OPERATIONS RESEARCH

The Journal of the Operations Research Society of America

VOLUME 5 — *October, 1957* — NUMBER 5

THE THEORY OF SEARCH

III. THE OPTIMUM DISTRIBUTION OF SEARCHING EFFORT

BERNARD O. KOOPMAN

Columbia University, New York, New York

(Received June 20, 1957)

This is Part III, concluding the series of three papers on *The Theory of Search*. Part I, *Kinematic Bases*, appeared in *Opns. Res.* **4**, 324–346 (1956) and Part II, *Target Detection*, in **4**, 503–531 (1956). The present paper, while similar in content and results to the third chapter of *Search and Screening* (O. E. G. Report No. 56, CONFIDENTIAL) represents a considerable change in methodology as well as in exposition. The material was first presented by the author at the fall meeting of the OPERATIONS RESEARCH SOCIETY OF AMERICA in 1953, where it was treated as a natural extension of his paper, *The Optimum Distribution of Effort* [this JOURNAL, **1**, 52–63 (1953)].

INTRODUCTION TO THE PROBLEM: AN EXAMPLE

SUPPOSE THAT an object is in an unknown position, but that its probabilities of being in the various possible positions are known. Suppose, further, that a limited total amount of searching effort (or time) is available. Assume, finally, that the law of detection is known, telling the chance of finding the object when a given amount of search is carried out in its vicinity. The practical problem is to find the optimum manner of distributing the available searching effort: the one which maximizes the chance of finding the object.

The answer can be given explicitly and can be represented by a simple graphical construction based on a logarithmic plot in the important class of cases in which the law of detection has the 'exponential-saturation' form, as in random search (*cf.* Part II, p. 519). The ideas and methods go far beyond the problems of search and constitute a branch of operations analysis, the power and practical value of which were amply shown by their naval applications in World War II.

To bring out the ideas, we shall first set up the problem and give its graphical solution without proof, in the special case in which the object searched for (the *target*) is a point on a line (taken as the x-axis), all parts of which are equally easy to search in (uniformity of visibility). Later on, the mathematical methods will be introduced, and, finally, various extensions and generalizations will be given.

The precise statement of the problem in the example is as follows: The probability density $p(x)$ is given, $p(x)\,dx$ being the probability (to first order quantities in dx) that the target be between x and $x+dx$. It has the usual properties of distributions

$$p(x) \geqq 0, \qquad \int_{-\infty}^{+\infty} p(x)\,dx = 1,$$

to which we add (merely for simplicity of exposition) that $p(x)$ is continuous except, possibly, for a few simple infinities and finite jumps.

The amount of search between x and $x+dx$ is supposed to be representable (to first order quantities in dx) by $\varphi(x)\,dx$, where the *search density* $\varphi(x)$ has the two properties

$$\varphi(x) \geqq 0, \qquad \int_{-\infty}^{+\infty} \varphi(x)\,dx = \Phi. \tag{1}$$

Here Φ is a positive constant which is the measure, in some appropriate unit, of the total amount of search available. Then, according to the exponential-saturation law being assumed here, the (conditional) probability of detecting the target (given to be at x) is $1-e^{-\varphi(x)}$. As mentioned above, this is an x-axis version of the law of random search.* Furthermore, if there is only a probability $p(x)\,dx$ that the target be in $(x,x+dx)$, the probability that it be detected *in this interval* is

$$p(x)\,[1-e^{-\varphi(x)}]\,dx.$$

By the usual summation and passage to the limit, the over-all probability

* This one-dimensional case can be explained by the following example: Let n glimpses be made from random positions $x_1, x_2, \ldots, x_n$ in an interval R; and suppose that in the x_i-glimpse the target is detected if and only if it is between x_i-r and x_i+r (r = definite detection range, thought of as much less than R), so that its probability is $2r/R$. If, as we assume, the n detection events are independent, the probability of finding the target is, as usual, $1-(1-2r/R)^n$, or $1-e^{-2rn/R}$ for large n. Now suppose that the glimpse positions are so numerous and so smoothly distributed that n/R can be represented approximately by a continuous function of x (the mid-point of R). Then we can also write $2rn/R=\varphi(x)$, so that the expression for the probability becomes $1-e^{-\varphi(x)}$.

In the case in which $R=dx$ is 'infinitesimal,' it might be supposed that the detection probability should be linear in the search effort $\varphi(x)\,dx$, rather than exponential in $\varphi(x)$. The explanation of this paradox is that the effectiveness of each glimpse is not entirely localized, but affects a finite span $2r$ of positions.

of detection is*
$$P=\int_{-\infty}^{\infty} p(x)\,[1-e^{-\varphi(x)}]\,dx=P[\varphi]. \tag{2}$$

The mathematical statement of the problem is now clear: *Find the distribution of effort function $\varphi(x)$ which, of all functions satisfying* (1), *maximizes the expression P given in* (2). This differs from the conventional problems in elementary calculus of variations in that one of the side conditions of (1) is an *inequality* and not an equation: if the method of setting $\delta P=0$ were used, much of the time an absurd answer giving a negative $\varphi(x)$ would be obtained. As will be shown later, a very simple method gives the correct answer.

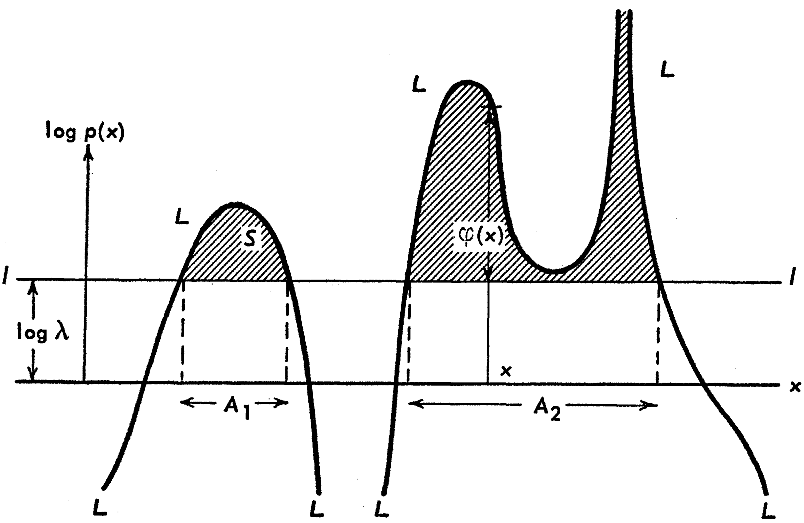

Figure 1

THE SOLUTION OF THE ILLUSTRATIVE PROBLEM

STEP 1. Graph the natural logarithm of the given probability $p(x)$ of the target's position for all x at which $0<p(x)<+\infty$. This graph of $y=\log p(x)$, L in Fig. 1, may be a single curve, or made up of several pieces separated by intervals where $p(x)=0$, or points where $p(x)=+\infty$.
STEP 2. Draw a horizontal line l, and, keeping it always parallel to the x-axis, move it up or down until the area S *above* this line and *under* the

* The use of the infinite limits of integration is perfectly consistent with the possibility that the target is actually known to be confined to a smaller interval $a\leqq x\leqq b$; it simply means that $p(x)=0$ outside of the latter. It is more convenient always to write the integration over the full axis (to be denoted, later on, by $\int$ without limits).

curve L has a value precisely equal to the total available quantity of searching effort Φ. Mark as $A=(A_1, A_2,$ etc.) the perpendicular projections onto the x-axis of the segments cut off from l by the graph L.

STEP 3: THE ANSWER. No search should be made *outside* the intervals A_1, A_2, etc. *Inside* these intervals, the density of search $\varphi(x)$ should be equal to the length cut off by l and L from the vertical line drawn for the fixed x in question.

These three steps solve the problem completely. This, and the fact that the constructions can always be carried out [supplementing L, if necessary, by appropriate vertical segments if it 'stops in mid-air,' i.e., at finite jumps of $p(x)$], will be proved later.

While still on the example, suppose that the full searching effort Φ has been used up according to the optimum schedule given above. What is the probability P of successful detection? What does the probability of position of the target become if the search has failed? (This is an a posteriori probability, $p'(x)$, given by Bayes' theorem.) The answers are found by proceeding as follows:

STEP 4. Take a (positive) number λ equal to the exponential of the constant y-value for points on l, so that this line has the equation $y=\log\lambda$. Graph the equation $y=p(x)$, as shown in Fig. 2. *The area* (shaded in the figure) cut from under this graph by the horizontal line $y=\lambda$ is the probability that the search be successful.

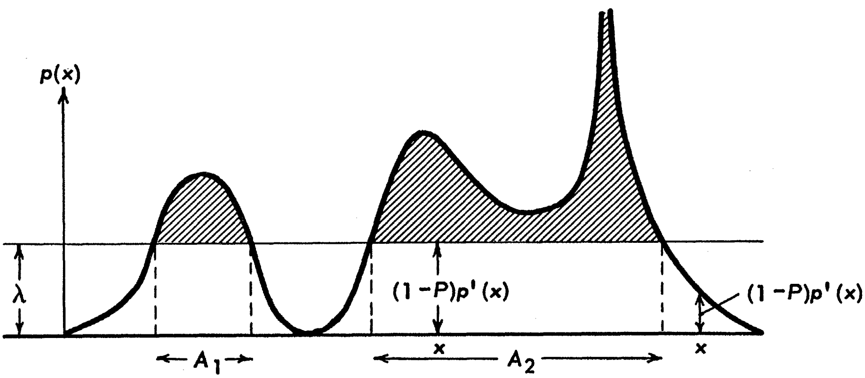

Figure 2

STEP 5. Decapitate the region below the curve $y=p(x)$ and above the x-axis by removing this shaded area above the line $y=\lambda$; the residual region will be bounded above by some horizontal segments of this line (projecting on the x-axis into A_1, A_2, etc.) and by remaining parts of the original curve $y=p(x)$ elsewhere; and, below, by the x-axis. Its area is $1-P$, the probability of failure of the search. The distance from any given point x on the x-axis to the upper boundary of the residual (decapitated) region is *proportional* to the new probability density $p'(x)$ of the target's position (given failure). When divided by area $1-P$, the distance becomes *equal* to $p'(x)$.

IF, AFTER the unsuccessful expenditure of effort Φ, a new quantity of searching effort Φ' becomes available, its optimum distribution is of course found as in the first case, but starting with $p'(x)$ instead of $p(x)$. Suppose that, on the other hand, it had been known in advance that we were to have the effort $\Phi+\Phi'$ available; could we have planned a better distribution of this total effort than in the case just described, in which first Φ and then Φ' are scheduled? The answer turns out to be in the negative: The successive scheduling is no worse—and, indeed, leads to the same over-all optimum search distribution—as the scheduling based on the greater knowledge in the latter case just mentioned. This very convenient state of affairs seems to be a characteristic property of the basic exponential law of search assumed throughout.

MATHEMATICAL TREATMENT

THE GRAPHICAL solution given above, while simple, requires proof of its validity. The mathematical methods involved, while elementary, are not trivial, nor can they be referenced in the existing literature. For this reason and because of their practical usefulness in other problems of this type, they will be set forth in detail in the case of the example, generalizations being noted later.

As stated above, the inequality in (1) (*unilateral condition*) excludes the ordinary variational methods. A more fundamental approach is in order, one that goes to the roots of the situation.

Instead of the ordinary variation $\delta\varphi(x)$ we shall use the function

$$\psi(x)=\psi(x;\, x_1,x_2,h,k),$$

graphed as shown in Fig. 3. This function has the following properties:

1. $\psi(x)$ is continuous.

2. $\psi(x)=0$ outside the (non-overlapping) intervals (x_1-h,x_1+h), (x_2-h,x_2+h).

3. $\psi(x)>0$ in (x_1-h,x_1+h); $\psi(x)<0$ in (x_2-h,x_2+h).

4. The graphs in these intervals are *congruent*, so that $k=\max\psi(x)$ in $(x_1-h,x_1+h)=-\max\psi(x)$ in (x_2-h,x_2+h) and

$$\int_{x_i-h}^{x_i+h}\psi(x)\,dx=-\int_{x_2-h}^{x_2+n}\psi(x)\,dx.$$

We wish now to obtain necessary conditions that the answer, the optimum $\varphi(x)$, must satisfy.*

Let x_2 be any point where $\varphi(x_2)>0$, and let x_1 be any other point. By the continuity of $\varphi(x)$, two positive constants, h, k may be found such that $\varphi(x)>k$ for all $x_2-h<x<x_2+h$. Furthermore, suppose h so small that the intervals $(x_1-h,\ x_1+h)$ and (x_2-h, x_2+h) do not overlap.

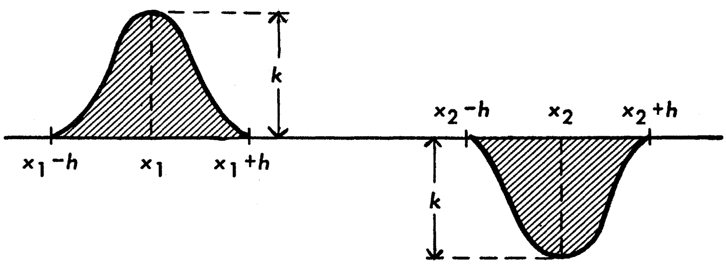

Figure 3

Using the above $\psi(x)=\psi(x;\,x_1,x_2,h,k)$, consider the function

$$\varphi(x)+t\,\psi(x),$$

where $0\leqq t\leqq 1$.

Clearly this function satisfies, as well as $\varphi(x)$, the conditions (1) above. If, then, $\varphi(x)$ is to maximize $P[\varphi]$, we must have

$$P[\varphi]-P[\varphi+t\psi]\geqq 0.$$

[Note that $\varphi+t\psi$ may be described as the result of transferring some of the searching (an available amount) from the neighborhood of x_2 to that of x_1). We have evidently (omitting all infinite limits of integration from the symbols)]

$$P(\varphi)-P(\varphi+t\psi)=\int p(x)\, e^{-\varphi(x)}\, [e^{-t\psi(x)}-1]\, dx$$

$$=-t\int p(x)\, e^{-\varphi(x)}\, \psi(x)\, dx+t^2\int p(x)\, e^{-\varphi(x)}\, G(x,t)\, dx,$$

where $G(x,t)$ represents the sum of high order terms in the expansion of $e^{-t\psi(x)}$.

Suppose that $\int p(x)\, e^{-\varphi(x)}\, \psi(x)\, dx>0$. Clearly, by taking t positive and sufficiently small, the right-hand members of the above relation can

* It is tacitly assumed that $\varphi(x)$ is continuous except for a few possible simple infinities and finite jumps (and is continuous at x_1, x_2). This is uniquely to maintain the simple elementary level of the discussion. Anyone familiar with modern measure theory will see at once how to alter the treatment in more general cases.

be made *negative*, contrary to what was seen before. We conclude that

$$\int p(x)\, e^{-\varphi(x)}\, \psi(x)\, dx \leqq 0.$$

Using the properties of $\psi(x)$, this gives

$$\int_{x_1-h}^{x_1+h} p(x)\, e^{-\varphi(x)}\, \psi(x)\, dx + \int_{x_2-h}^{x_2+h} p(x)\, e^{-\varphi(x)}\, \psi(x)\, dx \leqq 0.$$

Applying the law of the mean:

$$2h\, p(x')\, e^{-\varphi(x')}\, \psi(x') + 2h\, p(x'')\, e^{-\varphi(x'')}\, \psi(x'') \leqq 0,$$

where x' and x'' are in the intervals (x_1-h, x_1+h) and (x_2-h, x_2+h), respectively. Clearly, for all smaller h than the one here used, all these relations continue to hold, so we may let $h \rightarrow 0$ in these formulas. Then $x' \rightarrow x_1$ and $x'' \rightarrow x_2$ and

$$p(x_1)\, e^{-\varphi(x_1)}\, \psi(x_1) + p(x_2)\, e^{-\varphi(x_2)}\, \psi(x_2) \leqq 0.$$

But $\psi(x_1) = k$, $\psi(x_2) = -k$; hence, finally

$$p(x_1)\, e^{-\varphi(x_1)} \leqq p(x_2)\, e^{-\varphi(x_2)}. \tag{3}$$

The only special assumption* made in order to derive (3) was that

$$\varphi(x_2) > 0. \tag{4}$$

We shall derive a sequence of conclusions from the fact just established, namely, that (4) implies (3). For convenience of statement, let A be the set of points x for which $\varphi(x) > 0$:

$$A = \{x : \varphi(x) > 0\}.$$

Also let $\bar{A}$ be the set of points x for which $\varphi(x) \leqq 0$ (i.e., $=0$):

$$\bar{A} = \{x : \varphi(x) \leqq 0\} = \{x : \varphi(x) = 0\}.$$

Obviously A and $\bar{A}$ are complementary: no point is in both; every point is in one or the other.

Conclusion (a). If x is in A, $p(x)\, e^{-\varphi(x)} = \lambda$ (constant). For if x_1 and x_2 are in A, so that $\varphi(x_1) > 0$ and $\varphi(x_2) > 0$, we have both (3) and the same with x_1 and x_2 interchanged; hence, finally,

$$p(x_1)\, e^{-\varphi(x_1)} = p(x_2)\, e^{-\varphi(x_2)}.$$

* Other than the assumption of continuity at x_1, x_2. Clearly the same proof applies if $\varphi(x)$ is infinite at one or both points. And the results are extended to finite jump discontinuities by one-sided approach to the limit. These possibilities may require a trivial re-definition of $\varphi(x)$, $p(x)$, at jump-points. All this is without practical effect and will not be mentioned further.

In other words, $p(x)\,e^{-\varphi(x)}$ has the same value at all points of A; λ is *defined* as this value.

Conclusion (*b*). *If x is in A, $p(x)>\lambda$; if x is in $\bar{A}$, $p(x)\leqq\lambda$.* The first follows from $p(x)\,e^{-\varphi(x)}=\lambda$, so that $p(x)=\lambda\,e^{\varphi(x)}>\lambda$ (since $\varphi(x)>0$). The second follows from (3), with $x=x_1$ (so that $\varphi(x)=0$), and with x_2 any point of A.

Conclusion (*c*). *A is the set of points x for which $p(x)>\lambda$: $A=\{x:p(x)>\lambda\}$. $\bar{A}$ is the set of points x for which $p(x)\leqq\lambda$: $\bar{A}=\{x:p(x)\leqq\lambda\}$.* This is an immediate consequence of conclusion (*b*).

Conclusion (*d*). *$\lambda>0$ and the length of A (denoted by A) is finite.* From conclusion (*a*), $\lambda=p(x)\,e^{-\varphi(x)}\geqq 0$. If $\lambda=0$, $p(x)=0$ for all x in A; and by conclusion (*c*), $p(x)\leqq\lambda=0$ for all other x. Hence we should have $p(x)=0$ for all x, contrary to requirement $\int p(x)\,dx=1$. Therefore $\lambda>0$. It follows that $1=\int p(x)\,dx\geqq\int_A p(x)\,dx>\int_A\lambda\,dx=A\lambda$; therefore $A<1/\lambda<\infty$.

We can now calculate the function $\varphi(x)$: For x in A, conclusion (*a*) gives $p(x)\,e^{-\varphi(x)}=\lambda$ so that

$$\varphi(x)=\log p(x)-\log\lambda\,. \tag{5}$$

Now integrate through, using (1):

$$\Phi=\int\varphi(x)\,dx=\int_A\varphi(x)\,dx=\int_A[\log p(x)-\log\lambda]\,dx,$$

i.e.,

$$\int_A[\log p(x)-\log\lambda]\,dx=\Phi. \tag{6}$$

From this (since $\lambda>0$ and $A<\infty$), we find the value of $\log\lambda$ and, introducing this value into (5) we obtain

$$\varphi(x)=\log p(x)-\left[1/A\int_A\log p(x)\,dx-\Phi/A.\right]$$

If λ were known, then A would be also, since, from conclusion (*c*), $A=\{x:p(x)>\lambda\}$, and $p(x)$ is known. To find λ, we must solve (6). To examine this possibility we proceed graphically, plotting the curve $y=\log p(x)$ as in Fig. 1, and reasoning from the formulas (5), (6), (7) to the geometrical constructions and interpretations given in the solution described at the outset. The remainder of this section is but a step by step verification of these statements.

Denote by u any positive variable, and consider the set $A(u)$ of points [interval(s)] on the x-axis for which $p(x)>u$, or equivalently $\log p(x)>\log u$; for the length of $A(u)$ we have

$$A(u)=\{x:p(x)>u\}=\{x:\log p(x)>\log u\}.$$

It is like $A=(A_1, A_2,$ etc.) in Fig. 1. In fact, $A=A(\lambda)$ and is the projection on the x-axis of the region $S(u)$ under the curve $y=\log p(x)$ and above the horizontal line $y=u$. Denoting this area also by $S(u)$, we have $S(u)\leqq 1/u$ as can be seen as follows:

$$\begin{aligned}S(u)&=\int_{A(u)}[\log p(x)-\log u]\,dx\\&=\int_{A(u)}\log\frac{p(x)}{u}\,dx\leqq\int_{A(u)}\frac{p(x)}{u}\,dx\leqq\frac{1}{u}\int p(x)\,dx=\frac{1}{u}.\end{aligned}$$

We have used the fact that for x in $A(u)$, $p(x)/u$ is greater than unity and hence is greater than its logarithm. This inequality shows that $S(u)$ is finite and approaches zero as $u\rightarrow+\infty$. It follows at once that the length $A(u)$ is finite for $u>0$, since

$$1=\int p(x)\,dx\geqq\int_{A(u)}p(x)\,dx\geqq\int_{A(u)}u\,dx=u\,A(u)\,.$$

I say that $S(u)\rightarrow+\infty$ as $u\rightarrow 0$. To show this, note that if $u<u'$, $A(u)$ contains $A(u')$ (*cf.* the definition of these sets); consequently $A(u)$ and hence its length *grow* as *u decreases.* There must be some u_0 for which $A(u_0)>0$; otherwise $p(x)$ would differ from zero only on a set of zero length (measure), so that $\int p(x)\,dx=0\neq 1$. For such a u_0 and any smaller u we have

$$\begin{aligned}S(u)&=\int_{A(u)}[\log p(x)-\log u]\,dx\\&\geqq\int_{A(u_0)}[\log p(x)-\log u]\,dx\\&\geqq\int_{A(u_0)}[\log u_0-\log u]\,dx=A(u_0)\,\log\frac{u_0}{u}.\end{aligned}$$

Since the last expression approaches $+\infty$ as $u\rightarrow 0$, the same is true of $S(u)$, as was to be proved.

Finally, $S(u)$ is a continuous, monotonically decreasing function of u. For $u<u'$,

$$\begin{aligned}S(u)&=\int_{A(u)}\log\frac{p(x)}{u}\,dx\\&\geqq\int_{A(u')}\log\frac{p(x)}{u}\,dx>\int_{A(u')}\log\frac{p(x)}{u'}\,dx=S(u'),\end{aligned}$$

and the inequality $S(u) > S(u')$ gives the monotonicity. To prove the continuity, write

$$S(u)-S(u')=\int_{A(u)}\log\frac{p(x)}{u}\,dx-\int_{A(u')}\log\frac{p(x)}{u'}\,dx$$

$$=\int_{A(u')}\log\frac{p(x)}{u}\,dx-\int_{A(u')}\log\frac{p(x)}{u'}\,dx+\int_{A(u)-A(u')}\log\frac{p(x)}{u}\,dx$$

$$=\int_{A(u')}\log\frac{u}{u'}\,dx+\int_{A(u)-A(u')}\log\frac{u'}{u}\,dx=A(u)\log\frac{u'}{u}\,.$$

As $u \to u'$ or as $u' \to u$, $\log(u'/u) \to 0$, and hence the nonnegative

$$S(u)-S(u') \to 0.$$

To sum up, $S(u)$ decreases continuously and monotonically from $+\infty$ to 0 as u increases from 0 to $+\infty$. *Hence $S(u)$ takes on every given positive value such as Φ once and only once.* If the unique u for which $S(u)=\Phi$ is denoted by λ, and the corresponding $A(\lambda)$ by A, it is seen that (6) has a unique solution λ, A, and that it is given by the graphical construction set forth in the steps 1, 2, and 3 of the solution given at the outset. The remaining facts, incorporated into steps 4 and 5, are proved by an elementary application of Bayes' theorem to the material just derived—by the simple process of algebraic substitutions!

GENERALIZATIONS

THE FIRST order of generalizations keep the problem one-dimensional (on the x-axis) but consider modified conditions and objectives.

First, let the requirement of homogeneity be waved: suppose that the visibility varies from position to position. As can be seen in the random search formula (l.c., Part II), this means that $\varphi(x)$ in the exponential must be replaced by a product such as $g(x)\varphi(x)$, where $g(x)$ is a measure of the ease with which a unit of search conducted near x detects a target there. In the model given in our second footnote, $\varphi(x)$ could be used to denote n/R (the glimpse density) and $g(x)$ the quantity $2r$ (the double detection radius, a measure of visibility). Whatever its physical interpretation, the function $g(x)$ is to be regarded as *given*, while $\varphi(x)$ must be found [subject to (1)] so as to maximize

$$P[\varphi]=\int p(x)[1-e^{-g(x)\,\varphi(x)}]\,dx. \qquad (2')$$

Second, suppose that the *value* or usefulness of finding the target depends greatly on *where* it is found, and that it is this value that we wish to maximize. Under many conditions, this leads to the replacement of $p(x)$ by $v(x)\,p(x)$, where $v(x)$ is a weighting factor. But this makes a difference in the interpretation only and not in the mathematics.

As a further extension, suppose that a given amount of searching *costs more* (in some appropriate effort measurement terms) at some positions than at others. If, for instance, $\varphi(x)\,dx$ units of search at x cost an effort $h(x)\,\varphi(x)\,dx$, then the requirement that the total amount of effort be Φ (given) takes the form of the equation

$$\varphi(x) \geqq 0, \qquad \int h(x)\,\varphi(x)\,dx = \Phi, \tag{1'}$$

which now replaces (1). Here $h(x)$ is supposed known.

For example, suppose that the searcher has an unlimited *time* for search, but is constantly exposed to a *danger*, depending on position x. If he searches an amount $\varphi(x)\,dx$ near x, his chance of being destroyed is $h(x)\,\varphi(x)\,dx$. Let the given quantity Φ represent, not the total available effort, but the total permissible risk. An interesting case would be that in which the risk is that of detection by the enemy, whose search of us also obeys the law of random search, so that $h(x) = 1 - e^{-p(x)}$, or some obvious modification. A game-theoretic play with $\varphi(x)$ versus $p(x)$ expresses a sort of hide-and-seek, the results of which can be worked out with a fair degree of explicitness.

The second order of generalization is that in which the original problem—or any of the above generalizations—is set in a region of more than one dimension. In cases of interest to the Army and the Navy, search in a plane (or spherical) surface is of frequent concern. The Air Force is often occupied with three-dimensional search. But the 'space,' which is the seat of the searching operation, need not be that of geometry. In adjusting an instrument, it may be the space of adjustment parameters. In picking up an electromagnetic signal, it can be a space of frequencies and polarizations. There could be 'search' in phase space (position-momentum space) in certain physical situations; and so forth.

Moreover, the basic quantitative ideas and methods of this paper need not be confined to the case of *search*: Any other form of *effort*, provided merely that its pay-off is additive and that each unit is expressible by the exponential saturation formula, as above, is an appropriate field of application of the present material. Some applications in World War II were made to the optimum distribution of destructive effort; e.g., in bombardment.

The formulation of the problem in all these cases is easily indicated: we use (1) and (2), or else (1′) and (2′), with the symbol x standing for a point in the relevant space $x = (x_1, \cdots, x_n)$ and $\int f(x)\,dx$ standing for n-fold integration:

$$\int f(x)\,dx = \int \cdots \int f(x_1, x_2, \cdots, x_n)\,dx_1\,dx_2 \cdots dx_n.$$

EXTENSION OF THE MATHEMATICAL TREATMENT

THE MATHEMATICAL methods for solving all these generalizations of the problem are exactly the same, apart from obvious straight-forward modifications, as in the case of the original problem which we have treated at length. It is however possible to make a more direct transference, by changes of independent variables and of dependent variables, of the *results* just secured. In some but not in all cases, this represents the preferred practical method. And it always has the theoretical advantage of emphasizing the basic unity of the whole subject.

Before applying either method, however, certain general preliminary matters should be examined, bearing on the possibility that there be regions of vanishing of $g(x)$ or of $h(x)$.

Let G be the set of points x at which $g(x)>0$, and $\bar{G}$ its complement, at every point of which $g(x)=0$ [$g(x)$ is nonnegative!]. Similarly, let H be the set of points x at which $h(x)>0$, and $\bar{H}$ its complement, throughout which $h(x)=0$. Then clearly, in (1′) and (2′),

$$\Phi=\int h(x)\,\varphi(x)\,dx=\int_H h(x)\,\varphi(x)\,dx,$$

$$P[\varphi]=\int p(x)[1-e^{-g(x)\,\varphi(x)}]\,dx=\int_G p(x)[1-e^{-g(x)\,\varphi(x)}]\,dx.$$

The first equation shows that all effort expended in $\bar{H}$ *costs nothing*; the second shows that all effort expended in $\bar{G}$ is a *complete waste.* Let us therefore confine our effort to G, and expend infinitely great effort in those parts that also belong to $\bar{H}$ (denote their intersection by $G\bar{H}$), and as much as we can afford, subject to the limited Φ, to the intersection GH of G with H. Then we obtain

$$P[\varphi]=\int_{GH} p(x)[1-e^{-g(x)\,\varphi(x)}]\,dx+\int_{G\bar{H}} p(x)\,dx,$$

since, when $\varphi(x)=\infty$, $1-e^{-g(x)\,\varphi(x)}=1-0=1$. Our problem reduces itself, therefore, to choosing $\varphi(x)$ so as to maximize

$$P'[\varphi]=\int_{GH} p(x)[1-e^{-g(x)\,\varphi(x)}]\,dx,$$

subject to the conditions

$$\varphi(x)\geqq 0, \qquad \int_{GH} h(x)\,\varphi(x)\,dx=\Phi.$$

Thus, we have a similar problem to the original one, but with the region GH, at which $g(x)$ and $h(x)$ are both positive, used instead of the unrestricted region.

On the basis of these considerations, the method of change of variables can be given in the case in which $g(x)$ and $h(x)$ are both positive. We shall illustrate it by taking the one-dimensional case, and assuming for simplicity in exposition that $g(x)$ and $h(x)$, in addition to being positive, are continuous and have the following boundedness properties on any finite interval of the x-axis:

$$0<g_1\leqq g(x)\leqq g_2<\infty; \qquad 0<h_1<h(x)\leqq h_2<\infty.$$

Now introduce the new function $\varphi_1(x)=g(x)\varphi(x)$. Clearly its knowledge gives that of $\varphi(x)$, since $\varphi(x)=\varphi_1(x)/g(x)$ and $g(x)\neq 0$. Also, the conditions in (1′) become

$$\varphi_1(x)\geqq 0; \qquad \int [h(x)/g(x)]\,\varphi_1(x)\,dx=\Phi.$$

Now in any finite interval containing the origin, $h(x)/g(x)$ is continuous. Hence, setting

$$y=\int_0^x [h(x)/g(x)]\,dx=f(x),$$

we have defined y as a continuous function $f(x)$ of x, having a positive continuous derivative. Hence the change of variable formalism applies; and on writing, as the inverse,

$$y=f(x), \qquad x=f_1(y), \qquad \varphi'(y)=\varphi_1[f_1(y)],$$

(1′) becomes
$$\varphi'(y)\geqq 0; \qquad \int \varphi'(y)\,dy=\Phi.$$

As for (2′), it becomes

$$P'[\varphi']=\int p(x)[1-e^{-\varphi_1(x)}]\,dx=\int p'(y)[1-e^{-\varphi'(y)}]\,dy,$$

where we have set $p'(y)=p(x)\,[g(x)/h(x)]$ for $x=f_1(y)$ so that

$$p'(y)\,dy=p(x)\,dx.$$

Thus, the general problem (1′), (2′) has been reduced to the original problem (1), (2), in the new variables.

In the case of more than one dimension, changes of variables with preassigned jacobian determinants have to be applied. The theory remains simple, but the computation may or may not.

CONCLUDING REMARKS

THE CLASS of problems considered herein is more general than the problems of linear programming, since the expression to be optimized is nonlinear and involves integration; it is more special, since only two linear side-conditions are given. If the number of such conditions were increased, much of what has been done here could be extended, although not without going drastically beyond the present paper. The same is true, with even greater difficulties, if the detection law (or pay-off function) is not of the simple exponential type assumed. For all these extensions, only the general approach of the unilateral variational schemata remains. This approach, while familiar through its use in many similar situations in the present period, actually goes back to the work of Willard Gibbs, who applied it (in the case of finite sums, rather than integrals) in his theory of the equilibrium of heterogeneous substances, an epochal work of the last century.

But a much more fundamental question is in order: When can the present methods, or anything like them using the same general approach, have any hope of being applicable? The answer is that the pre-conditions of the present type of approach are that the pay-off P should be *additive* in the separate portions of effort (so that it can be expressed as an integral, in some sufficiently general sense); and, furthermore, that the separate returns of the portions of effort be functions of these portions, together with the local conditions.

The easiest way to see how very special these cases are, even among simple practical problems, is to consider the familiar case in which either of two types of 'gadget' A or B (e.g., weapons) can perform similar functions, and suppose their unit costs, C_A and C_B, are known. What is their cost/effectiveness relation? How does the cost of doing a given amount of the job with one compare with doing it with the other? This is a very familiar (and often simple) problem for operations research; but it can occur in a far from simple form: In many cases of crucial practical importance, the effectiveness comparison of A versus B—i.e., the marginal increment of effectiveness when the addition of a unit of A or of B to the system is considered—*depends on the previous make-up of the system.* Then the whole notion of cost/effectiveness in its familiar form falls to the ground. One is easily led to problems more complicated by an order of magnitude than the familiar ones of operations research.