by setting the fallback reward equal to the (dis)utility of henceforth doing without the item altogether, which can be normalized to zero.

[4] This stopping rule is well known and appears in many places. See, for example, Lippman and McCall [**8**] or Landsberger and Peled [**7**].

Another area of application concerns the optimal sequential research strategy for developing various uncertain technologies to meet the same or a similar purpose. The reward is the potential cost saving of the new technology, unknown until after it has been developed. Search fees are research and development expenditures. Search time is the anticipated length of the research and development process. The option of choosing to continue with the current known technology is represented by having a zero fallback reward.

There are several other possible interpretations of the model. Two of them are described in the section on applications.

## THE OPTIMAL STRATEGY

Any closed box is characterized by a fee for opening it, a time lag for discovering its contents, and a probability distribution for the reward it contains. Suppose all this information must somehow be compressed into a single index number, a kind of internal rate of return. One heuristic procedure would be to evaluate the intrinsic search value of a closed box by assigning it the hypothetical reward of that opened box to which it is in some sense equivalent.

Suppose for the moment there are just two boxes. One is the closed box $i$. The other is an already opened hypothetical box offering reward $z_i$. If the searcher elects not to open box $i$, she receives the "sure thing"

$$z_i. \tag{5}$$

If she opens box $i$, the searcher can expect a net benefit

$$-c_i+\beta_i\left[z_i\int_{-\infty}^{z_i} dF_i(x_i)+\int_{z_i}^{\infty} x_i\, dF_i(x_i)\right]. \tag{6}$$

The closed and opened boxes are heuristically "equivalent" if the searcher is just indifferent between opening box $i$ and not opening it. This will occur if (5) and (6) are equal to each other, a condition which can be written

$$c_i=\beta_i\int_{z_i}^{\infty}(x_i-z_i)\, dF_i(x_i)-(1-\beta_i)z_i. \tag{7}$$

An alternative interpretation of equation (7) is illuminating. Suppose, hypothetically, there are only type $i$ boxes and an infinite number of them. (In this situation it is irrelevant whether sampling is with or without recall.) Let $z_i$ represent the expected present discounted value of following an optimal search policy. Some reflection shows that $z_i$ must equal (6), viewed in this context as a dynamic programming expression.

The critical number $z_i$ which satisfies (7) is called the *reservation price* of box $i$.

Although the definition (7) has been motivated by heuristic considerations, it turns out there is a rigorous sense in which *all* relevant information about box $i$ is summarized by its reservation price $z_i$.

The following decision strategy, called *Pandora's Rule*, completely characterizes an optimal policy.

SELECTION RULE: *If a box is to be opened, it should be that closed box with highest reservation price.*

STOPPING RULE: *Terminate search whenever the maximum sampled reward exceeds the reservation price of every closed box.*

What is remarkable about this rule is that the entire structure of an optimal policy has been reduced to a simple statement about reservation prices. Furthermore, the reservation price of each box is calculated by equating a hypothetical gain of stopping (5) *not* with the full gain of opening the box and continuing on in an optimal manner, but rather with the myopic gain of opening the box and terminating (6). In other words, the reservation price of a box depends *only* on the properties of that box and is independent of all other search opportunities.

Note that if Pandora samples from $n$ identical boxes, the optimal policy is to continue search until she uncovers a reward greater than the common reservation price of each box. In this special case, it is comparatively simple to prove optimality.

The proof of Pandora's Rule, which is quite technical, is relegated to the final section.

## PROPERTIES OF RESERVATION PRICES

From (7), the reservation price of a box is completely insensitive to the probability distribution of rewards at the lower end of the tail. Any rearrangement of the probability mass located below $z_i$ leaves $z_i$ unaltered. It is important to understand this feature. Considering that a box could be opened at any time, the only rationale for opening it now is the possibility of terminating further search by drawing a relatively high reward. That is why the lower end of its reward distribution is irrelevant to the *order* in which box $i$ should be sampled even though it may well influence the *value* of an optimal policy by altering the likelihood that $x_i$ will end up being the largest reward drawn.

On the other hand, as rewards become more dispersed at the upper end of the distribution, the reservation price increases and so does the net benefit of search. Other things being equal, it is optimal to sample first from distributions which are more spread out or riskier in hopes of striking it rich early and ending the search. This is a major result of the present paper. Low-probability high-payoff situations should be prime candidates for early investigation even though they may have a smaller chance of ending up as the source ultimately yielding the maximum reward when search ends.

The standard comparative statics exercises performed on (7) yield anticipated results. Reservation price decreases with greater search cost, increased search time, or a higher interest rate. Moving the probability mass of rewards to the right (i.e., changing the distribution function $F_i(x_i)$ to $G_i(x_i) \leq F_i(x_i)$) makes $z_i$ larger.

Thus, although there is no necessary connection between the *mean* reward and the reservation price, there is a well-defined sense in which higher rewards increase the reservation price. Similarly, performing a mean preserving spread on the distribution function $F_i(x_i)$ makes $z_i$ bigger. In this sense a riskier distribution of rewards implies a higher reservation price.[5]

Note that acceptance levels decline with the duration of search, as the best opportunities are sampled first and the poorer ones later.

Because it is so easy to calculate reservation prices, sensitivity analysis is made especially simple. The effect on project ranking (and hence on an optimal policy) of changing such parameters as search costs, the probability distribution of rewards, search time, or the interest rate is easily determined. It is also easy to say how an optimal search strategy changes when certain opportunities are added to or deleted from the list of prospective candidates.

## APPLICATIONS

To illustrate the nature of the solution concept and indicate on what it depends, two explicit examples are calculated for interesting special cases.

In the first example, suppose that box $i$ contains one of two outcomes: either zero reward ("failure") with probability $1-p_i$, or positive reward $R_i$ ("success") with probability $p_i$. To keep things simple there is no discounting ($\beta_i = 1$) and the expected net gain $p_iR_i - c_i$ is positive.

Applying (7) to this special case yields the closed form expression

$$z_i = \frac{p_iR_i - c_i}{p_i}. \tag{8}$$

The reservation price of a box is the expected net gain divided by the probability of success. For the same expected net gain, that box is opened first which offers a *smaller* probability of success.

After ranking boxes to be opened in order of decreasing $z_i$, the searcher moves down the list until a success is encountered. At that point search ends because $R_i > z_i$.

Suppose, as a further restriction, search is for the *same* object, for example a new product with certain well-defined characteristics. Then all rewards $R_i$ are identical. In that case (8) reduces to

$$z_i = R - \frac{c_i}{p_i}.$$

The next opportunity sampled is the one offering the highest probability of success per dollar of search cost. This is also a well-known characterization of the best way to locate a lost or hidden object.[6]

[5] This feature has been analyzed by Kohn and Shavell [**6**] for the case of identical boxes.

[6] See, for example, De Groot [**3**], Kadane and Simon [**5**], or Stone [**12**].

A second example is the so-called "gold mining problem."[7] Suppose there is a movable gold mining machine. Mine $i$ contains amount $G_i$ of gold, but after digging it out the machine is liable to break down with probability $q_i$, preventing all further mining. In what order should the mines be exploited to maximize total expected gold?

Problems like this are a special (almost degenerate) case of Pandora's problem. Equation (2) is a valid dynamic programming formulation of the gold mining problem under the special assumptions $x_0=0$, $x_i=0$ (each box contains zero reward with probability one), $c_i=-G_i$, $\beta_i=1-q_i$.

Applying (7) to this special case, the integral vanishes when $z_i>0$, yielding as a solution the closed form expression

$$z_i=\frac{G_i}{q_i}.$$

The optimal policy is to exploit next that unopened mine with maximum gold per probability of machine breakdown, a classic result.

## SOME LIMITATIONS

The purpose of the model formulated in this paper is to sharply characterize optimal search among alternative sources with different characteristics. Naturally certain "other" aspects of the optimal search problem have been abstracted away.

Many of the underlying assumptions of the present formulation are unrealistic. There has been no provision made for: adaptive learning about correlated probability distributions; pay-as-you-go research (with the possibility of backing out of a project if prospects start looking unfavorable); parallel search activity; risk aversion; incomplete or no recall; collecting some reward before search is terminated; randomly generated new opportunities; a binding time horizon; uncertain search costs or search time; etc.[8] Yet the model as a whole captures enough essential aspects of reality that it should be useful in providing project rankings which might serve as a rough planning guide of sorts, a kind of pre-investment screening device, or a reference point for the numerical analysis of a more comprehensive dynamic programming type formulation.

The fact that it is possible to explicitly construct an optimal solution makes the problem analyzed here a natural preliminary to more general formulations. And the present model may even be a reasonable description of some situations.

That such an elementary decision strategy as Pandora's Rule is optimal depends more crucially than might be supposed on the simplifying assumptions of the model. There does not seem to be available a sharp characterization of an optimal

[7] There are several variants of this problem, having essentially the same underlying structure. Other names are the quiz show problem, the obstacle course problem, the least cost testing sequence problem. See, for example, Bellman [**1**] or Kadane [**4**]. With only minor changes in interpretation, the present framework can incorporate such features as possible breakdown before receiving the prize, waiting times, consolation prizes, etc.

[8] Some of these topics have been treated in the literature, most typically for the symmetric case where all boxes are identical. See the bibliography cited at the end of this paper.

solution when certain features of the present model are changed. Pandora's Rule does not readily generalize.

For example, Pandora's Rule does not determine an optimal ordering if she may only open $m$ of the $n$ boxes available to her. An example of this was provided in the second section, where alpha was preferable when only one opportunity could be searched, whereas omega was the better starting choice with the possibility of sequential sampling from both sources. In the general case $n > m \geqslant 2$, an involved permutational exercise would be required to determine which $m$ boxes should be potentially sampled. However, once *given* the list of $m$ boxes, Pandora's Rule applied to *this* subset would be the optimal decision strategy.

If reward distributions were not independent, the optimal search strategy could be very complicated. When a box is opened, the searcher would learn not only about its contents, but also about the reward distributions of alternative boxes.[9] It appears plausible that other things being equal it would be better to open a box whose reward is highly correlated with other rewards because this adds a positive informational externality. But translating such an effect into a simple search rule seems difficult except in the most elementary cases.

Parallel search efforts and pay-as-you-go research with the option of backing out are important features of the research and development scene omitted from the current formulation. They seem to be very hard to model well.[10] Perhaps it is wishful thinking, but my feeling is that the results of this paper might still constitute a useful guide here. Even though modified by more complicated and realistic considerations, something like Pandora's Rule should remain part of any optimal sequential search policy.

If some fraction of its reward can be collected from a research project before the sequential search procedure as a whole is terminated, that could negate Pandora's Rule in extreme cases. It might be optimal to start off with a cheap low-risk research project which promises to supply modest benefits through the period of sequential search for the best alternative. Such a project is unlikely to be chosen at the end of the search, but it is developed at the beginning because it can provide a stream of interim rewards while the results of further sampling are awaited.

The cases of sampling without recall, risk aversion, and randomly generated new opportunities have been treated to some extent in the literature.[11] An optimal policy is typically complicated, especially when there are different kinds of boxes. What to do next will depend in an intricate way on past results and future possibilities.

A binding time horizon is somewhat like a restriction on the number of boxes that can be opened. To some extent the intended effect may be captured by naming an appropriately high discount rate. As in the case of a curved utility function, an optimal policy will move toward sampling less risky distributions.

[9] Rothschild [**10**] contains an illuminating analysis of adaptive search policies for the case of identical boxes.

[10] See, for example, Marshak et al. [**9**].

[11] See especially Salop [**11**] who treats thoroughly the no recall case. The other two cases are briefly surveyed for a situation with identical boxes in Lippman and McCall [**8**].

If search costs or discount factors are randomly distributed independently of everything else, no changes in formulation are necessary so long as $c_i$ and $\beta_i$ are interpreted as mean values.

## PROOF OF OPTIMALITY

First it must be shown that the formula for determining the reservation price of a box is well defined.

Let

$$H_i(z) = \beta_i \int_z^\infty (x_i - z)\, dF_i(x_i) - (1-\beta_i)z. \tag{9}$$

It is easily verified that the function $H_i(z)$ is continuous and monotonic. By taking the appropriate limits, $H_i(-\infty) = \infty$, $H_i(\infty) = -\infty$ ($=0$ if $\beta_i = 1$). Thus, so long as $c_i > 0$ or $\beta_i < 1$, there exists a solution $z_i$ to the equation

$$c_i = H_i(z_i)$$

which is unique.

The proof of the main proposition is by induction on the number of closed boxes. Suppose Pandora's Rule is optimal with $m$ closed boxes remaining and any value of $y$ (representing the maximum return from the previously opened boxes and the initial fallback reward). For $m = 1$, the optimality of Pandora's Rule is easily demonstrated just by directly applying the definition of reservation price.

Henceforth we will be considering a situation with $m+1$ closed boxes (the set $\bar{S}$) and any value of $y$.

Let $j$ be a box with biggest reservation price in the collection of $m+1$ closed boxes

$$j \in \bar{S} \qquad z_j = \max_{i \in \bar{S}} z_i. \tag{10}$$

If $y \geqslant z_j$, it is simple to demonstrate the optimality of not opening any boxes. (After one box is opened, by Pandora's Rule applied to $m$ closed boxes it will be optimal to stop. Hence the question is whether opening exactly one box is better than not opening any, which is easily answered in the negative.) The stopping criterion of Pandora's Rule is thus proved for $m+1$ closed boxes.

If $y < z_j$, it is straightforward to show the nonoptimality of no further search. (Just opening box $j$ and then stopping would yield a higher expected present discounted value.) Thus, at least one box should be opened.

Suppose (by contradiction with Pandora's Rule) it is optimal to open $k$ first, where $k$ is any box in $\bar{S}$ having a lower reservation price than $j$:

$$k \in \bar{S}, \qquad z_k < z_j. \tag{11}$$

If box $k$ is opened first, by the induction assumption on Pandora's Rule for $m$ closed boxes there is an exact prescription of what to do in an optimal policy thereafter. Let the expected discounted present value of opening box $k$ and following Pandora's Rule thereafter, which is alleged to constitute the *best* strategy, be $B$.

Consider the following alternative. Open box $j$ first. Let $h$ be a box with second biggest reservation price in the collection of $m+1$ closed boxes

$$h \in \bar{S}-\{j\}$$
$$z_h = \max_{i \in \bar{S}-\{j\}} z_i. \tag{12}$$

If $x_j \geqslant z_h$, terminate. Otherwise, open box $k$ next. From then on proceed by Pandora's Rule. Let the expected present discounted value of this *alternative* policy be $A$.

The rest of the proof, although technical in its details, essentially consists of showing that $A > B$. From this it follows that the originally proposed policy of first opening box $k$ cannot in fact be best. It must be optimal to first open the box with biggest implicit worth in $\bar{S}$ and then, from the induction assumption, to proceed by Pandora's Rule for the remaining $m$ closed boxes. But this is just Pandora's selection rule for $m+1$ closed boxes, completing the induction step.

The following notation is employed (Figure 1 may be useful in providing a sort of mnemonic device).

$$\pi_j = \text{prob}\,(x_j \geqslant z_j), \qquad w_j = E[x_j | x_j \geqslant z_j],$$
$$\pi_k = \text{prob}\,(x_k \geqslant z_j), \qquad w_k = E[x_k | x_k \geqslant z_j],$$
$$\lambda_j = \text{prob}\,(z_h \leqslant x_j < z_j), \qquad v_j = E[x_j | z_h \leqslant x_j < z_j],$$
$$\tilde{v}_j = E[\max\,(x_j, y) | z_h \leqslant x_j < z_j],$$

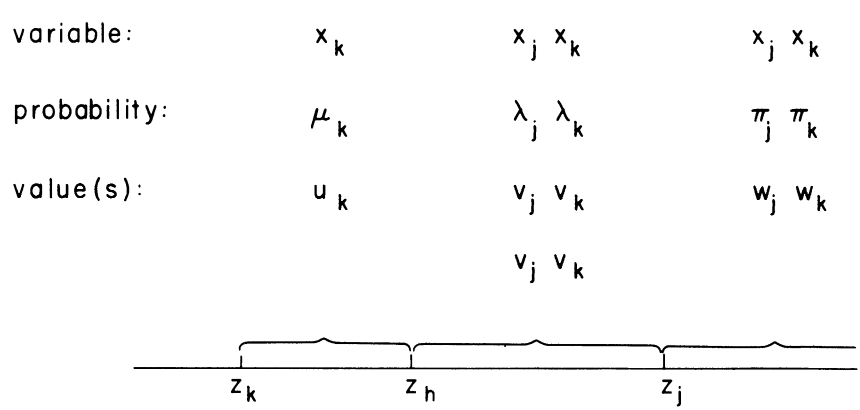


FIGURE 1

$$\lambda_k = \text{prob}\,(z_h \leqslant x_k < z_j), \quad v_k = E[x_k | z_h \leqslant x_k < z_j],$$
$$\tilde{v}_k = E[\max\,(x_k, y) | z_h \leqslant x_k < z_j],$$
$$\mu_k = \text{prob}\,(z_k \leqslant x_k < z_h), \quad u_k = E[x_k | z_k \leqslant x_k < z_h].$$

$$d = E[\max\,(x_j, x_k, y) | z_h \leqslant x_j < z_j;\, z_h \leqslant x_k < z_j], \tag{13}$$

$$\Phi = E[\Psi(\bar{S} - \{j\} - \{k\}, \max\,(x_j, x_k, y)) | x_j < z_h;\, x_k < z_h]. \tag{14}$$

The expected present discounted value of opening box $k$ and then proceeding by Pandora's Rule is

$$\begin{aligned} B = {} & -c_k + \pi_k\beta_k w_k + \lambda_k\beta_k[-c_j + \pi_j\beta_j w_j + \lambda_j\beta_j d + (1-\pi_j-\lambda_j)\beta_j\tilde{v}_k] \\ & + (1-\pi_k-\lambda_k)\beta_k[-c_j + \pi_j\beta_j w_j + \lambda_j\beta_j\tilde{v}_j] \\ & + (1-\pi_k-\lambda_k)(1-\pi_j-\lambda_j)\beta_k\beta_j\Phi. \end{aligned} \tag{15}$$

The proposed alternative is: open box $j$; if $x_j \geqslant z_h$, terminate; if $x_j < z_h$, next open box $k$ and then proceed by Pandora's Rule. The expected present discounted value of such a policy is

$$\begin{aligned} A = {} & -c_j + \pi_j\beta_j w_j + \lambda_j\beta_j\tilde{v}_j + (1-\pi_j-\lambda_j)\beta_j[-c_k + \pi_k\beta_k w_k + \lambda_k\beta_k\tilde{v}_k] \\ & + (1-\pi_j-\lambda_j)(1-\pi_k-\lambda_k)\beta_j\beta_k\Phi. \end{aligned} \tag{16}$$

Subtracting (15) from (16), cancelling some terms and grouping others,

$$\begin{aligned} A - B = {} & [(c_j - \pi_j\beta_j w_j)((1-\pi_k)\beta_k - 1) + (c_k - \pi_k\beta_k w_k)(1-(1-\pi_j-\lambda_j)\beta_j) \\ & + \lambda_j\beta_j\tilde{v}_j - \lambda_k\beta_k\lambda_j\beta_j d - (1-\pi_k-\lambda_k)\beta_k\lambda_j\beta_j\tilde{v}_j. \end{aligned} \tag{17}$$

From (7),

$$c_j = \beta_j\pi_j(w_j - z_j) - (1-\beta_j)z_j, \tag{18}$$

$$c_k = \beta_k[\pi_k(w_k - z_k) + \lambda_k(v_k - z_k) + \mu_k(u_k - z_k)] - (1-\beta_k)z_k. \tag{19}$$

Substituting in (17) for $c_j$ and $c_k$ from the above expressions yields something that can be manipulated into the form

$$\begin{aligned} A - B = {} & (z_j - z_k)[(\pi_j\beta_j + 1 - \beta_j)(\pi_k\beta_k + 1 - \beta_k)] \\ & + (v_k - z_k)[\lambda_k\beta_k(1 - \beta_j + \pi_j\beta_j)] + (\tilde{v}_j - z_k)[\lambda_j\beta_j(1 - \beta_k + \pi_k\beta_k)] \\ & + (u_k - z_k)[\mu_k\beta_k(1 - \beta_j + \pi_j\beta_j + \lambda_j\beta_j)] \\ & + (\tilde{v}_j + v_k - z_k - d)[\lambda_k\beta_k\lambda_j\beta_j]. \end{aligned} \tag{20}$$

From the definition (13),

$$(21)\qquad d = z_h + E[\max(\max(x_j, y) - z_h, x_k - z_h)/z_h \leqslant x_j < z_j;\, z_h \leqslant x_k < z_j]$$
$$\leqslant z_h + E[(\max(x_j, y) - z_h + x_k - z_h)/z_h \leqslant x_j < z_j; z_h \leqslant x_k < z_j]$$
$$= \tilde{v}_j + v_k - z_h$$
$$\leqslant \tilde{v}_j + v_k - z_k.$$

Using the above inequality and the fact that $0 < \beta_j \leqslant 1$, $0 < \beta_k \leqslant 1$, every term of expression (20) is seen to be nonnegative, with the first term strictly positive. Thus,

$$(22)\qquad A > B.$$

This concludes our proof of the form of an optimal policy.

Strictly speaking, we have proved the necessity of Pandora's Rule. That rule specifies a unique strategy for each state (except when there is a tie for the maximum reservation price of a closed box, in which case it can be shown, along the lines of the current proof, that how the tie is broken makes no difference to the value of the objective function). Thus, since an optimum policy exists, sufficiency of Pandora's Rule has also been demonstrated.

*Massachusetts Institute of Technology*

*Manuscript received August, 1977; final revision received March, 1978.*

REFERENCES

[1] BELLMAN, RICHARD: *Dynamic Programming.* Princeton: Princeton University Press, 1957.
[2] CHOW, Y. S., HERBERT ROBBINS, AND DAVID SIEGMUND: *Great Expectations: The Theory of Optimal Stopping.* New York: Houghton-Mifflin, 1971.
[3] DE GROOT, MORRIS H.: *Optimal Statistical Decisions.* New York: McGraw-Hill, 1970.
[4] KADANE, JOSEPH B.: "Quiz Show Problems," *Journal of Mathematical Analysis and Applications*, 27 (1969), 609–623.
[5] KADANE, JOSEPH B., AND HERBERT A. SIMON: "Optimal Strategies for a Class of Constrained Sequential Problems," *The Annals of Statistics*, 5 (1977), 237–255.
[6] KOHN, MEIR G., AND STEVEN SHAVELL: "The Theory of Search," *Journal of Economic Theory*, 9 (1974), 93–123.
[7] LANDSBERGER, MICHAEL, AND DAN PELED: "Duration of Offers and Price Realizations," *Journal of Economic Theory*, 14 (1977), 17–37.
[8] LIPPMAN, STEVEN A., AND JOHN J. MCCALL: "The Economics of Job Search: A Survey," *Economic Inquiry*, 4 (1976), 155–189.
[9] MARSHAK, T. T. GLENNON, AND R. SUMMERS: *Strategy for R. & D.* New York: Springer-Verlag, 1967.
[10] ROTHSCHILD, MICHAEL, "Searching for the Lowest Price When the Distribution of Prices is Unknown," *Journal of Political Economy*, 82 (1974), 689–711.
[11] SALOP, STEVEN C.: "Systematic Job Search and Unemployment," *Review of Economic Studies*, 40 (1973), 191–201.
[12] STONE, L. D.: *Theory of Optimal Search.* New York: Academic Press, 1975.