

LEVEL

12



DECISION SCIENCE CONSORTIUM, INC.

**ALTERNATIVE FRAMEWORKS FOR
THE RECONCILIATION OF
PROBABILITY ASSESSMENTS**

DTIC
SELECT
APR 8 1989
C

AD A 097584

Anthony N. S. Freeling

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

DTIC FILE COPY

November 1989

81 4 8 101

Unclassified

TH-

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER TH-80-4 ✓	2. GOVT ACCESSION NO. AD-A097584	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Alternative Frameworks for the Reconciliation of Probability Assessments		5. TYPE OF REPORT & PERIOD COVERED Technical (F-1)
7. AUTHOR(s) Anthony N. S./Freeling		6. PERFORMING ORG. REPORT NUMBER DSC Rpt. No. 80-4
9. PERFORMING ORGANIZATION NAME AND ADDRESS Decision Science Consortium, Inc. 7700 Leesburg Pike, Suite 421 Falls Church, Virginia 22043		8. CONTRACT OR GRANT NUMBER(s) N00014-80-C-0046 ✓
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research 800 N. Quincy Street Arlington, Virginia 22217		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (15) 16
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Same		12. REPORT DATE November-1980
		13. NUMBER OF PAGES 70
		15. SECURITY CLASS. (of this report) unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for Public Release; distribution unlimited. Reproduction in whole or in part is permitted for any purpose of the U.S. Government.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Decision analysis; probabilities; probability judgments; coherence; log-odds; least-squares; information; expert use.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The problem of a probability assessor who has, directly or indirectly, given two idfferent numbers for the probability of one event is addressed. In order to use these assessments in, for example, a decision analysis, these two numbers must be reconciled to give one value as the probability. The previous work on th's subject by Lindley, Tversky and Brown (1979), and Brown and Lindley (1979) is explored. In particular the paper by Lindley, Tversky and Brown is summarized and discussed in the light of its practi- cability. A philosophical discussion of our perspective when attempting such		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

reconciliation is given, together with an indication of how the present work relates to similar problems addressed in the past. We also discuss the potential role of an axiomatic extension to probability theory.

A detailed mathematical exposition is given concerning the calculation of the least-squares reconciliation proposed by Lindley, Tversky and Brown (1979), in both probability and log-odds metrics. It is shown that the method of reconciliation they have proposed is formally equivalent to that of taking a weighted average of log-odds, with weights proportional to the independent information content of each assessment. This method has the advantage of being simple in application. It is further argued that the motivation for taking multiple probability assessments is an attempt to obtain more information from the subject, and that the proposed method of reconciliation captures the essence of this motivation. The relationship between this research and the well-known "expert-use" problem is explored. Finally, a discussion of alternative potential approaches to the reconciliation problem is included.

Accession For	
NTIS GRA&I	
DTIC TAB	
Unannounced	
Justification	
By	
Distribution/	
Availability	
Avail. and/or	
Dist	Special
A	

ALTERNATIVE FRAMEWORKS FOR THE
RECONCILIATION OF PROBABILITY ASSESSMENTS

By

Anthony N. S. Freeling

Prepared For:

Office of Naval Research
Naval Analysis Program
Contract No. N00014-79-C-0209

November 1980

Decision Science Consortium, Inc.
7700 Leesburg Pike, Suite 421
Falls Church, Virginia 22043
(703) 790-0510

Technical Report 80-4

ABSTRACT

↙ The problem of a probability assessor who has, directly or indirectly, given two different numbers for the probability of one event is addressed.

In order to use these assessments in, for example, a decision analysis, these two numbers must be reconciled to give one value as the probability. The previous work on this subject by Lindley, Tversky and Brown (1979), and Brown and Lindley (1979) is explored. In particular the paper by Lindley, Tversky and Brown is summarized and discussed in the light of its practicability. A philosophical discussion of our perspective when attempting such reconciliations is given, together with an indication of how the present work relates to similar problems addressed in the past. We also discuss the potential role of an axiomatic extension to probability theory.

↘ A detailed mathematical exposition is given concerning the calculation of the least-squares reconciliation proposed by Lindley, Tversky and Brown (1979), in both probability and log-odds metrics. It is shown that the method of reconciliation they have proposed is formally equivalent to that of taking a weighted average of log-odds, with weights proportional to the independent information content of each assessment. This method has the advantage of being simple in application. It is further argued that the motivation for taking multiple probability assessments is an attempt to obtain more information from the subject, and that the proposed method of reconciliation captures the essence of this motivation. The relationship between this research and the well-known "expert-use" problem is explored. Finally, a discussion of alternative potential approaches to the reconciliation problem is included. ↗

Key words. Decision analysis; probability judgments; coherence; log-odds; least-squares; information; expert use.

TABLE OF CONTENTS

	<u>Page</u>
1.0 THE PROBLEM	1-1
1.1 Introduction	1-1
1.2 The Problem, and the Motivation for Studying It.	1-1
2.0 SUMMARY OF LINDLEY, TVERSKY AND BROWN, AND COMMENTS	2-1
2.1 Introduction	2-1
2.2 A Mathematical Formulation of the Problem.	2-1
2.3 Brief Summary of the Assumptions and Results of LTB.	2-1
2.4 Some Comments on LTB	2-5
2.5 The Underlying Motivation for RIJ.	2-9
3.0 FOUNDATIONS	3-1
3.1 Introduction	3-1
3.2 True Probabilities	3-1
3.3 The π 's as Parameters.	3-4
3.4 The Role of Bayesian Updating.	3-6
3.5 Extending the Axiomatic System	3-7
3.6 Philosophical Background	3-11
3.7 Expert Use	3-14
3.8 Summary of the Various Different Approaches Identified for RIJ, and Their Rationales	3-15
3.8.1 Bayesian updating	3-15
3.8.2 Least squares	3-16
3.8.3 "Information" approach.	3-16
3.8.4 Alternative simple techniques	3-17
3.8.5 Fuzzy set theory.	3-18
4.0 THE LEAST-SQUARES APPROACH.	4-1
4.1 Introduction	4-1
4.2 The Choice of Metric	4-1
4.3 The Linear Model	4-4
4.4 Use of the Lagrangean.	4-5
4.5 A Procedure for Optimizing in the Log-Odds Metric.	4-9
4.6 A Practical Example of the Methodology	4-10
4.7 The Importance of Correlations	4-10

	<u>Page</u>
5.0 ALTERNATIVE INTERPRETATIONS OF THE LEAST-SQUARES PROCEDURE	5-1
5.1 A Psychological Interpretation of the Metric.	5-1
5.2 Least Squares and a Weighted Average.	5-2
5.3 An Information Oriented Approach.	5-4
5.4 Assessments	5-8
5.5 Expert Use.	5-12
6.0 ALTERNATIVE RECONCILIATION TECHNIQUES.	6-1
6.1 Psychological Biases.	6-1
6.2 Satisficing	6-2
6.3 Degree of Incoherence	6-3
7.0 SUMMARY AND CONCLUSIONS.	7-1
8.0 REFERENCES.	8-1

FOOTNOTES

1.0 THE PROBLEM

1.1 Introduction

This report presents some research following from the work of Lindley, Tversky and Brown (1979), and Brown and Lindley (1979) on the reconciliation of incoherent judgments. We discuss the problem and the motivation for studying it and summarize the paper by Lindley, Tversky and Brown (LTB). We show some shortcomings of their work, in particular pertaining to the practicability of their proposed methodology, and then present some extensions of their ideas. We go quite a long way towards deriving an alternative operational procedure, but it is pointed out that further research is still necessary in this direction before the procedure is truly operational.

1.2 The Problem, and the Motivation for Studying It

When a subject, S , is asked to produce judgments (about utilities, probabilities, or even preferred decisions) in a variety of different ways, it is quite likely that the responses will contradict each other in some way, or fail to satisfy the computational constraints imposed by the probability calculus. Perhaps the most obvious example of this occurs when a decision analysis is carried out, and the selected alternative differs from the option that S had chosen via direct introspection. Another example occurs when S produces numbers for the probability of a "target event" A , $P(A)$, in two different ways, and these two numbers differ.

To provide a concrete example of this latter situation, suppose I am interested in the target event that Oxford University will win their next annual rowing race with Cambridge University. Taking all things into consideration, I decide that this has probability 0.7. Then I decide to make the assessment by conditioning on Oxford winning the toss. I decide that I believe they have a chance 0.8 if they win the toss, but only 0.5 if they lose the toss. Assuming the probability of winning the toss to be 0.5, these latter two assessments imply a probability for the target event of 0.65, and hence I have caught myself in an inconsistency.

Typically in such a situation, one of the judgments will be considered "better" than the other, and the other judgment simply ignored. So, for example, when a decision analysis has been performed for a decision, it is usually assumed that one should trust the analysis more than a non-analytical, intuitive, judgment. Again, if one has arrived at the probability of a target variable in two ways, once directly (holistic assessment) and once by obtaining probabilities conditional on another event (decomposed assessment), the decomposed assessment will typically be used and the holistic assessment disregarded. In fact, this selection is often made implicitly, before any assessments are made, and only a "minimally-specified" set of judgments is taken, e.g. only the decomposed assessment, so that there is no chance for incoherence to be discovered.

However, I may have a strong gut feeling that a decision analysis failed adequately to capture all my opinions about a decision, and that my direct choice really did have something extra to offer. Similarly, I feel that 0.65 is too low for the probability of Oxford winning, and that my holistic assessment may have captured aspects of my uncertainty left

untapped by the decomposed assessment. Had I not tried multiple approaches to the elicitation I would not have discovered this. The fundamental thesis of this paper is that there is something to be gained, in terms of digging into S's psychological field, by pursuing several alternative methods of eliciting the same judgment. For further discussion of this motivation, see Brown and Lindlev (1979). This paper looks only at inconsistent probability assessments, although this is just a small part of the much wider field of inconsistent judgment.

At present, if multiple assessments are used, and inconsistency discovered, S will typically have the inconsistency pointed out to him/her, and be requested to perform the reconciliation informally. The research described here attempts to provide a theoretical basis leading to a practical technique for a formal reconciliation procedure. Such a theoretical basis is desirable to aid S in the reconciliation. Perhaps more importantly, a theoretical foundation will raise these procedures to the same level of credibility and defensibility as the rest of decision analysis, and will, we hope, cause practicing decision analysts to view this seeking out of inconsistency as an integral part of a good decision analysis.

2.0 SUMMARY OF LINDLEY, TVERSKY AND BROWN, AND COMMENTS

2.1 Introduction

In this section we summarize and discuss the main points of LTB. These comments are then used as the starting points for the further research which forms the remainder of this paper.

2.2 A Mathematical Formulation of the Problem

We suppose there is a subject, S , whose probabilities q_i , $i=1, \dots, n$ we have elicited. These probabilities will typically be inconsistent. Our aim is to provide a reconciled set of probabilities π_i , $i=1, \dots, n$ which satisfy the constraints specified by the probability calculus, which can be stated in the form $f_j(\pi_1, \dots, \pi_n) = 0$, $j=1, \dots, n$. We shall use vector notation for simplicity, in which case the constraints can be stated as $f(\underline{\pi}) = 0$. As an example, suppose S provides probabilities for an event A and for its complement, $\sim A$, each of 0.4. Then $q_1 = 0.4$, $q_2 = 0.4$, and the single coherence constraint is $\pi_1 + \pi_2 = 1$. This can be viewed geometrically in Figure 1.

The assessed $\underline{q}$ is the point $(0.4, 0.4)$, and this is incoherent as it does not lie on the constraint set represented by the line $\pi_1 + \pi_2 = 1$. Our reconciliation task is to find one point on the constraint set which is in some sense "the best." This is $\hat{\underline{\pi}}$.

2.3 Brief Summary of the Assumptions and Results of LTB

The basic model that the authors use is a measurement error model. They assume that any subject who gives an incoherent set of probability

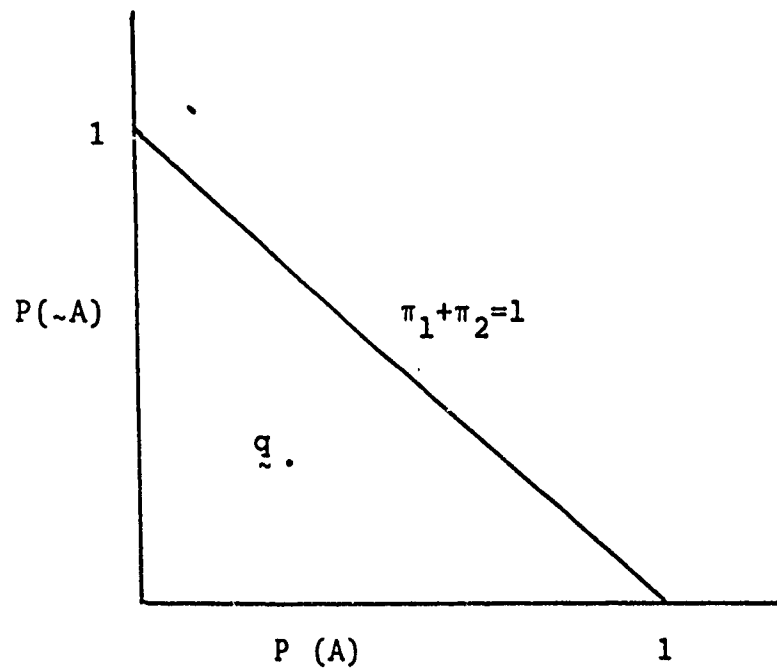


FIGURE 1
A GRAPHICAL ILLUSTRATION OF INCOHERENCE

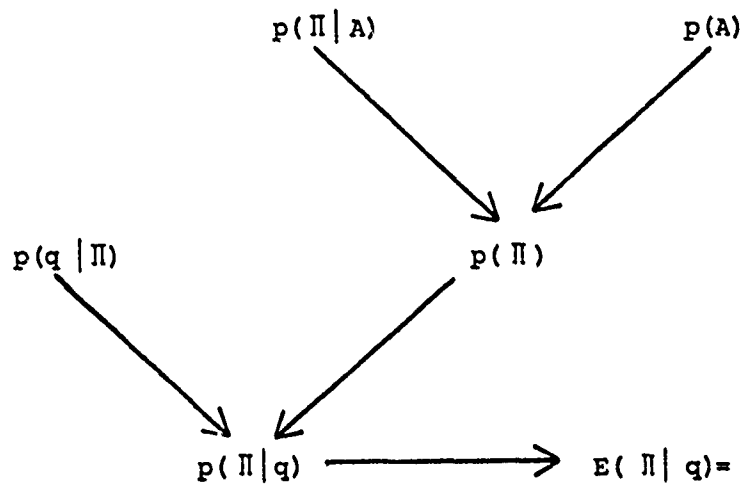
judgments in fact has a coherent set of probabilities, π , latent within him/her, but which can only be verbalized with a certain amount of error. By viewing this error as a random variable, we have a measurement model which is amenable to standard Bayesian statistical analysis.

The assessed q_i 's are viewed as readings on π together with a measurement error. LTB then introduce the concept of a coherent investigator, N , who provides information about this measurement error. They suggest two alternative Bayesian procedures to arrive at π . Both these procedures require the following three probability distributions from N .

- i) $p(A)$; N 's (coherent) probabilities corresponding to q
- ii) $p(\pi|A)$: N 's view of what S will have as true probabilities, if A in fact obtains.
- iii) $p(q|\pi)$: N 's opinion of what S will say, if the true probabilities are in fact π .

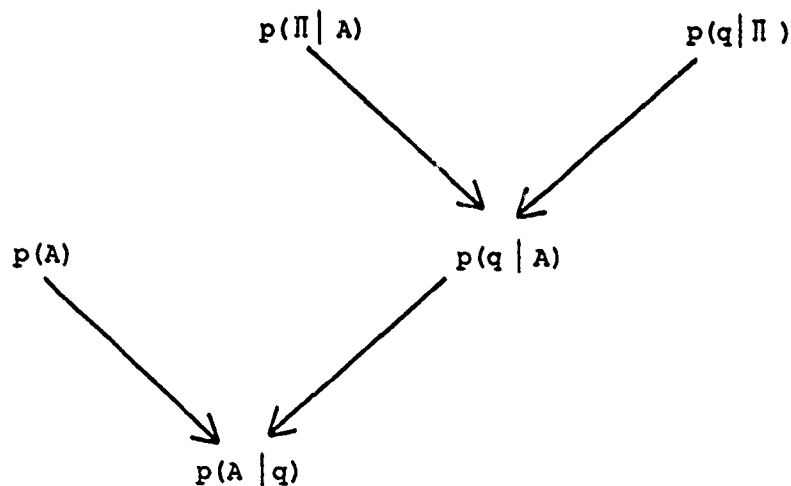
These three distributions can be viewed as representing, respectively, N 's own beliefs about A , N 's model of S 's knowledge acquisition, and N 's model of S 's performance as a probability appraiser. It should be noted that we assume that $p(q|\pi) = p(q|\pi, A)$, so that S 's measurement error does not depend on whether A in fact obtains or not. With these three distributions, the authors of LTB develop an internal and external approach. In the internal approach, N derives a probability distribution for π updated in the light of the elicited q , $p(\pi|q)$, and uses this to arrive at π , so here $\hat{\pi}$ is viewed as a "best" estimate of the true π . In the external approach, N updates his/her own probabilities for A , in the light of the information provided by q . So here, π is N 's revised view of the world, $p(A|q)$. Figure 2 shows diagrammatically the way N combines the probability distributions for each approach.

INTERNAL APPROACH



WE ELICIT $p(\Pi|A)$, $p(A)$ AND $p(q|\Pi)$. THESE ARE THEN USED CONSECUTIVELY TO DERIVE $p(\Pi)$, $p(\Pi|q)$ AND HENCE $\hat{\Pi}$.

EXTERNAL APPROACH



WE ELICIT $p(\Pi|A)$ AND $p(q|\Pi)$ AND $p(A)$. THESE ARE THEN USED, CONSECUTIVELY, TO DERIVE $p(q|A)$ AND $p(A|q)$.

FIGURE 2

THE INTERNAL & EXTERNAL BAYESIAN UPDATING APPROACHES

Within the internal approach, the authors develop a least-squares estimation procedure, based on the assumption that the measurement error is normally distributed. (They suggest that using log-odds increases the validity of this assumption.) Some specific examples and calculations are given, and results derived showing tremendous increases in precision of estimates, when several judgments are elicited and then reconciled. These results are all based on the assumption that the errors in the elicited probabilities are independent. A particular conclusion of the paper is that conditioning the probability of an event on an equiprobable partition of event space will provide the greatest increase in precision of an estimate. The authors call this procedure "extending the conversation."

2.4 Some Comments On LTB

My comments and criticisms lie in two areas, 1) philosophical and psychological and 2) mathematical. The comments in area 1) mainly concern the assumption of the existence of "true" probabilities, π . While the axiomatic systems leading to subjective probabilities do permit the postulation of such π 's, their psychological reality is, at best, very dubious. Hence one of the distributions required in the Bayesian approach, $p(q|\pi)$, may not be psychologically meaningful, and hence is very hard to assess. This throws doubt on the usefulness of the Bayesian approach as a practical reconciliation tool.

It is also the case that the π 's cannot be viewed as subjective probabilities if one looks at a rigorous mathematical analysis of the Savage axioms. This point is explored in an unpublished DSC manuscript by Robin Bromage. The difficulty lies in the fact that Savage probabilities can

only be assigned to events which are, in fact, resolved at some time, for only then will we receive the payoff about which we are making our judgment. However, the human ability to abstract allows us to make judgments about how we would feel, were the payoff achievable, and the probabilities thus derived ought not to be very different from the strict Savage probabilities. Hence this point is in fact not very worrisome. The status of the π 's is more fully explored in Section 3.0. The point we wish to make here is that the authors of LTB claim that the assumption of the π 's is vital to any reconciliation procedure. The recurrent theme of the present paper is that such a contention is false. By attempting to analyze the concepts underlying LTB, I develop an alternative approach to RIJ, which does not use the measurement model directly. This extends the least squares ideas of LTB, and is discussed in Section 5.0.

A second difficulty lies in the assumption of N. Who is this fully coherent investigator? LTB appear to view N as a part of our subject, in a more reflective mood. If one views decision analysis as a procedure in which all the judgmental inputs come from S rather than the analyst, then LTB's view of N is more satisfactory than assuming him/her to be the analyst. For the external approach would then view the subjects' judgments purely as evidence to update the analysts' opinions, and this takes the decision analyst far away from a supposedly neutral, purely analytical, role.¹ The internal approach also depends very heavily on N's judgment and again, this situation may be regarded as somewhat unsatisfactory. If, on the other hand, we view N as a part of the subject, we shall be asking some very strange questions. For example, to get $p(A)$, we shall have to ask:

"I'm afraid the q 's you have just given me were inconsistent. Could you please give me a coherent set of probabilities $p(A)$, so that I can use my method to find coherent probabilities, $\hat{\pi}$?"

Why would we not simply use $p(A)$ instead of $q(A)$? This raises the general question of second-order incoherence. $\underline{N}$ will typically be incoherent too, so his/her judgments will need to be reconciled. This is further discussed later. A third difficulty arises in the internal approach. Even if we can get the distribution $P(\underline{\pi} | \underline{q})$, how do we then arrive at $\hat{\underline{\pi}}$? LTB tell us to take $\hat{\underline{\pi}} = E(\underline{\pi} | \underline{q})$, but there is no reason why these $\hat{\underline{\pi}}$ should satisfy the constraints if these are non-linear.

This is because our analysis using Bayesian updating only permits us to arrive at the distribution $P(\underline{\pi} | \underline{q})$, and then to extract the marginal distributions $P(\pi_i | \underline{q})$. If we have a constraint of the form $\pi_1 = f(\pi_2)$ then our analysis assures us that $P(\pi_1 | \underline{q}) = P(f(\pi_2) | \underline{q})$. However, if we then take expectations, we may not be sure that $E(\pi_1 | \underline{q})$ (which equals $E(f(\pi_2) | \underline{q})$ by definition) will equal $f(E(\pi_2 | \underline{q}))$, unless $f(\cdot)$ is a linear function.

Our other criticisms of the mathematical structure of LTB are directed primarily at the independence assumptions in the least-squares procedure, and the effect these have on the conclusions about extending the conversation.

It can be shown that the dramatic increases in precision found by the authors' calculations arise mainly out of these independence assumptions. Such assumptions are equivalent to saying that each assessment contains totally new information. As information and precision are closely related, it is not surprising that the increases in precision appear. In fact, independence is a false assumption in most cases; for example if someone always gives optimistic assessments, then his/her "errors" will have strong positive correlations, as the probabilities will tend to be "too high." The importance of this fact, and the detailed calculations showing its effects, are discussed in Section 4.0.

One of the problems with RIJ is that the measurement error theory tends to pervade our thinking even when not being explicitly used. Concepts such as independence and correlations derive from a probabilistic interpretation of errors, which is precisely the measurement model. Such concepts are ordinarily used to describe connections between observations when no satisfactory explanatory description can be made. It is our contention that, given a set of inconsistent assessments, we can gain some idea of what the sources of this inconsistency are. In that case we ought to develop different models for reconciliation depending on the relevant sources. The measurement model should really be viewed as a last resort, when all else fails, and we are forced to consider incoherence as the result of a purely random process. The different possible sources of incoherence are examined by Detlof von Winterfeldt (1980).

Another observation which is perhaps relevant here concerns the concept of "error." When a subject produces inconsistent probabilities, the only

error we can point to, from a mathematical perspective, is that the axioms of probability are violated. We ought to be careful about saying, without some type of measurement model, that one estimate is "too high." This carries connotations of implying error from a "true" probability, whereas (if we are going to try and avoid the assumption of such true probabilities) we ought only to describe it as deviating from a number with some essentially arbitrary characteristics, e.g., that number with which the subject feels most comfortable. An examination of what characteristics we should look for in a reconciliation and what the implied definitions of "improving" are, is developed in more detail in Section 3.0.

2.5 The Underlying Motivation for RIJ

In order to decide upon a good methodology for RIJ, it should be made clear what our motivation for the work is. It is to enable us to get the best probability estimates possible. As pointed out before, it is unclear exactly what is meant by "best," but one desideratum surely is that the subject should take as much of his psychological field into account as possible. One way this can be achieved is by making as many different attempts at eliciting a probability as possible, rather than using only one minimally-specified set of readings. It is in this situation that the potential for incoherent judgments develops, and our extra elicitations are of course of no value if we have no method, however simplistic, of getting a reconciled value from the assessments. The underlying motivation for RIJ, then, is to allow us to dig as far as possible into the subject's psychological field, without getting results with which we cannot cope.

The necessary subsequent research is to devise a digging procedure, i.e., to enable us to formulate a strategy in a given situation which will permit us to decide a priori what sort of decomposed estimates to ask for. This methodology should also tell us a posteriori how well we have done, and what further digging may be necessary. In order to help towards this end, our RIJ research should have some quantitative indication of the precision of reconciled estimates.

The two primary elements which appear to lead to increased precision are:

- 1) Further consideration of items of information in the subject's data base which had formerly been only cursorily examined.
- 2) Improved mathematical manipulation of probabilities, by decomposing assessments and writing down explicitly the probability calculations involved.

The assumption which underlies the common belief amongst analysts that "decomposed is best," and indeed the whole DA paradigm, is that by splitting up a problem into smaller parts, and building formal, explicit probability models, we may gain improvements under both 1) and 2) above. An approach to RIJ which quantifies this concept, and explores the limitations of the above rationale is discussed in Section 5.0.

3.0 FOUNDATIONS

3.1 Introduction

In this section we discuss briefly the various alternative approaches to RIJ that have so far been identified and the different general philosophical viewpoints underlying each one. The reasons for suggesting each approach are examined in detail. The following sections then deal with the technical details of each approach, and discuss the extent to which each methodology can, in fact, achieve successful reconciliations.

3.2 True Probabilities

The concept of true or authentic probabilities is one that is beginning to recur in the decision-analytic literature (see e.g., LTB, or Tani, 1978). These articles reflect the growing realization that the normative decision analytic theory is often shown to be incomplete when real probability assessors become involved. In an attempt to counter the fact that people are sometimes unable to act in accordance with the theory, it is convenient to postulate the existence of numbers with which people would operate according to the probability calculus, if only these people were capable of accessing the numbers. There is certainly nothing in the mathematical theory of subjective probability which refutes this concept, but that must not be taken as a proof of their existence. We should also take great care, when introducing such a concept in a given situation, to insure that such an introduction is in fact useful, and does not produce more problems than it solves.

The major difficulty with the concept of true probabilities lies in the use to which these probabilities are put. The idea is that elicited

probabilities are simply imperfect approximations to the true probabilities. Then, since we do not know just how good an approximation we have made, we build a probability distribution describing the possible value of the true probability, given our elicited value. Then we may perform a decision analysis, using the information about the true probabilities, instead of a direct use of the elicited values. This may sound very plausible at first reading, but on further examination, there are some major problems with the above procedure.

First there is no evidence that such true probabilities have any actual existence. On the contrary, much of the psychological literature provides direct evidence discrediting the notion of true probabilities. (See Phillips contribution to the discussion of LTB.) In this case, as Phillips points out, the distribution $p(q|\pi)$, which is essential to the RIJ methodology of LTB, is not psychologically meaningful. There is a tendency to view π as that probability that a subject would arrive at, given infinite time for reflection (Brown and Lindley, 1978, and Tani, 1978). This would not help with the elicitation of $p(q|\pi)$, for how can a subject be expected to give a sensible assessment of what he would think if he spent longer thinking?

The author feels that the entire concept of "true" probabilities arises out of the tendency of Bayesians to believe that, within the field of decision-making, probability theory must be the appropriate modeling tool. So, in this situation, when we discover that the probability axioms are being violated, a second-level model is built, with hypothetical true probabilities included, so that we may once again use probability theory.

It is an ingenious mathematical trick, but there seems to be an element of getting the situation to fit the theory, rather than adapting the theory to fit the situation, which I would contend was the real aim of mathematical modeling. There is perhaps a suspicion here that we are clinging to old and trusted friends when they can really do nothing to help us. On the other hand, there is often a counter-tendency to throw out old methods as soon as we have difficulties with them (cf. the current high divorce rate in the U.S.) rather than trying to work out these difficulties within the existing system. It is better, whenever possible, to try out all the available alternatives, and base our choice of methodology on a full examination of these. LTB have developed the Bayesian method of performing reconciliation. In this paper we develop a heuristic based on an approximation to their technique, and in a forthcoming paper (Freeling 1980a) we develop another alternative, based on a different axiomatic theory (Fuzzy Set Theory, see Zadeh, 1965; Freeling, 1980b). These alternatives should be compared to decide upon their appropriateness in different situations.

When we are dealing with multiple experts, whose knowledge a DM is obtaining via probability assessments, in order to improve his/her own probability assessments, then the Bayesian updating paradigm is an elegant way of modeling the DM's problem. Then the π 's would have an interpretation as the DM's probabilities. When we have only one probability assessor, we have to postulate a hypothetical split personality, and a hypothetical "meta-DM" who wishes to update his/her own (true) probabilities. Not only is the postulate somewhat clumsy, but it also, of necessity, involves us in third- and higher-order probability

assessments. For, if we discovered incoherence in the initial probability assessments made by the DM, we are likely to discover incoherence in his/her second-order, more difficult, judgments. This means that we shall have to elicit further judgments in order to reconcile these second-order assessments, and so on, ad infinitum (see Section 3.5).

When faced with a situation where a normative model is proving inadequate, due to its being insufficiently descriptive, we have two alternative ways of proceeding. We may either accept the limitations of our present model as an approximation to reality and work within these limitations, or we can attempt to enrich our normative model by extending the axioms in such a way as to better model reality. If we take the first course, extra effort must go into transforming observed behavior into a form compatible with the normative model, whereas such transformation will hopefully be unnecessary if we enrich the model successfully.

In the context of RIJ, these two alternatives take, respectively, the form of 1) discovering a mathematical procedure which produces consistent, conventional probabilities from incoherent elicited probabilities, or 2) extending probability theory in such a way as to allow the elicited values to be direct inputs to the decision model.

3.3 The π 's as Parameters

What is the reason behind introducing a Bayesian updating, and the π 's? Surely it is not intended in any way to extend the theory by making it more behavioral, but rather it is a mathematical device which is being

used to cope with an awkward situation. In that case LTB may not intend to imply that a given subject really does, somehow, have a set of true probabilities. π is simply a parameter which can, we hope, enable us to gain a better understanding of the underlying process.

Parameters are very often introduced in this manner in statistical inference models. Justification for such an introduction must be made on the grounds of whether the parameter really helps the inference. Thinking in terms of the mean and variance of a Normal distribution is often very useful. Part of the reason for this is that spread and central tendency have intuitive meanings. However in the present situation, as argued above, π , even if interpreted as a true probability, does not.

Furthermore, a Bayesian updating model is typically of value when there is good reason to believe that a random process is underlying a situation. Where there are inconsistent probability assessments, there is often a causal explanation of the inconsistency available, and in this case, the Bayesian model is inappropriate. For this reason, the work by von Winterfeldt (1980) is of particular importance. In Section 6.0 we briefly examine alternative methods of reconciling judgments, relating them to the possible identified sources of inconsistency.

I must once again stress that it is in no way my intention to negate the value of the work described in LTB. It has brought forcefully to attention a potentially serious difficulty with applying decision-analytic methods and equally importantly it builds a mathematical structure for analyzing the problem. By examining that structure critically, we hope to provide practicing analysts the means to decide rationally what

reconciliation methodology is appropriate for them to use. The method advocated in this paper is intended to be simpler than the method of LTB and it is hoped sufficiently useful to apply in the majority of situations. The ideas based on Fuzzy Set Theory are part of the initial stages of an extended theory of belief, which the author feels is only partially developed, but which should be vigorously studied if analysts are to remain responsive to the needs of decision-makers.

3.4 The Role of Bayesian Updating

The technique of RIJ developed in LTB does not lie clearly in either of the two areas outlined at the end of Section 3.2. This is due at least partly to confusion over the interpretation of the entity π , as true probability, or as parameter. When π is viewed as a "true" probability, we are reinterpreting the rationale of subjective expected utility to mean that a DM attempts to maximize the expected utility calculated with his/her true probabilities, and that as we can never know these true probabilities we attempt to get the best estimate for them that we can.

In that case I would argue that what we are actually doing is extending the decision-theoretic framework in an attempt better to describe the workings of the mind of a DM, though we may use the tried and tested calculus of probability within this description. It should however be made clear that there is an implicit extension of the axioms involved here; viz., that we should attempt to maximize expected utility calculated with these true probabilities rather than with the elicited probabilities.

If, on the other hand, we interpret π as simply a parameter, then this is just a mathematical construct to help us provide consistent probability inputs to our decision-analytic model, cf. alternative 1 of 3.2.

As an axiomatic extension, we do not believe the concept of true probabilities to have been very successful. It is not acceptable as an improved descriptive theory, since it contravenes too much of the existing psychological literature. The work already mentioned using Fuzzy Set Theory (Freeling 1980a,b) should be compared with LTB viewed in this light. The author believes that some such extension of the calculus we use to model uncertainty will, in the long run, prove to be of the most value.

So far as the parametric interpretation of π goes, it should be compared to the heuristic discussed in Section 5.0. Both methods use mathematical manipulations to transform our elicited probabilities into acceptable inputs for a decision analysis. The selection of a particular technique in any given situation will depend on the context.

3.5 Extending the Axiomatic System

Brown and Lindley (1979), state that enriching the axiomatic structure is "unlikely to be successful," and support their claim with an analogy to a surveyor engaged in the measurement of angles. Such a surveyor does not extend the axioms of Euclidean geometry to explain differences in the measurements, but uses an error theory. However, this analogy may be inappropriate to the problem of RIJ. In this section we provide a counter-argument for the use of an axiomatic extension, and support it with a continuation of the above analogy.

Our view of the psychological reality of subjective probabilities may be roughly stated as follows. We view the DM as a "black box," as in systems theory, about whose internal mechanisms we know nothing; we can, however, observe the inputs to and outputs from the black box. In the case of a DM, the inputs are data from the real world, and the outputs are the decisions taken. The DA paradigm, then, is simply a model we make of the internal workings of the DM's mind. The value of the paradigm lies in the fact that, in most cases, the variables in the model (i.e. probabilities and utilities) have sufficient intuitive appeal (i.e. are sufficiently descriptive) to the DM for him/her to be able to place numerical values on them, and thus use the model as a normative decision aid.

However, in using this DM model, we do not (or at least, we ought not) claim that the actual workings of the black box are as in our model. The probabilities we use are, at best, an approximation to whatever actually occurs in the human mind. There is a vast amount of psychological literature showing that the Savage axioms are only approximately obeyed by real subjects (e.g. Slovic and Tversky, 1974; Kahneman and Tversky, 1979; Tversky, 1969). In this case, there is no such entity as a "true" probability that actually exists within a DM's mind, and probing deeper and deeper for such a number must inevitably lead to asking questions which do not have a clear intuitive meaning to the DM. In this case we can only expect further contradictions and confusion to arise.

Brown and Lindley (1979) are aware that with the Bayesian updating approach to RIJ, there is a potential for second-order incoherence, which they suggest could be reconciled in the same way as the initial

incoherence (i.e. using a Bayesian approach). This leads to the concept of an infinite regress, which the authors suggest may converge to a single value, π . In the context of our "black box" model, such convergence appears most unlikely. This is because we would be asking the DM to be ever more precise about a concept which is perforce only a vague one in his/her mind. Even if convergence occurred, we would have to expect that the convergence was a mathematical phenomenon, rather than an accurate model of a real process. We should be very careful of putting an intuitive explanation to the limit point achieved. The mathematics of the regress are, however worthy of attention, to see if convergence can in fact occur, and whether such a limit does have an intuitive explanation.

We may extend the analogy of the surveyor here. We are not, however, in the situation of taking several "readings" on a particular quantity. Rather, we are in a similar position to a surveyor who is measuring angles over a very large area of ground, so that the earth's curvature becomes a factor. The Euclidean result that the angles of a triangle always sum to 180° could be used to reconcile the surveyor's readings, but, in fact, we know that plane geometry is simply an approximation to spherical geometry. In this case, then, the correct way to proceed is to extend the axiomatic system by taking the earth's curvature into account, and work with spherical geometry. This exemplifies the fact that when the approximation with which we are working can be shown to be inadequate, it is correct to seek a better approximation. On the other hand, the fact that an extension of the axiomatic system may provide a more precise view of reality, as with spherical geometry, does not mean that the cruder approximation is to be discarded. In many situations, the simpler model

is satisfactory, and so should be used, e.g. 360 is a sufficiently accurate value of the number of degrees in the angles of a square room -- nobody would bother calculating the error due to the earth's curvature. Another example of a successful axiomatic extension which is usually not necessary for calculations is the extension of Newtonian dynamics to relativistic dynamics.

When dealing with RIJ, the types of axiomatic extensions at present available to us would replace probabilities by "approximate probabilities." These would, roughly speaking, replace the present precise probability values by ranges of permissible values. The previous work on such extensions was performed by Dempster (1968) and Smith (1965) who produced ranges for probabilities, by Shafer (1975) who has extended the work of Dempster and Smith to provide a new evidential theory of belief, based on entities he has termed "belief functions," and by Watson et al. (1979), and Freeling (1979, 1980b), who used the new ideas of Fuzzy Set Theory (Zadeh, 1965) to produce a richer extension. An example of how the fuzzy set ideas could be used to explain away inconsistent probability assessments is given in Freeling (1980a).

Although I believe that working with axiomatic extensions is a valuable direction in which to proceed, it is clear that we are not yet very close to an adequate extension. There is therefore a great deal of value in continuing the previous work on RIJ, and trying to produce consistent probabilities from elicited inconsistent ones. Furthermore, even if an adequate axiomatic extension is found, a good procedure for RIJ based on Savage's axioms may prove more useful in the majority of situations, as a

proxy for the more complete analysis. For this to be the case, it would be preferable for the RIJ procedure to be as simple to apply as possible. This desideratum is a motivation for much of the research described later in this paper.

3.6 Philosophical Background

It is worthwhile at this point to take a look with a very broad perspective at our purpose in pursuing research in this area. The discussion in this section draws partly on the work in the philosophical literature on Inquiring Systems (see for example, Churchman, 1971; Mitroff and Turoff, 1975; Mitroff, Betz and Mason 1970; Mitroff 1974).

The main thesis of this section is that, underlying any form of scientific inquiry, there must be a philosophical basis or theory about the nature of the world, on which that inquiry ultimately depends. The authors mentioned above classify an Inquiry System (IS) in accordance with the general philosophical view on which it appears to be based. The types of IS they define can be broadly placed into two categories -- those which rely essentially on one model to arrive at "truth" (be this model empirical or theoretical) and those which believe that truth can only be arrived at by taking into account several different models. The traditional method for dealing with potential inconsistencies appears to arise out of an IS in the first category. The procedure very often applied is simply to take only a minimally specified set of readings (i.e. to use only one model), so that there is no potential for incoherence. An alternative method is, if inconsistent probability assessments are discovered, to look at these, decide which the subject feels are "wrong"

and then simply discard them. This is effectively the same as using only one assessment, although we have used somewhat more information in choosing which assessment. If neither of the above methods are liked, the present practice is to ask the subject to choose a value which he/she finds easiest to "live with" (see e.g. Brown, Kahr, and Peterson, 1974). An attempt at formalizing this procedure is discussed in Section 5.0. It will be noted that such an idea still does not make use of the concept that having alternative models is a necessity to arriving at truth.

It is our contention that such a concept is the correct one for a foundation of our work in RIJ, i.e. that we should use an IS which falls into the second category discussed above. For, surely, the motivation for wishing to take more than a minimally specified set of assessments is that by using different models of the uncertainty, we can improve on the values achieved with only one model. (The alternative potential models here are, for example, holistic assessments, as opposed to conditioning on various events.) By achieving some form of synthesis of the results from the different models, we hope to have improved our understanding of the world.

The work in LTB seems to be based on the assumption that we have only one model, and that each assessment of the target probability is one reading taken with this model. It then makes sense simply to build a measurement theory to reconcile these readings. If we are thinking of the assessments as coming from different models, however, we should concentrate on finding the strengths and weaknesses of each, and using this information as the basis from which to arrive at an "improved" probability value. We thus are not thinking in terms of errors made by the subject--rather we

consider the assessments to be the results of different approximating models each of which provides us with information about the uncertainty. Our reconciliation procedure should thus attempt to quantify the accuracy of each approximation model, and formally arrive at a reconciled value. It is also hoped that such quantification will permit us to suggest which approximation model, or combination of models, would be most appropriate to use to gain a better understanding of the nature of the world. This is the same concept as the "design problem" posed in LTB, or the question of "digging" introduced by Brown and Lindley (1979).

The differences in separate assessments can be ascribed to the two aspects of the elicitation process mentioned in 2.4 (i.e. different information considered, and different ways of processing it), and also to errors such as biasing, optimism/pessimism etc. It is convenient to treat these latter problems separately from the first two, by identifying the errors and eliminating them as far as possible before a final reconciliation is performed. An example of how this might be done is given in Section 6.0. We shall, therefore, assume that the inconsistency arises only from discrepancies in data and in processing. It is, after all, such discrepancies we are seeking.

We are now moving closest to the Hegelian dialectical IS (see Mitroff et al. 1970). This takes the view that to arrive at "truth," we should take two models with opposing theoretical bases, but with access to the same information, and by examining the ways in which these disagree, arrive at a synthesis which represents an improved "Weltanschauung." With RIJ, the

common data base is the subject's psychological field, and we hope that our synthesis takes account of as much of this field as possible, whilst processing in the most relevant way. This of course is a philosophical view of the project, and we cannot expect total success, but it gives us a background against which to work. The procedure developed in Section 5.0 may be viewed as a first attempt at quantifying this concept.

3.7 Expert Use

There has been a number of methods suggested in the past for dealing with the expert use problem. This is the case of a decision maker who is using experts to help improve his/her decisions. If we have two experts, each assigning different probabilities to the same event, the DM is faced with an inconsistency which must be reconciled in order to arrive at his/her own probabilities. Since this problem has been extensively addressed, it is a natural place to commence our search for a procedure for RIJ.

The Bayesian updating approach of LTB is an adaptation of the methodology proposed by Morris (1974) to deal with expert use. To permit the use of the methodology for a single DM, the authors of LTB are forced to postulate a hypothetical division of the DM, into an expert and a user. The main difficulty with this hypothesis is that some of the concepts which make sense when there truly are different people involved, do not make sense when we have only a single DM. Morris states "the key idea (of his work) was the distinction between the meaning of an expert's probability assessment to the DM and to the expert himself: to the expert, the probability assessment is a representation of his state of information, to the DM, the probability assessment is information." It is

hard, if not impossible, for an individual to think of his/her own assessments as new information, separate from his/her previous state of information. It is this difficulty, presumably, that led to the assumption of the existence of π .

Similarly, the work of Mitroff et al. (1970), which uses the dialectical IS as a basis, addresses the problem of expert use. Again, the specific methods they postulate for quantification of their ideas, do not really carry through to the case where our two "experts" are an hypothetical construct. Such analogies with the expert use problem can provide a useful way to thinking about RIJ, and have undoubtedly been valuable in formulating an initial approach to the problem. However, it is only an analogy, and we should be sure that the analogy holds in any particular situation in which a reconciliation is necessary.

3.8 Summary of the Different Approaches Identified for RIJ, and their Rationales

In this section, we give a brief summary of each of the different approaches so far identified as being potentially useful for RIJ. The emphasis here is on the perspective of the problem underpinning each approach, and the way this has been translated into a quantitative tool. The similarities and differences of each approach are highlighted.

3.8.1 Bayesian updating. This is the approach already described, as in LTB. It is essentially a "measurement error" approach, and is based on the concept that a probability assessor is attempting to discover his/her "true probability," π , and that this is best achieved by taking several

different "readings" on π , each subject to random error. π may be interpreted merely as a parameter (see earlier) but this rationale remains -- π needs some intuitive meaning.

3.8.2 Least squares. This procedure is developed in LTB as an approximation to the internal approach of Bayesian updating. The basic idea is to discover $\hat{\pi}$, our "best" estimates of π , by finding the solution to

$$\text{Minimize } (q - \pi)^t W (q - \pi)$$

subject to the coherence constraints. The q_i and π_i may be transformed here, to log-odds, for example. The weights w_{ij} are defined in LTB as the elements of the inverse of the variance matrix of q_i , but alternative definitions are possible, and these are considered in Section 5.0. The basic idea of this approach is one of either "confidence" or "stability." By confidence we mean that our reconciled vector, $\hat{\pi}$, should be as similar as possible to q , with the q_i in which the subject has least confidence, changing the most. Alternatively, we may view this as a way of modeling the informal "jiggling" of the elicited values that a subject would perform if asked to reconcile the numbers, without aid. The reconciled vector, $\hat{\pi}$, would then represent that coherent set of values with which the subject felt most "comfortable," or which required the least "mental anguish" in order to adjust the elicited q_i .

3.8.3 "Information" approach. This approach is a heuristic developed as an approximation to the Bayesian approach, and is described in Section 5.0. Although this approach is not fully developed, we feel it is most likely to prove useful in reconciliation. It is based on an alternative

conception of what is truly underlying the least-squares approach. This is that we really wish to take a weighted average of our different elicited values for the target probabilities, with the weights being determined by the amount of "information" contained in each assessment - that with most information having largest weight, etc. Here information is being used to include a) parts of the subject's internal cognitive field, or data base, and b) different ways of processing this data so as to produce probabilities. This idea is based on the observation, made in LTB and extended in Section 5.0 of the present paper, that the least-squares approach actually produces a weighted average, and that the weights may be interpreted as representing information, but that this is achieved in a very round-about manner. The present approach improves on this.

The technique can also be viewed as arising from the dialectical inquiring system ideas discussed in Section 3.7--the underlying rationale of this approach is that the inconsistencies arise from considering different models of S's uncertainty with each assessment, and that the "best" assessment is that which considers all the possible models.

3.8.4 Alternative simple techniques. The information approach discussed in the previous subsection assumes that there are no consistent biases in the values elicited. If we have a good reason to believe that such biases are indeed present, then some other technique should be used first, before submitting the data to the information approach. This leads to the concept of developing several different reconciliation techniques, each simple, and each designed to address one (or more) of the possible sources of incoherence identified and discussed in von Winterfeldt (1980). In

this way one can envisage, for a given reconciliation problem, using those techniques from our selection which appear most appropriate. One of these techniques would be simply to ask the subject to reconcile the numbers informally, if such appeared appropriate. Another would be to use a form of "satisficing," rather than optimizing. With this concept, one would present different possible sets of reconciled values to the subject until one was presented which was considered the most acceptable. In this way, whilst not necessarily finding a "best" estimate, we find one that is "good enough," without entering into any mathematics. This idea is further discussed in Section 6.0, as are other simple techniques.

3.8.5 Fuzzy set theory. This technique of extending the underlying axiomatic system is not discussed any further in this paper, but is developed in Watson et al. (1979) and Freeling (1980b), and discussed with especial reference to the incoherence problem in Freeling (1980a).

4.0 THE LEAST-SQUARES APPROACH

4.1 Introduction

LTB, aware of the practical difficulties of the Bayesian approach, suggest the least-squares procedure as an approximation to the internal approach, which avoids most of the assessment difficulties posed by the assumption of π . They propose to take $\hat{\pi}$ as the solution to the following constrained minimization problem:

$$\text{Minimize } \sum_{i,j} w_{ij} (q_i - \pi_i) (q_j - \pi_j) \quad (4.1)$$

subject to the coherence constraints $f(\pi) = 0$. Or, in full vector notation, minimize $(q - \pi)^t W (q - \pi)$.

LTB take $W = V^{-1}$, where V is the variance-covariance matrix of the q 's. So, for example, if the q_i are independent, with variances σ_i^2 , the function to be minimized is

$$\sum_i \frac{1}{\sigma_i^2} (q_i - \pi_i)^2.$$

With this definition of the weights, w_{ij} , and under the assumption that $p(q|\pi)$ will be multivariate normal with mean π , and variance independent of π , and that N 's prior beliefs about π are diffuse, so that $p(\pi)$ is approximately constant, the solution to (4.1) is a good approximation to the internal approach. This is the motivation for developing the least-squares ideas.

4.2 The Choice of Metric

It will be noted that the normality assumption is far more reasonable if we are working with log-odds, which can take all values in $(-\infty, \infty)$,

rather than with probabilities which are constrained to the interval (0,1), since the normal distribution has an infinite range. The assumption of equal variance over all possible π does not make sense, working with probabilities, as for q_i close to 0 or 1 we may expect the absolute variance to be small. Such a consideration does not hold true when using log-odds. For some psychological work which lends support to this theory see Wheeler and Edwards (1975). In this chapter we examine the computational consequences of these assumptions, and look at possible choices for the W matrix, together with justifications of these choices.

Using log-odds ($\log$) clearly makes sense, but as we shall see, it also dramatically increases the difficulty of implementing the technique, as we are now faced with quite a complex constrained non-linear optimization problem, which, at best, is soluble only with a sophisticated computerized optimization package. However, there are some potential simplification techniques we now mention.

(If we are interested in assessments solely of the target variable, so that all the q_i are intended to represent the same probability, our constraint is $\pi_1 = \pi_2 = \dots = \pi_n$, and this is unaltered working with $r_i = \log \pi_i$. In this case, then, we are still faced with a quadratic optimization problem with simple linear constraints, and this situation can be easily handled (see next section). With any other constraints, however, the form of the constraints is made more complex by transforming to log-odds: e.g. $\pi_1 = \pi_2 k + \pi_3 (1-k)$ becomes a very nasty expression if we work with $r_i = \log \pi_i$.)

Let $r_i = \log q_i$ and $\rho_i = \log \pi_i$. Then we wish to minimize $(\underline{r} - \underline{\rho})^t W (\underline{r} - \underline{\rho})$ over all $\underline{\rho} \in \mathbb{R}^n$, subject to $f(\underline{\pi}) = 0$. By Taylor's theorem, however, we see that

$$(r_i - \rho_i) \approx \frac{(q_i - \pi_i)}{q_i (1 - q_i)}, \text{ since } \frac{d}{dx} (\ln(x/1-x)) = \frac{1}{x(1-x)}.$$

So if we set

$$m_i = \frac{1}{q_i (1 - q_i)} \quad \text{and} \quad u_{ij} = m_i m_j w_{ij} \quad \forall i, j,$$

we see that

$$\begin{aligned} \sum_{i,j} (r_i - \rho_i) w_{ij} (r_j - \rho_j) &\approx \sum_{i,j} (q_i - \pi_i) m_i w_{ij} m_j (q_j - \pi_j) \\ &= (\underline{q} - \underline{\pi})^t U (\underline{q} - \underline{\pi}). \end{aligned}$$

So we have reduced our non-linear optimization problem to a constrained optimization problem with a quadratic objective function. What we have in fact done is to approximate the log-odds metric on $\mathbb{R}^n$ by the best Euclidean metric at the assessed $\underline{q}$. An idea of the magnitude of the error of the approximation can be gained by looking at the next term in the Taylor expansion of $\log x$. This is

$$\frac{1}{2} (q_i - \pi_i)^2 \cdot \frac{(2q_i - 1)}{q_i^2 (1 - q_i)^2}$$

which is small if q_i is near 0.5, or if π_i is close to q_i . This conforms with our intuition that log-odds will only give radically different answers if the probabilities involved are extreme, and that we only run into problems if there is a large degree of incoherence. In fact, in all examples tested, the approximation gave satisfactory answers. We therefore suggest that this approach should be used, for, as we shall see, the computation is relatively straight-forward with a quadratic objective function.

4.3 The Linear Model

It may not at first glance be apparent that the least-squares procedure discussed here is in fact equivalent to the well known statistical procedure of least squares used in finding the parameters of a linear model, for example with analysis of variance (see e.g., Scheffe, 1959). The constraints may appear unfamiliar. However, in some cases our problem is in precisely that form, and hence the analysis of the linear model can be used.

The general linear model takes the form:

$$\begin{aligned} \tilde{q} &= \tilde{A} \tilde{\pi} + \tilde{\varepsilon} & E(\tilde{\varepsilon}) &= 0 \\ & & \text{Var}(\tilde{\varepsilon}) &= \tilde{V} \end{aligned}$$

where $\tilde{A}$ and $\tilde{V}$ are matrices.

Suppose we are dealing with n different estimates of a target probability.

Then we may state our problem as

$$\begin{pmatrix} q_1 \\ q_2 \\ . \\ . \\ q_n \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ . \\ . \\ 1 \end{pmatrix} \pi + \tilde{\varepsilon}. \quad \text{So here } \tilde{A} = \begin{pmatrix} 1 \\ 1 \\ . \\ . \\ 1 \end{pmatrix}.$$

Alternatively, if we have assessments for an event X and for its

complement $\sim X$, then we may transform $q_2 = q(\sim X)$ to $x_2 = q_2 - 1$ and then we have

$$\begin{pmatrix} q_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \pi + \tilde{\varepsilon},$$

$$\text{so here } \tilde{A} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Again, if we have the constraint $\pi_1 = \pi_2 k + \pi_3 (1-k)$

we have

$$\begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} = \begin{pmatrix} k & 1-k \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \pi_2 \\ \pi_3 \end{pmatrix} + \tilde{\epsilon},$$

$$\text{So } \tilde{A} = \begin{pmatrix} k & 1-k \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Then the solution of this problem is well known, being

$$\hat{\tilde{\pi}} = (\tilde{A}^t \tilde{V}^{-1} \tilde{A})^{-1} \tilde{A}^t \tilde{V}^{-1} \tilde{q} \quad \text{and} \quad \text{Var}(\tilde{\pi}) = (\tilde{A}^t \tilde{V}^{-1} \tilde{A})^{-1}.$$

With the aid of a computer, this can be easily calculated. With more complex, non-linear constraints the linear model is not applicable, however, but we may use the powerful mathematical tools made available to us by Lagrangian theory. In the next section we exemplify the use of this concept with fully worked out examples in the linear case.

4.4 Use Of The Lagrangian

The first thing that should be noted is that it is due to the constraints that we cannot perform the optimization simply by differentiating the objective function and setting it to 0, for differentiating

$(\tilde{q}-\tilde{\pi})^t \tilde{W}(\tilde{q}-\tilde{\pi})$ with respect to $\tilde{q}$ gives

$$2(\tilde{q}-\tilde{\pi})^t \tilde{W}.$$

But this can never equal 0, for $\tilde{W}$ is of necessity positive-definite², so it has an inverse. Thus

$$2(\tilde{q}-\tilde{\pi})^t \tilde{W} = 0 \Rightarrow (\tilde{q}-\tilde{\pi})^t \tilde{W} \tilde{W}^{-1} = 0 \cdot \tilde{W}^{-1} = 0 \Rightarrow (\tilde{q}-\tilde{\pi})^t = 0 \Rightarrow \tilde{q} = \tilde{\pi}$$

and this is not possible if q was incoherent. Instead, with a quadratic objective function, and the fairly simple constraints usually present we can use the concept of the Lagrangian function to discover an analytical solution. We shall consider two cases to exemplify the method.

1) A Partition

Suppose we have assessed the probabilities of a partition A_i , so that our required constraint is $\sum_i \pi_i = 1$.

Then to reconcile our incoherent q 's we wish to minimize

$$\sum_{i,j} (\pi_i - q_i) w_{ij} (\pi_j - q_j) \quad \text{subject to} \quad \sum_i \pi_i = 1.$$

Form the Lagrangian

$$L = \sum_{i,j} (\pi_i - q_i) w_{ij} (\pi_j - q_j) - \lambda (\sum_i \pi_i - 1).$$

Differentiate with respect to π , and set to 0:

$$\frac{\partial L}{\partial \pi_j} = 2 \sum_i (\pi_i - q_i) w_{ij} - \lambda.$$

Suppose $W^{-1} = V$.

Then we can multiply the above expression by V , to give

$$2(\hat{\pi}_k - q_k) = \lambda \sum_i v_{jk} \Rightarrow \hat{\pi}_k = q_k + \frac{\lambda}{2} \sum_j v_{jk}.$$

It now remains to find λ , which is done by using the constraint $\sum_k \pi_k = 1$.

$$\text{So } \sum_k q_k + \frac{\lambda}{2} \sum_k \sum_j v_{jk} = 1 \Rightarrow \lambda = 2 (1 - \sum_k q_k) / (\sum_k \sum_j v_{jk}).$$

Then

$$\hat{\pi}_k = q_k + (1 - \sum_k q_k) \sum_j v_{jk} / (\sum_k \sum_j v_{jk}).$$

Note that if we define $\tilde{W}$ as the inverse of the variance-covariance matrix, these v_{jk} are simply the variances and covariances. This means that we never have to perform the inversion of the matrix, and thus the computation is simplified.

As a special case, if all the correlations are taken to be zero, so that we are minimizing $\sum_i w_i (q_i - \pi_i)^2$ subject to the constraints, our expression reduces to:

$$\hat{\pi}_i = q_i + (1 - \sum_j q_j) / w_i \sum_j w_j^{-1}$$

which is equation 16 of LTB.

2) External conditioning

Suppose we have made two assessments of a target variable--one holistic and the other via decomposed estimates conditioned on an external event of known probability k . Then our constraint is

$$\pi_1 = \pi_2 k + \pi_3 (1-k)$$

where the notation should be clear.

Proceeding as before,

$$L = (\pi - q)^t \tilde{W} (\pi - q) + \lambda (\pi_1 - k\pi_2 - (1-k)\pi_3),$$

$$\frac{\partial L}{\partial \pi} = 2(\pi - q)^t \tilde{W} - \lambda [1 \quad -k \quad k-1],$$

and

$$(\pi - q)^t = \frac{\lambda}{2} [1 \quad -k \quad k-1] \tilde{W}^{-1} .$$

If $\tilde{W}^{-1}$ is the variance-covariance matrix

$$\begin{bmatrix} \sigma^2 & \rho\sigma\tau & \rho\sigma\tau \\ \rho\sigma\tau & \tau^2 & \delta\tau^2 \\ \rho\sigma\tau & \delta\tau^2 & \tau^2 \end{bmatrix} \quad (4.2)$$

where we assume q_2 and q_3 have the same statistical structure, (so that σ^2 is $\text{Var } q_1$, $\tau^2 = \text{Var } q_2 = \text{Var } q_3$, ρ is the correlation between q_1 and q_2 , and δ that between q_2 and q_3 .)

we have $(\pi - q)^t = \lambda$
$$\begin{bmatrix} \sigma^2 - \rho\sigma\tau \\ \rho\sigma\tau - k\tau^2 + k\delta\tau^2 - \delta\tau^2 \\ \rho\sigma\tau - k\delta\tau^2 + k\tau^2 - \tau^2 \end{bmatrix}$$

and again we find λ by substituting $\hat{\pi}$ into the constraint function, giving

$$\lambda = \frac{q_1 - kq_2 - (1-k)q_3}{(k^2-1)\delta\tau^2 - 2k^2\tau^2 + 2k\tau^2 - \tau^2 + 2\rho\sigma\tau - \delta^2} .$$

We see that already the expressions are very complicated, and in the more complex situations, an analytical solution cannot be found. For, although the Lagrangian ideas exemplified in this section are still

applicable, the expression $\frac{\partial L}{\partial \pi} = 0$ has no analytical solution with most non-linear objective functions.

4.5 A Procedure for Optimizing in the Log-odds Metric

When working with log-odds we are forced to use numerical techniques and a computer. Constrained non-linear optimization problems are in general hard to solve, but in most cases we shall be able to exploit a special structure in order to change our problem into an unconstrained one, which will be more tractable.

This is achieved simply by substituting from the constraints into the objective function. This makes the objective function more complicated, but in most cases with probability constraints this increase in complexity is minimal, and the saving in computation obtained by removing the constraints is well worth it. As an example, suppose we have external conditioning; then we must minimize:

$$\sum_{i,j=1}^3 (\log q_i - \log \pi_i) w_{ij} (\log q_j - \log \pi_j)$$

subject to $\pi_1 = k\pi_2 + (1-k)\pi_3.$

Then whenever π_1 appears we substitute $k\pi_2 + (1-k)\pi_3$ for it, thus eliminating the constraint and simultaneously reducing the dimensionality of the problem. This is another attractive feature of this approach.

The resulting unconstrained problem can be solved using a method such as conjugate gradients, which is well adapted to this type of problem, being both robust and efficient. However, a lot of work is required for the programming. We therefore suggest always approximating the log-odds metric by a

Euclidean one, if at all possible, as described in section 4.2. However, the observations of the next section may be used to simplify the computation.

4.6 A Practical Example of the Methodology

To exemplify the method, consider again the Oxford and Cambridge boat race. Then A, the target event, is "Oxford wins" and take X to be "Oxford wins the toss." Then $q_1 = q(A) = 0.7$; $q_2 = q(A | X) = 0.8$; $q_3 = q(A | \sim X) = 0.5$, and we will assume that $P(X)$ is known to be 0.5. Then, in general, the matrix will take the form of (4.2). Then if we assume that all the assessments q are independent and have equal variance, so $\sigma^2 = \tau^2 = 1$, $\rho = \delta = 0$, and W becomes the identity matrix, we find that $\hat{\pi}_1 = 0.67$ is the reconciled value. Furthermore, if we define precision as the inverse of the variance, as is often done, we find that the precision of $\hat{\pi}$ is three times that of the original assessments. This calculation appears in LTB, where it is used to indicate the dramatic increase in precision achieved by taking multiple assessments and reconciling them. There are several caveats about this procedure which should be considered. These are discussed in the next section.

4.7 The Importance of Correlations

I now wish to show that the results of LTB in fact arise largely from the assumptions of independence among assessments. Indeed, the increase in precision by a multiple of three is directly attributable to this assumption, as may be understood from the following heuristic argument. Precision as here defined is closely related to the statistical concept of

the amount of information described by an assessment. Also, independence among assessments is equivalent to saying that each donates entirely different information to the reconciliation. Hence, with three independent assessments, we have thrice the information, and equivalently, thrice the precision.

Such an analysis interpreting the quality of assessments in terms of their information content has a very intuitive appeal. It was after all an attempt to consider all the possible available information (in the form of searching the assessor's psychological field) that prompted this search for incoherence. However, the assumption of independence among assessments is clearly untenable. For much of the same information will be used in assessments directed towards the same target variable (e.g., $q(A)$ and $q(A|X)$). Alternatively, looking at the same situation from a different angle, if $q(A)$ is overestimated we might well expect $q(A|X)$ to be overestimated as well, since $\underline{S}$ might make the same mistake in each case, so the correlation would be non-zero.

LTB are aware of the falsehood of the independence assumption, and the effect this has on precision. They extend the previous example, in a calculation to be found in an earlier draft of LTB (Lindley, Tversky and Brown, 1978), by taking $\sigma^2 = \tau^2 = 1$, $\rho = \delta = 0.5$. The calculations again give 0.67 as the reconciled value, but the precision is only increased by a half.

However, LTB appear to ignore this fact when making one of the major conclusions of their paper. For they conclude that the following

procedure is a good one for increasing precision. Find a partition X_i ($i = 1, \dots, n$) of the sample space, such that $P(X_i) = P(X_j)$ for all i, j (where these probabilities are assumed known, e.g., as, perhaps, with a coin toss). Then "extend the conversation about A" to include X_i , by assessing $q(A|X_i)$ $i = 1, \dots, n$. Then we have two assessments of the target variable $P(A)$, in a direct, holistic assessment $q(A)$, and an indirect decomposed assessment $\sum_i q(A|X_i) P(X_i)$. These should be reconciled via the least-squares procedure.

LTB show that under the assumption that all assessments are independent and of equal variance, this procedure gives an increase in precision by a factor of $n+1$, and that using an equiprobable partition is optimal over all partitions of size n . They then suggest that we should thus always try to extend the conversation to include such an equiprobable partition. When correlations are included, this conclusion no longer holds true. (Indeed, we can see that precision is maximized by utilizing as much information as possible. This concept is made more explicit in a later section.) To take an absurd example, suppose in the boat race example, that I decide to condition not on the relevant coin toss, but on a coin I toss. Then the analysis would be identical to the real case if correlations are ignored, yet clearly there should be no increase in precision by considering irrelevant events. The point is that $q(A)$, $q(A|X)$ and $q(A|\sim X)$ should all be very similar, as they are really the same assessment, and so the correlations are very high.

LTB also state that correlations "have little effect on the probabilities," noting that in each of the above examples, the reconciled

value was 0.67. They assume that the correlations affect only the precisions, not the values. This is untrue, as can be seen by once again altering the variance-covariance matrix of the previous example. With $\sigma^2 = 4$, $\tau^2 = 1$, $\rho = \delta = 0.5$, we find that the reconciled value is 0.645. This is a reconciliation of 0.65 and 0.7 which is different from 0.67, and perhaps somewhat counter-intuitive. This is due to the fact that describing the relationship between assessments by correlations is difficult and not very intuitive. Whereas one can assess variances fairly well by asking for credible intervals for an assessment, assessing correlation coefficients is not so easy. With the boat race example, the author was not able even to produce a variance-covariance matrix for his own assessments which was positive definite. He thus has very little faith in direct methods of assessment for correlation coefficients. As these correlations have been shown in this section to be of paramount importance to the least-squares technique, we now proceed to look at alternative ways of interpreting the relationships causing non-zero correlations, and thus making indirect assessments of correlation coefficients.

5.0 ALTERNATIVE INTERPRETATIONS OF THE LEAST-SQUARES PROCEDURE

5.1 A Psychological Interpretation of the Metric

The expression (4.1) is an example of a generalized distance, or metric.

The matrix $\tilde{W}$ transforms the familiar Euclidean space R^n to a curvilinear space. In this section we interpret this curvilinear space as being the psychological field of the assessor, with respect to his/her assessments.

So, if $\underline{q}$ is the assessed probability vector, and $\underline{x}^i$ ($i = 1, \dots, m$) are other possible probability vectors, then that $\underline{x}^i$ minimizing

$(\underline{q} - \underline{x}^i)^t \tilde{W} (\underline{q} - \underline{x}^i)$ is the closest $\underline{x}^i$ to $\underline{q}$ in this psychological space.

An intuitive understanding of this distance is to consider it as measuring the unease of $\underline{S}$ at being forced to take some vector other than $\underline{q}$ as the probability vector. So the solution to (4.1) is that probability vector which satisfies the coherence constraints, which $\underline{S}$ is least unhappy using as his/her probability vector. This then gives us an alternative method for assessing the distance matrix $\tilde{W}$. If we can discover $\underline{S}$'s perceived distances between different points, we may then deduce the $\tilde{W}$ these distances imply.

Such a program appears very attractive. It does not depend on the assumption of hypothetical true probabilities and its definition of the "best" reconciliation is totally subjective, in the spirit of the theory of subjective probability. However, on a further examination, the method appears unworkable. This can be exemplified by a thought experiment.

If such a metric existed, one would be able to use the methods of multi-dimensional scaling (MDS) to find it. To take a concrete example, suppose

$\underline{S}$ assesses $q_1(A) = 0.5$ and $q_2(A) = 0.4$, so $\underline{q} = (0.5, 0.4)$ revealing incoherence. Suppose the analyst selected $\underline{x}^1 = (0.5 \ 0.5)$ and $\underline{x}^2 = (0.6 \ 0.4)$ for presentation to $\underline{S}$. Using MDS we would require $\underline{S}$ to answer questions of the form:

- a) Which of $\underline{x}^1$ or $\underline{x}^2$ is closer to $\underline{q}$?
- b) Which of $\underline{x}^1$ or $\underline{q}$ is closer to $\underline{x}^2$?

Question a) is answerable, but question b) is not. Both $\underline{x}^1$ and $\underline{x}^2$ are vectors invented by the analyst, and $\underline{S}$ may well find it impossible to assess his/her feelings of discomfort at being forced to move from $\underline{x}^2$ as his/her probability. The mental gymnastics required are too difficult. In fact, we see that the only feelings of discomfort $\underline{S}$ truly has concern moving from $\underline{q}$ to the various $\underline{x}^i$, but not moving between any two points, arbitrarily selected by the analyst. In this case there exists no matrix $\underline{W}$ with the interpretation of this section. (For, if there were, question b) would be answerable.) It is possible that some other method of using this interpretation of the metric may yield better fruit, but for the moment we are forced to look elsewhere for a practical and satisfactory reconciliation procedure.

5.2 Least Squares and a Weighted Average

For the rest of this section we concentrate on two estimates of the target variable, $p(A)$. So q_1 may be a holistic assessment, and q_2 the assessment logically implied by decomposed judgments. Hence from now on the coherence constraints take the form $q_1 = q_2$. We also assume that we are working in log-odds, for the reasons noted in Section 4.

If we use the least squares approach, with $\underline{y}$ equal to

$$\begin{pmatrix} \sigma^2 & \rho\sigma\tau \\ \rho\sigma\tau & \tau^2 \end{pmatrix}$$

then the reconciled value is

$$\hat{\pi} = \frac{(\tau^2 - \rho\sigma\tau)q_1 + (\sigma^2 - \rho\sigma\tau)q_2}{\tau^2 + \sigma^2 - 2\rho\sigma\tau} \quad (5.1)$$

(This is a consequence of the general result stated in 4.3, that

$$\hat{\pi} = (\tilde{A}^t \tilde{V}^{-1} \tilde{A})^{-1} \tilde{A}^t \tilde{V}^{-1} \tilde{q}$$

In this case $\tilde{A}^t = (1 \ 1)$, and the substitution easily gives equation 5.1).

Note that $\hat{\pi}$ is simply a weighted average of q_1 and q_2

$$\hat{\pi} = (Aq_1 + Bq_2)/(A + B) \quad (5.2)$$

with $A = 1/\sigma^2 - \rho/\sigma\tau$, $B = 1/\tau^2 - \rho/\sigma\tau$, though one of the weights may be negative. These weights have an appealing intuitive interpretation. For example $1/\sigma^2$ may be taken as a measure of how "good" an assessment q_1 is and may be viewed as a measure of the amount of information contained in q_1 , so A is the amount of information in q_1 reduced by a quantity due to the correlation. In the next section we interpret this quantity as the amount of information shared by both q_1 and q_2 .

It is appropriate to note here that (5.1) can also be derived in a different way. We may decide a priori to make our reconciled value a weighted average of the two elicited target probabilities. In this case, since our motivation is to increase precision, which we may equate with reducing variance, we wish to seek the minimum-variance weighted average.

Then it is easy to show that the optimal weights are as in (5.1). For, assuming $\hat{\pi} = kq_1 + (1-k)q_2$, we find that the variance of $\hat{\pi}$ becomes

$$k^2\sigma^2 + (1-k)^2\tau^2 + 2k(1-k)\rho\sigma\tau.$$

Then differentiating with respect to k and setting to 0, we find $k=A/(A+B)$ as defined above. This idea is discussed by Bunn (1978), with regard to pooling the results of different forecasting models. It provides an alternative motivation for using this approach, which may be more acceptable to some. We can also see from this interpretation of the least-squares technique that there is an underlying assumption that our elicited values are unbiased estimators, (i.e. that the error of q is expectationally zero; so $E(q|\pi) = \pi$). For, were this assumption false, $\hat{\pi}$ would not be an unbiased estimator, so $E(\hat{\pi}|q) = \pi + a$, where a is non-zero, and a better estimate would be $\hat{\pi} - a$. The implications of this assumption are discussed later.

5.3 An Information Oriented Approach

In this section we argue that the least-squares approach is in fact an attempt to quantify the information (in the broad sense discussed in 3.8) captured by an assessment, and to perform the reconciliation based on this quantification. Consider Figure 3. This diagram illustrates the information accessed by our two (log-odds) assessments q_1 and q_2 . So q_1 has information $|I_1|$ and q_2 has information $|I_2|$. Then $A = |I_1/I_2|$ quantifies the information accessed only by q_1 , $B = |I_2/I_1|$ quantifies the information accessed only by q_2 , and $C = |I_1 \cap I_2|$ quantifies the information common to both q_1 and q_2 . In this formulation, the total amount of information is $|I_1 \cup I_2| = A + B + C$.

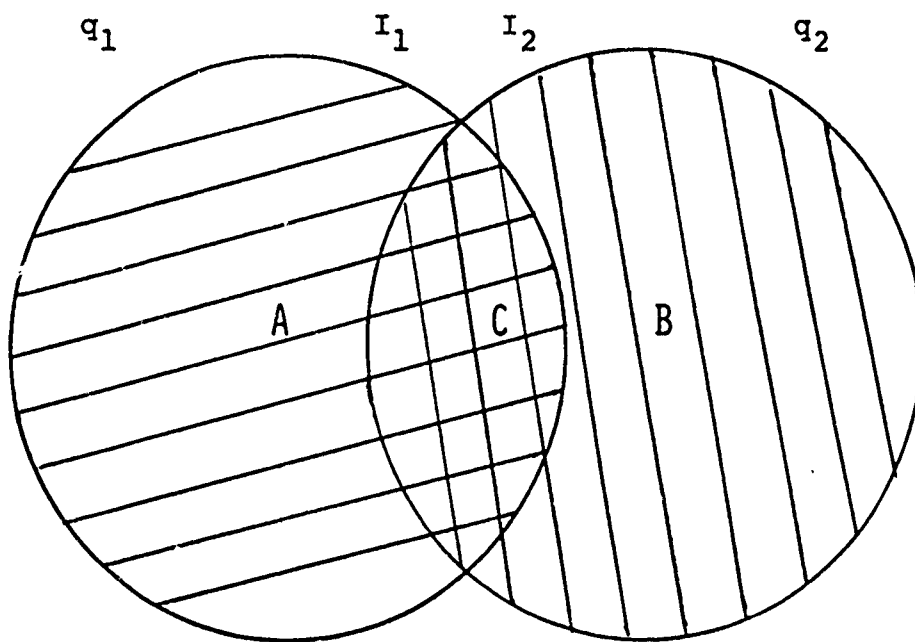


FIGURE 3

INFORMATION OVERLAP BETWEEN 2 ASSESSMENTS

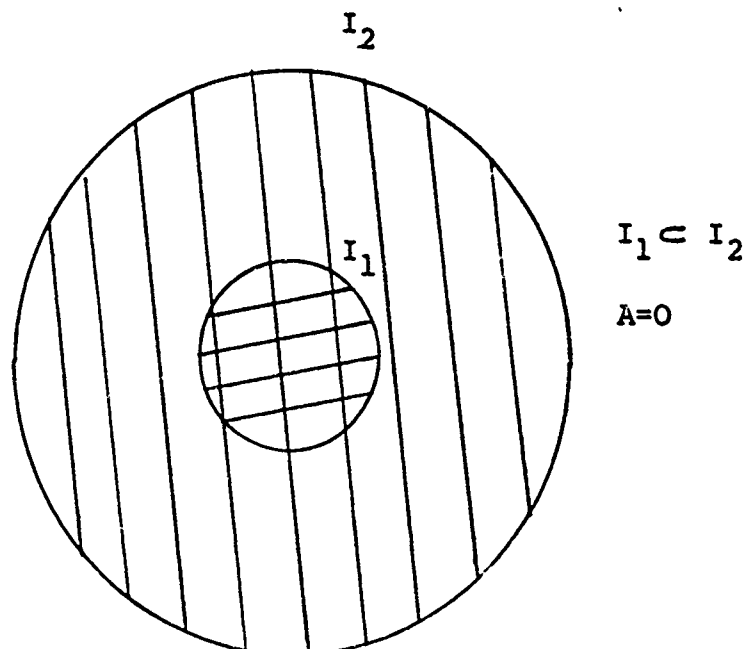


FIGURE 4

ALL THE INFORMATION OF 1 CONTAINED IN 2

A formal definition of "information" is really necessary in order to fully operationalize and quantify the concept. Such a definition is unfortunately very hard to produce.

However, there are two aspects of information; the degree to which S has been able to dip into his/her psychological field, and the extent to which that gleaned data has been correctly processed in accordance with Bayesian principles. The first aspect has the intuitive meaning of describing what of relevance to the target event was taken into consideration when making the assessment. The second refers to the ability of human beings adequately to process data; which, from the psychological evidence, is limited. I believe these concepts can be further explored and made explicit, but for now we must trust our intuition that such concepts have meaning.

We now have a model which is able to describe simply both our motivation for studying incoherence, and the weights that are "optimal" when using a weighted average. First, by eliciting both q_1 and q_2 , we have obtained more information from S than if we had only elicited one of them. It must be better to take account of all this information if possible, rather than using just some of it by using only one assessment of $p(A)$. The increase in quality of our result is measured by the additional information used. Second, an intuitively reasonable way of weighting the two assessments is in proportion to the information unique to them, the information common to both tipping the scales in favor of neither one nor the other. This then makes the natural reconciliation to use $(Aq_1 + Bq_2)/(A + B)$, as in the previous section.

This intuitive procedure has an obvious correspondence to the least squares procedure. $|I_1|$ corresponds with $1/\sigma^2$, $|I_2|$ with $1/\tau^2$, and $|I_1/I_2|$ with $\rho/\sigma\tau$. In particular, this provides a clear interpretation of the correlation coefficient, ρ , in this context. The two assessments are related to the extent that they each draw on the same information. I believe that it is this relationship we are attempting to quantify by including ρ in our analysis. However, quantifying the relationship in terms of information content is a more natural way of proceeding, as ρ is a non-intuitive entity. This explains both the difficulty involved in assessing a ρ that is coherent, and also the potential for unexpected (and unsatisfactory) reconciliations which arise from using a ρ which does not correctly capture one's belief.

As an example of the value of these information ideas, consider the classical situation, such as the boat race example, where q_1 is a holistic assessment, and q_2 a decomposed one. Then the assumption (often unspoken) of decision analysts has been that q_2 captures all the information of q_1 , and some extra as well (i.e. it is assumed that by decomposing the judgment we are able to take some aspects of the situation into account that previously we could not; and also that there has been an improvement in processing of the data by making explicit use of the equation $p(A) = \sum_1 p(A|X_1)p(X_1)$). Analysts will therefore often not bother with eliciting q_1 at all--it would not appear to have anything to offer. In the present formulation, the above argument means that $I_1 \subseteq I_2$ (see Figure 4) so that $A = 0$. Hence the weighted average (5.2) becomes $Bq_2/B = q_2$, confirming the heuristic reasoning above.

With the least-squares formulation however, to achieve such a reconciliation, we see from (5.1) that ρ must equal τ/σ . I very much doubt that such a value would be elicited from a subject who actually held the above beliefs. This example also makes explicit once again our motivation for seeking incoherence--if we do not agree with the above reasoning, but in fact believe that $I_1/I_2 \neq \emptyset$, then we gain by considering both q_1 and q_2 .

Another interesting consequence of the present formulation lies in the correct value of ρ to use in a statistical analysis when one has no information about its value. Lindley (1965) has suggested $\rho = 0.5$ is appropriate. From Figure 3, one could invoke a form of the principle of insufficient reason, and take $A = B = C$. In this case one can easily calculate the implied value of ρ to be 0.5.

5.4 Assessments

The concepts of "information" discussed above have a fairly intuitive interpretation, but it is rather difficult to obtain quantitative assessments for them. In this section we make some suggestions for quantification.

The first item to note is that if we have equal confidence in each of q_1 and q_2 , then we know that $A + C = B + C$, so $A = B$, and we may simply take the arithmetic mean of (log-odds) q_1 and q_2 . This illustrates the point that a quantification of C is of use only in assessing the precision of the reconciled estimate, and also that we can arbitrarily assign one of the values, e.g., A , as it is only relative quantities in which we are interested. It should also be noted that this explains the findings of

LTB (see Section 4.7) that correlations will not affect the probabilities. For, their calculations were performed with $\sigma^2 = \tau^2$, or $A = B$, and as we have noted, this then eliminates the effect of the correlation.

There is a standard statistical concept upon which we may draw, to aid our understanding of information, that of Fisher's information, but while such a concept is of value in providing a theoretical basis for the work, it does not aid the practical problem of assessment. Perhaps a more promising line of research would be to explore the use of Shannon and Weaver's information measure (Shannon and Weaver, 1949), but we have not had the opportunity to pursue this very far. The following suggestions are only tentative, and further work is necessary to extend some of the ideas.

One could simply assign the weights for the weighted average directly, without explicitly considering the information content. This and any other such attempt at quantification will need to be an interactive process between the analyst and $\underline{S}$, so as to capture the subjective feelings of $\underline{S}$ and the more objective knowledge the analyst has about the different assessment techniques.

A more satisfying method of direct elicitation is to use the intuitive idea of information, and ask the following two questions:

- a) How much extra information was gleaned by taking q_2 , when q_1 had already been assessed?

b) How much extra information would have been gleaned by assessing q_1 , had q_2 already been assessed?

Each of these answers should be made relative to the amount of information contained in q_1 . To exemplify the way A, B, and C could be calculated from these answers, suppose the answer to a) was "as much again" and to b) "half as much again." Then we deduce that:

$$B = A + C \text{ (from a)}$$

$$\text{and } 2A = (B + C) \text{ (from b).}$$

Hence $A = 2C = 2B/3$. So the weighted average is $0.4q_1 + 0.6q_2$, and the precision of the reconciled estimate measured by $A + B + C$ is twice that of q_1 and one and a half times that of q_2 .

One might instead suggest that the weight should be related directly to the confidence placed in the judgments. This is related to the accuracy we believe to be associated with each assessment--those in which we place greater confidence are those we consider to be more accurate. Again there appears to be no satisfactory definition at present allowing a quantification of what is nevertheless an intuitive concept. We could envisage using some psychological measurement procedures to permit such a quantification; perhaps allowing us to translate concepts such as "very confident" into an ordinal scale. If we have, for example, certain amounts of confidence in each of two assessments, it is likely that some of the reasons for our confidence are common to each. In that case, we shall say that the confidence arising from those reasons lies in the intersection of our Venn diagram (Figure 3.) Note that we are implicitly

using this "confidence" as a surrogate measure for the reasons of our uncertainty in our assessments. The use of some such surrogate measure appears to be the best way of proceeding, and is common to all our suggested reconciliation techniques.

In terms of actually making these assessments of confidence, we envisage displaying Figure 3 to S, and asking for an allocation of 100 coins between areas A, B and C, in proportion to his/her confidence judgments. We would then use these judgments to perform our calculations in the same manner as that discussed above.

Alternatively, one might use the concept of equivalent sample size to assess the information content of an assessment, by relating the extra information gained from an assessment to the number of extra observations from a binomial process that would have provided equivalent gain in information. Bunn (1978) has discussed ways of using this idea for assessing the parameters of a beta distribution, and an extension of those ideas might provide a good method of dealing with the present situation.

One could also use the ideas of LTB to help decide upon the weights--by assessing credible intervals for each assessment, we gain a good idea of the relative degrees of confidence in each assessment. The variance of an assessment may be taken as proportional to the square of the confidence interval. Assessing the information common to the two assessments is not so easy this way.

5.5 Expert Use

In this section we have discussed the situation of a single decision maker who gives inconsistent probability assessments. However, the technique of taking a weighted average of log-odds is also applicable to the problem of expert use, i.e. the situation when two or more experts each give probability assessments for a target variable. One would expect the experts to differ somewhat in their probability assessments so, in order for a decision-maker to make explicit use of the assessments, a reconciliation needs to be performed. Morris (1974) has developed a Bayesian procedure for performing this reconciliation that is similar to the method of LTB. The argument we have presented in previous sections can be used to show that the reconciliation should be a weighted average of log-odds. In this case the interpretation of the weights is much easier than before; they are the decision-maker's opinion of the relative expertise of the various experts. So, for example, the intersection $I_1 \cap I_2$ represents the shared expertise.

We are now in a position to offer an interesting perspective on the well known problem of what reconciliation to use for multiple experts of equal expertise. The arithmetic mean of the probabilities is an obvious candidate, but Norman Dalkey⁴ has suggested that the geometric mean is better than the arithmetic mean. From our work we can conclude that the arithmetic mean of the log-odds is the appropriate procedure. It will be recalled that log-odds were suggested because the assumption of normality necessary for the least-squares procedure was more valid.

We make the observation that taking the geometric mean is equivalent to taking the arithmetic mean of the log-probabilities. For, if the reconciliation $q' = (q_1 q_2)^{1/2}$, and letting $r_1 = \ln p_1$, $r_2 = \ln p_2$, then $r' = \ln q' = 0.5(r_1 + r_2)$. Thus taking the geometric mean would be our recommended procedure if we believed log-probabilities to be normally distributed. Such an assumption may be better than taking probabilities as normal since log-probabilities have infinite range, but log-probabilities are always non-positive, so the normality assumption can not be strictly true. Our work thus implies that taking the geometric mean is better than taking the arithmetic mean, in agreement with Dalkey, but that taking the mean of log-odds is better than either.

6.0 ALTERNATIVE RECONCILIATION TECHNIQUES

6.1 Psychological Biases

There has been a lot of work reported in the psychological literature aimed at discovering how good people actually are as probability assessors (e.g., Tversky and Kahneman, 1974). These studies have identified various biases in assessments, and have typically attempted to explain these biases behaviorally. If one has particular reason to believe, in a given situation, that one of these biases or heuristics is causing inconsistency, it makes sense to find a reconciliation methodology that addresses that particular bias. For example, if it appears that a subject is poorly calibrated, then using a calibration curve makes sense (see Lichtenstein, Fischhoff, and Phillips, 1977). It is only after all the apparent biases have been eliminated, yet we are still left with inconsistency, that the procedures discussed in previous sections should be applied.

We thus have come to the concept of developing several different reconciliation techniques, each, we hope, fairly simple and each designed to address one or more of the possible sources of incoherence. For a given reconciliation problem, we would use those techniques in our package which appeared most appropriate.

As an example of one such technique, suppose we have reason to believe that the only error being made by a subject is that he/she always over- (or under-) estimates probabilities. We might then assume that he/she is in fact operating with the numbers he/she produces according to the probability calculus, except that he/she is mistaking the first axiom of probability (that

$P(\Omega) = 1$, where Ω is the complete sample space) and instead of using $P(\Omega) = \alpha$, $\alpha \neq 1$. If $\alpha > 1$ he/she is being optimistic and if $\alpha < 1$ he/she is being pessimistic. It is then easy to show that his/her assessment of the probability of an event X , $q(X)$, will simply be $P(X)$ multiplied by a factor α . In different situations we shall have different ways of discovering α -- the simplest is when he/she assesses $q(X)$ and $q(\sim X)$, for then we may simply divide by their sum. This then provides a justification for using an intuitive technique, as proposed by Bartholomew in the discussion to LTB.

6.2 Satisficing

The techniques so far discussed have all been attempts at optimizing--we have in each case been attempting to discover the "best" reconciliation, $\hat{\pi}$, although we have used differing interpretations of what is best. However there is a fundamentally different way of approaching the whole problem, which may be likened to the concept of satisficing in economics. Instead of attempting to find the best reconciliation, we could simply look for one that is "good enough." The simplest way of doing this would be to inform the subject that he/she had been incoherent and then to present to him/her various coherent sets of values until he/she accepted one as being sufficiently descriptive of his/her true feelings of uncertainty.

This is probably not too different from what occurs at the present time if incoherence is discovered. One could visualize an interactive computer program which elicited the probabilities in various different ways, analyzed the responses for inconsistencies, and then presented a range of

potential reconciled values for the subject's consideration. It would be interesting to see if a reconciliation produced in this way in fact performed any worse in decision-making contexts than our optimizing techniques.

6.3 Degree of Incoherence

Another totally different approach to this problem would be achieved if some measure for the degree of incoherence could be developed. In the methodologies of the previous chapters, we have assumed that after the initial inconsistency has been discovered, we shall look only at vectors on the constraint set, and select our reconciled value through a search amongst these. One could have taken the alternative perspective of informing S that he/she has been incoherent, and giving him/her some indication of how incoherent, and perhaps in which directions his/her values should change to reach the constraint set. S would then produce a new set of values, which we could once again inspect, and tell him/her how incoherent, etc. this new vector was. In this way we could arrive at a sort of hill-climbing algorithm, where we would attempt to minimize the degree of incoherence (to a level of 0), by repeatedly trying different points.

At present we have no adequate measure of incoherence--it is unclear to what extent an objective measure could be produced. However some form of entropy measure might well be appropriate here. A more serious difficulty arises with the actual hill-climbing algorithm. We cannot be certain that S will in fact produce successive values that reduce incoherence, i.e., the algorithm might not converge. It is also possible that we might encounter the phenomenon of jamming, or of reaching a limit that was

suboptimal, as often occurs with hill-climbing algorithms. Another difficulty might be the motivation of S. He/she might get frustrated if repeatedly told to try again with another set of values, unless he/she could perceive that convergence to zero incoherence was fairly rapid. However the method might be successful, and it would certainly be an interesting experiment to try. Development of a measure for incoherence would be especially useful.

7.0 SUMMARY AND CONCLUSIONS

In this paper we have examined in detail the work of Lindley, Tversky and Brown, and further explored some of the consequences of that work. We have concluded that taking a weighted average of log-odds, with the weights proportional to the independent information content of each assessment, is equivalent to the procedure developed by LTB, while being simpler and of greater intuitive appeal.

The procedure of taking a weighted average may smack somewhat of "ad hocery," but it should be clearly understood that it has been derived as an approximation to a complete Bayesian analysis. A comment of de Finetti (1974) is appropriate here. The use of "ad hoceries" "may sometimes be an acceptable substitute for a more systematic approach . . . only, if--and in so far as--such a method is justifiable as an approximate version of the correct (i.e. Bayesian) approach. (Then it is no longer a mere "ad hocery.")"

It is hoped that we have adequately demonstrated in this paper that the procedure of using a weighted average of log-odds to reconcile inconsistent assessments is sufficiently simple to apply, and the justification for seeking out incoherence in order to increase the amount of information used sufficiently compelling, for this strategy to become a standard and useful part of the decision analyst's armory.

8.0 REFERENCES

- Brown, R. V., Kahr, A. S., and Peterson, C. R. Decision analysis for the manager. New York: Holt, Rinehart and Winston, 1974.
- Brown, R. V., and Lindley, D. V. Reconciling incoherent judgments (RIJ) - Toward principles of personal rationality (Technical Report TR-78-8-72). McLean, VA: Decisions and Designs, Inc., July, 1978.
- Bunn, D. The synthesis of forecasting models in decision analysis. Basel: Birkhauser Verlag, 1978.
- Churchman, C. W. The design of inquiring systems. New York: Basic Books, 1971.
- de Finetti, B. Theory of probability (2 vols.). London: Wiley, 1974.
- Dempster, A. P. A generalization of Bayesian inference. Journal of the Royal Statistical Society, Series B, 1968, 30, 205-232.
- Freeling, A. N. S. Decision analysis and fuzzy sets. Unpublished masters dissertation, Cambridge University, Engineering Department, 1979.
- Freeling, A. N. S. Fuzzy probability - An extended theory of belief. In preparation, Decision Science Consortium, Inc. 1980. (a)
- Freeling, A. N. S. Fuzzy sets and decision analysis. IEEE Transactions on Systems, Man, and Cybernetics, 1980, SMC-10, 341-354. (b)
- Kahnemann, D. and Tversky, A. Prospect theory: An analysis of decisions under risk. Econometrica, 1979, 47, 263-291.
- Lichtenstein, S., Fischhoff, B. and Phillips, L.D. Calibration of probabilities: The state of the art. In H. Jungerman and G. de Zeeuw (Eds.) Decision making and change in human affairs. Amsterdam: D. Reidel, 1977.
- Lindley, D. V. Probability and statistics. Cambridge University Press, 1965.
- Lindley, D. V., Tversky, A., Brown, R. V. On the reconciliation of probability assessments. Unpublished draft, 1978.
- Lindley, D. V., Tversky, A., and Brown, R. V. On the reconciliation of probability assessments. Journal of the Royal Statistical Society, Series A, 1979, 142, 146-180.
- Mitroff, I. I. A Brunswick lens model of dialectical inquiring systems. Theory and Decision, 1974, 5, 45-67.
- Mitroff, I. I., and Turoff, M. Philosophical and methodological foundations of Delphi. In H. A. Linstone, and M. Turoff (Eds.), The Delphi method: Techniques and applications. Reading, MA: Addison-Wesley, 1975.

Mitroff, I. I., Betz, F. and Mason, R. O. A mathematical model of Churchmanian inquiring systems with special references to Popper's measures for the severity of tests. Theory and Decision, 1970, 1, 155-178.

Morris, P. A. Decision analysis expert use. Management Science, 1974, 20, 1233-1241.

Savage, L. J. The foundations of statistics. Wiley, 1954.

Scheffe, H. The analysis of variance. New York: Wiley, 1959.

Shafer, G. A mathematical theory of evidence. Princeton, NJ: Princeton University Press, 1976.

Shannon, C. E., and Weaver, W. The mathematical theory of communication. Champaign-Urbana, IL: University of Illinois Press, 1949.

Slovic, P. and Tversky, A. Who accepts Savage's axiom? Behavioral Science, 1974, 19, 368-373.

Smith, C. A. B. Personal probability and statistical applications. Journal of the Royal Statistical Society, Series A, 1965, 125, 469-499.

Tani, S. N. A perspective on modeling in decision analysis. Management Science, 1978, 24, 1500-1506.

Tversky, A. Intransitivity of preferences. Psychological Review, 1969, 76, 31-48.

Tversky, A., and Kahneman, D. Judgment under uncertainty: Heuristics and biases. Science, 1974, 185, 1124-1131.

Watson, S. R., Weiss, J. J., and Donnell, M. L. Fuzzy decision analysis. IEEE Transactions on Systems, Man, and Cybernetics, 1979 SMC-9, 1-9.

Wheeler, G. E. and Edwards, W. Misaggregation explains conservative inference about normally distributed populations (SSRI Technical Report No. 75-11). Los Angeles: University of Southern California, Social Science Research Institute, August 1975.

Zadeh, L. A. Fuzzy Sets. Information and Control, 1965, 8, 338-353.

Footnotes

1. There is however a very strong case for arguing that we must always use our own beliefs to determine someone else's meaning. In this case, one could view a reconciliation procedure as part of the analyst's model for interpreting S's statements, and it would then be appropriate to take the analyst as N. In the traditional view of decision analysis, the analyst is portrayed as a logical machine, whose only function is to point out necessary logical implications to a decision-maker, without any of the analyst's own beliefs ever entering into the analysis. This complete neutrality of the analyst has been one of the big selling points of the methodology, but it is now becoming apparent that the judgmental inputs to an analysis come from the DM-analyst pair, viewed as a single entity. This is especially noticeable with the present problem for, as Savage (1954) noted, the logic of personal probability can only tell us we are inconsistent; it can make no recommendation towards remedying the situation. Hence a reconciliation methodology will of necessity include judgments of some form from others than just the subject. It is important that the true role of the analyst and his position of power be acknowledged and understood by practising analysts.

2. A positive definite symmetric matrix A is one satisfying the following condition:

$$\underline{x}^T A \underline{x} > 0 \text{ for all non-zero } \underline{x}.$$

It can be shown that a variance-covariance matrix must always be positive-definite. Being positive-definite is the matrix equivalent of being a positive number, and the condition that the variance-covariance matrix of

a multi-variate distribution be positive-definite is an extension of the condition that the variance of a univariate distribution be positive.

A practical check on whether a symmetric matrix is positive-definite is to discover the eigenvalues of the matrix. A theorem of linear algebra shows that a symmetric matrix is positive-definite if and only if all its eigenvalues are positive.

3. Here I_1 and I_2 denote the sets describing the information content of q_1 and q_2 . We use the modulus symbol $|\cdot|$ to denote the size of the set (in mathematical terms, its cardinality). So, for example if A is the set $\{1, 3, 5, 7, 9\}$, then $|A| = 5$.

4. Dalkey's point was made at the 18th Annual Bayesian Research Conference, held in Los Angeles, February 14-15, 1980.

The distribution list as per contractual requirements is as follows:

<u>ADDRESSEE</u>	<u>DODAAD CODE</u>	<u>UNCLASSIFIED/UNLIMITED</u>
Mr. J. R. Simpson (Scientific Officer)	NO0014	6
Ms. Betty Driskill (Administrative Contracting Officer)	S2101A	1
Director, Naval Research Laboratory, ATTN: Code 2627 Washington, D.C. 20375	NO0173	6
Defense Documentation Center Bldg. 5, Cameron Station Alexandria, VA 22314	S47031	12
Office of Naval Research Branch Office Bldg 114, Sec. D 666 Summer Street Boston, MA 02210	N62879	1

SUPPLEMENTAL

RIJ DISTRIBUTION LIST

Director, Research Administration
Naval Postgraduate School
Monterey, CA 93940

Dr. Edgar M. Johnson
Organization and Systems Research
Laboratory
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Craig Fields, Director
Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Perceptronics, Inc.
ATTN: Dr. Amos Freedy
6271 Variel Avenue
Woodland Hills, CA 91364

Decisions and Designs, Inc.
8400 Westpark Drive, Suite 600
McLean, VA 22101

Institute of Behavioral Science
University of Colorado
ATTN: Dr. Kenneth Hammond
Room 201
Boulder, CO 80309

Dr. William Howell
Rice University
Houston, TX 77001

D. J. White
University of Manchester
ENGLAND

Dr. Detlof von Winterfeldt
SSRI
University of Southern California
University Park
1042 W. 36th Street
Los Angeles, CA 90007

Peter Moore
London Business School
Sussex Gardens
Regent's Park
London NW3
ENGLAND

Office of Naval Research
Information Systems Program
Code 437
800 N. Quincy Street
Arlington, Virginia 22217

Dr. Bruce McDonald
Office of Naval Research
Scientific Liaison Group
American Embassy, Room A-407
APO San Francisco 96503

Air Force Office of
Scientific Research
Sciences Directorate, NL
Bolling Air Force Base
Washington, DC 20032

Air Force Office of
Scientific Research
Life Sciences Directorate
Building 410
Bolling Air Force Base
Washington, D.C. 20032

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the
Marine Corps
Code RD-1
Washington, D.C. 20380

Director, Applied Psychology
Unit
Medical Research Council
ATTN: Dr. A.D. Baddeley
15 Chaucer Road
Cambridge, CB 2EF
ENGLAND

Dr. Ron Howard
Stanford University
Stanford, CA 94305

Decision Analysis Group
Stanford Research Institute
ATTN: Dr. Miley W. Merkhofer
Menlo Park, CA 94025

Professor Howard Raiffa
Morgan 302
Harvard Business School
Harvard University
Cambridge, MA 02163

Technical Library
Naval Ordnance Station
Indian Head, MD 20640

Bureau of Naval Personnel
Department of the Navy
Technical Library
Washington, D.C. 20370

Library
Naval Ocean Systems Center
San Diego, CA 92152

Director, Operations Research
Programs, Code 434
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

Director, Statistics and Probability
Programs, Code 436
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

Director, Organizational Effectiveness
Research Programs, Code 452
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Director, Engineering Psychology
Programs, Code 455
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Director, Personnel and Training
Research Programs, Code 458
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Naval Research and Development
Center
Code 311
San Diego, CA 92152