AD-770 741

JUDGMENTAL METHODS FOR SCORING ASW
EXERCISES

Ward Edwards

Michigan University

Prepared for:

Office of Naval Research

26 November 1973

AD 770 741

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
010766-1-T		

4. TITLE (and Subtitle)	5. TYPE OF REPORT & PERIOD COVERED
JUDGMENTAL METHODS FOR SCORING ASW EXERCISES	Technical
	6. PERFORMING ORG. REPORT NUMBER
	None

7. AUTHOR(s)	8. CONTRACT OR GRANT NUMBER(s)
Ward Edwards	N00014-67-A-0181-0046

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Engineering Psychology Laboratory The University of Michigan Ann Arbor, Michigan 48105	NR 364-071

11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE
Office of Naval Research 800 North Quincy Street Arlington, Virginia 22217	26 November 1973
	13. NUMBER OF PAGES
	44

14. MONITORING AGENCY NAME AND ADDRESS (if different from Controlling Office)	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

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U S Department of Commerce
Springfield VA 22151

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

Bayes's theorem
Informational consequences
ASW
Likelihood ratios

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

An application of Bayesian judgmental methods to the development of an informational scoring procedure for ASW exercises is presented. The ASW exercise is conceptualized as an informational game, both sides wanting to acquire as much information as possible about the opponent while denying the opponent any useful information. Techniques for measuring, cumulating, and scoring the informational consequences of decisions made by each side in the course of an ASW exercise are discussed. Several obstacles remain between the present formulation and its application. Key problems are sketched and avenues for overcoming them are discussed.

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE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

Unclassified

JUDGMENTAL METHODS FOR SCORING ASW EXERCISES

Technical Report

26 November 1973

Ward Edwards[1]

Engineering Psychology Laboratory

The University of Michigan

Ann Arbor, Michigan

TABLE OF CONTENTS

iii.

Introduction

This report summarizes some work and thought looking toward the use of Bayesian judgmental methods for scoring ASW exercises. The work, though seriously incomplete, has progressed far enough so that both the bare bones of the idea and some of the obstacles that must be overcome can be sketched and discussed.

ASW can be thought of as an information game. The two sides have an objective in common. Each wants to acquire as much information as possible about the opponent while denying to the opponent any useful information. The weapons available are not highly dependent on actual position, in that the carrier's are air-borne and the sub's can be long-range. But they are critically dependent on knowledge of the opponent's position. Hence a goal of every move on both sides should be to obtain information or to deny it to the opposition, or both.

If so, every event can be scored according to its expected (or perhaps actual) success in obtaining and/or denying information. An act that contributes to the opponent's information is bad; an act that contributes to one's own information is good. Such indices of information value can, if expressed in suitable units, be cumulated over the sequence of actions that occurs during an exercise. Such a score might be useful both for overall assessment of performance and for a decision-by-decision critique.

This idea obviously depends for its usefulness on the availability of a suitable measure of information contribution. No objective measure is, in my opinion, possible. Information is a probabilistic concept, and the

relative frequencies that we usually think of as objective estimates for probabilities are, of course, not available for unique events. (And all events of realistic interest in ASW are unique.)

The Bayesian point of view in statistics (see for example Edwards, Lindman, and Savage, 1963) offers an approach to circumventing this problem. From that point of view, a probability is an orderly opinion, and hence is inherently subjective. Often such opinions may be based on relative frequencies or other objective data or calculations. Even more often, they are not.

"Orderly" here means that probabilities, even though they are opinions, must obey rules--in particular, the probabilities of an exhaustive set of mutually exclusive events must sum to 1. Since this rule implies all of the formal structure known as probability theory, the subjectivity implicit in the idea that probabilities are opinions is very severely limited, and becomes more so as evidence accumulates.

Bayes's theorem is an elementary, completely non-controversial fact about probabilities. Its importance arises because it is the appropriate formal rule that specifies how probabilities or opinions should be revised on the basis of new information.

Suppose you wish to know the probability of some hypothesis H_A subsequent to observing some relevant (or irrelevant) datum D. Bayes's theorem says:

$$P(H_A|D) = \frac{P(D|H_A)\ P(H_A)}{P(D)} \tag{1}$$

In Equation 1, $P(H_A)$ is the probability of H_A before datum D was observed, $P(H_A|D)$ is the revised probability of H_A after D has been observed, $P(D|H_A)$ is the probability that datum D would be observed if H_A were known to be true, and $P(D)$ is a normalizing constant that will drop out of subsequent applications.

Now, write Bayes's theorem again, this time for H_B, on the basis of the same datum.

$$P(H_B|D) = \frac{P(D|H_B)\ P(H_B)}{P(D)} \tag{2}$$

Next, divide Equation 1 by Equation 2. The result is

$$\frac{P(H_A|D)}{P(H_B|D)} = \frac{P(D|H_A)}{P(D|H_B)} \cdot \frac{P(H_A)}{P(H_B)} \tag{3}$$

Rewriting Equation 3 in simple notation, we have

$$\Omega_1 = L \quad \Omega_0 \tag{4}$$

Next, taking logarithms (base 10 is convenient, but it doesn't matter),

$$\log \Omega_1 = \log L + \log \Omega_0. \tag{5}$$

The quantities Ω_1 and Ω_0 are known at the racetrack as odds; of course Ω_0 is called the prior odds, since it is the odds before the datum is observed, and Ω_1 is called the posterior odds, since it is the revised odds after this

datum has been observed. Statisticians know L as a likelihood ratio.
If a second datum comes along, Ω_1 is the prior odds with respect to it,
and it generates a likelihood ratio that transforms Ω_1 into Ω_2. Equations
4 and 5 are the most useful versions of Bayes's theorem. Equation 5 is
formally more useful than Equation 4, since it is linear in form. However,
much research shows that people can estimate quantities like Ω and L better
than they can estimate logarithms of those quantities.

Highly trained estimators can estimate likelihood ratios directly.
However, for the application envisioned here, it is easier to estimate
Ω_n and Ω_{n-1} and then calculate

$$\log L_n = \log \Omega_n - \log \Omega_{n-1} \qquad (6)$$

Log L_n is a measure of the diagnosticity of the datum to which it refers
with respect to H_A and H_B. Diagnosticity, a technical term from Bayesian
statistical inference, refers to the impact that a datum has on the (log)
odds as between a pair of hypotheses; the name is chosen because the
intellectual process is essentially that of differential diagnosis in medicine.
The further away log L_n is from 0, in either direction, the more diagnostic
or informative datum n is with respect to that pair of hypotheses. Consequently,
log L_n is a natural basis for an information score to use in ASW exercises.

If more than two hypotheses are being considered, as will typically
be the case in ASW, it might be appropriate to take the largest value of

log L_n , with respect to all possible pairs of hypotheses, as the information score for the datum with which it is associated, or it might be appropriate to compare the correct hypothesis with the union of its alternatives, thus in effect reducing the list of hypotheses to two. The latter approach seems preferable, and is taken here.

Considerable experimentation and experience has shown that human estimates of odds can be the basis of the design of real world information processing systems. (See for example Edwards, Phillips, Hays, and Goodman, 1968; and Kelly and Peterson, 1971.) Knowledge of the technology of eliciting such judgments from experts is well developed.

The idea around which this work was structured, then, was that experts should estimate odds associated with appropriate hypotheses before and after the effects of decisions altered the information states of participants in ASW exercises, that from these estimates log likelihood ratios should be calculated, and that these log likelihood ratios should then be thought of as information scores for the decisions that gave rise to them. Most decisions by either side can be expected to have informational consequences for both sides. Roughly speaking, an act that provides information to the opposition must justify itself by providing even more information to the actor. This suggests a diagram like that in Figure 1, tracking gains of information by both sides in an exercise. In a typical exercise one would think of at least two such diagrams, one referring to the actions of

each side--or perhaps one for each decision maker. Of course it would
be perfectly feasible to aggregate these diagrams into a single one,
which would constitute a sort of running box score of the entire exercise.

Figure 1

Information for Blue Commander (log L)
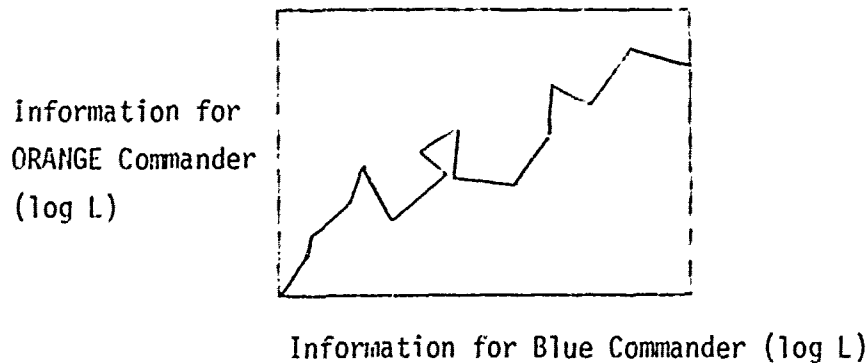
The very vague sketch in the preceding paragraphs conceals a
large set of difficulties; much of this report will be devoted to discussing
them. In order to gain some experience with possible applications of the
idea, I arranged to participate in an exercise of the UPTIDE series[2]. Most
of the remainder of this report is based on that experience.

Technical Issues in Information Scoring

Identifying hypotheses. When this approach to evaluating decisions
was originally proposed a serious problem was expected to result from the
fact that a log likelihood ratio requires two mutually exclusive hypotheses
for its specification. In an ASW exercise, what are the hypotheses?

In experience with the information scoring technique, this turned out
to be no problem. The hypotheses are all of the form: The HVU (or a submarine)

is located here; or, almost equivalently, lies in the following direction from my present location. The relative unimportance of geography to the weapons available means that such a hypothesis, if correct and if acted on, will lead to a phase of ASW in which the kinds of issues being discussed here are of relatively little importance.

The HVU and the submarines are in asymmetrical relation to each other with respect to this question. Since the HVU has the capability of dispersing its sensors and its weapons over a wide area, one can conceptualize its problem in looking for a submarine as a grid-search problem. Since the submarine, looking for the HVU, mounts both sensors and weapons at a central point, a polar representation of its problem is the natural one, and information about bearing is far more important than information about range. Moreover, for many purposes the submarine's bearing information need not be particularly accurate; $\pm$ 10° seems to be often entirely adequate in the early stages of hunting down a HVU[3]. So, it is often an adequate representation of the submarine's information acquisition to say that an initial uncertainty that produces a more-or-less flat distribution over 36 hypotheses has been reduced to a relatively peaked distribution over only 1 or 2. (If a boundary, shore, or other constraint on the HVU's location exists, the initial uncertainty may be less; 18 hypotheses would be a frequently encountered case.)

The HVU, searching for a submarine, can obtain, and may need, much more uncertainty reduction much quicker than that. It is not surprising that greater sensory and effector capability should lead to higher information

requirements--but it does present a problem for the information score, since it implies that there is a much higher upper bound on acquisition of _useful_ information for the HVU than for the submarine. In other words, it is not appropriate to compare directly the information gains by the HVU with those by the submarine, even though they are about the same sort of question and expressed in the same units.

Identifying information-producing acts. The most difficult technical problem in applying the information-scoring idea turns out to be the problem of identifying acts or decisions. The root of the problem is that the action of a ship is continuous, but the abstraction called "making a decision" implies an episodic, discontinuous character. For example, when the HVU changes course and/or speed, that is ordinarily an indication that a decision has been made. Sometimes, however, that decision was made hours before the change occurred. Sometimes it is preprogrammed. And of course decisions are quite often made _not_ to do something, and such decisions are difficult to recognize in any kind of log of observable events.

An obvious solution in principle is to ask the decision maker to identify his decisions as he makes them, and then simply to define a decision as having been made whenever the decision maker says he made one. This simply shifts the burden of recognizing decisions from the scorer to the decision maker. The shift is appropriate, since only the decision maker has access to his own mental processes. But he would need to be provided with criteria for recognizing that he has made a decision. The most obvious such criterion is that he makes a decision whenever he weights the advantages and disadvantages

of two or more courses of action (one of which typically is to do nothing
or to change nothing) and selects one of them. He would also need auxiliary
rules to guide him about such questions as whether two periods, ten minutes
apart, of considering a change of course, both resulting in the decision
not to change it, are one decision or two. Answers to questions of this
sort would of necessity be conventions rather than direct consequences
of the notion of a decision.

Asking the decision maker to label his decisions would work, and
would not even be difficult, if done intentionally and systematically in
an exercise. It is not feasible retrospectively, however. In UPTIDE 3B,
COMASWGRUTHREE was exceptionally helpful in discussing his decisions with
observers and his staff as he made them. Observers' logs were consequently
of great help in identifying his decisions. The observers' logs from
submarines are far less useful, and the CO's narratives don't help much
either.

Identifying the informational consequences of acts. A HVU decides
to proceed at a certain course and speed, and does so. Three hours later,
it comes into detection range of a submarine. In due course the submarine
detects, pursues, identifies, and sinks it. Was the detection and identifica-
tion an informational consequence of the original decision to adopt the
specified course and speed?

The problem here is that the mathematics of information scoring
behaves as though not only decisions but also information is episodic. It
cannot easily handle informational consequences of an action when they unfold

in time--especially when the alternatives to that action are ill specified. (After all, the HVU had to use some course and speed, or some sequence of them.)

The problem is sometimes easy to handle. A decision to speed up to above-cavitation speeds, followed by a detection at distances such that it would probably not have occurred at slower speeds, seems clearly to be a case in which the decision produced the informational consequence. The issue is one of the closeness of linkage between action and informational consequence.

It may be that this issue requires further development of the information score idea. Perhaps the informational "consequences" of an action should be weighted by the directness or inevitability of these consequences, in generating the information score. No experience yet exists with such judgments, though I did in fact use just this kind of thinking, in a less-than-precise way, in working over the UPTIDE 3B data.

A related question is: is it information if nobody notices it? In UPTIDE 3B there were several instances in which decisions by the HVU led to sensory consequences for a submarine, but these sensory inputs were dismissed as less likely to be useful (or less relevant to the submarine's current problems) than others. How should such potential consequences be scored when the potential doesn't actually materialize? The question has complex ramifications. Suppose that an informational consequence does not materialize because some sensor on an opposing vessel malfunctions, or is turned off, or something of the sort. What then?

Such questions must be answered by convention. A natural convention to explore is that the actual, not the potential, consequences should be scored. This enormously simplifies the scoring, and greatly reduces the need to ask and answer hypothetical "What if. . ." questions.[4]

Procedures for estimating likelihood ratios. A few conventions for estimating likelihood ratios were developed in working with UPTIDE 3B data; many more such conventions will be needed before the technique is ready for application.

Perhaps the most important of these conventions was sketched in the discussion above of hypotheses. Information that a submarine can have about a HVU location is defined in terms of 36 (or fewer) mutually exclusive hypotheses about bearing. Information that a HVU can have about a submarine location is defined in terms of a square grid pattern.

A further convention has been to treat the initial uncertainty as being equally distributed over the hypotheses being taken seriously, and to regard information sufficient to locate the ship as raising the probability of one hypothesis to a level where it at least equals the probability of the union of all other hypotheses. This implies that, for example, if the initial bearing uncertainty about the HVU is uniform over 18 hypotheses, and the HVU is localized to one 10° bearing, a likelihood ratio of at least 17:1 has occurred. A better way of getting such numbers would, of course, be to obtain from the decision maker prior and posterior probability distributions over the hypotheses. This is perfectly feasible--but it requires more time and effort from respondents than has been available so far.

Exploration of this sort of question in a setting less expensive than an
at-sea exercise is clearly a high priority requirement for further development
of the information scoring idea. (The American College of Radiology is
applying exactly this technique to evaluating the informational content
of x-rays, in a major national study for which Edwards is the technical
consultant. It is entirely appropriate to think of that study as the
first application of the fruits of this research program.)

An obvious convention governs a situation in which, for example,
a submarine is prosecuting a contact that may be the HVU. The submarine
commander may make a number of decisions that have informational consequences
for him during the prosecution. Clearly the sum of these consequences cannot
be greater than the net amount he would have received if he had localized
the HVU all at once. Fortunately, the log likelihood ratio is a nicely
additive measure that can be partitioned, so that credit for a detection
can be divided among several decisions.

Conventions concerning identification have not so far been studied,
but are clearly important. An obvious thought is that the greater the set
of possibilities, the more credit a correct identification deserves. An
interesting consequence is that the use of ADDS (which increases the number
of possibilities) increases the total score a submarine skipper could earn--
while presumably making it harder for him to earn it. This seems realistic;
a submarine skipper who successfully locates the HVU is doing better if he
does so in spite of ADDs than if no ADDs were present.

- 13 -

Gaming the score. Obviously the possibilities of gaming this
sort of information score are likely to be abundant. That is, the decision
maker has abundant opportunities to make himself look good, from an information
scoring point of view, by for example labelling as decisions only those decisions
that he feels relatively sure will have good informational consequences for
himself and/or bad ones for the other guy. (For example, the HVU commander
might try to label all deployments of ADDs as decisions, since they can only
have unfavorable informational consequences for the other side, while not
labelling his course-and-speed actions as decisions, since they are on the
whole unlikely to do much for his own information, while they can provide
significant information to the other side.)

Experience is lacking to indicate whether this is a problem, or to
think much about what to do about it. An obvious implication is that
until problems of gaming the score have been carefully worked through,
nothing important should be allowed to depend on information scores.

It seems likely that such problems can in fact be worked out. The
most important way in which a CO can game the score is by selective identifica-
tion of decisions. Perhaps a way around the problem would be to establish
a convention whereby time is divided into n-minute periods, and exactly one
course-and-speed decision is assumed to have been made in each period.
(Such a solution is not very well designed for active prosecution of a
contact, but the information scoring procedure doesn't fit active-prosecution
situations very well anyhow; it is better designed for the pre-engagement
search process.)

Conclusion

This report has summarized efforts to apply Bayesian ideas to the development of an information scoring procedure for ASW exercises. The rudimentary germs of such a procedure have been developed and trie_ ou_-- but a thorny host of problems remain to be looked into. A few conclusions seem to be reasonable.

1. Since ASW can be thought of as an information game, scoring procedures based on the informational aspects of the game are appropriate. Such procedures cannot be expected to capture all aspects of the ASW problem into one embracing score. All they can do is to quantify intuition concerning the impact of command decisions on deception, detection, localization, and identification. But such quantifications of intuition might be extremely useful for scoring exercises, real or simulated, for exploration of ASW tactics, and for training commanders of ASW forces and of submarines. The log likelihood ratio is a natural number to base such scores on.

2. A key problem is identification of decisions and of their consequences. A natural approach to identifying decisions is to let the commander identify them. Not much progress has been made in selecting conventions for tying together decisions with their consequences. The alternatives seem to be pure intuition or some combination of intuition and conventional rules. A few natural conventions suggest themselves (e. g. that the longer time intervenes between a decision and an informational consequence, the less of that consequence should be credited to the decision), but much more needs to be done.

3. Conventions for estimating likelihood ratios have been
explored, and at least in some examples seem to work reasonably well. A
better but more demanding approach would be to collect prior and posterior
distributions, and calculate likelihood ratios from them; these distributions
would need to be obtained on-line in real time from decision makers, which
would be a nuisance for the decision makers. Some combination of these
approaches may be feasible.

4. Further working over of old data, although necessary to further
development of the idea, is hardly sufficient. A more important future task
is to find a test bed for development of the information score idea which
will be less costly than real exercises. The obvious place to look is at
the simulations used for ASW training.

References

Edwards, W., Lindman, H., and Savage, L. J. Bayesian statistics for
psychological research. Psychol. Rev., 1963, 70, 193-242.

Edwards, W., Phillips, L. D., Hays, W. L., and Goodman, B. C. Probabilistic
information processing systems: Design and evaluation. IEEE Trans.
Syst. Sci. Cybernetics, 1968, 7, 248-265.

Kelly, C. W. and Peterson, C. R. Probability estimates and probabilistic
procedures in current-intelligence analysis. Report on Phase I.
Federal Systems Division, IBM, FSC 71-5047, 1971.

Footnotes

[1]Now at Social Science Research Institute, University of Southern California, Los Angeles. This report, written while the program it reports is seriously incomplete, summarizes the portion of that program for which I am responsible as it stood when I left the University of Michigan.

[2]I am grateful to COMASWFORPAC and in particular to CDR Fishburn, then UPTIDE Project Manager, for making it possible for me to ride on U.S.S. TICONDEROGA during UPTIDE 3B. I am also grateful to RADM Seiberlich, COMASWGRUTHREE, who gave his time generously to explain his decisions throughout the exercise, and to LCDR Sickman, of RADM Seiberlich's staff, who helped me formulate the idea of the information score. A memorandum written to CDR Fishburn reporting in detail on the result of an attempt to apply information scoring techniques to UPTIDE 3B data is included in this report as Appendix I.

[3]This somewhat controversial comment is based on interviews with experienced submarine commanders and on the contents of CO decision logs for UPTIDE 3B.

[4]This sort of question has a long, respectable philosophical history. Example: does the falling tree in the forest make a noise if no one is present to hear it? It obviously creates pressure waves in the air, but should those waves be called "a noise"? The notion of an unfelt pain is nonsense; the notion of an undetected vibration is easily understood; but such notions as unheard sounds and undetected or unused information have an ambiguous status, neither obvious nonsense nor obviously reasonable ideas.

Engineering Psychology Laboratory

5 February 1973

MEMORANDUM TO: CDR Fishburn

COPIES TO: Mr. Page, LCDR Sickman, Dr. Simpson, Mr. Webster

FROM: Ward Edwards

SUBJECT: Information scoring of UPTIDE 3B

This informal memo summarizes the result of two weeks spent trying to apply the idea of an information score to the UPTIDE 3B data. The task has not progressed far enough to produce a publishable result (I only worked with COMASWGRP3 decisions, and only for the first two transits), but I am left feeling encouraged about the feasibility of the idea--albeit somewhat appalled about the difficulties that I had not anticipated, many still unattacked. Still, I got far enough to give a reasonable picture of what can be done, what some of the problems are, and what might be done about them--and that's as much as I might have hoped for under the circumstances. I could have done a lot more if LCDR Sickman had been here.

What follows is a rather rambling exposition of what I did, thought, and found. I apologize for its prolixity; I am too much in haste to condense it.

The basic idea. The next few paragraphs summarize part of what LCDR Sickman and I said in our previous memorandum on this topic; the reason for including them here is to make this memo a bit self-contained.

One point of view about ASW--especially in its UPTIDE version--is that it is an information game. The reason for the thought is that the weapons

available to both sides are long-range, but depend on reasonably accurate localization of the enemy. The sensors are also long-range, in a sense. Consequently, it makes relatively little difference where either side is; what matters is how well his opponent knows where he is. In consequence, the informational consequences of an action are often more important than their geographical consequences.

From a Bayesian point of view (see for example Ref. 1) the logarithm of the likelihood ratio represents the extent to which a datum (item of evidence) helps to discriminate between two hypotheses, and so measures the informativeness of that datum. Perhaps, then, one could score the actions taken in an exercise in accordance with their informational consequences, measured as logarithms of likelihood ratios, and thus evaluate actions. From this point of view, actions that deny information to the enemy are good, actions that mislead him are better, actions that inform him are bad. Similarly, actions that gain correct information for you are good, actions that do not are bad. Clearly many other considerations enter into ASW, but an information score at least might capture that aspect of the total picture, leaving other aspects to be captured by other scores. Also, such an information score could be a debriefing and training tool, and perhaps eventually a tactical decision-making aid.

Elaboration of the basic idea. That was the basic idea with which I approached the masses of UPTIDE 3B data. The idea still, two weeks later, looks good to me--but it sure was naive! I spent most of the first week sorting out some issues that had to be thought about before the task of implementing the idea could even be begun.

1. <u>Identification of decisions.</u> What is a decision, and when is it made? When some major change of behavior, such as a change of course or speed, occurs, one is tempted to call it the result of a decision (and sometimes that temptation should be resisted). But what about occasions when the decision-maker considers whether to change course and chooses not to? On any reasonable definition, that is as much a decision as its opposite. But what a bucket of worms this opens! DRUM (transit 1) proceeds for 10 hours at more or less the same course and speed. How many times did the skipper consider changing course and decide not to? When? How seriously? In TICO, the Admiral had an audience on whom he liked to try out ideas about tactical decisions in prospect before making those decisions. The CCC **observer's** log records such instances. But the DRUM observer's log records far fewer of them. Because the skipper in fact made fewer decisions, or because he didn't have as interested, responsive, and large an audience?

The best solution, of course, would be for the decision-maker himself to label his decisions, and I recommend that that be tried in a future UPTIDE exercise. Since that was not done in 3B, I chose instead to work with the detailed narratives--CCC observers' log and CCC decision log for TICO, and CO's narratives supplemented by observers' logs for the submarines. Every time a significant action was taken, I looked through the narrative information to see if it was the result of a conscious on-the-spot decision, rather than of an adventitious set of circumstances or of a pre-transit plan. If so, I labelled it as a decision. (Of course pre-transit plans are themselves decisions, but

their analysis presents especially difficult problems, and I chose to confine myself to on-the-spot modifications of pre-transit plans for the time being.) I also looked through the logs for decisions that did not result in significant new actions, and labelled them as decisions also. I did not find this procedure particularly difficult or ambiguous, and I suspect that it would yield high inter-judge reliability.

The number of decisions per transit thus generated is not very large-- 6 to 10 for TICO, and perhaps slightly larger for the submarines. One reason for this small number is that I excluded decisions that were obvious implementations either of a previously formed tactical plan or of some standard pre-programmed set of tactics. Thus the separate course changes of a saw-tooth search pattern are not decisions. Neither are the decisions made by VP or VS pilots in prosecuting a contact.

Such rules are arbitrary, and alternatives to them could and should be tried. Yet I like these rules fairly well. They do capture the flavor of what happened on TICO as I remember it. And they permit the information score to be applied only to decisions important enough so that there would be no doubt in anyone's mind of its relevance to them.

2. <u>Informational consequences of decisions</u>. This topic turned out to be the most difficult by far. To illustrate the difficulty, consider TICO's first transit. At 311130U, the Admiral considered increasing speed, but decided against it. At 311330, TICO speeded up from 8 to 12.5 kt. At 010445 she went to over 17 kt. And at about 101630, she was sunk. DRUM gained contact at about 010200,

a 2 CZ contact, and had little difficulty prosecuting after that, though the 010445 speed-up helped considerably. It is extremely likely, considering the tracks involved, that TICO would not have been detected (at least, not then) if she had continued at 8 kt. (Of course, she would have been many hours behind PIM! An unfortunate ambiguity of the exercise is that no information was given to any commander that might help him decide the relative values of being behind PIM, of being sunk, and of sinking various units of the enemy. I have a recommendation about this for future exercises.)

DRUM's eventual acquisition of TICO was clearly the informational outcome of some decision. But of which one? I see no approach to this problem other than the establishment of conventions. I chose in this example to allocate most of the (dis)credit for DRUM's acquisition of TICO to the first decision to speed up. Potential for gaming the score is rife in this convention. Thus, every time a commander considers speeding up and decides not to, he is likely to win points. For this reason, cumulation of scores over the run of an exercise is not very meaningful. This kind of analysis is most useful as a method of critiqueing individual decisions (or of making them), not of overall scoring. That should not be too surprising; so far as I can see, matters not under the decision-maker's control, such composition and capabilities of enemy forces, so control any kind of overall score for such exercises that no overall score can be very meaningful. Still, this is a change from the original Sickman-Edwards idea. (I have a half-baked idea about a way out of this difficulty, but won't go into it here.)

What kinds of informational outcomes are feasible, and with what actions
are they associated? A great difference exists between TICO and the submarines.
Both can do three kinds of things: acquire information, deny information to the
enemy, mislead the enemy. TICO acquires information by laying sonobuoy fields
(directly or via VP), denies or fails to deny information to the enemy by course-
and-speed decision, and deceives the enemy by means of ADD and other-ship actions.
A submarine probably has no capability of misleading the enemy, except perhaps
during close prosecution. It accomplishes the other two functions by means of
course, speed, and depth actions. The fact that each such action by the sub-
marines has informational implications for both sides makes the analysis problem
more difficult than it is for TICO. The difficulty is not fundamental, and can
fairly easily be overcome. However, its existence combined with my relative
ignorance of submarine activities and my lack of access to more knowledgeable
people led me to confine my analysis during this trip to TICO decisions (except
that I did identify decisions for DRUM and SCULPIN during the first transit, just
to make sure I could; it wasn't difficult). Even at that I only got through
Transit 2. This was unfortunate. During the first two transits, most ADDs were
non-functional, so I could not score ADD deployment decisions. Transit 3 was
the most important one for ADDs.

3. Hypotheses. The original Sickman-Edwards memorandum had supposed that
identification of hypotheses would be the most difficult part of the problem of
information scoring, but in fact they now seem to be the easiest. From the
submarine point of view, the only (or, to be precise, the primary) question of
interest is the current bearing of TICO. In a way, range is of secondary

importance, both because the available sensors give relatively poor range information and because submarine tactics are reasonable range-independent in a one-HVU situation. (They are of course highly dependent on information, and that depends on range, but that is irrelevant to the formulation of hypotheses.) Moreover, for the initiation of submarine tactics a 10° bearing accuracy is plenty good enough. So, the submarine is typically interested in at most 36 mutually exclusive hypotheses. More frequently, prior knowledge permits the effective elimination of half or more of these hypotheses; typically there are large quadrants which the submarine would be willing to rule out as a location for TICO. (That is probably a special feature of exercises, which have defined boundaries; in a hot-war situation submarines would be less willing to assume the absence of CVSs on some bearings.)

Of course the course and speed of a potential TICO target is also of interest, but more for classification purposes than for purposes of initial elicitation of interest. Moreover, classification is often unimportant, especially in an UPTIDE environment. If the submarine is holding only one target, he will probably prosecute it, if geography makes it interesting. If the submarine is holding several targets, he will prosecute one of them, or perhaps will adopt an intermediate or compromise strategy. Since the TICO may sound like almost anything, thanks to turn count masking, use of only two screws, EMCON, and the like, he is likely to prosecute whatever target is easiest to hold on sonar. In fact, on several occasions in 3B, a submarine commander decided to prosecute an ADD, while holding it and TICO simultaneously, because the ADD was easier to hear than TICO.

All of these reasons led me to be fairly content with a formulation in which the submarine commander is considering, at any given time, only hypotheses about TICO's bearing, and those only to the nearest 10°.

TICO's hypotheses are somewhat more complicated. However, again the
distinction between detection and prosecution is helpful. For detection purposes,
location of the sonobuoy, bearing (if DIFAR is used), and whether the contact
is direct path or 1st, 2nd, 3rd, . . . , CZ are the crucial questions. Course
and speed are also of considerable importance, since they control subsequent
actions. But they are seldom crucial in initial contact. Moreover, they are
extremely evanescent information, since both can change at a moment's notice,
and probably will if the submarine has any idea it has been spotted. Other
questions, such as allocation of operational areas to submarines, are of
considerable tactical importance, but on the whole decisions by TICO don't have
much impact on obtaining that information.

For this first cut, I was able to use simple localization of the sub-
marine as the relevant set of hypotheses for TICO, treating that localization
in the familiar divide-the-ocean-into-squares way.

4. <u>Conventions about likelihood ratios</u>. The preceding discussion of
hypotheses pretty well settles the questions about magnitudes of likelihood
ratios. All the submarine commander requires is that he has a contact, or, if
he has more than one, that this is the most promising one. In other words,
information that reduces his uncertainty from a uniform distribution over 36
(or 18, or whatever) hypotheses to one in which some hypothesis is at least as
probable as the union of the others, is more than enough. A sonar contact
usually gives far more information than that; the bearing accuracy is apparently
far more than is needed for the present purpose. This sets an upper bound of
36:1 (or 18:1, or whatever) on the information gained by the submarine.

TICO's information gain can be much greater from a single action, though typically it is much smaller than that. A buoy field laid on top of the submarine could generate a 1000:1 likelihood ratio easily. It all depends on prior information and on geography.

A particular problem arises when several decisions made in sequence all contribute to a particular detection. It is crucial not to allow the cumulative impact of them all to sum to more than the log likelihood ratio that would have been scored had the detection been made all at once. A fractional partitioning of the log likelihood ratio for the complete detection is the appropriate procedure.

An interesting unsolved problem is whether the sum of log likelihood ratios associated with an eventual complete detection can be less than that associated with an immediate complete detection. My tentative answer is yes. Exclusion of the "decisions" involved in following up an initial contact from the information score implies that information gained as a result of them is not counted.

The result. Appendix I contains the information scores and enough information to explain how they were calculated for TICO's decisions in Transits 1 and 2. The respondent was Jim Webster (and to some extent myself). Of course RADM Seiberlich would have been a preferable respondent--but he was not here, and the technique is not yet sufficiently fully developed to justify obtaining his judgments.

Suggestions for the future. I plan to continue to think about this topic. I would like, if possible, to finish the other four TICO transits and to do the

corresponding job for DRUM and SCULPIN. Now that I know how the data are organized, I think I could do the job in Ann Arbor, if the relevant materials could be sent there. (I have suitable storage facilities.) I would need the following:

Observers' logs for CCC, DRUM, and SCULPIN.

COs' logs for the submarines, and decision log for TICO.

Location plots for all ships for the last four transits.

Course and speed logs (preferably not the raw ones) for TICO, and if available for the submarines also.

Information about ADD and sonobuoy field locations, last four transits.

VP debriefs, last four transits.

Sonar contact information and contact correlation information for both submarines.

Of course, all that is a lot. I could probably do something with only the first three items, though I have in fact been using all of these items, plus others, during the last two weeks.

Or, I could come back to Hawaii during spring vacation. Given my preferences, I would much rather work in Ann Arbor.

I have two other suggestions for the future. One concerns future UPTIDEs, including perhaps ROPEVAL 173. It would be extremely desirable for key COs to keep decision logs on forms provided to them by UPTIDE. The instructions would be to fill out a form every time a significant decision is made (with emphasis that not all decisions are significant, and that as many as one significant decision every three hours is rare). The key questions on the form would include: What is the problem that makes a decision necessary? What courses of

action have you considered? Which did you select? Why? What informational

consequences do you expect it to have, and how do they compare with the

informational consequences of other courses of action? What other con-

sequences do you expect it to have? Of course I would be delighted to help

prepare both the form and instructions for its use, both by COs and by sub-

sequent analysts.

The advantage of this form, of course, is that it would define most of

the questions that I found it necessary to define this time. It would identify

decisions, options considered, and best guesses about the consequences of each.

These guesses can of course later be evaluated in the light of the tactical

situation at the time, both as the CO knew it and as it actually was.

My other suggestion for the future concerns the problem of conflicting

goals in UPTIDE exercises. In 3B, TICO was supposed not to be sunk, but also to

more or less keep up with PIM. And she had a secondary goal of sinking submarines.

To what extent should she risk being sunk in order to keep up with PIM? The

Admiral had no real basis for making that decision. Similarly, the submarines

were to sink TICO, remain unsunk themselves, and as a subsidiary goal to sink

escorts. To what extent should they expose themselves in order to have a better

chance of sinking TICO? Again, the submarine COs had little basis for decision.

The obvious solution to this problem is to provide a scoring system for

each side which explicitly places all these good and bad outcomes on the same

scale. (I often hear the objection that fixed linear scoring systems distort

the real nature of the problem. It may be much more than 60 times as bad to be

10 hours late as to be 10 minutes late. The obvious solution: if the values

are non-linear, use a non-linear scoring system.)

This suggestion is far more difficult to carry out than the one about decision logs, since it implies an explication of things usually left implicit. Still, I think it would help, and, at least technically, it is extremely easy to do. The political problems are another matter. Again, I'd be glad to help.

TICO - Key Decisions - Transit 1

Decision 1

Time: 311025U

Actions: The Admiral ordered TICO's track moved 10 mi. N. The obvious
alternative was to continue on planned track.

Surrounding circumstances: At 0945, VS obtained a contact in the S corner of
their sonobuoy field. Also, another at 1010, again S of intended track. VP
also was holding a DIFAR contact, S of track. The contacts were in fact
SCULPIN, and she was S of intended track. However, she did not hold a contact
on TICO, and would probably not have gotten one.

Informational outcomes: For BLUE, none. For ORANGE, it seems likely that the
displacement N had the effect of helping DRUM to make contact with TICO sooner
than if she had stayed on planned track.

Log likelihood ratio: .389. (Calculation: a reasonable upper bound for the
likelihood ratio would be 36, since the original detection was remote. Give
the course displacement 1/4 credit. 1/4 (log 36) = .389.) This number is credited
to TICO, with negative sign (indicating that the informational outcome was
unfavorable for BLUE).

TICO - Key Decisions - Transit 1

Decision 2

Time: 311125U

<u>Actions</u>: The Admiral considered increasing speed from the 8 kt TICO was

making at the time to an above-cavitation speed. He decided to continue at

present speed until at least 1250.

<u>Surrounding</u> <u>circumstances</u>: Same as for Decision 1, except that VS had a madman

contact.

<u>Informational</u> <u>Outcomes</u>: For BLUE, none. For ORANGE, as it happened, it made

no difference either way. All log likelihood ratios are 0.

TICO - Transit 1

Decision 3

Time: 311255U

Actions: The Admiral decided not to send surface ships to investigate SCULPIN. It would take them 2 hours to get there and he feels it best to leave them where they are to lure the SCULPIN S away from TICO.

Informational Outcome: Not relevant, since SCULPIN was OOA at 1257. In a score concerned with intended rather than with actual outcomes, this decision would probably have received a positive score, since at the time it was made there was no way of knowing that SCULPIN would be OOA.

TICO - Transit 1

Decision 4

Time: 311259U

Actions: TICO increased speed from 8 kt to about 12.5 kt.

Surrounding circumstances: SCULPIN was OOA; nothing had been heard of or from DRUM. SCULPIN was expected to reenter in about 10 hours at a location roughly 140 mi. SW of OOA position.

Informational consequences: For BLUE, none. For ORANGE, DRUM got a 2 CZ contact on TICO at 010200U, and ultimately localized and sunk her. The geography was such that TICO would have been much farther away from DRUM, as well as much quieter, if she had not sped up. (Of course, she would have been many hours behind PIM, but that is not relevant to an information score.)

Log likelihood ratio: A total of log 36 is available to be partitioned as a result of this contact. This gets half of it, or .778. This is credited to BLUE with negative sign.

TICO - Transit 1

Decision 5

Time: 311355U

<u>Action</u>: Relocated VP pattern from S of track to N of track.

<u>Informational consequences</u>: None, for either side. Again, this would have been an informationally desirable decision if the score were concerned with intentions rather than results.

TICO - Transit 1

Decision 6

Time: 010445U

Action: TICO sped up from 12.5 kt to more than 17 kt.

Surrounding circumstances: No contacts were active at the time, VP or VS. TICO was well behind PIM.

Informational consequences: DRUM already had and was prosecuting a contact on TICO. This strengthened it, and encouraged prosecution. Prior to this time DRUM was also interested in other contacts; after it, she was interested only in the one that turned out to be TICO.

Log likelihood ratio: .389, credited to BLUE with negative sign.

TICO - Transit 1

Decision 7

Time: 010600U

Action: At the 0600 briefing, the Admiral felt that DRUM was behind and to the North (actually, she was ahead and to the South). So he decided to move a VS field away from the goal line ahead and put it North and behind. This decision was fully formed at 0630--the same time at which DRUM killed TICO.

Informational Outcome: Irrelevant.

TICO - Transit 2

Decision 1

Time: 012235U

<u>Action</u>: The originally planned location of the first VP field was changed.
The original location was N of SCULPIN; the new location was on top of him.

<u>Surrounding</u> <u>circumstances</u>: COMEX was 012100. The change was made because
a SOSUS SPA was reported,and the field was laid to cover the SPA. VP
obtained a DP contact right away, incorrectly evaluated it as a CZ contact,
prosecuted, discovered their error, reclassified, obtained attack criteria,
attacked, missed, and in due course had to go home. However, this initial
contact resulted in SCULPIN's being more or less continually harrassed from
then on, and eventually in her being killed.

<u>Log</u> <u>likelihood</u> <u>ratio</u>: The SOSUS SPA included roughly 10,000 sq. mi. of area.
The VP field contact resulted in roughly a 10x10 mi. uncertainty. This is a
100:1 reduction--call it a 2:1 log likelihood ratio. This is credited to
BLUE with a positive sign.

TICO - Transit 2

Decision 2

Time: 020300U

<u>Action</u>: TICO's speed was reduced from 16 to 12 kt.

<u>Surrounding circumstances</u>: VP had a DIFAR fix straight ahead of track, 50-60 mi. The alternative (actually recommended to the Admiral) was to change course N; another alternative obviously was to continue at planned course and speed. Course at that time was 090. SCULPIN was being very actively prosecuted until 0135, but much less actively after that time. She was attempting evasion till 0452, and was in no condition to exploit any contact she might have made. I choose to lump this decision together with its successor for log likelihood ratio estimation purposes.

TICO - Transit 2

Decision 3

Time: 020515

Action: TICO reduced speed further from 12 to 8 kt. Again, the alternatives.
were to change course N or to continue at prior course and speed.

Informational outcome: Failure to take some action might have enabled
SCULPIN to detect TICO. By this time, she had some leisure, though VP were
still listening for her.

Log likelihood ratio: On the assumptions that SCULPIN had a 50-50 chance of
detecting TICO, and that her bearing uncertainty about her was uniform over
$180°$, the log likelihood ratio is $1/2$ (log 18) = .628, credited to BLUE with
positive sign.

TICO - Transit 2

Decision 4

Time: 020630U

Action: TICO came to course 000, speed 16.5 kt.

Surrounding circumstances: The Admiral got up. Best guess was that SCULPIN was 45 mi. ahead at 090, the then-current course. Even at 8 kt, TICO might have been detected, especially if the range was overestimated. (It wasn't).

Log likelihood ratio: Full credit for this decision is implied in the score for Decision 3. No additional credit for this one.

TICO - Transit 2

Decision 5

Time: 020640U

Action: The Admiral ordered MIDGETT to proceed on course 090, speed 14 kt, intending to have her turn on her NADC noisemaker later when she would be close to SCULPIN.

Surrounding circumstances: SCULPIN did detect MIDGETT at 1159 for 13 min; did not classify; was too busy to pay much attention.   TICO was detected but not identified at 1325 for 26 min.   MIDGETT was again detected at 1333.   SCULPIN was OOA at 1405.

Log likelihood ratio: A likelihood ratio of 1.8:1 yields a logarithm of .255, credited to BLUE with positive sign.

Note: A number of conflicting decisions were made about MIDGETT during the morning, as the prosecution of SCULPIN developed.   This credit is the assessment of the net outcome of them all, taken together.   I will not list subsequent decisions about MIDGETT separately.

TICO – Transit 2

Decision 6

Time: 020700

<u>A</u>ction: Decreased speed to 12 kt, holding course 000.

<u>Informational</u> <u>outcome</u>: None.

43

- 26 -

TICO - Transit 2

Decision 7

Time: 020730

Action: A NYVO then airborne was redirected to a point N of SCULPIN's

estimated position. Also, TICO went to course 045, 12 kt. At 0735 she went

to course 065. Informational consequences: None, in view of the active prosecu-

tion of SCULPIN by VP at this time. This decision (to change course and speed)

might have had negative consequences otherwise. It is difficult to know what

the decision about the NYVO might have done.

TICO - Transit 2

Decision 8

Time: 020905U

Action: TICO changed course to 325 at 0909.

Surrounding circumstances: The Admiral had previously changed course N,

then went back to PIM on the belief that SCULPIN was going N. Now a VS hot

contact implies that SCULPIN is going S, so TICO turns to N and beyond.

Informational consequences: This decision may have prevented a detection

of TICO by SCULPIN.

Log likelihood ratio: A likelihood ratio of 1.8:1 has a logarithm of .255,

credited to BLUE with positive sign.

Note: This suggests that the zero llr assigned to decision 7 should be

reassessed, but there is no time as this is being typed, and the

inconsistency was not noticed earlier.