CRC-357

LEVEL

AD A063845

A NEW LOOK AT THE RELATION BETWEEN INFORMATION THEORY AND SEARCH THEORY

CENTER FOR NAVAL ANALYSES
1401 Wilson Boulevard
Arlington, Virginia 22209

Exploratory Research Division

N00014-76-C-0001

By: John G. Pierce

June 1978

DDC FILE COPY

Prepared for:

OFFICE OF NAVAL RESEARCH
Department of the Navy
Arlington, Virginia 22217

OFFICE OF THE CHIEF OF NAVAL OPERATIONS (Op96)
Department of the Navy
Washington, D.C. 20350

02 035700.00

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: RC 357	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): A New Look at the Relation Between Information Theory and Search Theory	5. TYPE OF REPORT & PERIOD COVERED
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): John G. Pierce	8. CONTRACT OR GRANT NUMBER(s): N00014-76-C-0001
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Center for Naval Analyses, 1401 Wilson Boulevard, Arlington, Virginia 22209	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Department of the Navy, Arlington, Virginia 22217	12. REPORT DATE: June 1978
	13. NUMBER OF PAGES: 62
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Office of the Chief of Naval Operations (Op96), Department of the Navy, Washington, D.C. 20350	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

This Research Contribution does not necessarily represent the opinion of the Department of the Navy.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

detection, entropy, exponential detection functions, information theory, probability, search theory, target detection

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Some apparent contradictions in the operations research literature of search theory and information theory are reviewed, and the source of the conflict is isolated. The information processes connected with a search operation are examined by studying the geometric properties of information surfaces in search-allocation space. Some new theorems are developed which show that the connection between search and information theory is much stronger than previously believed.

It is demonstrated that the amount of search effort, C, is a critical parameter

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20 in the relation between search and information gain. For very small values of C, there is no unique relation between detection probability and information gain; for very large values of C, the optimal whereabouts search policy produces the greatest information gain; for a broad intermediate range of C (including many cases of practical interest) the optimal detection search policy produces the greatest information gain.

Center for Naval Analyses

an affiliate of the University of Rochester

1401 Wilson Boulevard *Arlington, Virginia 22209* *703/524-9400*

An Equal Opportunity Employer

18 December 1978

MEMORANDUM FOR DISTRIBUTION LIST

Subj: Center for Naval Analyses Research Contribution 357

Encl: (1) CRC 357, "A New Look at the Relation Between Information Theory and Search Theory," by John G. Pierce, June 1978

1. Enclosure (1) is forwarded as a matter of possible interest.

2. Research Contributions are distributed for their potential value in other studies and analyses. They do not necessarily represent the opinion of the Department of the Navy.

Andrew P. Borden

ANDREW P. BORDEN
Executive Vice President
and Acting President

Distribution List:
Reverse page

Subj: Center for Naval Analyses Research Contribution 357

DISTRIBUTION LIST

Department of the Navy

SNDL Part I:

21A1	CINCLANTFLT
21A2	CINCPACFLT
21A3	CINCUSNAVEUR
26F1	COMOPTEVFOR
26F2	DEPCOMOPTEVFORPAC

SNDL Part II:

A1	ASST SECNAV (RE&S)
A2A	ONR
B2	DNA
B2	DCA
B2	DIA
B3	NDU
B3	AFSC
E3A	NRL
FF38	USNA, Nimitz Library
FKA6A1	NAVAIRDEVCEN
FKA6A2	NAVWPNCEN
FKA6A3A	DTNSRDC
FKA6A3B	NAVCOASTSYSLAB
FKA6A9	NAVSURFWPNCEN
FKA6A12	NAVOCEANSYSCEN
FKA6A15	NUSC NPT
FT73	NAVPGSCOL
FT73	NAVPGSCOL (Prof. Forrest)
FT75	NAVWARCOL

OpNav:

Op-09BH

Op-96

Other

Under Secretary of Defense, Research and Engineering
Ass't Secretary of Defense, Program Analysis & Evaluation (2 copies)
Defense Advanced Research Projects Agency, Attn: Cybernetics Technology (2 copies)
Defense Documentation Center (12 copies)
Commandant, Army War College
Department of the Army (Adj Gen'l)(6 copies)
Air University
Department of the Air Force (AFXOD)
Central Intelligence Agency
National Science Foundation
Aerospace Corporation
Analytic Services, Inc.
Applied Physics Laboratory/University of Washington
Applied Physics Laboratory/Johns Hopkins University
Applied Research Laboratory
Calspan Corporation
Electromagnetic Compatability Analysis Center
Institute for Defense Analyses
Human Resource Research Organization
Jet Propulsion Laboratory
Lincoln Laboratory
Logistics Management Institute
Mathematics Research Center
The Mitre Corporation
The Rand Corporation
Sandia Laboratories
Scripps Institute of Oceanography
Stanford Research Institute
System Development Corporation

TRW Defense and Space Systems Group
Woods Hole Oceanographic Institute
National Academy of Sciences
Arthur D. Little, Inc
Arthur D. Little, Inc (Attn: Dr. B. Koopman)
J. Wagner Assoc. (Attn: Dr. Stone)
J. Wagner Assoc. (Attn: Dr. Barker)
Ketron (10 copies)

ACKNOWLEDGEMENT

The author is indebted to Dr. L.D. Stone of Daniel H. Wagner Associates, and Dr. R.W. Johnson of the Naval Research Laboratory for careful review of this paper and for helpful comments.

TABLE OF CONTENTS

1. INTRODUCTION

Background

The relation between search theory and information theory has had a thorny history. A 1967 view was summarized concisely by Koopman (reference 1): "Ever since the mid-nineteen-forties when the theories of information and of search became subjects of general interest, attempts have been made to apply the theory of information to problems of search. These have proved disappointing; neither the formulas nor the concepts of the former theory have found a place in clarifying the problems of the latter." These views doubtless reflected much of Koopman's own research on the subject.

An independent contribution to the generally pessimistic view of the matter was the 1961 paper by Mela (reference 2). He presented numerical examples for simple search models, demonstrating that a search policy that is designed to maximize detection probability does not necessarily maximize either the gain in expected information or the probability of correctly committing forces on the basis of search outcomes. Mela concluded: "...it does not seem likely that there is any intimate connection between search theory and information theory." And: "...search theory should be considered in connection with the general theory of statistical decisions rather than with information theory."

This negative view was reinforced by a later paper by Pollock (reference 3) in 1971. He presented yet another numerical example, in which different search policies were necessary to maximize either: the probability of detection as a result of the search, or the probability of guessing the position of the target after the search, or the information gained during the search.

These negative findings had a clearly inhibiting effect on research, and relatively little effort has been devoted to the connections between information and search for the past fifteen years. Nonetheless, the intuitive appeal of information theory remains strong, and more recently the tide of pessimism has been stemmed somewhat. In 1973, for instance, Richardson (reference 4) used Monte Carlo simulation to explore alternative surveillance policies in a false-target environment. Target motion based on a Markov process was also included. Richardson's policy options included: (a) an optimal single-stage look-ahead policy (about the same as maximizing the probability of a right guess after every search stage); (b) a policy to maximize information gain; (c) a policy to search the cell with the highest a priori probability; (d) a uniform surveillance policy (cycling systematically through all cells).

Richardson's results showed that the maximum-information-gain policy was generally (but not uniformly) best. In his summary, he said: "The principal conclusion...is that the maximum information gain policy...appears to have desirable characteristics in the idealized surveillance scenario considered. Among these characteristics...are good initial behavior in the early stages, and good asymptotic behavior in the later stages. The initial behavior is measured principally by comparison with the optimal single-stage look-ahead policy. . . which is designed to be good in the early stages. The asymptotic behavior is measured principally by comparison with the uniform surveillance policy, which, for a stationary target, is guaranteed to converge to 1. . . ."

Finally, inspired by Richardson's simulation findings, Barker (reference 5) proved these two theorems analytically:

> "I: Suppose that the detection function is given by $b(u) = 1 - e^{-u}$ for $u \geq 0$. Then, subject to a constraint on total search effort, the allocation of search effort which maximizes the probability of detection also maximizes the entropy of the posterior distribution."

And the converse:

> "II: Let b be a regular detection function. Suppose that for every prior target location distribution and for every constraint on total search effort, any allocation of search effort which maximizes the probability of detection also maximizes the entropy of the posterior distribution. Then, $b(u) = 1 - e^{-au}$ for some positive constant a."

These findings, taken as a whole, appear to be inconsistent. In particular, Mela's numerical example seems, at least superficially, to be in clear contradiction to Barker's theorems.

If science teaches us anything, it should be that where contradictions abound, there lies a fertile field for research. Accordingly, this paper reexamines the relation of search theory to information theory and and attempts to reconcile past results, to clarify the current state of knowledge, and to remove some impediments to further work on the subject.

Plan of the Paper

Section 2 establishes the mathematical framework of the problem. A model for the search process is constructed. A description of the problem in terms of information theory is provided, and distinctions are drawn among several possible definitions of information. These are then related to quantities discussed in the earlier literature.

Section 3 reexamines the examples of Mela and Pollock and extends their results numerically to the case of a large amount of search effort. Section 4 treats analytically the case when a large search effort is applied; section 5 treats analytically the case in which the search effort is infinitesimal.

Section 6 summarizes the findings and places the results of this paper in a framework with past work.

2. PROBLEM STATEMENT

The Search Model

We assume a simple model for the search process. A single target is in one of J cells. Our prior knowledge of its position is described by a set of probabilities, p_j, $j \in J$, $\sum_{j=1}^{J} p_j = 1$. We conduct a search operation in which various amounts of search effort, z_j, are applied to each of the J cells. The total amount of search effort, C, is fixed, so that the z_j amounts of search are subject to a constraint:

$$C = \sum_{j=1}^{J} z_j \quad . \tag{2.1}$$

For our general treatment, the z_j need not be quantized. Special cases, for which the search effort is applied in multiples of a fixed unit (the "look"), appear in the examples of Mela and Pollock.

We assume that the detection process is governed by an exponential detection function. This means that the conditional detection probability for cell j, given that the target is in cell j and that an amount of search effort, z_j, is applied there is:

$$\Pr(\det j | j, z_j) = 1 - \exp(-\alpha_j z_j) \quad . \tag{2.2}$$

α_j is a characteristic parameter of the detection process in cell j.

This formulation subsumes the discrete-look cases of Mela and Pollock. If q_j is the single-look conditional detection probability, then, for n_j looks in cell j:

$$\Pr(\det j | j, n_j) = 1 - (1-q_j)^{n_j} \quad , \tag{2.3}$$

which is equivalent to eq. 2.2 when n_j is identified with z_j, and α_j is identified with $-\ln(1-q_j)$.

Finally, we assume that no false alarms occur.

Within the framework of this basic model, a specific allocation of search effort may lead to J+1 possible outcomes: either a single detection in any of the J cells or no detections at all. Because we have assumed no false alarms, a detection in cell k produces a posterior probability distribution:

$$p_k' = 1; \; p_j' = 0, \; j \neq k \in J \; . \tag{2.4}$$

Failure to achieve detection in any cell produces a posterior distribution:

$$p_j' = p_j \; \exp(-\alpha_j z_j) \Big/ \left[\sum_{k=1}^{J} p_k \quad \exp(-\alpha_k z_k) \right] \; . \tag{2.5}$$

For convenience of notation, we define

$$\varphi_j = \varphi(z_j, \alpha_j, p_j) \equiv p_j \; \exp(-\alpha_j z_j) \; . \tag{2.6}$$

(These φ_j functions correspond to the "failure densities," $\beta(j)$, used by Barker.)

For nondetection, the posterior distribution is:

$$p_j' = \varphi_j / \sum_{k=1}^{J} \varphi_k \; . \tag{2.7}$$

The unconditional probability of detection in the jth cell is, in this notation:

$$PD_j = p_j \; (1 - \exp(-\alpha_j z_j)) = p_j - \varphi_j \; , \tag{2.8}$$

and the overall detection probability is:

$$PD = \sum_{j=1}^{J} PD_j = 1 - \sum_{j=1}^{J} \varphi_j \; . \tag{2.9}$$

The Information Description

It is possible to define many "information" quantities related to search operations. Values for self information and mutual information, for conditional information and unconditional information, for information as a random variable and for ensemble average information -- any of these can be computed at each step of a search. To avoid ambiguity, we proceed with caution in defining the specific "information" we wish to discuss in this paper and in relating that definition to the quantities discussed in the references.

First, we define two probability spaces, X and Y , that are concerned, respectively, with the true position of the target and the observed response of the search sensor. Let the event x_j represent the presence of the target in cell j. The space X has dimension J, and the discrete event probabilities are given by $p_X(x_j) = p_j$, $j \in J$. Let the event y_0 be the null response to search, and let the event y_k represent a detection in cell k . The event probabilities in the Y space can be constructed from the appropriate probabilities in the X space, together with the conditional detection probabilities: $p_Y(y_k) = \sum_{j=1}^{J} p_{Y|X}(y_k|x_j)\, p_X(x_j)$. In our present example, the dimension of the Y space is J + 1; it could be larger if multiple detections or false alarms were considered.

Our principal concern in this paper is the ensemble average mutual information between the X and Y ensembles. By definition (reference 6), the mutual information between the events x_j and y_k is:

$$I_{jk} = \ln\left[(p_{X|Y}(x_j|y_k))/p_X(x_j)\right] \quad . \tag{2.10}$$

This quantity is a random variable in the product space X • Y. Its ensemble average is

$$I_E = \sum_{j,k} p_{X,Y}(x_j,y_k) \quad I_{jk} \; ,$$

$$= \sum_{j=1}^{J} \sum_{k=0}^{J} p_{X|Y}(x_j|y_k)\, p_Y(y_k) \;\ln\; (p_{X|Y}(x_j|y_k))/p_X(x_j) \quad . \tag{2.11}$$

To compute this ensemble average, we use the following relationships. Implicit in all of these is an allocation of search effort, $\{z_j\}$:

$$p_X(x_j) = p_j \;, \text{ the prior probabilities;} \tag{2.12}$$

$$p_Y(y_0) = \sum_{j=1}^{J} \varphi_j \;, \text{ from (2.9);} \tag{2.13}$$

$$p_Y(y_k) = p_k - \varphi_k \text{ , from (2.8); } \quad k = 1,\ldots,J \tag{2.14}$$

$$p_{X|Y}(x_j|y_0) = \varphi_j / \sum_{k=1}^{J} \varphi_k \; ; \tag{2.15}$$

$$p_{X|Y}(x_j|y_k) = \delta_{jk} \; . \qquad k = 1,\ldots,J \tag{2.16}$$

The latter two equations are derived by appropriate applications of Bayes' rule. When these are substituted into eq. 2.11, we obtain:

$$I_E = -\sum_{j=1}^{J} p_j \ln p_j - \left[-\sum_{j=1}^{J} \varphi_j \ln \varphi_j + \left(\sum_{k=1}^{J} \varphi_k\right) \ln\left(\sum_{\ell=1}^{J} \varphi_\ell\right)\right] \tag{2.17}$$

The first term is recognized as the entropy of the prior ensemble, X, which we will designate H_O. By a general result of information theory, the mutual information between two ensembles, X and Y , can be written:

$$I_E = H(X) - H(X|Y) \; . \tag{2.18}$$

Thus, we can identify the bracketed term in eq. 2.17 as the conditional entropy of X, given Y. For reasons discussed below, we designate this as H_E. Thus:

$$H_E = -\sum_{j=1}^{J} \varphi_j \ln \varphi_j + \left(\sum_{k=1}^{J} \varphi_k\right) \ln \left(\sum_{\ell=1}^{J} \varphi_\ell\right) \; . \tag{2.19}$$

The mutual information is then:

$$I_E = H_O - H_E \; . \tag{2.20}$$

For a given prior distribution, H_O is a constant; H_E, and consequently I_E , are functionals of the allocation of search effort $\{z_j\}$, as well as functions of the prior probabilities, p_j, and the set of detection parameters, α_j .

We will eventually wish to investigate the extrema of I_E, but H_E is more convenient to work with. We note, therefore, that because of the minus sign in eq. 2.20, a maximum of H_E will mean a minimum of I_E, and vice versa.

Throughout this paper, we will discuss the ensemble average mutual information, I_E and the conditional entropy of X given Y --- $H_E \equiv H(X|Y)$ --- almost interchangeably. The reader should therefore bear their relationship (eq. 2.20) in mind as he proceeds.

An alternative approach is to consider the self-entropy of the posterior ensemble. If detection does not occur, posterior probabilities p'_j are as given by eq. 2.5 or 2.7; we can, therefore, define a posterior self-entropy, conditioned on nondetection of the target, as:

$$H_{ND} \equiv -\sum_{j=1}^{J} p'_j \ln p'_j \tag{2.21}$$

$$= -\sum_{j=1}^{J} \left(\varphi_j / \sum_{k=1}^{J} \varphi_k\right) \ln \left(\varphi_j / \sum_{\ell=1}^{J} \varphi_\ell\right)$$

$$= \left[-\sum_j \varphi_j \ln \varphi_j + \left(\sum_k \varphi_k\right) \ln \left(\sum_\ell \varphi_\ell\right)\right] / \sum_{m=1}^{J} \varphi_m \tag{2.22}$$

Conversely, if a detection occurs (and by our assumption there are no false detections) eq. 2.4 holds for the posterior distribution, and the posterior entropy, conditioned on detection in any $k \in J$, is:

$$H_D = -p'_k \ln p'_k - \sum_{j \neq k} p'_j \ln p'_j \tag{2.23}$$

$$= -1 \ln 1 - (J-1)\ 0 \ln 0 \equiv 0 \ . \tag{2.24}$$

By computing these two conditional entropies with the appropriate probabilities of occurrence, we can obtain the expected entropy of the posterior ensemble:

$$H_E = P_D\, H_D + P_{ND}\, H_{ND} \tag{2.25}$$

$$= P_D\, H_D + (1-P_D)\, H_{ND} \ .$$

Using equations 2.9, 2.22, and 2.24, we find an expression for H_E that is identical to eq. 2.19. The two approaches are thus equivalent, and the expected entropy, H_E, defined in terms of the expectation over the posterior distribution, is the same as that arising from computing the average mutual information between the two ensembles. The latter is the more fundamental approach, however; it should be followed in more complex cases such as those involving false alarms.

Having defined our basic terms, we can now return to the references to see what, in fact, the previous authors have discussed. None of those authors recognized explicitly the possibilities for confusion arising from the existence of various information and entropy values. Each, however, was consistent in doing calculations with the specific value he felt to be appropriate. Both Mela and Pollock used I_E. Richardson did his calculation with H_E and inferred max I_E based on min H_E; despite the consistency of his calculations, he did not draw a careful enough distinction between H_E and H_{ND} in presenting his conclusions. Barker's theorems refer to H_{ND}.

Because Mela and Barker were discussing different quantities, no inference of direct contradiction can immediately be drawn. The ramifications of their findings will now be discussed.

3. THE EXAMPLES CITED BY MELA AND POLLOCK

The examples devised by Mela and Pollock have two features in common: Both consider discrete looks only, and both choose the total number of looks to be small--no more than the number of search cells, J. They differ in that Mela considers situations in which both the prior probabilities and the conditional detection probabilities are identical in all search cells; Pollock, on the other hand, treats the case in which both prior probabilities and conditional detection probabilities vary. Tables 1-3 list the particulars of their examples, expressed in the notation of this paper.

TABLE 1

MELA'S FIRST EXAMPLE

$$J = 2;$$

$$p_1 = p_2 = .5;$$

$$\alpha_1 = \alpha_2 = .693$$

SEARCH POLICY	P_D	$I_E = H_O - H_E$
(1,1)	.5*	.347
(2,0)	.375	.380*

The starred entries in the tables indicate the respective maxima of P_D and I_E for the various search policies. It is obvious that the same policy leads to maxima of P_D and I_E in Mela's second example, and that different policies are required to maximize P_D and I_E in Mela's first example and in Pollock's example. From this it is self-evident that a given search policy does not always maximize both detection probability and information gain at the same time.

What is not self-evident -- yet is often asserted -- is that this fact further implies that there is no necessary connection between the extrema of P_D and I_E. As we show below, this assertion is, in fact, false. The existence of counterexamples, based on a small number of policy options and a small amount of search effort, shows that the connection between the extrema of P_D and I_E is not universal. In a wide range of circumstances, however, the extrema of P_D and I_E do indeed depend on identical search policies.

TABLE 2

MELA'S SECOND EXAMPLE

$J = 3;$

$p_1 = p_2 = p_3 = 1/3;$

$\alpha_1 = \alpha_2 = \alpha_3 = .693$

SEARCH POLICY	P_D	$I_E = H_O - H_E$
(1,1,1)	.5*	.549*
(2,1,0)	.417	.541
(3,0,0)	.292	.478

TABLE 3

POLLOCK'S EXAMPLE

$J = 3;$

$p_1 = .1,\ p_2 = .3;\ p_3 = .6;$

$\alpha_1 = +\infty,\ \alpha_2 = .511,\ \alpha_3 = .357$

SEARCH POLICY	P_D	$I_E = H_O - H_E$
(1,0,0)	.1	.14*
(0,1,0)	.12	.07
(0,0,1)	.18*	.04

We examine these first by relaxing the limitation of a small amount of search effort and by examining P_D and I_E as surfaces in a (J-1)-dimensional search allocation space. Although Mela and Pollock do not deal with H_{ND}, we shall examine that quantity numerically also, as a basis for analysis relating to Barker's work.

For Mela's first example, we consider a large number of discrete looks, C, allocated between the two cells; m looks are applied in cell 1, and C-m in cell 2. We then calculate P_D, H_{ND}, and H_E as functions of m for $0 \leq m \leq C$. Analytically:

$$P_D = 1 - [2^{-m} + 2^{-(C-m)}]/2 \tag{3.1}$$

$$H_{ND} = \ln\{[2^{-m} + 2^{-(C-m)}]/2\} + \{[2^{-m} + 2^{-(C-m)}]/2\}^{-1}\{(\ln 2)/2\} \cdot \{(m+1)2^{-m} + (C-m+1)2^{-(C-m)}\} \tag{3.2}$$

$$H_E = \{[2^{-m} + 2^{-(C-m)}]/2\} \ln [2^{-m} + 2^{-(C-m)}] + [(\ln 2)/2] \cdot [m2^{-m} + (C-m)2^{-(C-m)}] . \tag{3.3}$$

These quantities are plotted in figures 1-3 for the case C=30 . H_{ND} is plotted on a linear scale (figure 2); P_D and H_E are plotted logarithmically, as $-\log(1-P_D)$ and $\log H_E$, respectively (figures 1 and 3).

The obvious content of figures 1-3 and their supporting computations is that each of the three quantities -- P_D, H_{ND}, and H_E -- has a maximum at the same position, m=C/2=15. Equal allocation of search effort to the two cells produces the maximum detection probability, in keeping with Mela's finding for C=2, and the maximum value of H_{ND} as predicted by Barker's first theorem. The surprising finding is that the expected information gain, I_E, shows a <u>minimum</u> (max H_E) for equal allocation of search effort. Though this is consistent with Mela's calculation, it could not be anticipated on the basis of previously published results.

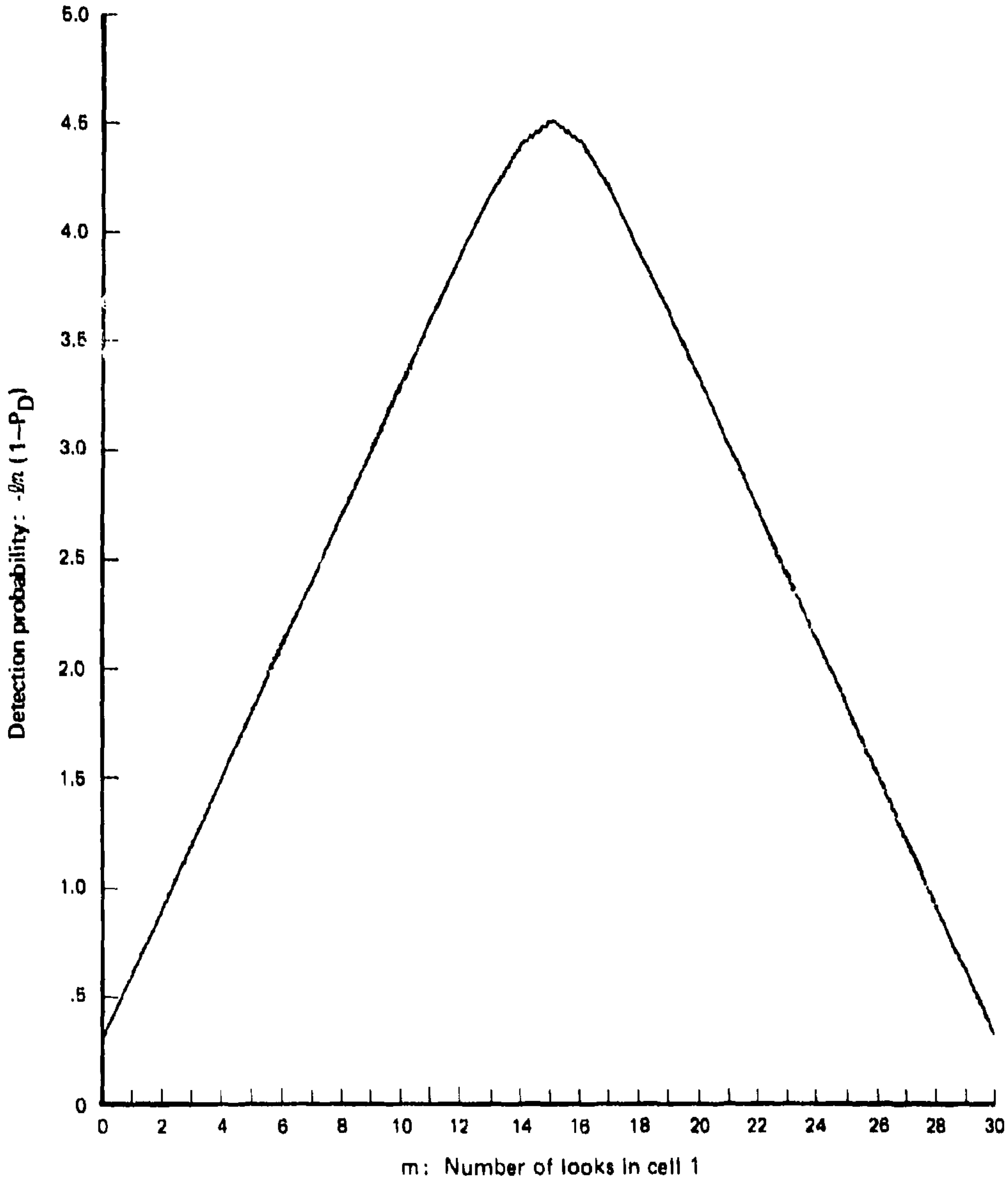

FIG. 1: MELA'S FIRST EXAMPLE – DETECTION PROBABILITY vs. SEARCH ALLOCATION

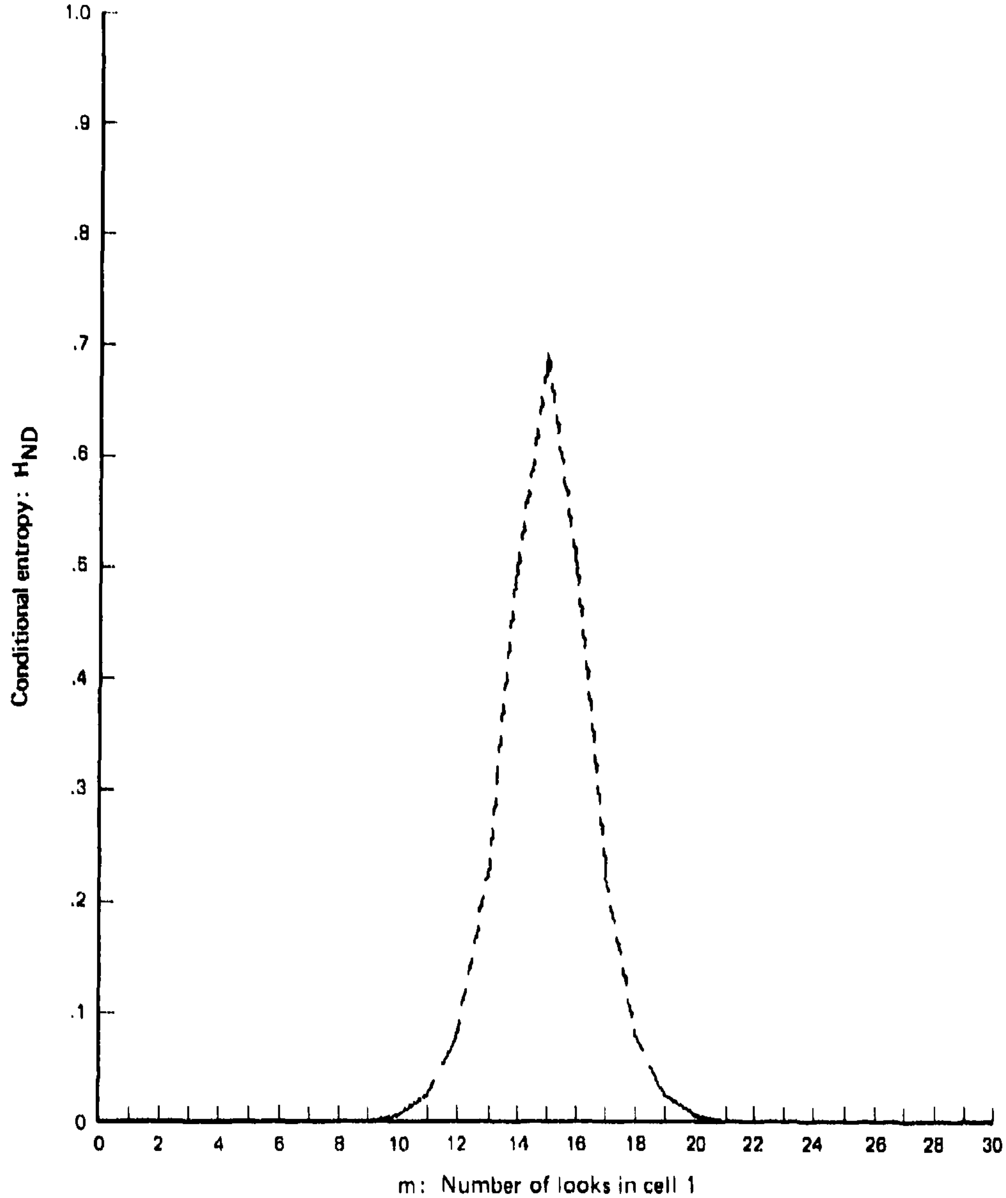

FIG. 2: MELA'S FIRST EXAMPLE – CONDITIONAL ENTROPY vs. SEARCH ALLOCATION

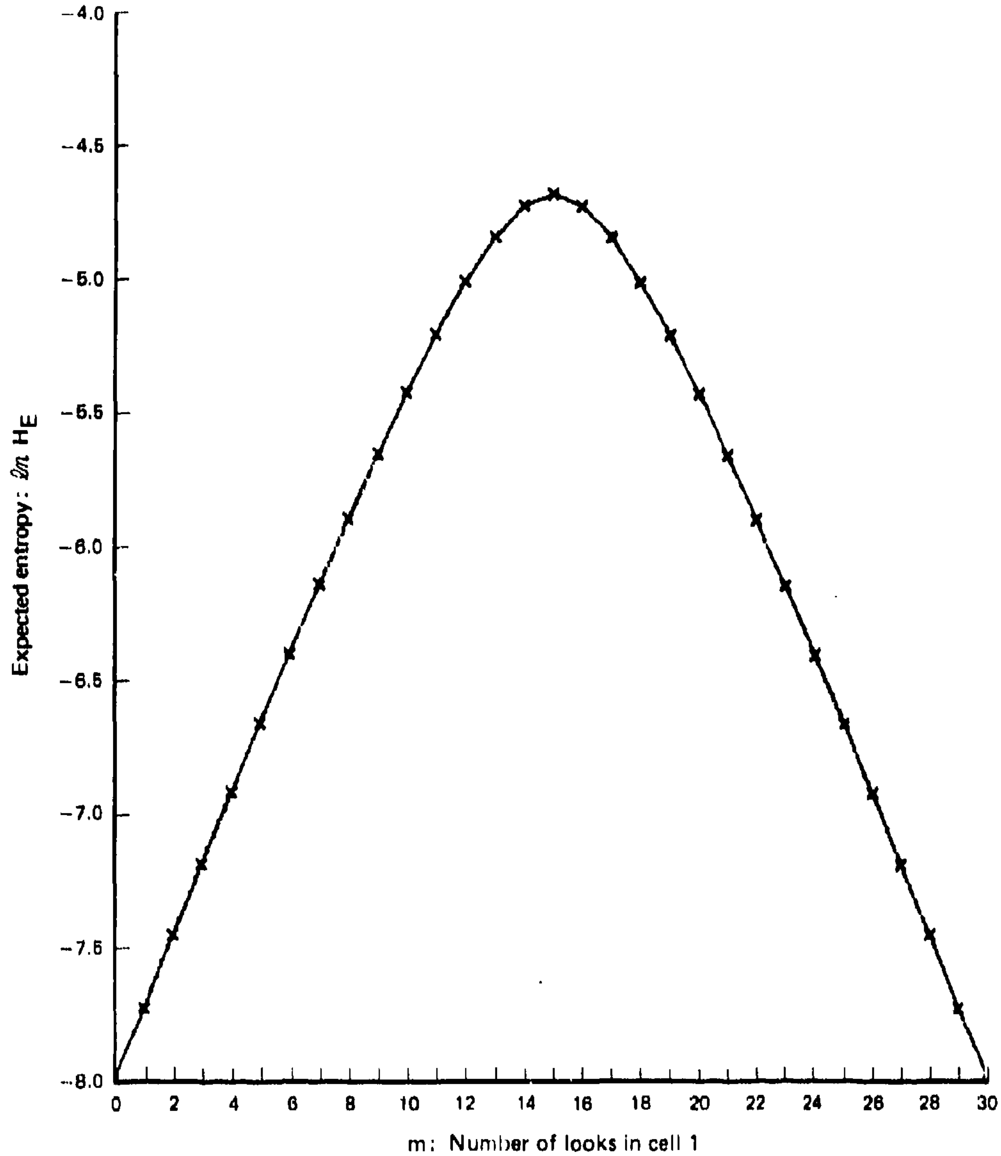

FIG. 3: MELA'S FIRST EXAMPLE – EXPECTED ENTROPY vs. SEARCH ALLOCATION

We use a similar approach to Mela's second example, with ℓ looks allocated to the first cell, m looks to the second, and C-ℓ-m to the third. The expressions for P_D (ℓ,m), H_{ND} (ℓ,m), and H_E (ℓ, m) are completely analogous to those of equations 3.1 - 3.3 (but too unwieldy to reproduce here). Figures 4 - 6 show the three-dimensional plots of P_D, H_{nd}, and H_E, for C=30, $0 \leq m, \ell \leq 30$, $0 \leq m+\ell \leq 30$. As before, P_D and H_E are plotted logarithmically; the H_{ND} scale is linear.

Figures 4 - 6 give a general idea of the shape of the surfaces. The perspective of the figures is so oblique, however, that precise conclusions must rest on the back-up computations. These say that the uniform search allocation (10, 10, 10) produces the maximum of P_D and H_{ND}, and the minimum of H_E (maximum I_E). The maxima of P_D and H_{ND} are both local and global maxima. But the minimum of H_E is a local minimum only. Smaller values occur on the boundaries, at (15, 15, 0), (15, 0, 15), and (0, 15, 15). Again, these findings are consistent with Mela's C=3 case and with Barker's theorem. But they are significantly different from Mela's first example, in that the uniform allocation now produces a maximum information gain rather than a minimum and that this maximum is relative, not absolute.

Pollock's example is modified slightly here for graphical presentation. The essence of Pollock's basic example is the variation of both α_i and p_i from cell to cell. To achieve extreme variation from cell to cell, Pollock chooses a value of 1 for q_1, the conditional detection probability in the first cell. The assumption $q_1 = 1$ is equivalent to $\alpha_1 \to +\infty$ in our notation. This has the unfortunate effect of introducing a singularity into our formalism, because, in computing φ_1, we are faced with the fact that:

$$\lim_{\alpha_1 \to \infty} \lim_{z_1 \to 0} p_1 e^{-\alpha_1 z_1} \neq \lim_{z_1 \to 0} \lim_{\alpha_1 \to \infty} p_1 e^{-\alpha_1 z_1} .$$

To avoid this singularity, we take $q_1 = .99$, which implies $\alpha_1 = 4.605$. This constitutes "Pollock's Example-Modified," which is presented in figures 7 - 9, in a format comparable to that of Mela's second example. The effect of the cell-to-cell variation is preserved by the modified example, while the added complications of the singularity are circumvented.

As we shall see in a later section, the cell-to-cell variation of either α_i or p_i will destroy the symmetry of the surfaces in (ℓ,m) space. Variation of α_i alone will

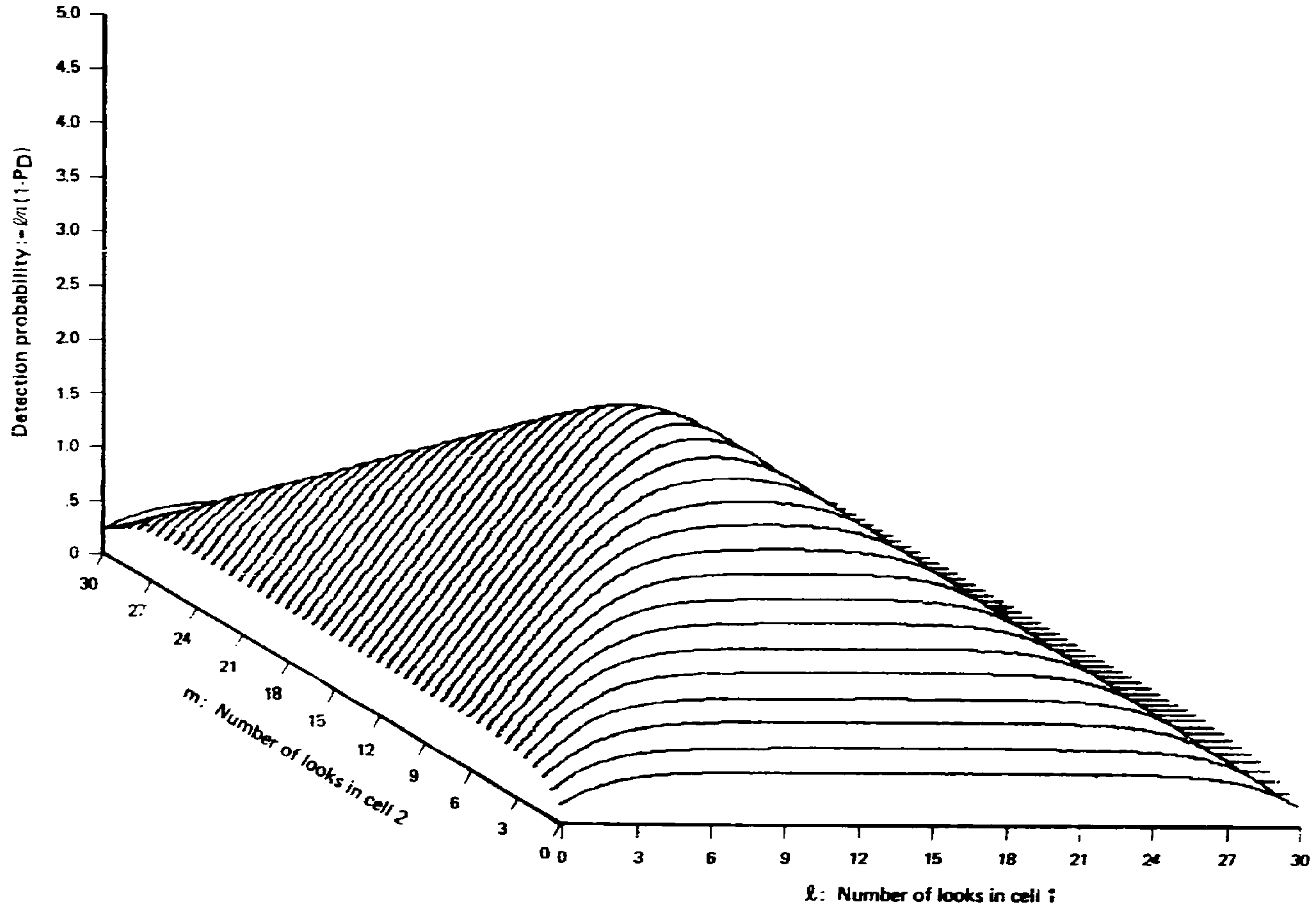

FIG. 4: MELA'S SECOND EXAMPLE – DETECTION PROBABILITY vs. SEARCH ALLOCATION

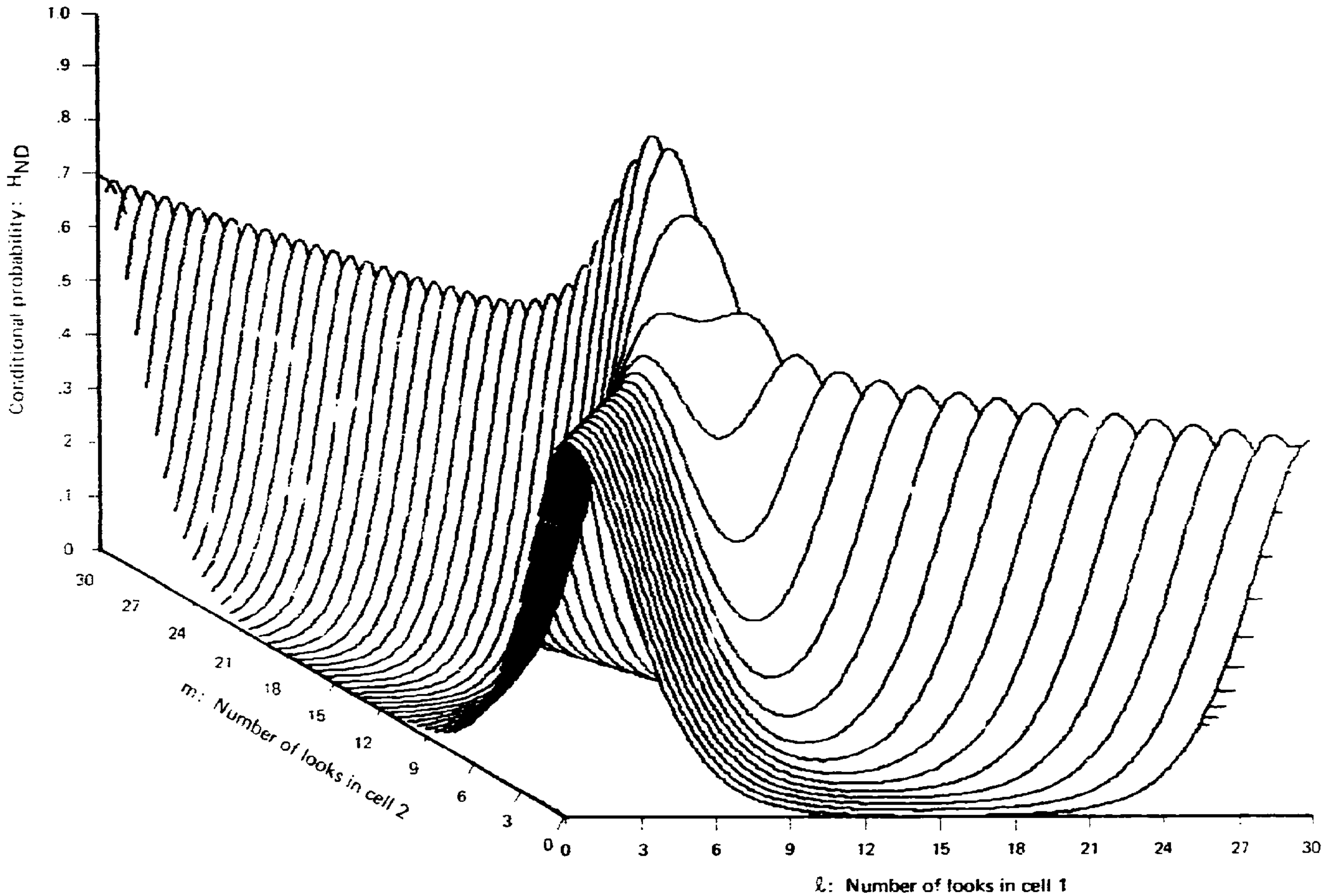

FIG. 5: MELA'S SECOND EXAMPLE – CONDITIONAL ENTROPY vs. SEARCH ALLOCATION

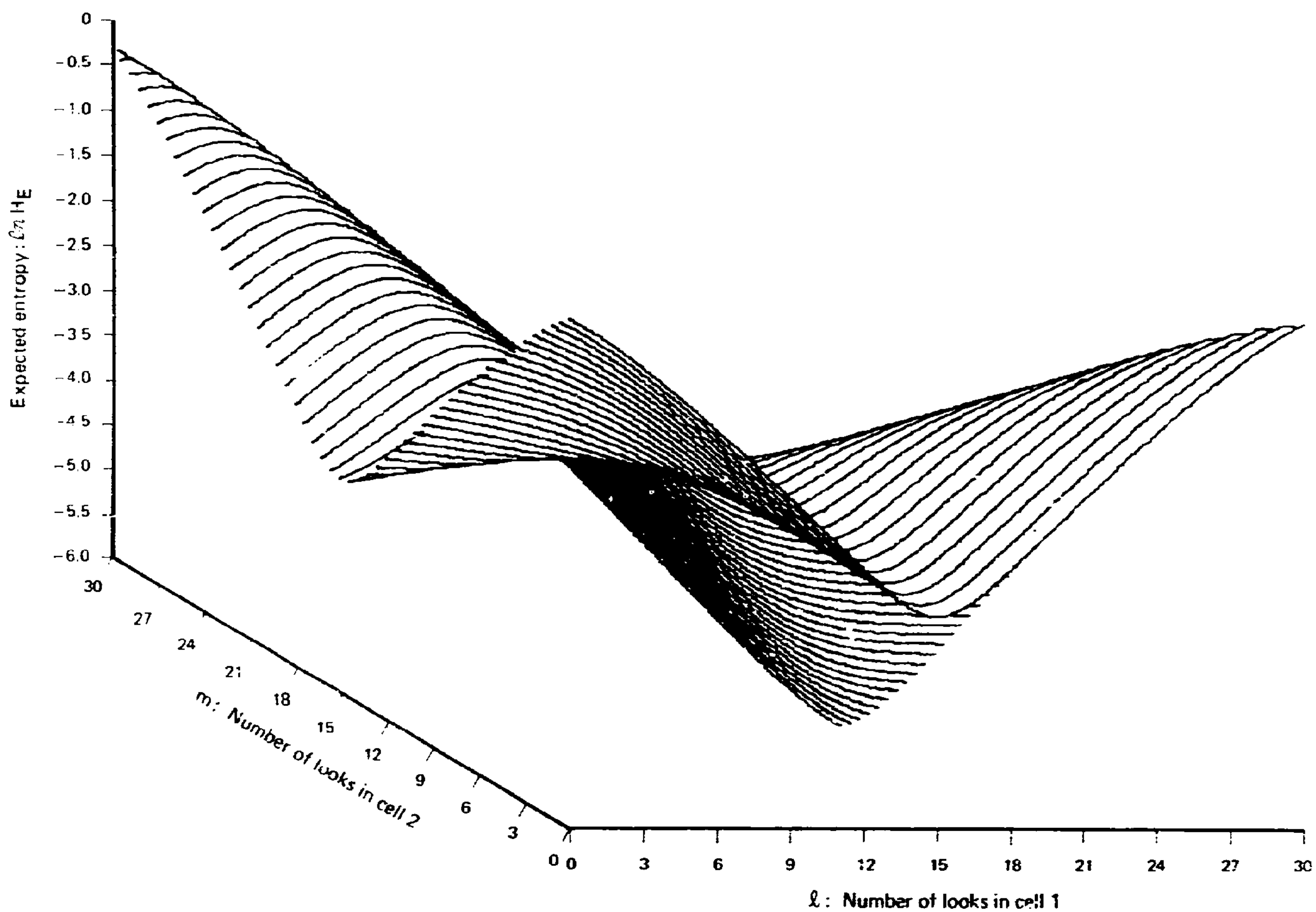

FIG. 6: MELA'S SECOND EXAMPLE — EXPECTED ENTROPY vs. SEARCH ALLOCATION

displace the relative locations of the stationary points of P_D, H_{ND}, and H_E, and can in fact change the whole character of the surface of H_E.

These points are not all obvious from figures 7 - 9. It is clear that the symmetry has been destroyed in all three figures. H_E no longer has a local extremum within the surface. A saddle point appears at (2, 10, 18), but the extreme values of H_E occur at edg points. P_D and H_{ND} retain local maxima, at (2, 11, 17) and (1, 11, 18), respectively.

The different positions for the maxima of P_D and H_{ND} could be inferred from Barker's second theorem; the different positions for the maximum of P_D and the minimum of H_E are consistent with Pollock's findings. But the totally different character of the H_E surface (a saddle point in place of a local extremum) could not have been anticipated from published work.

The numerical experiments presented in this section were designed to be suggestive, not definitive, and to point the way to fruitful lines of analysis. What the experiments have suggested can be summarized concisely:

a. When p_i and α_i are the same for all search cells:
   - P_D, H_{ND}, and H_E (or I_E) all have local extrema at a common point.
   - The local extrema of P_D and H_{ND} are maxima.
   - The local extremum of H_E may be either a maximum or a minimum.
   - The local maxima of P_D and H_{ND} are also global maxima, but the local extremum of H_E is not necessarily a global extremum.

b. When the p_i and α_i differ from cell to cell:
   - H_E does not necessarily have a local extremum.
   - P_D and H_{ND} have local maxima. The positions of their respective maxima are close, but not identical.

The next section of this paper consists of analysis based on these observations.

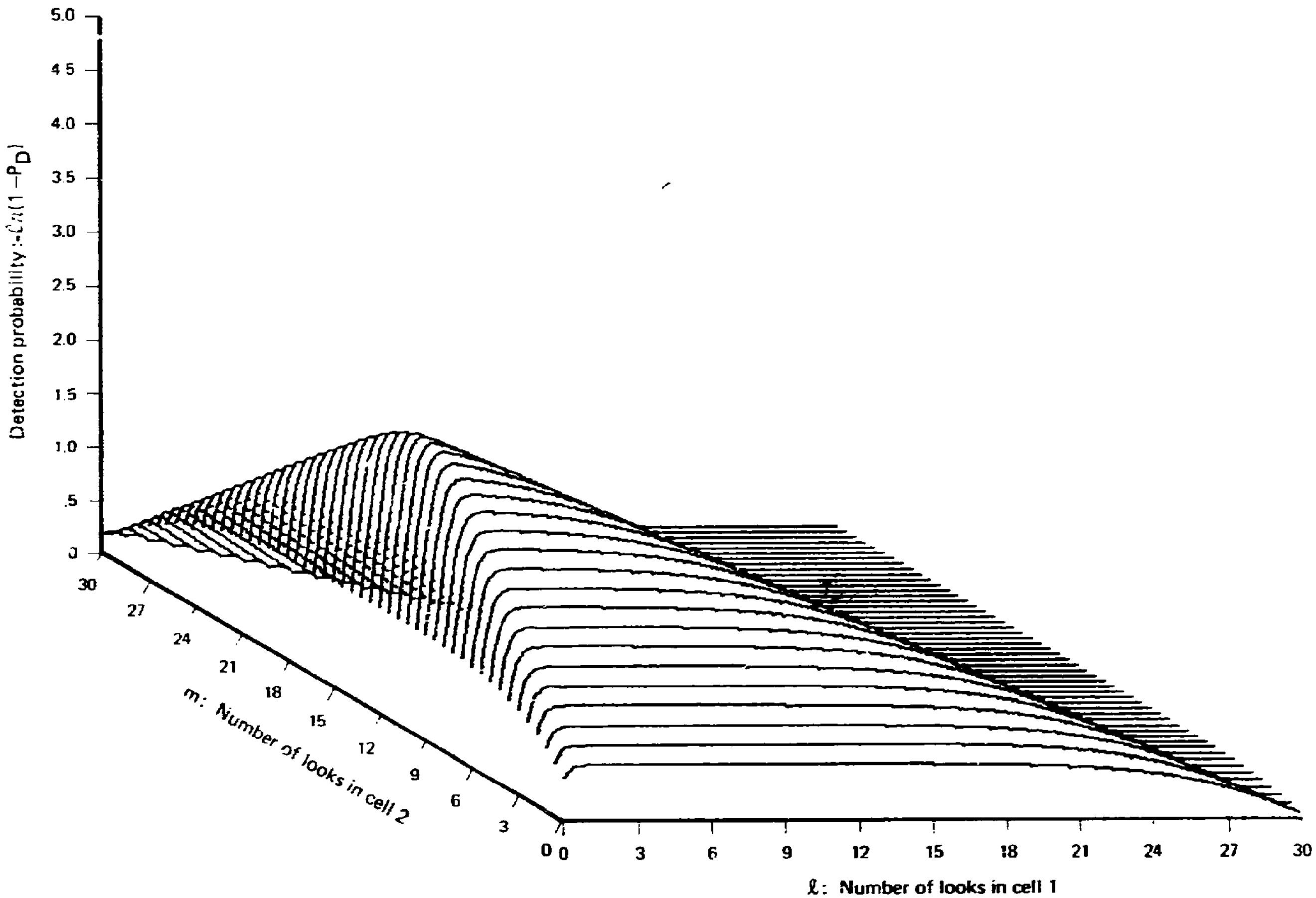

FIG. 7: POLLOCK'S EXAMPLE – MODIFIED DETECTION PROBABILITY vs. SEARCH ALLOCATION

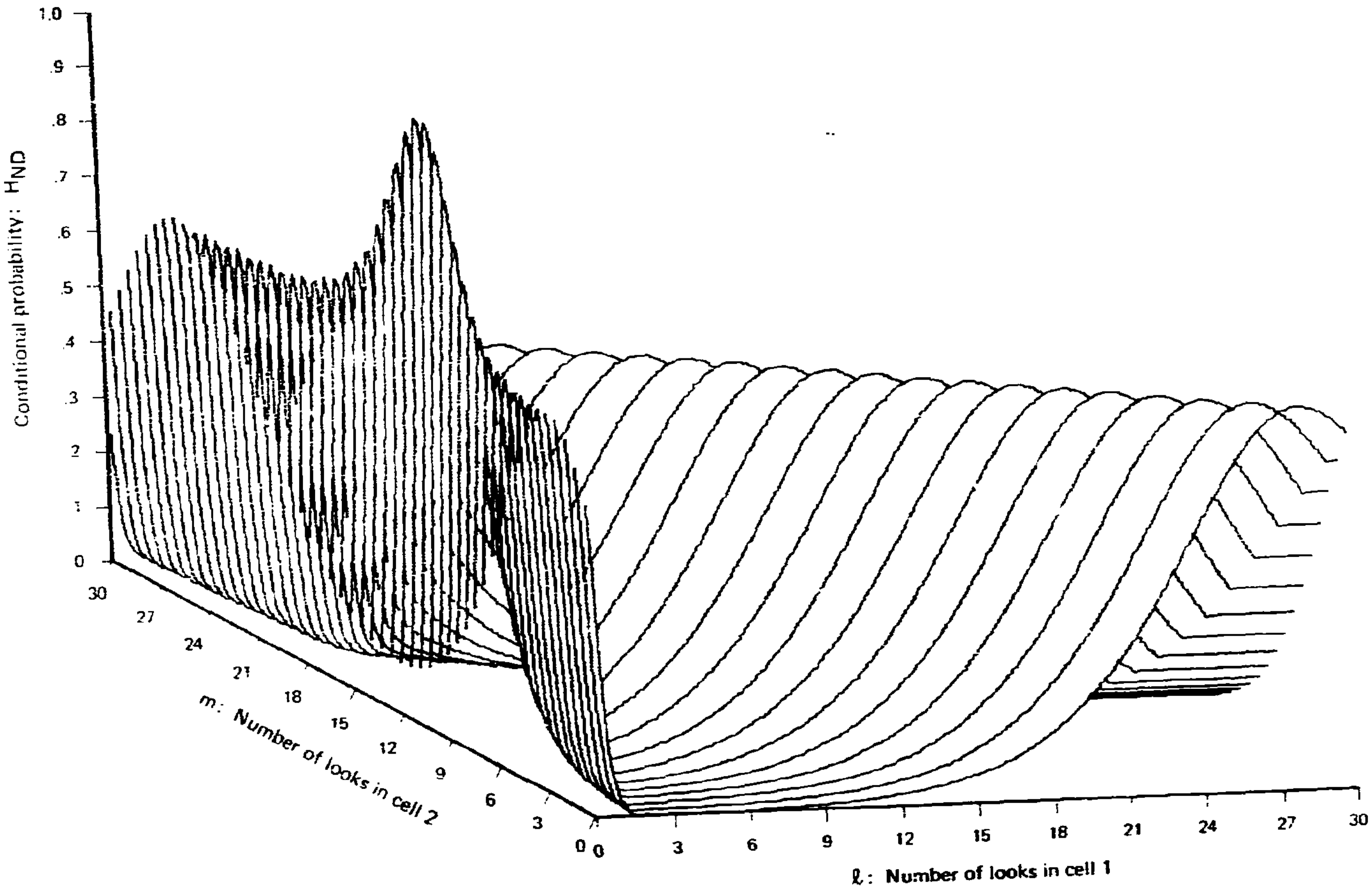

FIG. 8: POLLOCK'S EXAMPLE – MODIFIED CONDITIONAL ENTROPY vs. SEARCH ALLOCATION

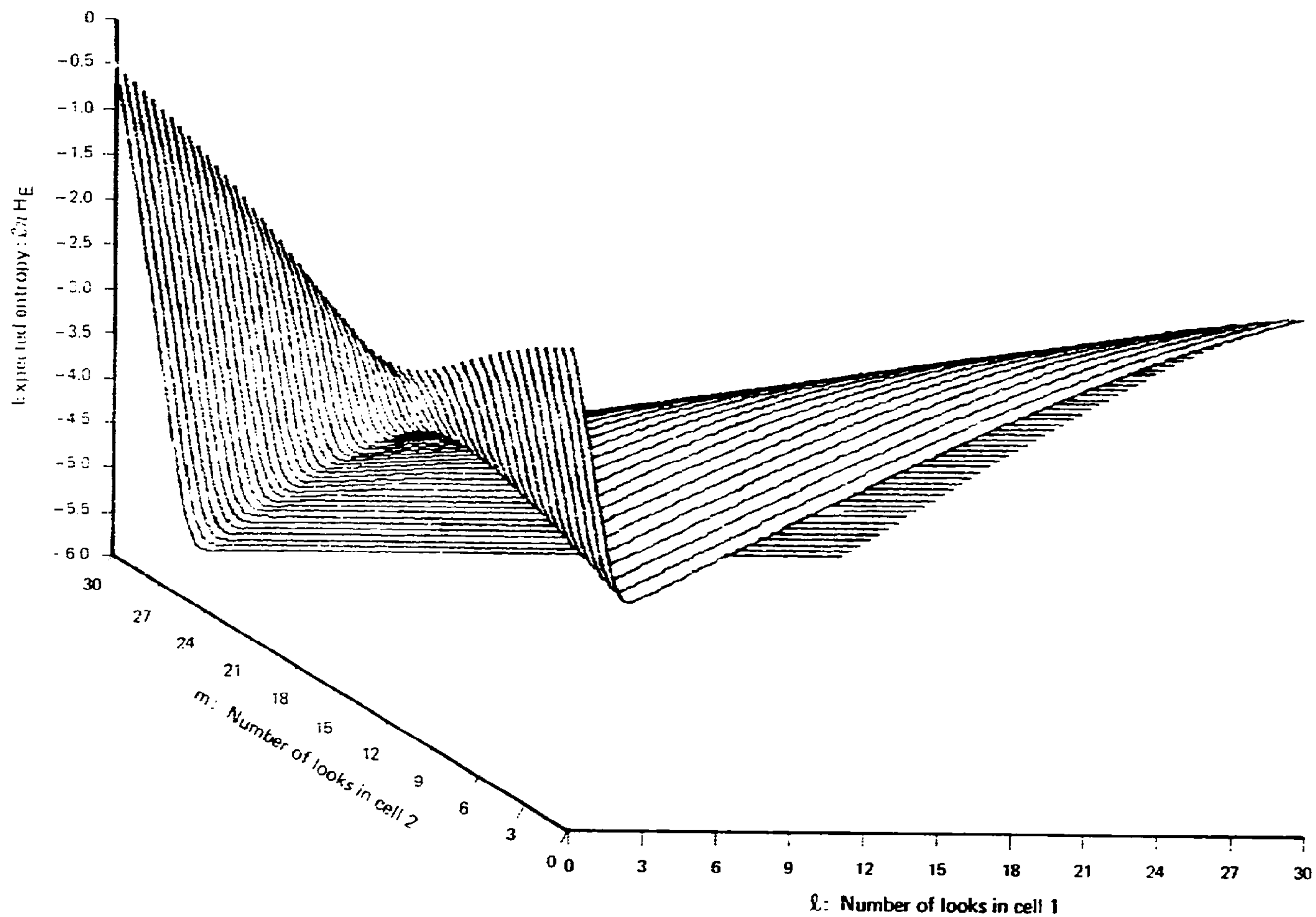

FIG. 9: POLLOCK'S EXAMPLE — MODIFIED EXPECTED ENTROPY vs. SEARCH ALLOCATION

4. ANALYSIS--LARGE SEARCH EFFORT

In this section we pursue analytically some of the ideas suggested in the previous section by the numerical extension of the Mela and Pollock examples to the limit of large search effort. The mathematical framework is that established in section 2. We take the three functionals of the search allocation $\{z_j\}$: P_D, defined in eq. 2.9; H_{ND}, defined in eq. 2.22; and H_E, defined in eq. 2.19. We seek to determine the extrema of these functionals, subject to the constraint on total search effort (eq. 2.1).

We defer temporarily the precise specification of a "large amount of search effort." We assert that for $C > C^*$, for some C^*, the conventional calculus of variation technique using Lagrange multipliers to incorporate constraints is applicable and will lead us to the desired extrema. The determination of C^* in terms of the p_j and α_j will follow as part of the analysis.

First, it is important to note that the three quantities P_D, H_{ND}, and H_E depend on only two independent functionals of the $\{z_j\}$:

$$A \equiv \sum_{j=1}^{J} \varphi_j(z_j) \ln \varphi_j(z_j) \tag{4.1}$$

and

$$B \equiv \sum_{j=1}^{J} \varphi_j(z_j) \quad . \tag{4.2}$$

Specifically:

$$P_D = 1 - B \tag{4.3}$$

$$H_{ND} = \ln B - A/B \tag{4.4}$$

$$H_E = B \ln B - A \quad . \tag{4.5}$$

We can obtain some general insights using this simplified representation. Let the operator δ signify constrained variation; that is, for a function f, δf refers to the set of J quantities $\partial f/\partial z_j$, subject to $C = \sum_{j=1}^{J} z_j$. With this notation:

$$\delta P_D = -\delta B \tag{4.6}$$

$$\delta H_{ND} = \delta A(-(1/B)) + \delta B \left[(1/B) + (A/B^2)\right] \tag{4.7}$$

$$\delta H_E = -\delta A + \delta B(1 + \ln B) \ . \tag{4.8}$$

From this set of equations we see that if there is an allocation of search effort $\{z_j^*\}$, such that δA and δB are both zero, then all three functionals P_D, H_{ND}, and H_E have extrema for that particular allocation $\{z_j^*\}$. Conversely, if δA and δB are not simultaneously equal to zero, then extrema of the three functionals will occur for different search allocations, given respectively by:

$$\delta B = 0 \tag{4.9}$$

$$\delta A = \delta B(1 + (A/B)): \quad (\delta A \neq \delta B \neq 0) \tag{4.10}$$

$$\delta A = \delta B(1 + \ln B): \quad (\delta A \neq \delta B \neq 0) \ . \tag{4.11}$$

Returning to the specific functional form of A and B , we find that:

$$\begin{aligned} \delta A &= (\partial\varphi_k/\partial z_k)(1 + \ln \varphi_k) - \lambda \\ &= -\alpha_k\varphi_k(1 + \ln \varphi_k) - \lambda \end{aligned} \tag{4.12}$$

for all $k \in J$. λ is a Lagrange multiplier. Similarly:

$$\delta B = -\alpha_k\varphi_k - \lambda \ . \tag{4.13}$$

If δA and δB are to be simultaneously zero, we must have:

$$\alpha_k \varphi_k (1 + \ln \varphi_k) = \alpha_j \varphi_j (1 + \ln \varphi_j) \tag{4.14}$$

$$\alpha_k \varphi_k = \alpha_j \varphi_j \tag{4.15}$$

for all $j, k \in J$. Eliminating φ_j and φ_k between these equations, we obtain:

$$\alpha_j = \alpha_k \text{ , for all } j,k \in J \text{ .} \tag{4.16}$$

Thus, δA and δB are simultaneously zero only if α is the same for each of the j search cells in J.

When all the α_j are equal, eqs. 4.14 and 4.15 reduce to:

$$\varphi_k (1 + \ln \varphi_k) = \varphi_j (1 + \ln \varphi_j) \tag{4.17}$$

and

$$\varphi_k = \varphi_j \text{ , for all } j,k \in J \text{ .} \tag{4.18}$$

The only solution for these sets of equations is:

$$\varphi_k = \text{const} = K \text{ ,} \qquad \text{for all } k \in J \text{ .} \tag{4.19}$$

This equation cannot be satisfied for small values of the total search effort, C. By requiring it to be satisfied, we can determine C^*, the lower bound on the search effort for which this method of analysis is valid. Using the definition of φ_k, (eq. 2.6), we infer from eq. 4.19:

$$p_1 e^{-\alpha_1 z_1} = p_2 e^{-\alpha_2 z_2} = \ldots = p_J \text{ .} \tag{4.20}$$

In this formulation we have assumed that the p_j values are arranged in nonincreasing order, so that $p_J = \mathrm{Min}_j \{p_j\}$. From eq. 4.20, with $\alpha_j = \alpha_k$:

$$z_j = (1/\alpha) \ln (p_j/p_J) \ ; \qquad j = 1,2 \text{ -- } J-1 \tag{4.21}$$

or:

$$C^* = (1/\alpha) \left(\sum_{j=1}^{J} \ln p_j - J \ln p_{min} \right) . \tag{4.22}$$

For $C > C^*$, eq. 4.19, leads to

$$z_k = (1/\alpha) \ln (p_k/K) . \tag{4.23}$$

Further, because of the constraint on the z_k, this constant K can be evaluated in terms of the total search effort; thus:

$$z_k = (C/J) + (1/\alpha) \left[\ln p_k - (1/J) \sum_{j=1}^{J} \ln p_j \right] \tag{4.24}$$

This a well known result, closely related to that originally derived by Koopman, (reference 8) and expounded in Stone (reference 9). The optimum allocations for Mela's two examples follow from this trivially.

We may now summarize these results in two theorems:

Theorem I: For the assumptions of our model (section 2.1) with exponential detection functions and with $C > C^*$, the allocation of search effort $\{z_k^*\}$ that leads to an extremum of P_D also leads to an extremum of H_{ND} only if the α_k values are the same for all $k \in J$. (This is the alternative statement of one of Barker's results.)

Theorem II: For the assumptions of our model (section 2.1) with exponential detection functions and with $C > C^*$, the allocation of search effort $\{z_k^*\}$ that leads to an extremum of P_D also leads to an extremum of H_E only if the α_k values are the same for all $k \in J$. (New result.)

The allocation $\{z_k^*\}$ that leads to simultaneous extrema of all three functionals P_D, H_{ND}, and H_E is given by eq. 4.24.

We must now find out whether the extrema thus determined are maxima or minima. To do this, we examine the second derivatives. Let the second constrained variation, denoted by $\delta^2 f$, be defined by the quadratic form $\sum_{i,j} \zeta_i \zeta_j (\partial^2 f/\partial z_i \partial z_j)$, subject to $C = \sum_j z_j$.

To compute $\delta^2 f$, it is more convenient to include the constraint explicitly, rather than use the Lagrange multiplier technique. This is done by noting that only J-1 of the z_j are independent. The first J-1 of the φ_j are taken as explicit functions of their respective z_j, and the remaining φ_J is taken as a function of $C - \sum_{j=1}^{J-1} z_j$. In computing derivatives with respect to a specific z_k, therefore, both φ_k and φ_J must be considered.

With this convention, the various second partial derivatives can be written, after some manipulation, as:

$$\partial^2 P_D/\partial z_k^2 = -\left[\alpha_k^2 \varphi_k + \alpha_J^2 \varphi_J\right]$$

$$\partial^2 P_D/\partial z_k \partial z_\ell = -\alpha_J^2 \varphi_J \qquad (4.25)$$

$$\partial^2 H_{ND}/\partial z_k^2 = (1/\sum_j \varphi_j)^2 (\alpha_k \varphi_k - \alpha_J \varphi_J)^2$$

$$- (1/\sum_j \varphi_j)(\alpha_k^2 \varphi_k + \alpha_J^2 \varphi_J) \qquad (4.26)$$

$$\partial^2 H_{ND}/\partial z_k \partial z_\ell = -\alpha_J^2 \varphi_J \left[1 + \ln \varphi_J - (\sum_{j=1}^{J} \varphi_j \ln \varphi_j)/(\sum_{j=1}^{J} \varphi_j)\right] / (\sum_{j=1}^{J} \varphi_j)$$

$$+ (-\alpha_k \varphi_k + \alpha_J \varphi_J)(-\alpha_\ell \varphi_\ell + \alpha_J \varphi_J)$$

$$\partial^2 H_E/\partial z_k^2 = -(\alpha_k^2\varphi_k + \alpha_J^2\varphi_J) + (1/\sum_j \varphi_j)(\alpha_k\varphi_k - \alpha_J\varphi_J)^2$$

$$-\alpha_k^2\varphi_k(\ln \varphi_k - \ln \sum_j \varphi_j)$$

$$-\alpha_J^2\varphi_J(\ln \varphi_J - \ln \sum_j \varphi_j) \qquad (4.27)$$

$$\partial^2 H_E/\partial z_k \partial z_\ell = -\alpha_J^2\varphi_J\left[1 + \ln \varphi_J - \ln \left(\sum_{j=1}^{J} \varphi_j\right)\right]$$

$$+ (-\alpha_k\varphi_k + \alpha_J\varphi_J)(-\alpha_\ell\varphi_\ell + \alpha_J\varphi_J)$$

These are the second variations evaluated at the extrema, where the relationships obtained by equating the first variation to zero have been used to simplify the expressions, especially in the case of H_{ND}.

For the case discussed above, $\alpha_j = \alpha$ and $\varphi_j = K$ for all j, the second variations simplify to:

$$\delta^2 P_D = -\alpha^2 KD \qquad (4.28)$$

$$\delta^2 H_{ND} = -((\alpha^2)/J)D \qquad (4.29)$$

$$\delta^2 H_E = -\alpha^2 K(1 - \ln J)D \ . \qquad (4.30)$$

D is the residual quadratic form:

$$D = \sum_{j=1}^{J} \zeta_j^2 + \left(\sum_{j=1}^{J} \zeta_j\right)^2 \geq 0 \ .$$

Noting that K, by definition, is a positive constant between zero and one, we have:

Theorem III: For the assumptions of our model (section (2.1)) with exponential detection functions, with $C > C^*$ and all α_j equal, the extrema of P_D and H_{ND} are always maxima. The extremum of H_E is a maximum for $J=2$ and a minimum for all $J > 2$.

This result clarifies a great deal of the mystery and confusion resulting from Mela's paper. His two cases, $J=2$ and $J=3$, gave contradictory indications about H_E and its relation to P_D, and produced the false impression that there was no connection between detection probability and information gain I_E (Mela's I_E is equal to $(\ln J - H_E)$ in our terminology). What we have shown here is that for the important case in which the conditional detection probability is uniform throughout the search area, optimizing the detection probability is identical with optimizing the change in the amount of information. Moreover, in all instances except the anomalous case $J=2$, maximizing the detection probability is identical with maximizing the information gain.

The procedures leading to Theorems I-III guarantee only that we have found local extrema of the expected entropy surface. To complete the analysis, we must further investigate:

- Whether there are additional stationary points on the H_E surface.
- Whether there are in general points on the boundaries of the surface for which the value of H_E exceeds the value at the local extremum, as suggested by the numerical results in section 3.

We note first the values of the pertinent quantities at their local stationary points. Using the solution (eq. 4.24), we find that:

$$\text{Max}\,\{P_D\} = 1 - \exp\left[\ln J + (1/J)\sum_{i=1}^{J} \ln p_i\right] \exp(-\alpha C/J) \quad (4.31)$$

$$\text{Max}\,\{H_{ND}\} = \ln J$$

$$\text{Extr}\,\{H_E\} = \ln J\left[\exp(\ln J + (1/J)\sum_{i=1}^{J} \ln p_i)\right] \exp(-\alpha C/J)\,. \quad (4.32)$$

For equal prior probabilities, $p_i = 1/J$, these reduce to:

$$\text{Max}\ \{P_D\} = 1 - \exp(-\alpha C/J) \tag{4.33}$$

$$\text{Max}\ \{H_{ND}\} = \ln J \tag{4.34}$$

$$\text{Extr}\ \{I_E\} = H_O - \text{Extr}\ \{H_E\} = \ell n J(1 - \exp(-\alpha C/J)) \tag{4.35}$$

To check for the existence of other stationary points on the H_E surface, we return to eq. 4.11. That equation can be put in the form:

$$\alpha\varphi_k\ (\ln \varphi_k - \ln \sum_{i=1}^{J} \varphi_i) = \lambda \tag{4.36}$$

or, by further manipulation:

$$-\gamma_k \ln \gamma_k = G\ , \tag{4.37}$$

where

$$\gamma_k = \varphi_k / \sum_{j=1}^{J} \varphi_j\ ;\ G = \lambda/\alpha \sum_{j=1}^{J} \varphi_j\ . \tag{4.38}$$

The γ_k must satisfy $\sum_{k=1}^{J} \gamma_k = 1$, and, for a solution to exist:

$$0 \leq G \leq 1/e\ , \tag{4.39}$$

as can be seen from figure 10.

The nature of the solutions can be argued from the shape of the curve in figure 10. Assume, first, that $G = 1/e$. Then, $\gamma_k = 1/e$, and $\sum_{k=1}^{J} \gamma_k = J/e$, which is less than 1 for J=2, and greater than 1 for J greater than 2. To drive $\sum_{k=1}^{J} \gamma_k$ toward 1, we decrease G. For J=2, the points must move down the right branch of the curve,

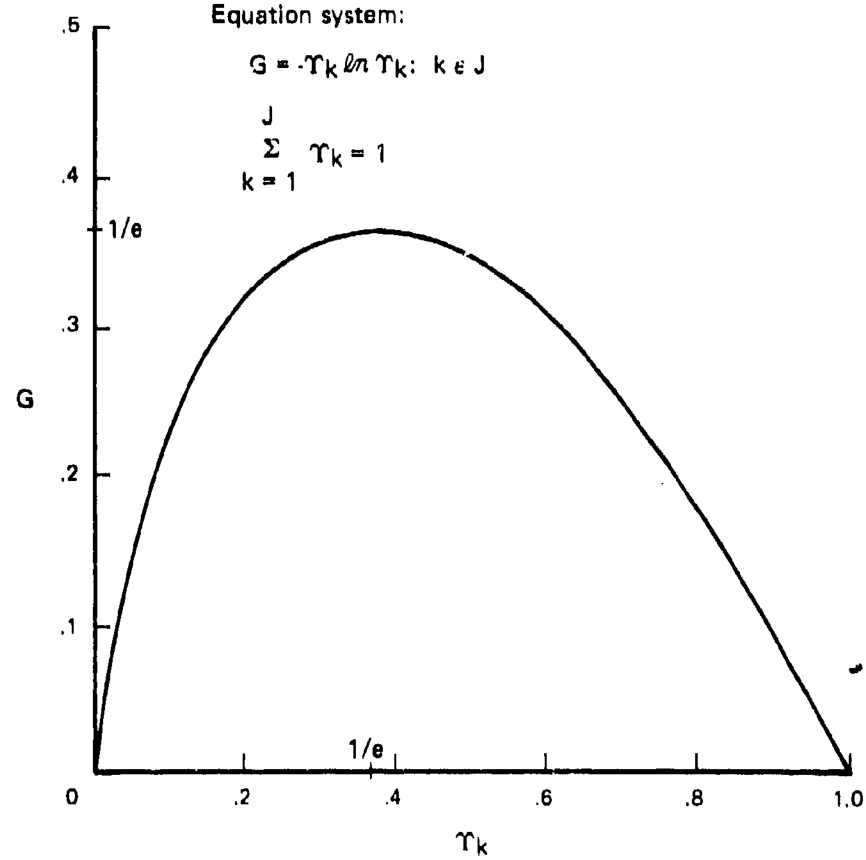

FIG. 10: GRAPHICAL ILLUSTRATION OF EQUATIONS 4.37 AND 4.38

increasing γ_k, until $\gamma_k = 1/2$, $G = 1/2 \ln 2$. For all $J > 2$, the points must move down the left branch of the curve, decreasing γ_k, until $\gamma_k = 1/J$, $G = (\ln J)/J$. These represent the solutions $\varphi_k = K$, already discussed in equation 4.19 and those which follow.

If, however, we continue to decrease G, with all γ_k on the left branch of the curve, $\sum_{k=1}^{J} \gamma_k$ approaches zero in such a way that for some G*, we can again make $\sum_{k=1}^{J} \gamma_k = 1$ by switching any one of the γ_k to the right branch of the curve. The condition for this is $\varphi_k = K1$; $\varphi_j = K2$; $j \neq k \in J$, for any specific k. Because of the shape of the curve, $\gamma_k \sim 1$, $\gamma_j \sim 0$, and thus $K1 >> K2$.

Let $K1 = \mu K2$, and $n = J-1$. Then, from eq. 4.37:

$$[\mu/(\mu+n)] \ln [\mu/(\mu+n)] = [1/(\mu+n)] \ln [1/(\mu+n)] \qquad (4.40)$$

The second equation for determining the two free parameters, K1 and K2, is the constraint on total search effort. Using the definitions of the φ_k :

$$\alpha C/J = (1/J) \left[\sum_{k=1}^{J} \ln p_k - \ln \mu \right] - \ln K2 \quad . \qquad (4.41)$$

Eq. 4.40 cannot be solved for $\mu(n)$, but it can be solved for $n(\mu)$:

$$n(\mu) = -\mu + \exp [(\mu \ln \mu)/(\mu-1)] \quad . \qquad (4.42)$$

This equation is plotted in figure 11, with $\ln\mu$ as a function of n. For large μ :

$$n \sim \ln \mu + (1/\mu) [(\ln \mu)^2/2 + \ln \mu] + O(\mu^{-2}) \quad . \qquad (4.43)$$

For $n \geq 8$ ($J \geq 9$), $n \sim \ln \mu$, $\mu \sim e^n$ within a few percent of accuracy. It is important to note from figure 11 that this type of solution exists only for $n > e-1$ ($J > e$). This means that for $J = 2$, the only solution is the one with both γ points on the right branch of the curve 10; for $J \geq 3$, however, the two classes of solutions exist: a single solution (Type 1)

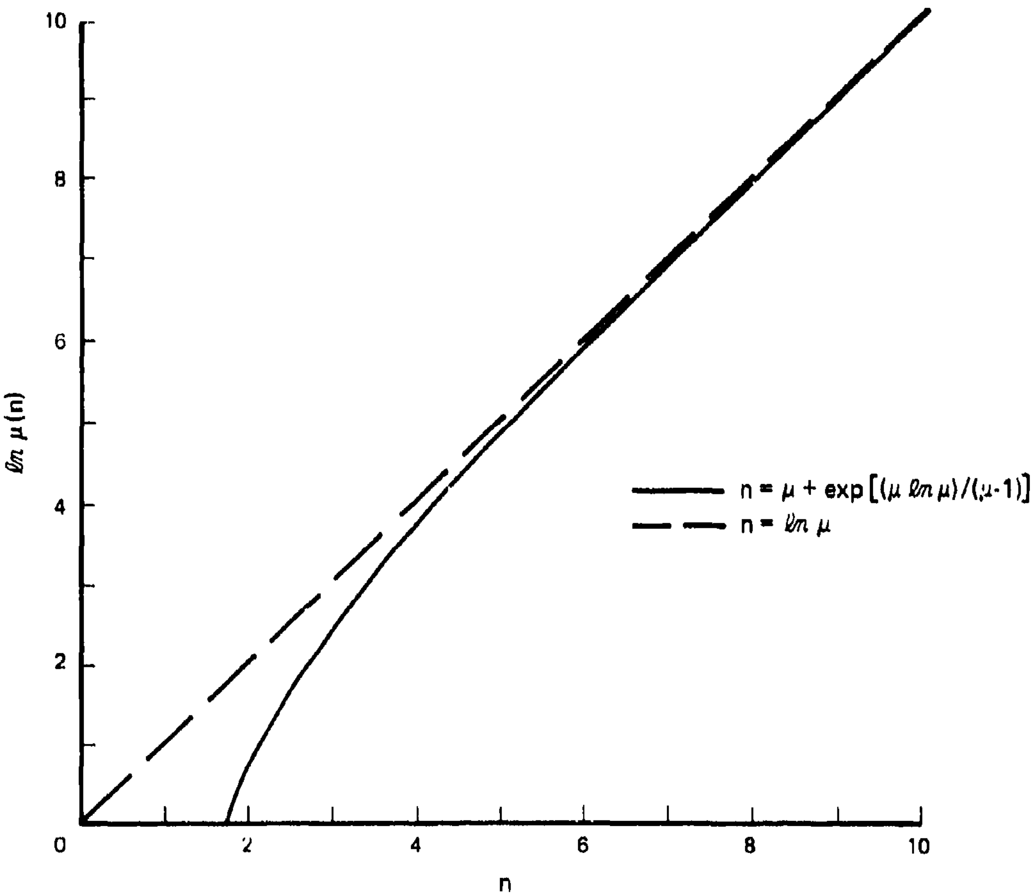

FIG. 11: GRAPHICAL ILLUSTRATION OF EQUATION 4.42

with all γ points on the left branch of figure 10, and J solutions (Type 2) with any one of the J γ points on the right branch of the curve and the other J-1 on the left branch.

There is an additional condition on the amount of search effort, C, however. Just as there was a C* threshold for the existence of the Type 1 solution, there are J additional thresholds for the appearance of each Type 2 solution. This is best seen when the search allocation solutions in the asymptotic limit are calculated:

$$z_k \cong (C/J) + (1/\alpha)\ \ln p_k - (1/\alpha J) \sum_{\ell=1}^{J} \ln p_\ell - (J-1)^2/\alpha J \qquad (4.44)$$

$$z_j \cong (C/J) + (1/\alpha)\ \ln p_j - (1/\alpha J) \sum_{\ell=1}^{J} \ln p_\ell + (J-1)/\alpha J \qquad (4.45)$$

Since the $z_k \geq 0$, eq. 4.44 provides the threshold criteria for C:

$$\alpha C_k^* \cong (J-1)^2 + \sum_{j=1}^{J} \ln p_j - J \ln p_k \qquad (4.46)$$

$$k = 1 \ . \ . \ . \ J$$

This, together with eq. 4.22, provides the values of C for which each of the J+1 stationary points of the H_E surface first appears. In the case of equal prior probabilities, $C^*=0$ from (4.22), and αC_k^* equals $(J-1)^2$ for all k.

Having found the approximate locations of the J additional stationary points of the H_E surface, we now find out what type of stationary points they are. Using eq. 4.27, we find:

$$\partial^2 H_E/\partial z_k^2 = -\alpha^2\ K2\ (1 - \ln J) \qquad (4.47)$$

for those coordinates such that $\varphi_k = \varphi_J = K2$. Thus, for $J>2$, we have a minimum of H_E in J-1 of the coordinates. For the remaining case, $\varphi_k = K1$, $\varphi_j = K2$.

$$\partial^2 H_E/\partial z_k^2 = -\alpha^2 K2\ [\ (\mu+1) - 2(\mu \ln \mu)/(\mu-1) - (\mu-1)^2 \exp(-(\mu \ln \mu)/(\mu-1))\ ] \qquad (4.48)$$

where eq. 4.42 has been used to eliminate n. Direct calculation for $1 \leq \mu \leq \infty$ shows that $\delta^2 H_E \geq 0$ for all μ. This result seems to suggest, paradoxically, that all J+1 stationary points are minima.

To resolve this apparent paradox, we must be considerably more careful with our coordinates. We seek, in particular, at least one linear combination of the z_i, such that its second derivative has the opposite sign at the stationary point.

Let us introduce a variable x, $0 < x \leq 1$, such that:

$$z_k = (1-x)\left[(C/J) + (1/\alpha)\ \ln p_k - (1/\alpha J)\sum_{\ell=1}^{J} \ln p_\ell\right] \tag{4.49}$$

$$\begin{aligned} z_j = &\left[(C/J) + (1/\alpha)\ \ln p_j - (1/\alpha J)\sum_{\ell=1}^{J} \ln p_\ell\right] \\ &+ (x/(J-1))\left[(C/J) + (1/\alpha)\ \ln p_k - (1/\alpha J)\sum_{\ell=1}^{J} \ln p_\ell\right] \end{aligned} \tag{4.50}$$

$j \neq k \in J$.

For ease of notation, define:

$$\beta_k/\alpha = (C/J) + (1/\alpha)\ \ln p_k - (1/\alpha J)\sum_{\ell=1}^{J} \ln p_\ell \quad . \tag{4.51}$$

Then equations 4.49 and 4.50 become:

$$z_k = (1-x)(\beta_k/\alpha) \tag{4.52}$$

$$z_j = (\beta_j/\alpha) + (x/(J-1))(\beta_k/\alpha) \quad . \tag{4.53}$$

By introducing this parameter, we achieve several results. Note that when x=0:

$$z_k = \beta_k/\alpha \quad ; \quad z_j = \beta_j/\alpha \quad . \tag{4.54}$$

These are the solutions for optimal detection search from eq. 4.24. When $x=1$, $z_k=0$, and:

$$z_j = C/(J-1) + (1/\alpha)\ \ln p_j - (1/\alpha(J-1)) \sum_{\ell \neq k}^{J-1} \ln p_\ell \tag{4.55}$$

The latter are recognized as the optimal detection search allocations in J-1 cells, when cell k is not searched. These represent the class of whereabout searches, whose aim is to locate an object by searching all cells but one, then guessing the position on the basis of the search outcome. Of this class, the one for which $p_k = \max_J (p_j)$ in the unsearched cell is called the "optimal whereabouts search." This search maximizes the probability of correctly guessing the object's location (reference 10).

Finally, for $x = (J-1)^2/J\beta_k$ we find:

$$z_k = \beta_k/\alpha - (J-1)^2/\alpha J \tag{4.56}$$

$$z_j = \beta_j/\alpha + (J-1)/\alpha J \quad . \tag{4.57}$$

These are identical to the asymptotic forms of the positions of the stationary points of the H_E surface, determined in equations 4.44 and 4.45.

Thus, by use of the parameter x, we have introduced a set of J hyperplanes. Each of these hyperplanes connects (a) the Type 1 stationary point corresponding to the allocation for optimal detection search; (b) one of the J Type 2 stationary points; and (c) one of the J boundary points corresponding to a whereabouts search. We now investigate the curve formed by intersection of these J hyperplanes with the H_E surface.

Using equations 4.52 and 4.53, we can put H_E in this form:

$$H_E = \left(\exp - \left[(C/J) - (1/\alpha J) \sum_{\ell=1}^{J} \ln p_\ell \right]\right) \Big\{ - \beta_k x \left[\exp(\beta_k x) - \exp(-\beta_k x/n) \right] + \left[\exp(\beta_k x) + n \exp(-\beta_k x/n) \right] \ln \left[\exp(\beta_k x) + n \exp(-\beta_k x/n) \right] \Big\} \; . \tag{4.58}$$

By computing $dH_E/dx = 0$, it can be shown that an equation identical to eq. 4.42 is obtained, with μ now identified as:

$$\exp\ [(\beta_k x)(n+1)/n] \quad . \tag{4.59}$$

Its solution is again the curve shown in figure 11, and the asymptotic form leads to:

$$x \simeq n^2/(n+1)(\beta_k) \quad . \tag{4.60}$$

This is indeed the position of the Type 2 stationary point as determined earlier.

The second derivative of H_E at the stationary point is:

$$d^2H_E/dx^2 \sim [(\mu \ln \mu)/(\mu-1) + \mu^2 \exp[-(\mu \ln \mu)/(\mu-1)] - (\mu+1)] \quad . \tag{4.61}$$

In contrast to eq. 4.48, this is negative for all $1 \leq \mu < \infty$. Thus, along the curves generated by the J hyperplanes, H_E has a max at the stationary point, and we may safely describe the J Type 2 stationary points as saddle points.

Since the entropy has a max at the type 2 stationary point along the x path, and then turns downward again, we may ask whether it later reaches a value below the minimum at the optimal detection search allocation. In fact, it does, if C is large enough. To find this threshold of C , we simply equate $H_E(O) = H_E(1)$, yielding:

$$J \ln J = [\exp(\beta_k) + n \exp(-\beta_k/n)] \ln [\exp(\beta_k) + n \exp(-\beta_k/n)] - \beta_k[\exp(\beta_k) - \exp(-\beta_k/n)] \quad . \tag{4.62}$$

In the dual limit $J >> 1$, $\beta_k >> \ln (J-1)$, this equation can be solved approximately:

$$\alpha C^k_{EQ} \cong \sum_{\ell=1}^{J} \ln p_\ell - J \ln p_k + J(J-1) \ln J \quad . \tag{4.63}$$

The minimum of this over the set J is:

$$\alpha C^{MIN}_{EQ} \cong \sum_{\ell=1}^{J} \ln p_\ell - J \ln p_{MIN} + J(J-1) \ln J \quad . \tag{4.64}$$

For any $C > C_{EQ}^{MIN}$, there is a whereabouts search that yields more expected information than does the optimal detection search.

We can now give a complete description of the expected information surface in allocation space, with the total search effort C as a parameter:

1. For $C < C^*$ (eq. 4.22), I_E has no stationary points. The extreme values of I_E occur on the boundaries.

2. For $C^* < C < C_k^*$ (eq. 4.46), and for $J > 2$, I_E has only one stationary point. That point is both relative and absolute maximum, and it occurs at the allocation of search effort corresponding to optimal detection search.

3. As C increases through the series C_k^* (k = 1, 2, -- J), J new stationary points appear on the I_E surface. All J are saddle points, whose coordinates are given approximately by equations 4.44 and 4.45. The one maximum point remains both relative and absolute maximum.

4. When C reaches C_{EQ}^{MIN} (eq. 4.64), the relative maximum is no longer the absolute maximum. The new absolute maximum of I_E occurs for the search allocation corresponding to the optimal whereabouts search. As C increases still further, new border regions are added to the surface, for which I_E exceeds the local maximum. The optimal whereabouts search allocation remains the absolute maximum of the surface, however.

Finally, it is interesting to compute the values of P_D and of the expected information at C_{EQ}^{MIN}, where the crossover occurs.

These are, for optimal detection search allocation:

$$P_D = 1 - p_{MAX} J^2 \exp(-J \ln J) \; ; \tag{4.65}$$

for optimal whereabouts search allocation:

$$P_D = (1 - p_{MAX}) - p_{MAX} (J-1) \exp(-J \ln J) \; ; \tag{4.66}$$

and, for the information in either case:

$$(I_{MAX} - I_E)/I_{MAX} = H_E/H_O = (p_{MAX}/H_O) J^2 \ln J \exp(-J \ln J) \quad . \qquad (4.67)$$

From these it can be seen that for even modest J ($J = 5$, say) P_D is extremely close to 1, and I_E is extremely close to I_{MAX}, by the time the crossover occurs.

We return now to the case in which the α_j are not equal. This corresponds to Pollock's example. In this case, the allocations $\{z_j\}$ that determine the extrema of P_D, H_{ND}, and H_E are given by the solutions of equations 4.9 - 4.11. They are, of course, not identical.

Using the Lagrange multiplier notation, we obtain the following equations. For $\delta p_D = 0$:

$$\alpha_k \varphi_k = \lambda_P \ , \qquad \text{all } k \in J \ ; \qquad (4.68)$$

for $\delta H_{ND} = 0$:

$$\alpha_k \varphi_k \left\{ \left[(\ln \varphi_k) / \sum_j \varphi_j \right] - \left(\sum_j \varphi_j \ln \varphi_j \right) \Big/ \left(\sum_j \varphi_j \right)^2 \right\} = \lambda_{ND} \ , \qquad (4.69)$$

for $\delta H_E = 0$:

$$\alpha_k \varphi_k \left[\ln \varphi_k - \ln \sum_j \varphi_j \right] = \lambda_E \ . \qquad (4.70)$$

The solution to eq. 4.68 is straightforward and well known:

$$z_k^{*P} = (\ln \alpha_k p_k)/\alpha_k - \left[\left(\sum_{j=1}^{J} (\ln \alpha_j p_j)/\alpha_j \right) \Big/ \alpha_k \sum_{j=1}^{J} (1/\alpha_j) \right] + C/\alpha_k \sum_{j=1}^{J} (1/\alpha_j) \qquad (4.71)$$

Equations 4.69 and 4.70 are transcendental equations and appear more formidable, but, in point of fact, only eq. 4.70 is so formidable. If we divide eq. 4.69 by α_k and then sum on the index k, we find that:

$$\lambda_{ND} \sum_{k=1}^{J} (1/\alpha_k) = \left[\left(\sum_{k=1}^{J} \varphi_k \, \ell n \, \varphi_k\right) / \sum_{j=1}^{J} \varphi_j\right] - \left[\left(\sum_{k=1}^{J} \varphi_k \sum_{j=1}^{J} \varphi_j \, \ell n \, \varphi_j\right) / \left(\sum_{j=1}^{J} \varphi_j\right)^2\right] = 0 \quad . \tag{4.72}$$

Since the $1/\alpha_k$ are all nonzero quantities of the same sign, we must have:

$$\lambda_{ND} = 0 \quad . \tag{4.73}$$

Putting this back into eq. 4.69, we get a much simplified equation. Since $\alpha_k \varphi_k$ is not zero, we must have:

$$\ln \varphi_k = \left(\sum_{j=1}^{J} \varphi_j \ln \varphi_j\right) / \sum_{j=1}^{J} \varphi_j \quad . \tag{4.74}$$

Because the right side is independent of k, the only solution to this set of equations is:

$$\varphi_k = \text{const} = K \, , \qquad \text{for all } k \in J \; . \tag{4.75}$$

This is identical in form to the result found in the case of equal α_k. But the solution for z_k^* is somewhat different:

$$z_k^{*ND} = (\ell n \, p_k)/\alpha_k - \left[\left(\sum_{j=1}^{J} (\ell n \, p_j)/\alpha_j\right) / \alpha_k \sum_{j=1}^{J} (1/\alpha_j)\right] + C/\alpha_k \sum_{j=1}^{J} (1/\alpha_j) \quad . \tag{4.76}$$

The difference between the solutions (equations 4.71 and 4.76) is a function of the α_k only, and does not depend on the prior probabilities p_k. Specifically:

$$\Delta_k \equiv z_k^{*P} - z_k^{*ND} = (1/\alpha_k) \left[\ell n \, \alpha_k - \left(\sum_{j=1}^{J} (\ell n \, \alpha_j)/\alpha_j\right) / \sum_{j=1}^{J} (1/\alpha_j)\right] . \tag{4.77}$$

Note that when all α_k are equal, both z_k^{*P} and z_k^{*ND} reduce to z_k (eq. 4.24), and the $\Delta_k = 0$ for all k.

Equations 4.71 and 4.76 can be used to calculate the optimum allocation of search effort in Pollock's modified example. The analytic solutions are, for P_D:

$$z_1 = 1.497$$
$$z_2 = 11.334$$
$$z_3 = 17.169$$

and, for H_{ND}:

$$z_1 = .996$$
$$z_2 = 11.126$$
$$z_3 = 17.878 \quad .$$

These values should be compared with those found in section 3 to be the maximum points on the integer grids: (2, 11, 17) and (1, 11, 18), respectively.

It is significant that the analytically determined maxima of P_D and H_{ND} are fairly close. With α_1 large, Pollock's example represents a major departure from a uniform probability of conditional detection. Nevertheless, the maxima show only a small displacement from each other, although they show a large displacement from their locations when all the α_j values are equal. (A representative example, using Pollock's p_j, but taking $\alpha_j = \alpha = .693$ for all j, yields $z_1 = 8.61$, $z_2 = 10.19$, $z_3 = 11.20$). Thus, in practical situations, optimizing H_{ND} will lead to search allocation nearly equal to that required for optimum detection probability.

These solutions, too, are valid for large search efforts only. The thresholds for their validity are determined from equations 4.68 and 4.75, using arguments analogous to those previously given. The results are, for P_D:

$$C^* = \sum_{j=1}^{J} (\ln \alpha_j p_j)/\alpha_j - \left(\sum_{j=1}^{J} (1/\alpha_j)\right) \ln [\alpha p]_{min} \tag{4.78}$$

and, for H_{ND}:

$$C^* = \sum_{j=1}^{J} (\ln p_j)/\alpha_j - \left(\sum_{j=1}^{J} (1/\alpha_j)\right) \ln p_{min} \tag{4.79}$$

Of greater theoretical significance than the z_k are the relative fractions of the total search effort assigned to each cell:

$$\zeta_k = z_k/C \quad .$$

These are trivally determined from equations 4.71 and 4.76:

$$\zeta_k^{*P} = \left(1/\alpha_k \sum_{j=1}^{J} (1/\alpha_j)\right) + (1/C) \left[(\ln \alpha_k p_k)/\alpha_k - \left(\sum_{j=1}^{J} (\ln \alpha_j p_j)/\alpha_j\right)/\alpha_k \sum_{j=1}^{J} (1/\alpha_j)\right] \tag{4.80}$$

$$\zeta_k^{*ND} = \left(1/\alpha_k \sum_{j=1}^{J} (1/\alpha_j)\right) + (1/C) \left[(\ln p_k)/\alpha_k - \left(\sum_{j=1}^{J} (\ln p_j)/\alpha_j\right)/\alpha_k \sum_{j=1}^{J} (1/\alpha_j)\right] \tag{4.81}$$

These clearly converge to the same asymptotic limit as $C \to \infty$

$$\zeta_{k\infty}^{*P} = \zeta_{k\infty}^{*ND} = 1/\alpha_k \sum_{j=1}^{J} (1/\alpha_j) \tag{4.82}$$

No simple strategem is available for dealing with equation (4.70), and closed-form solutions for the optimum search allocation do not appear possible in the case of H_E . Some limited statements can be made about the nature of the solutions, however. If we eliminate λ_E from any two equations in eq. 4.70, and use ζ_i rather than z_i , we get:

$$\left[\alpha_i p_i \exp(-\alpha_i \zeta_i C)\right] \ell n \left[\left(p_i \exp(-\alpha_i \zeta_i C)\right) / \sum_{j=1}^{J} p_j \exp(-\alpha_j \zeta_j C)\right]$$
$$= \left[\alpha_k p_k \exp(-\alpha_k \zeta_k C)\right] \ell n \left[\left(p_k \exp(-\alpha_k \zeta_k C)\right) / \sum_{j=1}^{J} p_j \exp(-\alpha_j \zeta_j C)\right] . \tag{4.83}$$

This can be further manipulated into the form:

$$(\alpha_i/\alpha_k) = \left\{\left[(p_i/p_k) \exp\left(-C(\alpha_k\zeta_k - \alpha_i\zeta_i)\right)\right] \cdot \ell n \left[\sum_{j=1}^{J} (p_j/p_k) \exp\left(-C(\alpha_j\zeta_j - \alpha_k\zeta_k)\right)\right] \right.$$
$$\left. / \ell n \left[\sum_{j=1}^{J} (p_j/p_i) \exp\left(-C(\alpha_j\zeta_j - \alpha_i\zeta_i)\right)\right]\right\} . \tag{4.84}$$

Since the left side of this equation is independent of C, the right side must be, as well, if the equation is to remain valid in the asymptotic limit $C \to \infty$.

Because the α_k values are given and independent of C and because, by definition, $0 \le \zeta_i \le 1$, then $\alpha_i \zeta_i$, must be $O(C^0)$ in its leading time. The most general form that satisfies these requirements is:

$$\zeta_i = \left[1/\alpha_i \sum_{j=1}^{J} (1/\alpha_j)\right] + (f_i/\alpha_i C) \tag{4.85}$$

This renders eq. 4.47 independent of C, to all orders, but still makes ζ of order C^0. The functions f_i are functions of the α_j, and p_j; they must satisfy:

$$\left[\alpha_i p_i \exp(-f_i)\right] \ell n \left[\left(p_i \exp(-f_i)\right) / \sum_{j=1}^{J} p_j \exp(-f_j)\right] = \lambda \tag{4.86}$$

and

$$\sum_{j=1}^{J} (f_j/\alpha_j) = 0 \tag{4.87}$$

An explicit solution for the f_i of eq. 4.86 is no more feasible than an explicit solution for the φ_i of eq. 4.70. But, by the artifice of eq. 4.85, we have separated out the asymptotic part of the solution. This leads to the very important Theorem IV:

Theorem IV: Under the assumptions of section 2.1, the allocations of the fraction of search effort among the J search cells that produce stationary points of P_D, H_{ND}, and H_E have a common asymptotic limit as the amount of search effort C approaches infinity. That limit is:

$$\zeta_i^{\infty} = \left[1/\alpha_i \sum_{j=1}^{J} (1/\alpha_j) \right] \qquad (4.88)$$

The final topic to address in this section is whether the extrema are maxima or minima in the case of differing α_j. This is easily answered in the case of P_D and H_{ND}. Combining equations 4.25 and 4.68, we get:

$$\delta^2 P_D \sim -\lambda_P (\alpha_k + \alpha_J) \qquad \text{for all} \quad k \in J \; . \qquad (4.89)$$

λ_P, α_k, and α_J are all positive, and P_D, therefore, always has a maximum.

Similarly, when we combine equations 4.26 and 4.75:

$$\delta^2 H_{ND} \sim -2\alpha_k \alpha_J / J^2 - (\alpha_k^2 + \alpha_J^2) \left[(1/J) - (1/J^2) \right] \qquad (4.90)$$

This is always negative for positive integer J; therefore, H_{ND}, too, always has a maximum.

No general rule appears appropriate for application to $\delta^2 H_E$. Though analytical results are not forthcoming, we have already seen in the numerical examples that maxima, minima, and saddle points are all possible as the stationary points of the H_E surface.

5. ANALYSIS--INFINITESIMAL SEARCH EFFORT

We turn now to the opposite limit, in which the total amount of search effort approaches zero. In particular, we consider the case in which an infinitesimal amount of search effort Δz_k is applied in only one cell, $k \in J$. We then apply the usual limiting processes of calculus to compute the rates of change of P_D, H_{ND}, and H_E as a result of searching in cell k.

The quantities A and B, defined in equations 4.1 and 4.2 are, for search in a single cell:

$$A = -H_O - (p_k \ln p_k)[1 - \exp(-\alpha_k \Delta z_k)] - \alpha_k \Delta z_k p_k \exp(-\alpha_k \Delta z_k) \quad (5.1)$$

$$B = 1 - p_k [1 - \exp(-\alpha_k \Delta z_k)] \quad (5.2)$$

where H_O has been defined in eq. 2.17. When A and B are expanded to first order in Δz_k and substituted in equations 4.3, 4.4, and 4.5, we obtain:

$$P_D = (\Delta z_k)\alpha_k p_k + O(\Delta z_k^2) \quad (5.3)$$

$$H_{ND} = H_O + (\Delta z_k)\alpha_k p_k (H_O + \ln p_k) + O(\Delta z_k^2) \quad (5.4)$$

$$H_E = H_O + \Delta z_k \alpha_k p_k \ln p_k \quad . \quad (5.5)$$

Thence:

$$\lim_{\Delta z_k \to O} (P_D - P_O)/\Delta z_k = dP_D/dz_k = \alpha_k p_k \quad (5.6)$$

$$\lim_{\Delta z_k \to O} (H_{ND} - H_O)/\Delta z_k = dH_{ND}/dz_k = \alpha_k p_k (H_O + \ln p_k) \quad (5.7)$$

$$\lim_{\Delta z_k \to O} (H_E - H_O)/\Delta_k = dH_E/dz_k = \alpha_k p_k \ln p_k \quad . \quad (5.8)$$

Our basic concerns will be with the magnitude of these rates of change, and with the extreme values of the magnitudes over the set J; we shall determine these extreme values. First, however, some observations about the signs are significant. Recalling that $0 \leq \alpha_k < \infty$, and $0 \leq p_k \leq 1$ for all $k \in J$, we note immediately that:

$$dP_D/dz_k \geq 0 \; ; \qquad \text{all} \quad k \in J \tag{5.9}$$

$$dH_E/dz_k \leq 0 \; ; \qquad \text{all} \quad k \in J \quad . \tag{5.10}$$

The corresponding result for dH_{ND}/dz_k is less obvious. However, the reader may persuade himself by selected numerical examples that dH_{ND}/dz_k can be either positive or negative, depending on the set of the prior probabilities p_j and on the searched cell, k. (Alternatively, we may note that $\sum_{k=1}^{J} (1/\alpha_k)\, dH_{ND}/dz_k = 0$, implying that the signs of dH_{ND}/dz_k are not the same for all k.

The observation concerning dP_D/dz_k is scarcely surprising. But those related to the entropies are significant:

Theorem V: For any search operation conducted according to the assumptions of section 2.1, the expected entropy never increases (and the expected information never decreases), regardless of the cell chosen for search. The entropy that is conditioned on nondetection of the target, however, may either increase or decrease, depending on local conditions.

It is worth pointing out that the rates of change of P_D and H_E depend entirely on local conditions, i.e., on the prior probability p_k in the cell being searched; the rate of change of H_{ND}, on the other hand, depends globally on the entire prior ensemble, through H_O. It is therefore conceivable that for two distinct prior ensembles, the rates of change of H_{ND} could vary in both magnitude and sign, even though the individual cells chosen for search in the two cases had identical prior probabilities.

We now determine whether the results obtained in section 4 still apply in the limit $\Delta z_k \to 0$. Specifically, does the allocation that maximizes P_D also maximize H_{ND} and produce an extremum (either maximum or minimum) in H_E?

We have already noted that in the case in which the α_k are different in different cells, no results of useful generality were obtained. We therefore restrict ourselves to the case $\alpha_k = \alpha$, all $k \subset J$, and -- for simplicity in this section -- let $\alpha = 1$. Then, the rates of change for P_D, H_{ND}, and H_E are:

$$p_k, p_k(H_O + \ln p_k), \text{ and } p_k \ln p_k ,$$

respectively.

Let us assume that the cells are numbered in order of nonincreasing magnitude of p_k: $p_1 \geq p_2 \geq - - - \geq p_J \geq 0$. By this device, we make sure that infinitesimal search in cell 1 will produce the highest rate of increase in P_D, and, hence, maximize P_D in the limit $\Delta z \to 0$. We then investigate the extrema over the set J for the rates of change of H_{ND} and H_E.

$\underline{H_{ND}}$

We noted earlier that the sign of dH_{ND}/dz_k might be either positive or negative, depending on the prior ensemble and on the searched cell, k. This general observation can be made somewhat more precise. Some preliminary observations:

1. For a given prior ensemble, H_O is a fixed positive constant.
2. The factor p_k is monotonically nonincreasing with the index k because of the assumed ordering. It is always > 0.
3. The factor $H_O + \ln p_k$ is monotonically nonincreasing with the index k from some initial value, and may approach $-\infty$ for a p_k that is small enough. Whether the maximum value of $H_O + \ln p_k$ is positive or negative remains to be determined.

We shall later prove that the max of $H_O + \ln p_k$ is in fact nonnegative for any prior ensemble. For the moment, we assume that this is true and derive the consequences.

By the assumption, there is a range of k, $1 \leq k \leq K < J$, for which $H_O + \ln p_k \geq 0$. Then, in this range, $p_k(H_O + \ln p_k)$ is a positive, monotonically nonincreasing function of k, because it is a product of two positive, monotonically nonincreasing functions of k.

In this range:

$$p_1(H_O + \ln p_1) \geq p_k(H_O + \ln p_k) \tag{5.11}$$

for $1 < k \leq K$. Outside this range $H_O + \ln p_k < 0$, and the monotonicity of the product is no longer assured. Since the product is then negative, however, it is obviously less than $p_1(H_O + \ln p_1)$. Thus, if $p_1 \geq p_k$ for any $1 < k \leq J$, then $p_1(H_O + \ln p_1) \geq p_k(H_O + \ln p_k)$ for any $1 < k \leq J$.

Now we prove the assumption. We need to show that for any prior distribution, there is at least one p_k such that $H_O + \ln p_k$ is nonnegative. This is easily done by noting that:

$$H_O + \ln p_k = \sum_{j=1}^{J} p_j(-\ln p_j + \ln p_k) = \sum_{j=1}^{J} p_j \left(\ln(p_k/p_j)\right) . \tag{5.12}$$

If p_k is $\underset{J}{\text{Max}}\ p_j$, then each term in the sum is nonnegative, and

$$H_O + \ln p_{max} \geq 0 . \tag{5.13}$$

One final comment about dH_{ND}/dz_k. We have seen that, for the largest p_k in J, $dH_{ND}/dz_k \geq 0$; in addition, for sufficiently small p_k values, $dH_{ND}/dz_k \to 0^-$. This sort of behavior suggests that for some intermediate value of p_k, dH_{ND}/dz_k has a relative minimum < 0 (which, in fact, is an absolute minimum, too). This is best illustrated by a specific model.

Consider a countably infinite set of search cells, in which the (ordered) prior probabilities are:

$$p_k = (1 - \gamma)\ \gamma^{k-1} . \tag{5.14}$$

Clearly $\sum_{k=1}^{\infty} p_k = 1$ when the sum converges ($0 \leq \gamma < 1$).

This model is a useful tool for investigating some situations involving finite J. General analytical results can be obtained with the infinite model, yet that model can be made to approximate a finite model reasonably well by truncation of the series after J-1 terms and lumping all the remaining probability $\sum_{k=J}^{\infty} p_k$ together as the Jth term. For $\gamma < 1/2$, the total residual probability in (J, ∞) is less than the J-1 term, and the decreasing ordering is preserved.

For this model, with $\alpha_k = \alpha = 1$, we can show that:

$$dH_{ND}/dz_k = (1/\gamma)|\ln \gamma| \exp(-k|\ln \gamma|)[1 - k(1-\gamma)] \quad . \tag{5.15}$$

This expression is positive for small k and negative for large k. The crossover point is given by:

$$k^{+}(\gamma) \approx 1/(1-\gamma) \tag{5.16}$$

which is monotonically increasing from 0 to ∞ as γ goes from 0 to 1^{-}. Thus, by choice of the parameter γ, we can make an arbitrarily large number of cells have positive values of dH_{ND}/dz_k.

Treating k as a continuous variable and differentiating, we find that:

$$\partial(dH_{ND}/dz_k)/\partial k = 0 \qquad \text{for} \quad k = k^{*} \ , \tag{5.17}$$

where:

$$k^{*}(\gamma) = 1/(1-\gamma) + 1/(|\ln \gamma|) \quad . \tag{5.18}$$

This can be shown to be a minimum of dH_{ND}/dz_k. Like $k^{+}(\gamma)$, $k^{*}(\gamma)$ is also a monotonically increasing function of γ, going from 1 to $+\infty$ as γ goes from 0 to 1^{-}. The position of the minimum can also be situated at arbitrarily large values of k, depending on the choice of γ, although clearly the choices of $k^{+}(\gamma)$ and $k^{*}(\gamma)$ are not independent of each other.

$\underline{H_E}$

An analysis of the relative magnitudes over $\{k\}$ of $p_k \ln p_k$ is closely related to a problem already solved by Browning (reference 11). Browning treats a slightly more complicated case, in which the amounts of search effort are finite ("looks") rather than infinitesimal. That makes his proofs more complex than the present case requires, but it does not alter their validity or their relevance to the infinitesimal limit.

We shall simply state the necessary results with qualitative justifications. The reader may consult Browning for analytical proofs.

We shall consider the rate of change of the expected information, rather than the expected entropy: $dI_E/dz_k = -p_k \ln p_k \geq 0$. Figure 12 shows the function $y(p) = -p \ln p$ in the domain $0 \leq p \leq 1$. This curve is asymmetrical, having a maximum at $(1/e, 1/e)$. We must visualize along this curve the set of points $y_k = y(p_k)$, subject to $\sum_{k=1}^{J} p_k = 1$, and to the ordering $p_1 \geq p_2 \geq p_3 \geq \cdots \geq p_J \geq 0$.

Browning has shown that the following cases should be distinguished:

1. $1/e \geq p_1$. In this case all the y_k are on the increasing part of the curve $y(p)$; therefore, the ordering of the p_k guarantees the ordering of the y_k: $y_1 \geq y_2 \geq y_3 \geq \cdots \geq y_k \geq 0$.
2. $1/e \leq p_2$. In this case, which can still satisfy the constraint $\sum_{k=1}^{J} p_k = 1$, y_1 is on the decreasing segment of $y(p)$; therefore, $y_2 \geq y_1$.
3. $p_2 \leq 1/e \leq p_1$. This case is problematical and is best resolved by numerical calculation.

Before turning to relevant calculations, we must resolve the question whether y_3 could be the largest of the y_k under any conditions. Reference to the figure shows this is obviously not possible if $p_2 \leq 1/e$, since then $y_3 \leq y_2$. On the other hand, if both p_1 and p_2 are $\geq 1/e$, then the maximum value possible for p_3 is $(1-2/e) < 1/e$, attained when $p_1 = p_2 = 1/e$. There again $y_3 < y_1, y_2$. Any increase in p_1 and p_2 above $1/e$ can occur only at the expense of p_3, which further decreases y_3 at a rate faster than either y_1 or y_2 because of the steeper slope for $p < 1/e$. Thus, there are no conditions

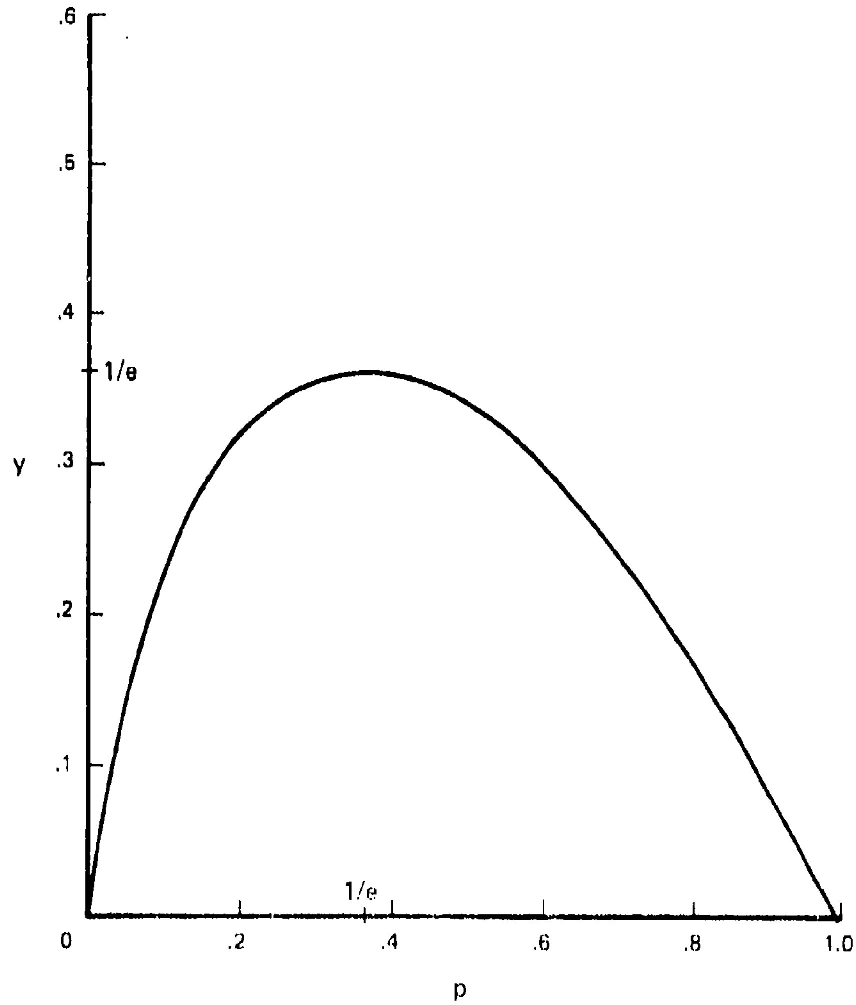

FIG. 12: GRAPHICAL ILLUSTRATION OF THE FUNCTION:
$y = -p \ln p$

for which $y_3 = \max_k y_k$; we need therefore consider only cells 1 and 2 when allocating search effort to produce the maximum rate of information increase.

Figure 13 shows the relevant portion of the p_1, p_2 plane. The boundary curves are $p_1 + p_2 = 1$ (the limiting case of $\sum_{k=1}^{J} p_k = 1$), $p_1 = p_2$ (the limiting case of the assumed ordering $p_1 \geq p_2$), and $y(p_1) = y(p_2)$. In region I, $y_1 > y_2$; in region II, $y_2 > y_1$.

It is important to note that the line $p_1 + p_2 = 1$ is a boundary of region II only. This implies that for the case $J=2$, $y_2 > y_1$, except for the endpoint $y_2(0) = y_1(1) = 0$. When there are only two search cells, the maximum rate of increase of information is always attained by searching the cell with the lower prior probability, and the minimum rate of increase of information is always attained by searching the cell with the higher prior probability. This is in agreement with the results obtained for the asymptotic case $C \to \infty$.

When $J > 2$, however, a new feature enters. There is a significant region in the p_1, p_2 plane, where the maximum rate of information increase is attained by searching in the cell with the second highest prior probability. The cell with the highest prior probability, which is searched to maximize the growth of P_D, does not represent either a maximum or a minimum in the rate of information growth. The results of the asymptotic analysis, therefore, do not carry over into the infinitesimal limit, and we cannot assert universally that the allocation that maximizes P_D also produces an extremum in H_E (or I_E).

Reference to our infinite-number-of-cells model will illustrate some of these points. For that model:

$$dI_E/dz_k = -(1/\gamma)(1-\gamma)\exp(-k|\ell n\ \gamma|)[\ell n(1-\gamma) + (k-1)\ \ell n\ \gamma] , \qquad (5.19)$$

which is, of course, always ≥ 0 for any k. Again, treating k as a continuous variable, we find that dI_E/dz_k has a maximum at:

$$k_E^*(\gamma) = 1 - [\ell n(1-\gamma)]/\ \ell n\ \gamma - 1/\ \ell n\ \gamma \quad . \qquad (5.20)$$

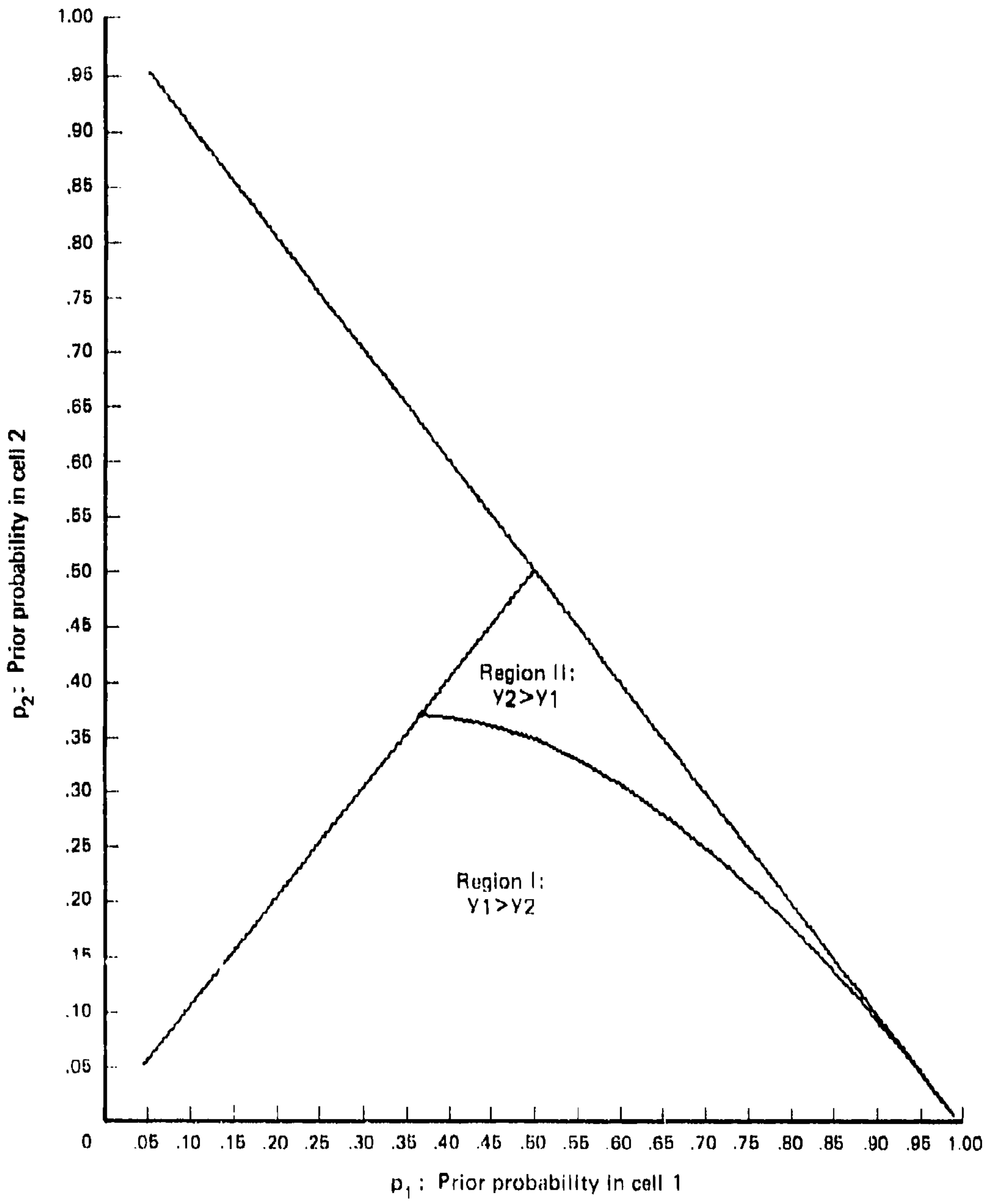

FIG. 13: DELINEATION OF REGIONS IN THE SPACE OF PRIOR PROBABILITIES

The behavior of $k_E^*(\gamma)$ differs significantly from that of $k_{ND}^*(\gamma)$. As γ goes from 0 to 1^-, $k_E^*(\gamma)$ starts at 1, increases slightly to a maximum of ~ 1.5422 at $\gamma \cong .351575$, then decreases monotonically toward $-\infty$. Of the positive integers, only k = 1 or 2 are closest integers to this curve for any γ, as suggested by our previous discussion. Thus, only cells 1 or 2 need be considered candidates for $\max_k \{dI_E/dz_k\}$.

The results of this section are summarized as follows. For ordered probabilities $p_1 \geq p_2 \geq p_3 \geq \cdots \geq p_J \geq 0$:

a. The rate of change of P_D is always positive. It has a maximum for search in cell 1 and a minimum for search in cell J.

b. The rate of change of H_{ND} may be either positive or negative. It has a nonnegative maximum for search in cell 1, and a negative minimum for search in some intermediate cell m: $1 < m < J$.

c. The rate of change of I_E is always positive. If J=2, it has a maximum for search in cell 2 and a minimum for search in cell 1. If $J > 2$, then, either (1) it has a maximum for search in cell 1 and a minimum for search in cell J, or (2) it has a maximum for search in cell 2 and a minimum for search in either cell J or cell 1.

6. SUMMARY

The relation between search theory and information theory remains complex.

Previous attempts to attack the problem have focused on only a narrow part of the relationship. Their narrow focus and the lack of clear definition of "information" or "entropy" created an unwarranted impression that the body of work was contradictory, if not incorrect. It is not.

The work reported in the earlier references is both correct and, when viewed from a broader perspective, consistent. The individual results reported earlier have been useful in establishing this broader perspective.

In the present paper, some of these earlier results have been rederived from a different viewpoint. Other new results have been added. Taken together, these findings are enough to allow us to sketch a coherent, though still incomplete, picture of the relation between search theory and information theory.

Equally important, we can now identify the gaps somewhat more clearly and can direct future research toward those gaps with the confidence that there is an underlying body of theory to be completed, rather than a mere collection of isolated observations.

Before summarizing the findings, it is important for us to recapitulate the assumptions. These are the most critical assumptions of our model, which are used consistently throughout this paper: There is a single target; it is stationary; no false detections occur; and the detection process is governed by an exponential detection function. Relaxation of any of these assumptions would add considerable complexity to the theory and might change the conclusions significantly. Within the framework of these general assumptions, this paper has examined subcases in which (a) the amount of search effort is either very large or very small, and (b) the conditional detection probability is either uniform or variable throughout the search cells.

Summary of Findings

a. Uniform conditional detection probability.

1. The allocation of search effort that maximizes the posterior entropy conditioned on nondetection H_{ND} also maximizes the probability of detection P_D. This result was first established by Barker and was confirmed by the present author in both the large and small search effort limits. Since Barker's proof is not restricted to these limiting cases, the result is valid in the intermediate range as well.

2. In the large search effort limit, the allocation of search effort that maximizes the probability of detection, P_D also produces a local extremum in the expected information I_E. For two search cells, this extremum is a local minimum of I_E. For any number of search cells greater than two, this extremum is a local maximum of I_E. This is consistent with Mela's calculations. (Although the number of looks is only equal to J in Mela's examples, this still corresponds to the large search effort limit, since $C^* = 0$ by eq. 4.22.

3. The local extremum of I_E is not necessarily the global extremum. For an intermediate range of C, $C^* \leqq C \leqq C_{EQ}^{MIN}$, the local extremum is also the global extremum. For $C_{EQ}^{MIN} < C$, there is always at least one whereabouts search (and there may be as many as J), corresponding to an edge point of the I_E surface, which produces more expected information than the optimal detection search. In practical cases the distinction appears unimportant, since $P_D \simeq 1$ and $I_E \simeq I_{MAX}$, when C reaches C_{EQ}^{MIN}.

4. In the small search effort limit, the exact correspondence between the maximum of P_D and the extremum of I_E breaks down. We achieve the maximum rate of increase of I_E by searching in the cell that has either the largest or second largest rate of increase of detection probability. In the two-cell case, this is at least consistent with the large search effort limit. For more than two cells, however, this is a distinct phenomenon, constituting another demonstration that the connection between optimum search and optimum information gain is not universal. Working with P_D and I_E directly, rather than with rates of change, Browning has obtained comparable results in the intermediate search effort region for cases J=2 and J=3. There remains an open question: precisely how the transition between the two limits takes place and whether the C^* determined in equation 4.22 is in fact the minimum value for which the large search effort limit is valid.

5. The rate of change of I_E is always positive for any search allocation. The rate of change of H_{ND} can be either positive or negative, depending both on the search allocation and on the ensemble of prior probabilities in the search cells.

b. Variable conditional detection probability.

1. In the large search effort limit, both P_D and H_{ND} have local maxima. The search allocations leading to those maxima are not the same, as is demonstrated by eq. 4.77. This finding is consistent with Barker's second theorem.

2. In the large search effort limit, no closed-form solution for the stationary points of I_E can be determined. Even the nature of the stationary points cannot be established analytically, but the numerical examples have shown that maximum, minimum, and saddle point are all possibilities. In the asymptotic limit ($C \to \infty$), the positioning of the stationary points of P_D, H_{ND}, and I_E <u>all</u> converge to a common limit:

$$(C/\alpha_i) \ (1/\sum_{j=1}^{J} (1/\alpha_j))$$

3. In the small search effort limit, no systematic relationships have been established among the maxima of P_D, H_{ND}, and I_E.

4. That the rate of change of I_E is positive for any search allocation, even when the conditional detection probabilities vary, remains valid.

<u>Interpretation</u>

Only in the case of uniform conditional detection probability do we have enough results to attempt some interpretation. There, however, some simple statements can do much to dispel the legacy of past confusion.

First, using Barker's theorems and our present results on the rates of change of P_D and H_{ND}, we have noted that identical search policies produce the maximum increases of both detection probability and posterior conditional entropy. Like all entropies, H_{ND} is a negative information. Thus, maximizing its rate of increase is equivalent to maximizing the rate of decrease of an information. The information that is being decreased is that contained in the ensemble up to the time of detection: the information about the target location that is expressed by the prior probabilities and that changes as the search progresses. With this view, we can state one general conclusion, already correctly anticipated by Richardson:

- The search policy that maximizes the probability of detection is the one that uses up the information contained in the prior ensemble at the maximum rate.

Second, we have shown in this paper that in most cases the optimum policies for P_D and I_E are equivalent. Here, however, information is being created rather than used up. Because this is expected information, visualizing the overall process of information creation may be harder. Consider a repeated search in one of the J cells. If that cell has a prior probability p_k of containing the object, its initial self information is $\ln p_k$. At each stage of the search, if detection does not occur, p_k decreases; consequently, the actual self information of the cell also decreases. The probabilities of the other J-1 cells, however, increase simultaneously (because of the constraint $\sum_{j=1}^{J} p_j = 1$), as do their values of self information. If detection does occur in the kth cell after some number of unsuccessful searches, its self information undergoes a sudden increase; the self information of the other cells then drops.

When we speak of ensemble averages, as we have throughout this paper, the competing effects of these information-creating and -destroying processes are combined, with a probabilistic weighting. Until the time of detection, the loss of information caused by failure to detect in the cell being searched is weighed against the inferred gain in information in the other cells. The net change of information may be either positive or negative, as we noted in our discussion of the rate of change of H_{ND}. If detection occurs, the positive and negative information contributions from that process must also be added in, with the proper weighting. Then, as indicated by the rate of change of I_E, the overall information gain is net positive.

This interpretation of the creation and destruction of information leads to our second general conclusion:

- In a wide range of practical cases, the search policy that maximizes the detection probability is the one that creates expected information at the maximum rate. We achieve that maximum rate of creation by using the existing information of the prior ensemble as fast as possible, thereby gaining an early detection, with its concomitantly large increase of expected information.

Except for the anomalous cases, these two statements contain the essence of the interpretation of optimum search: We consume existing information at the maximum rate in the expectation of gaining still more information, also at the maximum rate.

The anomalous cases are of three types: (1) the case of two search cells, in which maximum information gain is achieved by search in the cell yielding the lower detection probability, regardless of the amount of search effort applied and regardless of the probabilities in the prior ensemble; (2) the case of three or more search cells, in which maximum information gain is achieved by whereabouts search rather than optimal detection search for very large amounts of search effort; and (3) the case of three or more search cells, in which maximum information gain is achieved by searching the cell that yields the second highest detection probability, but only for small amounts of search effort, and only for special regions in the space of prior probabilities.

The anomaly in the first two cases appears to be caused by the two competing methods of increasing information: by actual detection in the searched cells, or by inference that the target is in the other cell, based on failure to detect in the searched cells. These two mechanisms are at work regardless of the number of cells. When there are only two cells, however, failure to detect in one leads to a large increase in the probability that the target is in the other, and, hence, much more information about the location of the target.

This mechanism of inferred information is dominant in the two-cell case, so that actual detection is not necessary to achieve the maximum information gain, even with small amounts of search effort. The two-cell search is thus an exemplary case of the whereabouts search. For three or more cells, the inferred information mechanism of the whereabouts search is not normally dominant; it becomes important only at large values of the search effort.

In the second case, the anomaly appears to be more dependent on the mathematical structure of the ensemble average information -- on the form $p \ln p$. In those areas of the (p_1, p_2) plane where the anomoly exists (figure 13), p_1 is large, and the amount of self-information to be gained by a detection in cell 1 is small. Detection in cell 2, although less likely, provides more self-information. When incorporated in the ensemble average, the greater but less likely expected information contribution from cell 2 -- $p_2 \ln p_2$ -- dominates the smaller but more likely contribution from cell 1 -- $p_1 \ln p_1$.

Future Research

Finally, we need to set out the areas in which further research may prove most fruitful.

The principal case treated here (uniform conditional detection probability) has broad practical applications. Its basic structure has been laid out, and only minor holes remain. Clearly, there are unsolved questions relating to the transition between the small and large search effort limits in the case of I_E. These are important theoretical questions, but less so in a practical sense.

The major extensions needed for use in real world search theories are, first, to false targets, second, to multiple targets, both real and false, and finally, to moving targets. Each of these cases is of immense practical significance, and any theory that does not include them can have little claim to completeness.

Relaxation of the assumption of exponential detection functions, and any further considerations of the case of varying conditional detection probabilities can be deferred on the grounds of lower practical priority, despite their considerable theoretical interest and complex mathematical challenge.

REFERENCES

1. B.O. Koopman, "Search and Information Theory: Part of Final Report on Stochastic Processes in Certain Naval Operations," Columbia University, Dept. of Mathematics Report, 1967

2. D.F. Mela, "Information Theory and Search Theory as Special Cases of Decision Theory," Operations Research 9, 907-909, 1961

3. S.M. Pollock, "Search Detection and Subsequent Action: Some Problems on the Interfaces," Operations Research 19, 559-585, 1971

4. H.R. Richardson, "ASW Information Processing and Optimal Surveillance in a False Target Environment," Daniel H. Wagner, Assoc. Report, Oct 1973

5. W.H. Barker, II, "Information Theory and the Optimal Detection Search," Operations Research 25, 304-314, 1977

6. R.G. Gallagher, "Information Theory and Reliable Communications," John Wiley and Sons, Inc., New York, p. 16, 1968

7. Gallagher, op. cit., p. 21

8. B.O. Koopman, "Search and Screening," OEG Report No. 56, Center for Naval Analyses, 1946

9. L.D. Stone, "Theory of Optimal Search," Academic Press, New York, 1975

10. J.B. Kadane, "Optimal Whereabouts Search," Operations Research 19, 894-904, 1971

11. W.J. Browning, "Maximum Information Gain Search Policies," Daniel H. Wagner, Assoc. Report, Jul 1976