LEVEL

(10)

IMPLEMENTATION OF SUBJECTIVE PROBABILITY ESTIMATES IN ARMY INTELLIGENCE PROCEDURES: A CRITICAL REVIEW OF RESEARCH FINDINGS

AD A093675

Ruth H. Phelps, Stanley M. Halpin,
Edgar M. Johnson, and Franklin L. Moses

DTIC
ELECTE
JAN 1 2 1981

C

HUMAN FACTORS TECHNICAL AREA

ari

U. S. Army

Research Institute for the Behavioral and Social Sciences

March 1980

Approved for public release; distribution unlimited.

DDC FILE COPY

81 1   09 025

U. S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the
Deputy Chief of Staff for Personnel

JOSEPH ZEIDNER
Technical Director

FRANKLIN A. HART
Colonel,   US Army
Commander

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
Research Report 1242	AD-A093 675	

4. TITLE (and Subtitle)

IMPLEMENTATION OF SUBJECTIVE PROBABILITY ESTIMATES IN ARMY INTELLIGENCE PROCEDURES: A CRITICAL REVIEW OF RESEARCH FINDINGS

5. TYPE OF REPORT & PERIOD COVERED

Research Rept.

6. PERFORMING ORG. REPORT NUMBER

--

7. AUTHOR(s)

Ruth H. Phelps, Stanley M. Halpin, Edgar M. Johnson, and Franklin L. Moses

8. CONTRACT OR GRANT NUMBER(s)

--

9. PERFORMING ORGANIZATION NAME AND ADDRESS

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, VA 22333

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

2Q762722A765

11. CONTROLLING OFFICE NAME AND ADDRESS

Deputy Chief of Staff for Personnel
Washington, DC 20310

12. REPORT DATE

March 1980

13. NUMBER OF PAGES

34

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

--

15. SECURITY CLASS. (of this report)

Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

--

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

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

--

18. SUPPLEMENTARY NOTES

--

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

Subjective probability
Intelligence estimate
Data evaluation
Tactical intelligence

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

The purpose of this report is to critically examine the formulation and expression of uncertainty in both the tactical intelligence estimate of enemy threat and the evaluation of information contained in spot reports. A study was made of both the doctrinal procedures for expressing uncertainty (FM 30-5) and the research conducted to detect problems in implementing these procedures. Findings showed that current procedures which focus on the use of terms such

(Continued)

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Item 20 (Continued)

such as "probable" and "unlikely" to express uncertainty are ambiguous com-
municators for both the user and the recipient. The use of numerical sub-
jective probability estimates as an alternative procedure (e.g., ".70"
instead of "likely") is explored by relating the psychological research on
the use of subjective probability estimates with the need of Army intelli-
gence analysts to unambiguously express uncertainty. It is concluded that
there is sufficient evidence indicating the superiority of the numerical
estimates to recommend a trial implementation of the use of numerical sub-
jective probabilities. General guidelines are discussed for the incorpora-
tion of numerical estimates in the analysis and communication of Army
intelligence.

Accession For

NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification

By
Distribution/

Availability Codes

Dist | Avail and/or
     | Special

A

IMPLEMENTATION OF SUBJECTIVE PROBABILITY ESTIMATES IN ARMY INTELLIGENCE PROCEDURES: A CRITICAL REVIEW OF RESEARCH FINDINGS

Ruth H. Phelps, Stanley M. Halpin,
Edgar M. Johnson, and Franklin L. Moses

Submitted by:
Edgar M. Johnson, Chief
HUMAN FACTORS TECHNICAL AREA

Approved by:

James D. Baker, Acting Director
'ORGANIZATIONS AND SYSTEMS
RESEARCH LABORATORY

iii

The Human Factors Technical Area of the Army Research Institute (ARI) is concerned with the human resource demands of increasingly complex battlefield systems used to acquire, transmit, process, disseminate, and utilize information. Current research focuses on human performance problems related to interactions within command and control centers, as well as issues of system development. Specific areas of work include software development, topographic products and procedures, tactical symbology, user-oriented systems, decision making, systems integration, and utilization.

One issue of special concern in tactical intelligence is the formulation and expression of uncertainty. The current report (a) critically reviews problems with current procedures, outlined in FM 30-5, for expressing uncertainty in both the intelligence estimate and in the evaluation of intelligence information; and (b) delineates the steps necessary for using subjective probability estimates to communicate uncertainty. Questions about the most effective implementation procedures can be answered only after subjective probability estimates have been incorporated routinely into tactical intelligence communications.

Research in the area of intelligence systems and procedures is conducted as an in-house effort augmented by organizations contracted for their unique capabilities and facilities for research in this area. This effort is responsive to the requirements of Army Project 2Q762722A765 and related to requirements of the U.S. Army Combined Arms Combat Development Activity expressed in HRN 79-145 (Processing and Problem Solving Aids in Tactical Automated Systems).

JOSEPH ZEIDNER
Technical Director

v

IMPLEMENTATION OF SUBJECTIVE PROBABILITY ESTIMATES IN ARMY INTELLIGENCE
PROCEDURES: A CRITICAL REVIEW OF RESEARCH FINDINGS

BRIEF

Requirement:

To critically analyze the potential utilization of subjective probability
estimates both in the intelligence estimate and in intelligence data evaluation.

Procedure:

The investigation encompassed two areas. First, an examination was made
of doctrinal procedures currently used for expressing uncertainty in the in-
telligence estimate and spot report data evaluation (FM 30-5), and research
on the current use and problems with these procedures was analyzed. Second,
preliminary steps were delineated for the implementation of numerical subjec-
tive probability estimates to express uncertainty as an alternative to the
present methods.

Findings:

Using current procedures, uncertainty is expressed by verbal probability
phrases (e.g., possible, unlikely) in the intelligence estimate, whereas in-
telligence information is evaluated by two 7-point rating scales (Information
Accuracy and Source Reliability) based on verbal phrases. Available research
indicates verbal probability phrases in the intelligence estimate are inter-
preted extremely ambiguously and that the current data evaluation system is
deficient.

The use of numerical subjective probability estimates is a feasible al-
ternative for expressing uncertainty in both the intelligence estimate and in
data evaluation. Although some questions still remain concerning the details
of training and implementation, current knowledge provides a sufficient base
to begin incorporating subjective probability estimates into Army doctrine
and practice.

Utilization of Findings:

The implementation of numerical subjective probability estimates is ex-
pected to decrease the ambiguity in communicating intelligence estimates to
commanders and other users and in evaluating spot reports. Once numerical
probabilities have been incorporated into intelligence procedures, an addi-
tional advantage will be the ease with which estimates and evaluations can be
compared among personnel, over time, or be used as inputs for automated
decision aids.

IMPLEMENTATION OF SUBJECTIVE PROBABILITY ESTIMATES IN ARMY INTELLIGENCE
PROCEDURES: A CRITICAL REVIEW OF RESEARCH FINDINGS

CONTENTS

LIST OF TABLES

LIST OF FIGURES

IMPLEMENTATION OF SUBJECTIVE PROBABILITY ESTIMATES IN ARMY INTELLIGENCE PROCEEDINGS: A CRITICAL REVIEW OF RESEARCH FINDINGS

INTRODUCTION

The efficiency and effectiveness of intelligence systems are continuing military concerns (e.g., Williams, 1972, 1974; Graham, 1973). One area of special concern is the formulation and expression of the uncertainty inherent in intelligence information and estimation. The impact of this uncertainty on the quality of intelligence is often compounded by the loose, ambiguous language used to communicate uncertain intelligence information (Brown & Shuford, 1973). For example, intelligence information is often communicated by terms such as "report X is very likely to be true while report Y is only probably true" or "the enemy is most likely to counterattack, but there is still some chance they may delay for another day or so." Although the recipient of such intelligence may have a general understanding of the situation, the use of the terms "probably," "likely," and "chance" to communicate uncertainty makes the exact interpretation ambiguous.

This report demonstrates how vague phrases could be replaced by more precise numerical estimates of uncertainty, called subjective probability estimates. For several years, the idea of using subjective probability estimates in intelligence communication has been discussed, and some commanders have actually made a few attempts to use numerical estimates. However, there has been no effort to systematically organize and summarize the current research and knowledge about the use of subjective probability estimates within the Army intelligence context. This report (a) summarizes and critically evaluates current research on the use of subjective probability estimates for expressing uncertainty in Army intelligence and (b) identifies necessary steps for incorporating subjective probability estimates into Army practice. The intention of this paper is to summarize relevant research and relate it to intelligence procedures, not to provide the detailed specifics necessary for implementation at a particular agency or G2/S2 section.

Since the structure of the intelligence system is hierarchical and sequential, the impact of vague communication of uncertainty may be compounded at every level. As shown in Figure 1, there are at least three, and in many cases more, phases in intelligence analysis, with each phase dependent on the previous one. In Phase 1, from the barrage of potentially important tactical information, a subset is selected, recorded, and evaluated in a spot report. The spot report, with an accompanying evaluation of the quality and uncertainty of the information, is eventually forwarded to the division G2 section. The G2 staff analyzes, condenses, and integrates numerous spot reports (as well as information from other sources) to formulate a predicted enemy threat in the intelligence estimate. This estimate provides the commander, who integrates it with other relevant information, with the basis for a tactical decision.
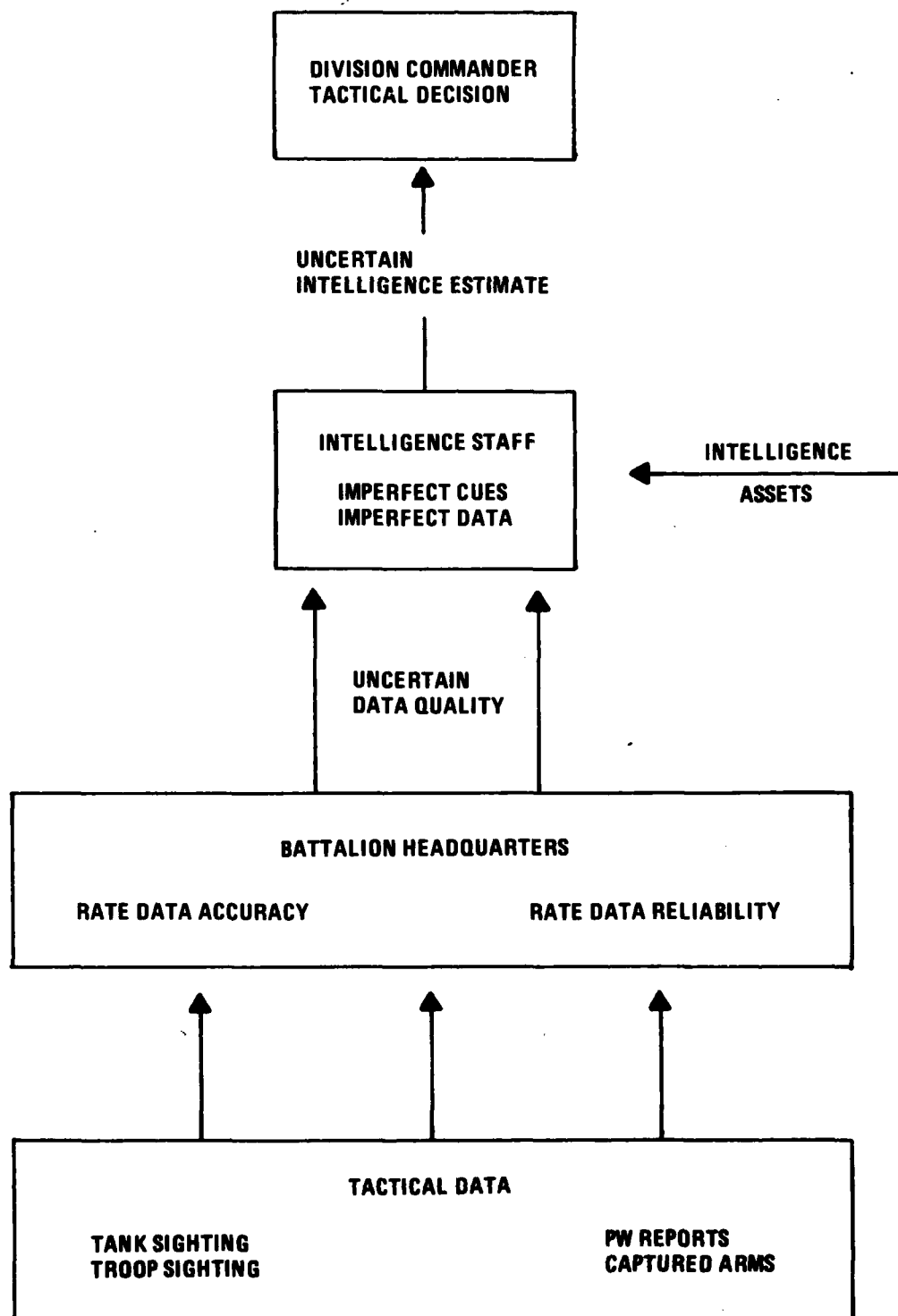
1

Figure 1. The Intelligence Hierarchy. As data are evaluated and forwarded up the hierarchy, uncertainty enters at (1) the communication of the intelligence estimate and (2) the evaluation of data quality.

Thus, the commander or other user of intelligence must rely on intelligence that has been evaluated and analyzed at least twice by the supporting staff. Even assuming careful and accurate evaluation of information, serious distortions may occur if the uncertainties in the quality of basic information and analyses are imprecisely communicated among the various staff sections; in other words, critical degradation may occur in the intelligence reaching the commander.

At two points in the intelligence system, as diagrammed in Figure 1, the current procedures and language used to communicate uncertainty are especially ambiguous. The first is in the language used to communicate the likelihood associated with the predicted enemy threats, i.e., intelligence estimation. The second is in the expression of the evaluation of the quality of information contained in spot reports, i.e., intelligence data evaluation. In both cases, more precise communication should increase the quality of intelligence available to a commander for making a tactical decision.

This report will analyze and critically evaluate the use of numerical subjective probability estimates for expressing uncertainties at the two points in the intelligence analysis system described above. The report is organized into four major sections: the first section summarizes the current doctrine for expressing uncertainty in intelligence estimation and in data evaluation, as well as critical research on the current procedures; the second section presents background information on the definition of subjective probability, research findings on the previous use of probability estimates, and research findings on the ability of personnel to be trained to assign accurate probabilities; section three outlines the steps that should be considered for incorporating subjective probability estimates in Army intelligence procedures; and the final section identifies several unresolved questions that may provide direction for evaluating the implementation of subjective probability estimates.

CURRENT DOCTRINE

Intelligence Estimation

The purpose of the intelligence estimate is to formally anticipate and predict possible actions and/or reactions of the enemy; for example, "there will be an attack on Camp X by noon tomorrow," "the enemy will delay," etc. The usual procedure for these predictions is to gather relevant data, formulate potential enemy courses of action, and list the courses of action in the order of their perceived likelihood of occurrence. This list is presented by the intelligence staff to the G2, who in turn briefs the commander (FM 30-5, 1973). Although these steps are discussed in current doctrinal materials, no uniform procedure is apparent for executing the steps or communicating the perceived likelihood of the various courses of action. Because of this lack of standardization, intelligence staffs vary considerably in their interpretation and execution of doctrine.

One common method for communicating the relative likelihoods of alternative courses of action in the intelligence estimate is to use verbal phrases such as "somewhat likely," "remote," and "probable." The commander and the intelligence staff may feel that more information is communicated in this way,

but dangerous ambiguities are inherent in such language. In a study designed to assess how verbal probabilities are interpreted in Army intelligence communications, intelligence personnel assigned numerical values to 15 probability phrases (Johnson, 1973). As shown in Table 1, the range of values assigned to the phrases was excessively large. Clearly, the verbal phrases are interpreted very differently by the different personnel, e.g., both "very probable" and "highly improbable" cover roughly the same range of probabilities. Such a diversity of interpretations could lead to serious misunderstandings and to degradation of the quality of intelligence available to the commander.

Recognizing the need to make relative likelihood assessments and the need for unambiguous communication of likelihoods, the Army signed the NATO Standardization Agreement (NATO STANAG, 1976), which says in part,

> In order that commanders and intelligence staffs should be able to express the probability of the enemy's adopting any one out of a number of possible courses of action in a more exact manner than can be conveyed by verbal expressions which are open to more than one interpretation, and also in order to permit the interchange of assessments with no loss of accuracy, degrees of probability should be expressed in percentage form.

For example, the statement, "The enemy is most likely to counterattack, but there is some chance they may continue to delay" could be restated as "there is an 80% chance of an enemy counterattack and 20% chance of a delay." Clearly, the numerical estimates provide a less ambiguous communication of the staff officer's evaluation of the threat situation. However, despite the NATO agreement, no systematic attempt has been made within the Army to adopt and promote the use of numerical probability estimates.

Data Evaluation

A second intelligence area in which the expression and communication of uncertainty could be improved is the evaluation of data contained in spot reports. According to current doctrine, in Phase 1 of the intelligence analysis system (Figure 1) an item of tactical information is recorded in an individual spot report. The quality of the information contained in the report is assessed by the originating headquarters by rating the accuracy of the information as well as the reliability of the source of the information (Army FM 30-5, 1973). The present standardized system used to make the evaluations is comprised of two scales listed in Table 2 (FM 30-5, 1973). For example, information assessed as being from a "fairly reliable" source and deemed "possibly true" should be rated "C3." The basis for determining the reliability rating appears to be previous experience with the source, while the basis for assessing the accuracy of the information is the degree to which it is compatible with and/or confirmed by other pieces of information. The ratings made on the two scales are to be independent; that is, the assessment of the source reliability should not influence the evaluation of the information accuracy, and vice versa. The expressed purpose of these evaluations is to provide the staff section receiving the information with a basis for deciding its importance or weight.

Table 1

Numerical Interpretation of Probability Phrases

Phrase	Median	Range
Highly Probable	85	20-99
Very Probable	80	5-98
Very Likely	80	10-99
Quite Likely	73	15-99
Fairly Likely	63	2-90
Likely	60	10-95
Probable	60	10-99
Possible	50	4-80
Fair Chance	50	1-100
Unlikely	20	0-70
Fairly Unlikely	20	0-65
Improbable	10	0-70
Very Unlikely	10	0-60
Quite Unlikely	10	0-50
Highly Improbable	10	0-90

Source. Johnson, 1973

Table 2

Doctrinal Rating Scales for Evaluating Spot Reports

Source Reliability	Information Accuracy
A – Completely reliable	1 – Confirmed by other sources
B – Usually reliable	2 – Probably true
C – Fairly reliable	3 – Possibly true
D – Not usually reliable	4 – Truth doubtful
E – Not reliable	5 – Improbable
F – Reliability cannot be judged	6 – Truth cannot be judged

Source. FM 30-5, 1973.

Over the many years that this evaluation system has been doctrine, sufficient dissatisfaction has accumulated to warrant scientific investigations into its use and effectiveness (e.g., Baker, McKendry, & Mace, 1968; Samet, 1975). Appendix A summarizes the details of this research. One finding of these studies is that only about half of all spot reports were ever evaluated. This failure to use the current evaluation scales and procedures may be due to the following deficiencies identified by the research:

1. Ratings of reliability and accuracy are not, in fact, made independently; intelligence personnel give corresponding ratings on both scales, A-1, B-2, C-3, etc. Such a correspondence indicates that the scales are viewed as redundant or at least as correlated.

2. Personnel do not use the full range of the scales; the majority of all tested spot reports were assigned a rating of B-2, "probably true, usually reliable."

3. The scales are unnecessarily complex.

4. Even when ratings are assigned, they are inconsistently interpreted by both different users and recipients.

Additional research (Miron, Patten, & Halpin, 1978; Halpin, Moses, & Johnson, 1978) examined the relationship between an individual's subjective evaluation of intelligence data and use of the current standard rating scales. This research indicates that the current rating scales do not allow users to express their complete evaluation of the information. Thus, a simple change in training procedure or a clarification of the scale definitions, for example, would not be adequate to improve the communication of the evaluation significantly. A different form of rating scale is required.

One alternative to current procedures is a scale based on subjective probability estimates, where each report would be assigned a probability (ranging from 0 to 1.00) corresponding to the rater's subjective estimate that the report is accurate. Additional probability ratings could be made for reliability or any other dimensions describing the quality of the report.

BACKGROUND: SUBJECTIVE PROBABILITY

Definition of Subjective Probability

In contrast to an objective probability, which is based on formal logic, probability theory, and frequency of events (e.g., probability of a head in a coin toss), a subjective probability reflects a person's degree of belief about an event. For example, after studying all relevant information, an intelligence analyst may feel there is a .80 subjective probability of a political riot in Chile during the next 6 months. Or, at the tactical level, a G2 officer may feel that there is only a .25 probability of an enemy attack on Camp X. The probability assigned summarizes and communicates the person's degree of belief in the likelihood or uncertainty of the occurrence of specified events.

7

Although a subjective probability estimate represents a person's best estimate and thus can never be strictly wrong, subjective probabilities assigned according to probability axioms can be used to communicate the user's beliefs accurately and unambiguously. The four most important axioms or rules are as follows:

1. The events must be stated such that they can be confirmed as true or false within some specified time period.

2. All reasonably possible events must be listed; that is, the events under consideration should be exhaustive.

3. The events must be stated such that they are mutually exclusive.

4. Subjective probabilities assigned to several events concurrently must sum to 1.0.

Appendix B contains an elaboration of these rules with examples.

Precedents in the Use of Subjective Probability

The use of subjective probabilities is widespread in nonmilitary settings. For example, with the advent of advanced technology, subjective estimates of the likelihood of nuclear power plant accidents have become crucial factors in policy decisions (Slovic, Fischhoff, & Lichtenstein, 1976). Weather fore-casters estimate the likelihood of precipitation in terms of subjective prob-abilities (Murphy & Winkler, 1974). Decision aids such as those based on Bayes' theorem or multi-attribute utility theory require users to estimate subjective probabilities; a variety of these aids have been employed in areas as diverse as land management (Gardiner & Edwards, 1975), conflict resolution among public officials (Hammond, Rohrbaugh, Mumpower, & Adelman, 1977), and medical diagnoses (Einhorn, 1972). In addition, there is a large body of research investigating the consistency, reliability, and accuracy of subjec-tive probabilities (Lichtenstein, Fischhoff, & Phillips, 1977).

Within military settings, the use of subjective probability estimates has been investigated, but only on a limited basis. Fifteen years ago the Air Force began exploring the use of probabilistic information processing proce-dures based on Bayes' theorem (Edwards, Lindman, & Phillips, 1965). More recently, subjective probability estimates have been incorporated in defense analyses on a trial basis or as a component in a decision aiding system (e.g., Kelly & Peterson, 1971; Decisions and Designs, 1977; Brown, 1978; Kibler, Watson, Kelly, & Phelps, 1978). In addition, exploratory work has been done with probabilities by Army intelligence image interpreters (Evans & Swensen, 1979). However, in none of these cases have subjective probabilities been incorporated into established procedures or doctrine.

Psychological Issues

Two psychological issues are critically important to implementing subjective probability estimates in either intelligence estimation or in intelligence data evaluation. The first issue is the basic concern about a probability assessor's ability to assign unbiased subjective probabilities; that is, do the estimates accurately reflect the person's true degree of uncertainty? Given that probability estimates can be biased, the second issue concerns whether people can be trained to assign unbiased, accurate estimates. If, in fact, the probability assessments are irrevocably biased, then their use will not reduce the current ambiguities in intelligence communication. However, if accurate probability estimates can be made, then the practical issues of incorporating subjective probability-based scales into Army intelligence doctrine can be addressed.

Are Subjective Probability Estimates Unbiased? There is a large body of research assessing people's ability to use subjective probabilities (for a full review, see Adams & Adams, 1961; Lichtenstein, Fischhoff, & Phillips, 1977). In the majority of studies, participants were presented with two alternative answers to a question; their task was to select one of the answers as correct and then assign a subjective probability corresponding to the confidence they felt in the correctness of the chosen alternative. The probabilities could range between .5 and 1.0 because only the chosen alternative was rated. If the assessor's confidence ratings matched reality then, for a large number of ratings made with confidence p, about p% should, in fact, be true. If confidence p = p% occurrence, the rater is said to be well calibrated. Deviations from perfect calibration can occur in two directions: (a) if the rater assigns subjective probabilities consistently lower than the percentage that occurs (confidence p < p% occurrence), the rater is underconfident; and (b) if the rater assigns subjective probabilities consistently higher than the percentage that occurs (confidence p > p% occurrence), the rater is overconfident. Figure 2 shows these biases.

The results of this research are consistent (Adams & Adams, 1961; Lichtenstein & Fischhoff, 1977); data obtained from American students, British students, military image interpreters, and research employees all indicate people are poorly calibrated. That is, while confidence ratings are indeed correlated with occurrence or accuracy (e.g., r = .59; Andrews & Ringel, 1964) there are large and consistent errors. The deviations from calibration appear as overconfidence when the alternatives are not easily discriminated and as underconfidence when alternatives are clearly discriminated. In other words, when there is a great deal of uncertainty about which alternative is correct, assessors assigned inappropriately high subjective probabilities. But when there was little uncertainty about the correct alternative, the probabilities assigned were too low. Thus, although assessors were poorly calibrated, their biases were clearly systematic. Such regularity in bias indicates that probability training could improve calibration.
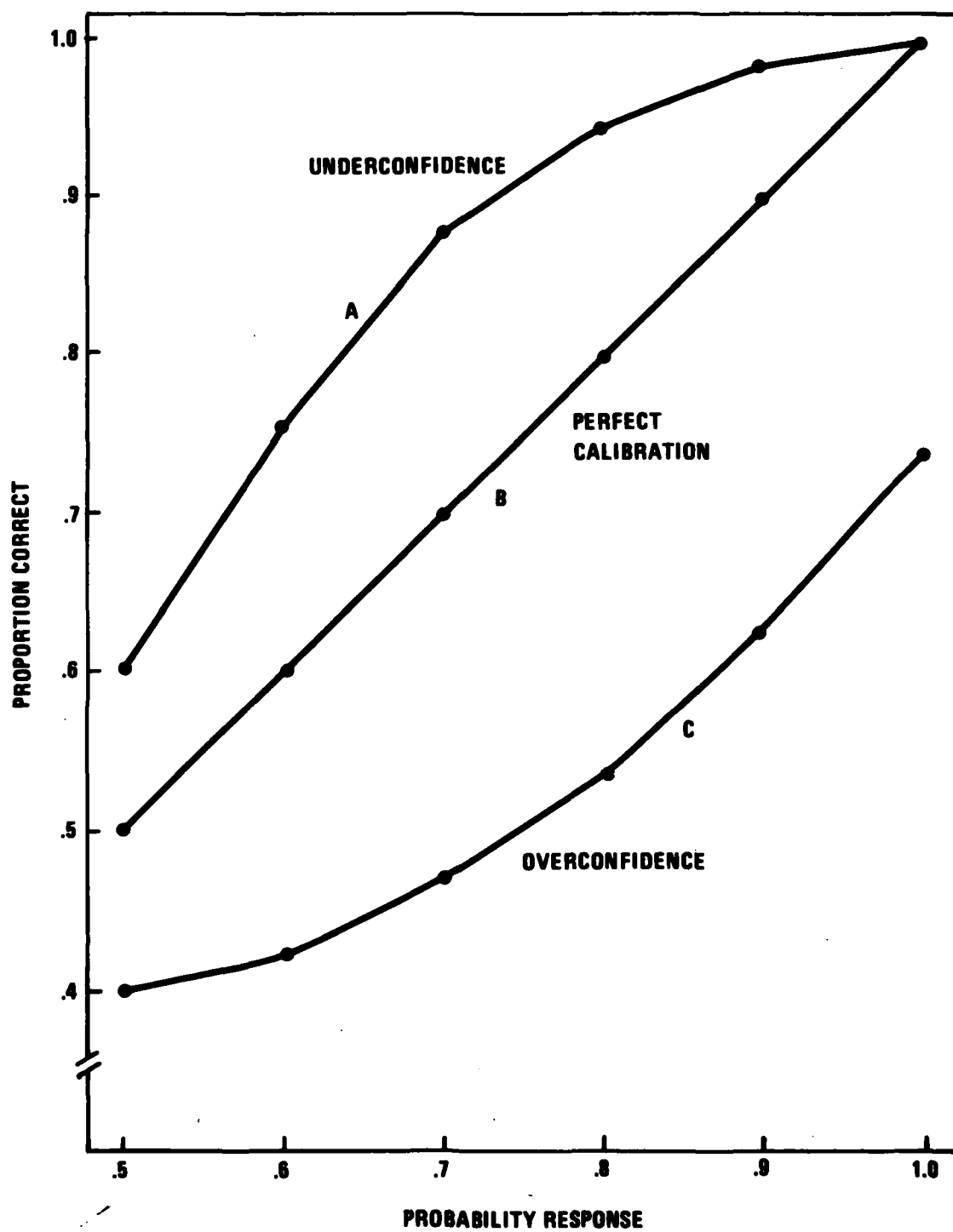
Figure 2. Idealized samples of perfect calibration, underconfidence, and overconfidence.
Source. Lichtenstein and Fischhoff, 1978.

10

Can Probability Assessors Be Trained? At least two professional groups have been shown to assign reasonably accurate subjective probabilities. In a study of 15 military analysts from the Defense Intelligence Agency (DIA), subjective probability estimates of the likelihood of 1,450 militarily relevant events were made over an 18-month period. The subjective probabilities assigned were quite accurate, but there was a small but consistent bias toward overconfidence (6%), as shown in Figure 3 (Kelly, Peterson, Brown, & Barclay, 1974).

Studies of meteorologists made before probability forecasting was regularly used show that the subjective probabilities assigned were reasonably accurate; however, consistent with the research findings for confidence judgments and military intelligence estimates, a widespread, fairly constant, overconfidence was found (Williams, 1951; Sanders, 1958; Root, 1962). In contrast, however, more recent data based on more than 60,000 predictions (U.S. Weather Bureau, 1969) showed that calibration was excellent. Apparently, through experience and training meteorologists have overcome their initial bias to overestimate the probability of precipitation.

Although the accurate performance of DIA analysts and weather forecasters demonstrates that well-calibrated probability assessments can indeed by assigned, the unstructured training and long intervals between training and assessment preclude determining exactly what was responsible for the learning and if the learning occurred within a reasonable time. To determine more precisely the feasibility of training probability calibration, ARI supported a controlled laboratory investigation.

Two experiments investigated the effectiveness of training probability assessors to be well calibrated, i.e., to make accurate, unbiased, probability estimates (Lichtenstein & Fischhoff, 1978). In both experiments assessors made subjective probability estimates corresponding to their degree of confidence that a selected answer was correct. In the first experiment, after participants made 200 subjective probability estimates, the accuracy of the participants' judgments was calculated and provided as feedback. This feedback helped assessors see the direction and magnitude of error. Following 10 sessions of 200 judgments each, assessors who initially showed considerable underconfidence and overconfidence biases were all well calibrated. In the second experiment, the training procedures were modified and abbreviated to only three training sessions. Even with this short training period, assessors learned to become accurately calibrated. Questions remain concerning the most efficient training methods, but these studies demonstrate the feasibility of teaching probability assessors to make accurate estimates.

In summary, a subjective probability estimate is a number that reflects a person's degree of belief or confidence in the certainty of an event or information. Psychological research has shown that although the subjective probability estimates assigned do not accurately reflect the estimator's degree of uncertainty, the errors made are systematic and identifiable. The evidence also shows that in both field and laboratory settings, subjective probability estimators learned to overcome their errors and assign accurate probabilities. Based on this research, it appears feasible to train intelligence personnel to assign accurate subjective probabilities to their feelings of uncertainty.
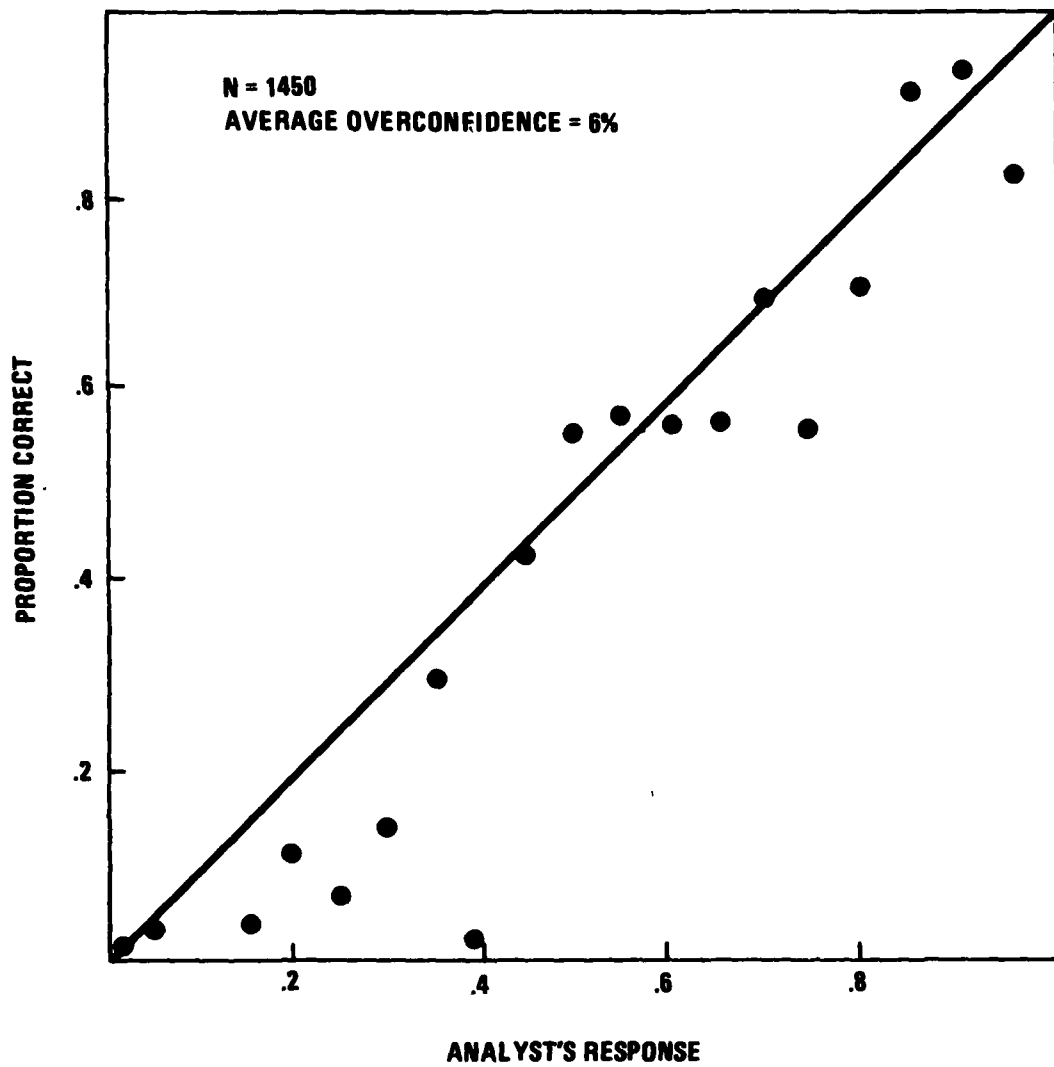
N = 1450
AVERAGE OVERCONFIDENCE = 6%

PROPORTION CORRECT

ANALYST'S RESPONSE

Figure 3. Calibration data of intelligence analysts.
Source. Kelly, Peterson, Brown, and Barclay, 1974.

12

IMPLEMENTATION OF SUBJECTIVE
PROBABILITY-BASED EVALUATIONS

Compared with current procedures used to express uncertainty in both intelligence estimation and spot report evaluation, numerical subjective probability estimates have several advantages. First, they should reduce the ambiguity of the expression of uncertainty because a numerical scale is easier to standardize, and therefore interpret, than are verbal labels (e.g., .60 vs. "probable"). Second, the 0-1.00 probability scale is continuous rather than a series of discrete categories; thus, the rater may be freer to use the full range of the scale. Third, several dimensions of information quality can be rated on the same numerical probability scale. Fourth, evaluations from different intelligence personnel can be unambiguously compared and/or combined to give a composite estimate. Finally, numerical estimates can be used easily as inputs for a variety of decision making aids. However, in order to incorporate subjective probabilities into intelligence communication effectively, at least three issues must be addressed: training, maintenance of high-level performance, and evaluation of the subjective probability program.

Training

While there are many unanswered research questions concerning the most effective training procedures for subjective probability estimates, it is clear that the G2, S2, and their supporting staffs must at least be trained to (a) follow the rules or axioms of probability theory, as listed in Appendix B; and (b) be well calibrated, that is, make accurate subjective probability estimates. The most effective and appropriate procedures for teaching the application of the probability axioms to Army intelligence personnel have yet to be determined. On the other hand, the methods used by Lichtenstein and Fischhoff (1978) for training calibration are well documented and could easily be automated for self-instruction and practice. However, Lichtenstein and Fischhoff (1978) found only poor to moderate transfer of training to other tasks. That is, people who were trained to be well calibrated on Task A may not be well calibrated on Task B or Task C. Calibration training, and probably training in the probability axioms, should therefore be conducted within the tactical intelligence context. Such instruction could easily be implemented in an appropriate school curriculum, in the field, or both.

Performance Maintenance

After intelligence personnel have been trained to assign accurate subjective probabilities and have incorporated the probabilities into the intelligence estimate, it is necessary to reassess periodically the accuracy of the estimates. Investigations of the long-term accuracy of weather forecasters show that when subjective probability estimates are made continuously and repeatedly, calibration remains excellent. At this point, however, there is little research to indicate how much use is necessary to maintain good calibration; therefore, at least initially, subjective probability estimates assigned by intelligence personnel should be evaluated periodically.

13

Calibration could be assessed in several ways. One approach would be to maintain a routine track record of all estimates made; the estimates could then be compared with the actual proportion correct at some later date. The obvious advantage of a continuous track record is that calibration can be assessed at any time. In addition, checks could be made to insure that the estimates are consistent with the axioms of probability. An alternate approach is to conduct calibration tests using hypothetical scenarios at various intervals. Testing could be easily automated or administered manually to large groups. While testing is convenient, some personnel might score well on such tests and still assign inaccurate probabilities when making intelligence predictions. Thus, whenever possible, a track record assessment would provide the most informative feedback; in situations where such record keeping is not feasible, or where too few estimates have been made, calibration testing would be necessary. The periodic evaluation of the intelligence personnel would serve as feedback indicating the development of any biases, in addition to providing a reminder of the rules for subjective probability assignment. The evaluation would also benefit both the operational unit or activity and the research and development community by either validating the training procedures or indicating areas needing new or additional training.

Program Evaluation

Program evaluation refers to an assessment of the effectiveness and usefulness of incorporating subjective probabilities into the intelligence estimate. The major question to be answered by such an evaluation is, "Has the ambiguity in the communication of intelligence decreased since incorporating subjective probability estimates?" While there are many approaches for evaluating this issue (e.g., Guttentag & Streuning, 1975), most require a comparison of the quality of communication under the current system with the quality under the new system. This requirement of program evaluation must be recognized so that the appropriate measures can be obtained before subjective probability estimates are implemented, as well as after. Such an evaluation will provide feedback not only on the overall effectiveness of the subjective probability-based estimates, but also on the aspects that could be modified to improve the program as well as those that could be eliminated without degrading the quality of intelligence.

Potential Problems

Problems exist in incorporating subjective probability-based scales for spot report evaluation and tactical estimation. These problems are not overwhelming, but they should be presented as cautions. Perhaps the greatest danger of using numerical ratings is the accompanying feeling of precision and accuracy. Simply replacing an ambiguous verbal phrase with a number does not alone increase precision. Since most people exhibit systematic biases, the use of numerical estimates must be accompanied by a validation of those estimates and supplemental training if necessary. After the validity of the estimates has been established and maintained, a sense of increased precision in communication is indeed warranted.

14

Another potential problem involves people's reluctance to commit them-selves to a specific number. Apparently, some psychological safety is present in the ambiguity of verbal phrases that is absent from numerical estimates. While this may cause anxiety for some probability assessors, it is expected that continued use, calibration training, and estimate validation should grad-ually reduce such fears. However, it is critical to anticipate such anxieties since, if nurtured, they may cause intelligence personnel to neglect the sub-jective ratings altogether.

Finally, Military Intelligence personnel have reported informally that distrust of estimates and evaluations made by other echelons, verbal or numer-ical, is severe and widespread. The result of this distrust is an unnecessary repetition of evaluations and analyses at different echelons. While this re-dundacy of effort is possible with the current paper and telephone message sys-tem, it will not be practically feasible with the automated high capacity in-formation systems of the near future. Because the automated systems will store several times the amount of information currently handled manually, verifica-tion of the accuracy or source reliability of each spot report will not be practical. In addition, if threat estimates are based on the integration of many times the current levels of information, condensed summaries of analysis will become necessary. Increased capability to handle more information faster will produce a trade-off between the benefits of additional information and inability of any one G2 or commander to evaluate the relevant raw data.

In summary, any program proposing to incorporate subjective probability estimates into intelligence communication must consider three issues: (a) procedures for training personnel to assign accurate estimates, (b) procedures for maintaining accurate estimation abilities, and (c) procedures for evaluat-ing the usefulness of the subjective probability estimates. In addition, psychological resistance to the use of numerical estimates may be encountered and should be actively considered in the implementation program.

UNRESOLVED RESEARCH QUESTIONS

Although the implementation of subjective probability estimates to express uncertainty in intelligence communication is feasible, several questions re-quire further research. Some of these can be addressed with laboratory experi-mentation, but others can be answered only after subjective probability esti-mates have been incorporated into Army doctrine.

Question 1: What Is the Minimal Effective Training? The Lichtenstein and Fischhoff (1978) experiments showed that training could be reduced from 10 to 3 sessions without loss, but the minimum effective amount has not been determined. Also, the number of questions that are necessary per session must be determined. In addition, the minimal frequency and quality of feedback are unknown. Answers to such questions are needed to develop cost-effective training.

Question 2: Once Trained, How Long Is Good Calibration Maintained? Data collected for weather forecasters indicate that with daily practice in assign-ing subjective probabilities, calibration remains excellent. However, informal data show that for DIA analysts who did not practice assigning probability es-timates, calibration deteriorated within the first 6 months after training.

15

Questions concerning the amount of practice necessary to maintain calibration
will help training personnel establish the maximum effective training-to-
implementation time interval. It simply may not be practical to train person-
nel much in advance of their use of subjective probability estimates.

Question 3: Are Intelligence Personnel Equally Well Calibrated at Vari-
ous Levels of Analysis? Intelligence estimates are global predictions about
the likelihood of various courses of action; however, intelligence analysts
also deal with uncertainty of the component parts of the prediction. For ex-
ample, a specialized analyst may be required to assess the likelihood that the
enemy can cross a particular river under a variety of weather conditions. Is
an analyst who is well calibrated for intelligence estimates also well cali-
brated for more detailed component-part judgments?

Question 4: Should Training Be on General Probability Estimates or
Within the Specialization Area? Current data show that even when participants
are well calibrated on one type of task, they may not be calibrated for
another; that is, calibration training does not always generalize to other
tasks (e.g., Lichtenstein & Fischhoff, 1978). Thus, the most effective train-
ing would probably occur for tasks within the specific area in which estimates
are to be made; in this case, tactical intelligence. Additional research is
necessary to determine if these results are valid for intelligence personnel.
In addition, the search for training techniques that do generalize to new
tasks should be continued.

Question 5: What Is the Relationship Between Expertise in Specialty
Area and Calibration? Are personnel who have greater knowledge and experience
in their particular specialty area also more accurate in assigning subjective
probabilities? If this is true, then training should jointly emphasize prob-
ability and substantive knowledge. Research on this issue could compare the
degree of calibration possessed by students and experienced intelligence ana-
lysts after similar amounts of probability training.

Question 6: Are Numerical Subjective Probability Estimates Accurately
Interpreted and Combined by Recipients? When numerical estimates become
widely used, analysts will be receiving such ratings from several sources.
In some cases the estimates received will represent a summary of estimates
made at lower echelons. Analysts and other personnel may adopt heuristics
for combining the various estimates that could possibly cause distortions and
biases in the final intelligence estimates. At present, we have no research
to indicate the type and extent of such errors, if any, within the intelligence
community.

Although these six issues and questions, briefly outlined here, would
require much laboratory and field research, the answers will provide explicit
guidelines for training and implementing subjective probability estimates in
intelligence communication.

SUMMARY AND CONCLUSIONS

This paper has summarized and critically reviewed psychological research on implementing and using subjective probabilities in Army tactical intelligence. Current procedures result in ambiguous communication of uncertainty in both intelligence estimation and the evaluation of intelligence data. Available research supports the feasibility of incorporating subjective probabilities in intelligence estimation in accord with NATO STANAG 2118 as well as in intelligence data evaluation. Although some questions still remain concerning the details of training and implementation, current knowledge provides a sufficient base to begin incorporating subjective probability estimates into Army doctrine and practice. Although new problems will arise, they can best be answered within an operational Army context.

REFERENCES

Adams, J. K., & Adams, P. A. Realism of confidence judgments. *Psychological Review*, 1961, 68, 33-45.

Andrews, R. S., & Ringel, S. Certitude judgments and accuracy of information assimilation from visual displays. ARI Technical Note 145, 1964.

Baker, J. D., McKendry, J. M., & Mace, D. J. Certitude judgments in an operational environment. ARI Technical Research Note 200, November 1968.

Brown, R. V. On the credibility of estimates: its evaluation and improvement. Technical Report TR 78-6-72. McLean, Va.: Decisions and Designs Incorporated, 1978.

Brown, T. A., & Shuford, G. H. Quantifying uncertainty into numerical probabilities for the reporting of intelligence. Report No. R-1185-ARPA. Santa Monica, Calif.: The Rand Corporation, July 1973.

Decisions and Designs, Inc. Advanced Decision Technology Program: Annual progress report PR 77-17-30. McLean, Va.: DDI, December 1977. (Supported by the Defense Advanced Research Projects Agency and monitored by the Office of Naval Research under Contract N00014-76-C-0074)

Department of the Army. *Combat intelligence field manual.* (FM 30-5). Washington, D.C.: 1973.

Edwards, W., Lindman, H., & Phillips, L. D. Emerging technologies for making decisions. In T. M. Newcomb (Ed.), *New directions in psychology II.* New York: Holt, Rinehart, and Winston, 1965.

Einhorn, H. J. Expert measurement and mechanical combination. *Organizational Behavior and Human Performance*, 1972, 7, 86-106.

Evans, L., & Swensen, R. Vectors of probable accuracy of image interpreter performance. ARI Working Paper HF-79-05, June 1979.

Gardiner, P. C., & Edwards, W. Public values: multiattribute utility measurement for social decision making. In M. F. Kaplan & S. Schwartz (Eds.), *Human judgment and decision processes.* New York: Academic Press, 1975.

Graham, D. O. Estimating the threat: A soldier's job. *Army*, 1973, 23, 14-18.

Guttentag, M., & Streuning, E. (Eds.) *Handbook of evaluation research.* Volumes I and II. Beverly Hills, Calif.: Sage, 1975.

Halpin, S. M., Moses, F. L., & Johnson, E. M. A validation of the structure of combat intelligence ratings. ARI Technical Paper No. 302, August 1978.

Hammond, K., Rohrbaugh, J., Mumpower, J., & Adelman, L. Social judgment theory: Applications in policy formation. In M. F. Kaplan & S. Schwartz (Eds.), *Human judgment and decision processes in applied settings.* New York: Academic Press, 1977.

Johnson, E. Numerical encoding of qualitative expressions of uncertainty. ARI Technical Paper 250, December 1973.

Kelly, C. W., & Peterson, C. R. Probability estimates and probabilistic procedures in current intelligence analysis: Report on Phase I. Report FSC-71-5047. Gaithersburg, Md.: International Business Machines (IBM), Federal Systems Division, 1971.

Kelly, C. W., Peterson, C. R., Brown, R. V., & Barclay, S. Decision theory research. Technical Progress Report 2, CONFIDENTIAL. McLean, Va.: Decisions and Designs, Inc., January 1974.

Kibler, A. W., Watson, S. R., Kelly, C. W., & Phelps, R. H. A prototype aid for evaluating alternative courses of action for tactical engagement. ARI Technical Report TR-78-A38, November 1978.

Lichtenstein, S., & Fischhoff, B. Do those who know more also know more about how much they know? The calibration of probability judgment. Organizational Behavior and Human Performance, 1977, 20, 159-183.

Lichtenstein, S., & Fischhoff, B. Training for calibration. ARI Technical Report TR-78-A32, November 1978.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. Calibration of probabilities: The state of the art. In H. Jungermann & G. de Zeeuw (Eds.), Decision making and change in human affairs. Amsterdam: D. Reidel, 1977.

Miron, M. S., Patten, S. M., & Halpin, S. M. The structure of combat intelligence ratings. ARI Technical Paper No. 286, September 1978.

Murphy, A. H., & Winkler, R. L. Subjective probability forecasting experiments in meteorology: Some preliminary results. Bulletin of the American Meteorological Society, 1974, 55, 1206-1216.

NATO Standardization Agreement (STANAG 2118). March 3, 1976.

Root, H. E. Probability statements in weather forecasting. Journal of Applied Meteorology, 1962, 1, 163-168.

Samet, M. Subjective interpretation of reliability and accuracy scales for evaluating military intelligence. ARI Technical Paper No. 260, January 1975.

Sanders, F. The evaluation of subjective probability forecasts. Scientific Report No. 5, Cambridge, Mass.: Institute of Technology, Department of Meteorology, 1958.

Slovic, P., Fischhoff, B., & Lichtenstein, S. Cognitive processes and societal risk taking. In J. S. Carroll & J. W. Payne (Eds.), Cognitive and social behavior, Potomac, Md.: Lawrence Erlbaum Associates, 1976.

U.S. Weather Bureau. Report on weather bureau forecast performance 1967-8 and comparison with previous years. Technical Memorandum WBTM FCST 11. Silver Spring, Md.: Office of Meteorological Operations, Weather Analysis and Prediction Division, March 1969.

Williams, P. The use of confidence factors in forecasting. *Bulletin of American Meteorological Society*, 1951, *32*, 279-281.

Williams, R. W. Commanders and intelligence. *Army*, 1972, *22*, 21-24.

Williams, R. W. Surprise: The danger signals. *Army*, 1974, *24*, 10-16.

APPENDIX A

RESEARCH SUMMARY: RATING SCALES FOR
SPOT REPORT EVALUATION

This summary of research on the use of the current 7-point rating scales
for evaluating spot report information focuses on two topics. The first issue
concerns two types of problems with the present rating scales. The second
topic addresses the question of how to modify the scale dimensions presently
used to be more consistent with those dimensions actually used by intelligence
personnel. Together these two topics summarize the current research and its
implications for developing a new scale for evaluating spot reports.

Problems with the Current Activity and Reliability Scales

Two categories of inadequacies of the spot report evaluation scales have
been identified. The first contains problems associated with the assignment
of the ratings, and the second category encompasses ambiguities in the inter-
pretation and use of the ratings by recipients.

Problems in Rating Assignment

An examination of all messages filed by two divisions of a corps during a
7-day training exercise revealed that 70% of more than 2,000 messages were
spot reports (Baker, McKendry, & Mace, 1968). However, 50% of these spot re-
ports did not contain evaluation ratings. In addition, in those reports which
were evaluated, the two ratings were not independent of each other; for 87% of
the ratings there was exact correspondence of the levels of the accuracy and
reliability scales, e.g., A-1, B-2, C-3 (See Table A-1). The fact that the
ratings are not independent implies that the scales are viewed by the partici-
pants as redundant or at least highly correlated with each other.

This interdependence of the scales was empirically investigated by Samet
(1975). Recent graduates of the Army Intelligence Career School completed
several experimental tasks designed to determine the relative importance and
independence of the two scales. Analyses based on the techniques of linear
multiple regression and analysis of variance showed that the interdependence
between the scales found by Baker et al. (1968) was replicated. In addition,
the accuracy scale was identified as being approximately four times as impor-
tant as the reliability scale; the overwhelming importance of the accuracy
dimension has been confirmed by Miron, Patten, and Halpin (1978).

An examination of the phrasing and structure of the scales reveals at
least two additional problems which may contribute to difficulties. First,
the phrases used to describe both of the scales are ambiguous; that is, inter-
pretation of terms such as "probably," "possible," and "usually" is unclear.
There is a large body of empirical psychological research demonstrating that
such verbal quantifiers are widely and generously interpreted (misinterpreted?)
by different raters; for example, when asked to assign a numerical probability
to the phrase "highly improbable," military personnel gave responses ranging

Table A-1

Distribution of Ratings Obtained During Field Exercise

Reliability of the source	Accuracy of the Information						Row Total
	1 Confirmed by other sources	2 Probably true	3 Possibly true	4 Doubtfully true	5 Improbable	6 Truth cannot be judged	
A. Completely reliable	43[a]	11	2	0	0	0	56
B. Usually reliable	11	518	57	2	0	0	588
C. Fairly reliable	0	0	5	1	1	0	7
D. Not usually reliable	0	0	0	8	0	0	8
E. Unreliable	0	0	0	0	3	1	4
F. Reliability cannot be judged	0	0	1	0	0	31	32
Total	54	529	65	11	4	32	695

Source. Baker, McKendry, and Mace (1968).

[a] Note the disproportionate number of ratings along the diagonal indicating an interdependence between the two scales. Also note the concentration of B-2 ratings.

24

from 0 to .90 (Johnson, 1973). Obviously, such ambiguous terms contribute to confusion in using the rating scales. The second problem is that the scales are not described by a single continuum. The accuracy scale is composed of at least two categories of dimensions rather than a single continuous accuracy dimension. The reliability scale is based on a continuum from high (A) to low (E), plus a second category (F). Given the time pressures of intelligence operations, it is necessary for the scales to be as intuitive and simple to use as possible. Complexity in such a situation will only foster misapplication or even elimination of the spot report evaluation. Since efforts to reduce the complexity by providing users with aids, such as the decision flow-chart shown in Figure A-1, failed to improve the quality of the ratings (Baker, McKendry, & Mace, 1968), a simplification of the scales themselves may be necessary.

Problems in Interpretation of Ratings

Difficulties encountered in the application of the current evaluation system are compounded by inconsistencies in the interpretation of the evaluation ratings by the recipient. When asked to assign a confidence rating to spot reports bearing various accuracy-reliability ratings, different participants assigned very disparate confidence ratings to the same spot report evaluation; for example, when a report was assigned a reliability rating of "E" (Unreliable), the confidence of various participants ranged from .05 to .53 (Samet, 1975).
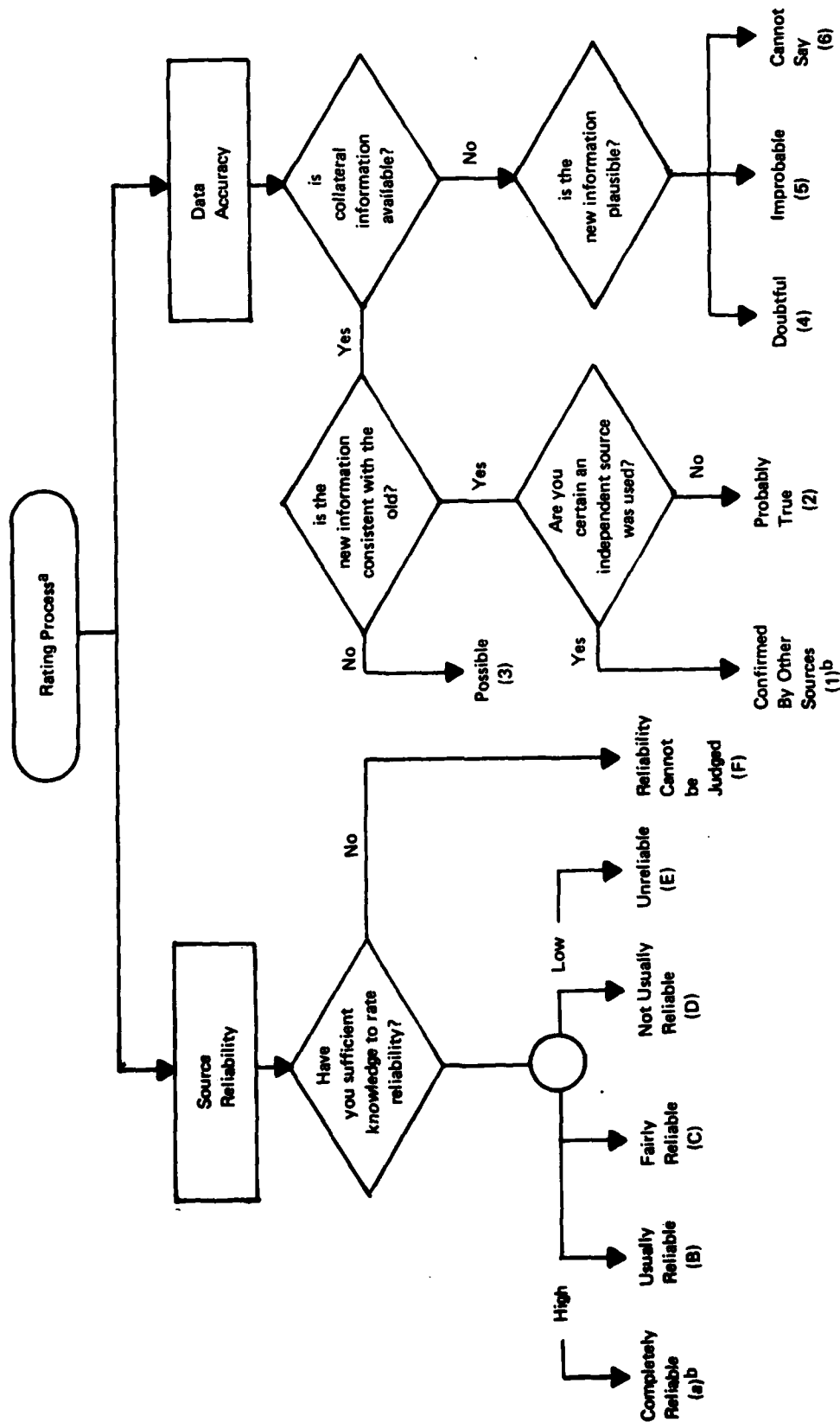
Modification of the Scales: Identification of Dimensions

The research investigating the application of the current accuracy and reliability scales clearly showed there is considerable confusion over the interpretation and apparent redundancy of the scales. Two additional experiments were conducted specifically to investigate the scale dimensions actually used by intelligence personnel for evaluating the quality of spot report information.

Procedure

In order to restructure the evaluation of information quality to allow analysts to communicate their judgments effectively, it is necessary to thoroughly understand dimensions of information value which are important to the analyst. What qualities of information does the analyst attend to in producing a valid, integrated, intelligence picture? The accuracy of information? The timeliness of information? The relationship between information received and enemy doctrine?

Value dimensions were sought by asking intelligence analysts to use 50 quality rating scales to evaluate the information in a series of messages. An examination of the relationships among the ratings made across many messages, using factor analytic techniques, made it possible to draw inferences about the "structure" underlying the value judgments. For example, we might find that the judgment structure underlying ratings of the quality of new homes was based on dimensions of the size of house, location of the property, and cost.

Rating Process[a]

Data Accuracy

is collateral information available?

Yes → is the new information consistent with the old?

No → is the new information plausible?

is the new information consistent with the old?
- No → Possible (3)
- Yes → Are you certain an independent source was used?
  - Yes → Confirmed By Other Sources (1)[b]
  - No → Probably True (2)

is the new information plausible?
- Doubtful (4)
- Improbable (5)
- Cannot Say (6)

Source Reliability

Have you sufficient knowledge to rate reliability?

No → Reliability Cannot be Judged (F)

High
- Completely Reliable (a)[b]
- Usually Reliable (B)
- Fairly Reliable (C)

Low
- Not Usually Reliable (D)
- Unreliable (E)

[a]Based on definitions and instructions in FM 30-5

[b]The letters in parentheses represent the value for the square reliability scale and the numbers in parentheses represent the numerical value for the data accuracy scale

Figure A-1. Decision table for determining source reliability and accuracy of information.
Source. Baker, McKendry, and Mace, 1968.

26

An initial experiment established the basic dimensions for judgments of the quality of intelligence data (Miron, Patten, & Halpin, 1978). A second experiment was conducted to validate the initial findings and to test an application of those findings to the development of new rating procedures (Halpin, Moses, & Johnson, 1978). The experiments involved 20 to 40 messages in one of two tactical scenarios. Participants were Army intelligence personnel with a variety of backgrounds.

Experiment I. Two groups in the first experiment (Miron, Patten, & Halpin, 1978) rated messages selected from the files of the 28th Infantry Division for the period just prior to the German Ardennes counteroffensive in 1944 (Battle of the Bulge). One group of enlisted personnel, called the untrained group, was just entering the U.S. Army Intelligence Center and School course for intelligence analysts (96B); the other group (the trained group) was just completing the same course. The rating scales included the standard Accuracy and Reliability scales, two repetitions of a 0 to 100 scale (Global Validity), and 46 bipolar adjectival scales (e.g., garbled/clear, true/false) developed to represent many possible facets of the analysts' judgment task.

Since there were only minor variations in ratings between the two groups, the data were combined. The analysis of the combined ratings showed that two, or at most three, dimensions were sufficient to account for essentially all of the variation in ratings as shown in Table A-2. The strongest dimension is labeled ACCURACY, which subsumes the standard Reliability and Accuracy scales, the Global Validity scales, and bipolar ratings such as True/False, and Probable/Improbable. The second dimension is related to ratings of RELEVANCE, such as ratings on bipolar scales Heavy/Light, Large-scale/Small-scale, and Many/Few. The third dimension was tentatively identified as DIRECTNESS.

Experiment II. A second experiment (Halpin, Moses, & Johnson, 1978) replicated the previous research with a group of experienced officers in the Intelligence Officers Advanced Course at the Intelligence School. In addition, a second group of students in the Advanced Course made similar ratings of 40 messages from a scenario set in modern day Central-Europe (Hof Gap).

There were no major differences in the results for the two scenarios, and there were strong similarities between these results and Experiment I. The most important dimension in the judgment structure for the officers dealing with either the Battle of the Bulge messages or the Hof Gap messages was ACCURACY; this dimension is related to essentially the same scales as the ACCURACY dimension found in the first experiment. A second judgment dimension from the ratings of the Battle of the Bulge messages primarily reflected considerations of information IMPORTANCE. The second and third dimensions from the Hof Gap ratings reflected judgments of THREAT and SCOPE. Taking these results together we see that a secondary judgment of RELEVANCE/IMPORTANCE/ THREAT is represented by the participants' ratings. Thus, the general finding of Experiment I concerning the structure of such ratings was validated using a different population of raters and a different scenario.

Table A-2

Summary of Derived Intelligence Factors

Factor I (57%) ACCURACY		Factor II (10%) RELEVANCE		Factor III (6%) DIRECTNESS	
Scale component attribute	Factor coefficient	Scale component attribute	Factor coefficient	Scale component attribute	Factor coefficient
Accurate - Erroneous	.98	Heavy - Light	.96	Interpreted - Uninterpreted	.76
Truthful - Deceptive	.97	Many - Few	.92	Implied - Unimplied	.75
True - False	.97	Large Scale - Small Scale	.92	Understandable - Confusing	.45
Acceptable - Unacceptable	.97	Dangerous - Safe	.90	Constant - Changing	.37
Accuracy	.96	Risky - Routine	.89		
Global 1	-.96	Massive - Insignificant	.88		
Dependable - Undependable	.95	Widespread - Local	.84		
Faultless - Faulty	.95	Precarious - Imprecarious	.80		
Probably - Improbable	.94	Extraordinary - Ordinary	.80		
Reliable - Variable	.92	Hazardous - Unhazardous	.78		
Possible - Impossible	.92	Volatile - Inert	.70		
Global 2	-.92	Active - Inactive	.48		
Likely - Unlikely	.92				
Direct - Indirect	.92				
Factual - Theoretical	.91				
Reliability	.90				
Confirmed - Unconfirmed	.90				
Pertinent - Extraneous	.89				
Consistent - Inconsistent	.89				

Source. Miron, Patten, and Halpin (1978).

The results of Experiment I suggested that a few scales, using a 0 to 100 format, could capture the essence of raters' judgments of information quality. An additional task was presented to the participants in Experiment II to evaluate four such scales: Truth, Relevance, Predictability, and Importance. Participating officers reevaluated 20 of the 40 messages on these four scales and the data were analyzed in the context of their other ratings.

The new single scale rating of Truth was strongly related to the judgment dimension of ACCURACY; the Predictable scale was also somewhat related to this dimension. Judgments of Relevance and Importance were related to the RELEVANCE/IMPORTANCE/THREAT dimension.

Discussion

The present research suggests several important guidelines for the development of more effective ratings of information quality. First, analysts with different backgrounds and levels of experience evaluate spot report information using similar multidimensional judgment dimensions. Second, these structures do not correspond to the evaluation structure embodied in current doctrine, and they are not modified to any significant extent by formal training. Thus, to correct the deficiencies in current evaluation procedures it is appropriate to modify doctrine and procedures to provide a closer match between the requirements for ratings of information quality and the actual judgment dimensions used by intelligence personnel.

The experiments on the evaluation of information quality clearly indicate a strong component related to the perceived accuracy or truth of the rated information. This feature of information is already stressed in present doctrine, and there is a clear functional requirement for its continued use. However, there is no apparent need for ratings of accuracy to be coupled exclusively to the presence or absence of confirming information. A 0 to 100 scale for ratings of information accuracy was tested in the present research and was shown to be an effective indicator of analysts' perceptions of the information. If this or a similar scale were adopted, it would allow a more general judgment of information accuracy than the current scale which is tied to a degree of confirmation. This in turn should reduce the confusion concerning the application and interpretation of accuracy ratings and increase the effective use of available information.

A second component of analysts' perceptions of information quality was related to the perceived relevance and/or importance of the information. This feature of information is not explicitly treated in present doctrine; information processors simply attend to information or ignore it depending on their implicit evaluation of its relevance. However, the development of data storage and retrieval systems within tactical data systems requires explicit ratings of relevance to permit filtering, purging, and selective retrieval of data. The scales designed and tested in this research were not totally successful in capturing this aspect of analysts' judgments, and further research will be required to provide a scale for this purpose.

One component of information quality, explicitly treated in current
doctrine, but which did not emerge in the present research results is the
"reliability of the source." This concept was not relevant in the context of
the scenarios used and is apparently not relevant in the majority of informa-
tion processing situations. It may be desirable to explore this rating further
in more realistic field studies to determine whether analysts have an implicit
appreciation of its meaning and application, and to develop a new scale that
more effectively represents judgments in this area. However, previous research
strongly suggests that ratings of source reliability are tied closely to per-
ceptions of information accuracy, and that an independent rating or reliability
would be of little value except to those few individuals directly involved in
management of collection assets.

APPENDIX B

THE AXIOMATIC RULES OF SUBJECTIVE PROBABILITY

1. The events must be stated such that they can be confirmed as true or false within some specified time period. An example of a confirmable event from the area of intelligence estimation is: the enemy will attack Camp X by noon tomorrow; a nonconfirmable statement is: the enemy will attack Camp X. In the first case, a subjective probability could be assigned and the truth of the statement be determined within the specified time period. However, in the latter case, since there is no time at which the truth can be assessed, the probability would be 1.0; i.e., there is no uncertainty since the enemy has an infinite amount of time in which to attack; such a probability estimate would be of little use to a commander.

2. The number of events to be assessed simultaneously must be determined. In the case of intelligence data evaluation, usually each report is evaluated separately; that is, the subjective probability estimate represents the rater's degree of belief in the accuracy of that report, independent of reports on other information. However, in intelligence estimation, several alternative events or enemy courses of action may be assessed at the same time. To assess several courses of action concurrently, the likelihood is expressed as a vector of probabilities, each probability corresponding to an alternative. For example, course of action A (enemy will attack Camp X by noon tomorrow) may be assigned a probability of .5, course of action B (enemy will attack Camp Y by noon tomorrow) may be assigned a probability of .3, while C (enemy will not attack at all) is assigned a probability of .2.

3. Multiple events or alternatives must be stated such that they are mutually exclusive; e.g., the enemy will attack Camp X by noon tomorrow, or, the enemy will not attack at all. The critical point is that there be no implicit overlap between events or alternatives being assessed; if, in fact, there is some chance that more than one alternative could occur within the same period, then this possibility must be formulated as a separate alternative, e.g., probability of A = .4, probability of B = .5, probability of both A and B = .1.

4. All mutually exclusive events should be assigned a probability corresponding to the perceived likelihood of each event. These probabilities need to be adjusted such that the sum of the probabilities equals unity. In other words, there is a 100% probability that something will occur; the assessor's task is to distribute the 100 percentage points among all possible mutually exclusive events.