NUMERICAL ENCODING OF QUALITATIVE EXPRESSIONS OF UNCERTAINTY

Edgar M. Johnson

Army Research Institute for the Behavioral and Social Sciences
Arlington, Virginia

December 1973
NOTICES

DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U. S. Army Research Institute for the Behavioral and Social Sciences, ATTN: RDMR-P, 1300 Wilson Blvd., Arlington, Virginia 22209.

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U. S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.
Title: Numerical Encoding of Qualitative Expressions of Uncertainty

Author(s): Edgar M. Johnson

Performing Organization: U.S. Army Research Institute for the Behavioral and Social Sciences

Program: DA/Office, Chief of Research & Development

Report Date: December 1975

Number of Pages: 27

Distribution Statement: Approved for public release; distribution unlimited

Subjective probability, Numerical encoding, Probability estimation, Semantics, Scaling, Information processing, Cluster analysis

Abstract: Judgments of the probable accuracy of intelligence data and products are integral to the intelligence process. Accurate communication of the probability of uncertainty implicit in these judgments is requisite to the effective production and utilization of intelligence. Twenty-eight subjects, 14 U.S. Army enlisted men and 14 extension college students, numerically encoded on a 0 to 100 scale each of 15 probability phrases in each of three sentence contexts. The 15 phrases systematically covered a wide range of probabilistic meaning and the three sentence contexts involved a weather forecast, a prediction...
of personal success and an intelligence report. There were no significant differences in the encoding of probability phrases into numerical equivalents among the three sentence contexts, between enlisted personnel and college students or as a function of age, sex or educational level. Subjects were relatively consistent in their own encoding of given phrases, but differed, often radically, from other subjects. Cluster analysis of the numbers assigned by subjects indicated an underlying asymmetric probability scale comprised of a small number of intervals. Further research should investigate the use of standardized lexicons with a small number of expressions or the direct use of numerical scales. Numerical scales appear to be the most promising and would facilitate the use of the tools of probability theory and decision theory in intelligence analysis.
NUMERICAL ENCODING OF QUALITATIVE EXPRESSIONS OF UNCERTAINTY

Edgar M. Johnson
Army Research Institute for the Behavioral and Social Sciences

Robert S. Andrews, Supervisory Project Director

SYSTEMS INTEGRATION AND COMMAND/CONTROL TECHNICAL AREA
Cecil D. Johnson, Chief

Submitted By:
Joseph Zeidner, Director
ORGANIZATIONS & SYSTEMS RESEARCH LABORATORY

Approved By:
J. E. Uhlaner
TECHNICAL DIRECTOR

U. S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
Office, Chief of Research and Development
Department of the Army
1300 Wilson Boulevard, Arlington, Virginia 22209

December 1973
ARI Research Reports and Technical Papers are intended for sponsors of R&D tasks and other research and military agencies. Any findings ready for implementation at the time of publication are presented in the latter part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.
The Intelligence Systems Work Unit within the U. S. Army Research Institute for the Behavioral and Social Sciences (ARI) is concerned with the functions of human information analysis, processing, aggregation, integration and consequent product utilization in intelligence systems. One of the major objectives is to provide research findings by which performance of these functions can be enhanced. One resulting requirement is to determine how human capabilities can be utilized to enable the intelligence information processing system to function with increased effectiveness. The entire research effort is responsive to requirements of RDTE Project 2016210A754, "Intelligence Information Processing," FY 1973 Work Program and to special requirements of the U. S. Army Intelligence Center and School.

The U. S. Army currently has under development intelligence information processing systems designed to maximize combat effectiveness by optimal utilization of human capabilities augmented by computer support. The present publication describes one effort which provides data for more effectively evaluating man’s capabilities and limitations in intelligence processing.

J. E. OHLANER
Technical Director
NUMERICAL ENCODING OF QUALITATIVE EXPRESSIONS OF UNCERTAINTY

BRIEF

Requirement:

Judgments of the probable accuracy of intelligence data and products are integral to the intelligence process. Accurate communication of the probability or uncertainty implicit in these judgments is requisite to the effective production and utilization of intelligence. Factors which may impact on the accurate communication of these values must be identified and their degree of impact ascertained to provide a rational basis for the development of improved procedures. Determination of whether the encoding of qualitative expressions is influenced by context or group membership and the form of the probability scale used to encode uncertainty is directly relevant to that goal.

Procedure

Twenty-eight subjects, 14 U.S. Army enlisted men and 14 extension college students, numerically encoded on a 0 to 100 scale each of 15 probability phrases in each of three sentence contexts. The 15 phrases systematically covered a wide range of probabilistic meaning and the three sentence contexts involved a weather forecast, a prediction of personal success, and an intelligence report.

Findings

1. There were no significant differences in encoding of probability phrases into numerical equivalents among the three sentence contexts, between enlisted personnel and college students, or as a function of age, sex, or educational level.

2. Individuals were relatively consistent in their encoding of given phrases, but differed, often radically, from other individuals.

3. Individuals’ numerical encodings indicated the use of an underlying asymmetric probability scale comprised of a small number of intervals.

Application of Findings:

The findings indicate the use of qualitative expressions to communicate the accuracy or relative likelihood of occurrence of intelligence data and products will often result in a high degree of misunderstanding. Personnel involved in the production and use of intelligence should be extremely wary of attempting to infer numerical values from qualitative expressions of uncertainty. Differences in the encoding of qualitative expressions do not appear dependent upon any of the general factors evaluated in this study other than the ambiguities of qualitative phrases themselves as influenced by individual differences of unknown sources.

Further research should investigate the use of standardized lexicons with a small number of expressions or the direct use of numerical scales. Numerical scales appear to be the most promising and would facilitate the use of the tools of probability theory and decision theory in intelligence analysis.
NUMERICAL ENCODING OF QUALITATIVE EXPRESSIONS OF UNCERTAINTY

CONTENTS

INTRODUCTION 1
Objectives 4

METHOD 4
Subjects 4
Experimental Materials 4
Procedure 5

RESULTS 6

DISCUSSION 12

CONCLUSION 13

REFERENCES 14

APPENDIX 15

DISTRIBUTION 18

DD Form 1473 (Report Documentation Page) 21

TABLES

Table 1. Table of Kent's chart showing expressions of degrees of uncertainty 2
2. Statistics on probability phrases 7
3. Mean percent of encoding agreements across sentence contexts 11

FIGURES

Figure 1. Dendrogram representing results of unweighted pair-group cluster analysis 9
INTRODUCTION

The weather forecast states "rain is likely within the next 24 hours." An intelligence estimate states "an enemy attack is likely within the next 24 hours." What do these statements mean? If the source of each statement were to express the forecast in terms of a numerical probability, would the numbers they provide be identical? Would our interpretation of these statements agree with the interpretation intended by the source? The answer to both questions is "likely" to be no. Yet reliable and valid communication of the degree of uncertainty in forecasted events can be of great practical importance.

Intelligence is seldom perfect and evaluations containing inadequate data and doubtful conclusions can often be extremely useful. The user of an intelligence evaluation will naturally be influenced by the degree of certainty which has been attached to it. It is the responsibility of the intelligence officer to determine the degree of uncertainty of a given statement and then to communicate this to the users of the intelligence. It should be noted that the problem of communicating uncertainty along such dimensions as source reliability and information accuracy is a problem within the intelligence section as well as for users.

However, there is no standard terminology in use to describe the probability of occurrence of forecasted events. Sherman Kent in the mid 1950's proposed a list of words and phrases to be associated with specified ranges of probability to ensure their understanding (Table 1). The motivation behind Kent's proposal was his observation of large differences in the probability values assigned to given phrases by different intelligence analysts. Kent's chart was never formally adopted by any intelligence agency.

More formal studies of the assignment of numerical probability values to probability phrases or words have also found large individual

---


Table 1

TABLE OF KENT’S CHART SHOWING EXPRESSIONS OF DEGREES OF UNCERTAINTY*

<table>
<thead>
<tr>
<th>Probability Range</th>
<th>Odds</th>
<th>Phrase</th>
<th>Synonyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>Certain</td>
<td>No estimate</td>
<td></td>
</tr>
<tr>
<td>85-99</td>
<td>9:1 or more</td>
<td>Almost certain</td>
<td>Almost certain</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Believe</td>
<td>Highly likely</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Evident</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Little doubt</td>
<td></td>
</tr>
<tr>
<td>60-84</td>
<td>3:1</td>
<td>Chances are good</td>
<td>Should be</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Probable</td>
<td>Is expected</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fairly certain</td>
<td>Logical to assume</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Likely</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Appears to</td>
<td>Reasonable to conclude</td>
</tr>
<tr>
<td>40-59</td>
<td>1:1</td>
<td>Chances about even.</td>
<td></td>
</tr>
<tr>
<td>15-39</td>
<td>1:3</td>
<td>Chances good that--</td>
<td>Appears not to</td>
</tr>
<tr>
<td></td>
<td></td>
<td>not</td>
<td>Should not be</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Probable</td>
<td>Is not expected</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not probable</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fairly uncertain</td>
<td>Not logical to assume</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unlikely</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not likely</td>
<td>Not reasonable to conclude</td>
</tr>
<tr>
<td>1-14</td>
<td>1:9 or less</td>
<td>Almost certain</td>
<td>Almost certainly not</td>
</tr>
<tr>
<td></td>
<td></td>
<td>that--not</td>
<td>Highly unlikely</td>
</tr>
<tr>
<td>0</td>
<td>Impossible</td>
<td>No estimate</td>
<td></td>
</tr>
</tbody>
</table>

* Modified from Platt (1957) pp 708-710
differences among subjects in intelligence. An informal study conducted by NATO\(^1\) provided similar results for intelligence statements. Twenty-three officers of 10 different nationalities, all fluent in English and with a background in intelligence, described in terms of "chances out of 100" what a series of statements meant to them. The size of the range of numerical values assigned to specific phrases varied from 25 to 80. Another informal study found differences as large as 50 between the numerical values assigned to an intelligence evaluation by its two authors\(^1\).

An understanding of the basis for the large individual differences in the numerical encoding of qualitative expressions of uncertainty should aid in designing better methods for communicating degrees of uncertainty. These methods may take the form of a glossary as suggested by Kent, a change to direct numerical estimation\(^5\) or some combination of these.

---


12. Ibid.

The present study was designed to explore the potential sources of individual differences in numerical encoding: sentence context and group membership of individuals. Previous work\textsuperscript{14} suggests that context influences the encoding of probability phrases, but did not assess the significance of context effect. Group membership may also play a role in numerical encoding due to the experience and training which differentiates groups of individuals. An additional interest in the present study is the nature of the scale used by individuals in mapping probability phrases into numerical equivalents.

Objectives

1. To assess the effect of context on the numerical encoding of qualitative statements of probability.

2. To compare numerical assignments to probability phrases made by military personnel and by evening college students.

3. To determine the consistency and the form of the probability scale used in assigning numerical estimates to probability phrases.

\textbf{METHOD}

\textbf{Subjects}

Two groups of subjects were used. The first group consisted of 14 U.S. Army enlisted men who had recently completed training as image interpreters. All had scored above the mean for all enlisted men on the Army's general technical aptitude test. The second group consisted of 14 students in an introductory psychology course at the Graduate School of the U.S. Department of Agriculture. All had a high school diploma, and the mean educational level was 14.2 years of school with a range of 12 to 20 years.

\textbf{Experimental Materials}

A questionnaire with 45 sentences, the factorial combination of 15 probability words or phrases and three sentence contexts, was prepared. The 15 phrases were chosen from among those used in prior studies\textsuperscript{15-19} to systematically cover a wide range of probabilistic meaning. (Appendix).

\textsuperscript{14} Cohen, De~rnley, a. Hansel, 1958.

\textsuperscript{15} Levine & Eldridge, 1970.

\textsuperscript{19} Lichtenstein & Newman, 1967.
Eight of these phrases consisted of a systematic variation of adverbs combined with the root "likely" and four were combinations of adverbs with the root "probable." The three sentences were selected to provide different contexts as illustrated below:

The official weather forecast says that rain is highly probable for tomorrow.

You tell someone that it is very likely you will win a contest.

The CIA reports that from satellite photographs it is very probable that anti-missile sites are being constructed around Moscow.

The sentences were presented in stimulus-response pairs composed of a stimulus sentence using one of the 15 probability phrases and a response sentence in which the subject encoded the probability phrase into the number of chances out of 100 which most clearly reflected the degree of uncertainty implied by the sentence.

The questionnaire was arranged in a booklet with each page containing three sentences. The three sentence contexts appeared in a random order on each page, and a given probability phrase appeared only once on any page. Each subject received a different booklet, i.e., a different random order of the 45 sentences.

Procedure

Subjects were given the questionnaire in three groups: the 14 college students, and two groups of enlisted men of 8 and 6 subjects, respectively. Each subject was given a copy of the printed instructions:

This is a study to determine the meaning of some common words for certainty. In the booklets you've received, you will find pairs of sentences like the following set:

The official weather forecast states that rain is somewhat likely tomorrow.

This means there are ___ chances out of 100 of rain tomorrow.

In the second sentence you should place a number from 0 to 100 describing the degree of certainty you think the sentence indicates. For example, in the sentence above I would put "79," indicating the sentence means to me that there are 79 chances out of 100 of rain tomorrow. The weather forecaster may have intended to indicate there are 70 chances out of 100 of rain tomorrow. Your answer may not agree with either mine or the forecaster's.
There are no right answers. You should consider each sentence separately and choose a number which best describes the certainty you think the sentence indicates. If you are not sure what number to use, use the first number between 0 and 100 that comes to your mind.

After the experimenter read the instructions aloud and answered any questions, each subject completed the questionnaire. The entire session required about 30 minutes for each group.

RESULTS

A three-way analysis of variance of subjects' responses, subject group x probability phrase x sentence context, revealed no significant main effects between military and college subjects or among the three sentence contexts. A significant three-way interaction of subject groups, probability phrases, and sentence context, F(28,728) = 1.70, p < .05, reflects the large differences between probability phrases and some small differences in the encoding of specific probability phrases in particular sentence contexts by military and college subjects. There were significant differences among the 15 probability phrases, F(14,364) = 197.06, p < .001. The differences among probability phrases account for over 68% of the total variance while the significant interaction accounts for less than one percent of the total variance. Using Scheffe's procedure for post hoc mean comparisons, the critical difference between mean responses on probability phrases is 12.7 at the 0.05 level of significance. The fifteen probability phrases fall into three clusters using Scheffe's criterion: phrases ranked 1-3, 5-9, 10-15 (Table 2). The phrase ranked fourth, "quite likely," falls between the first two clusters and is not significantly different from phrases in either cluster. With this latter phrase as an exception, phrases within a cluster are not significantly different from each other, but are significantly different from any phrase in another cluster.

Descriptive statistics summarizing the data on probability phrases are shown in Table 2. There is close agreement between mean and median numerical assignments. For only three probability phrases do the mean and median differ by more than 5. Two of these phrases, "fairly unlikely" and "probable" are strongly skewed to the left. The remaining phrases except "fairly likely," "fair chance" and "possible" also have skewed distributions—within the skew toward the lower end of the scale. The skewed distributions indicate that, depending upon the error criterion used, different "best estimates" (mean, median or mode) of an individual's numerical encoding of specific phrases will be generated.

---

Table 2

STATISTICS ON PROBABILITY PHRASES

<table>
<thead>
<tr>
<th>Rank</th>
<th>Phrase</th>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
<th>Std. Dev.</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Highly Probable</td>
<td>82.0</td>
<td>85.0</td>
<td>90</td>
<td>14.3</td>
<td>20-99</td>
</tr>
<tr>
<td>2</td>
<td>Very Probable</td>
<td>78.8</td>
<td>86.0</td>
<td>80</td>
<td>15.7</td>
<td>5-98</td>
</tr>
<tr>
<td>3</td>
<td>Very Likely</td>
<td>73.8</td>
<td>80.0</td>
<td>30</td>
<td>19.2</td>
<td>10-99</td>
</tr>
<tr>
<td>4</td>
<td>Quite Likely</td>
<td>68.5</td>
<td>72.5</td>
<td>80</td>
<td>18.9</td>
<td>15-99</td>
</tr>
<tr>
<td>5</td>
<td>Likely</td>
<td>60.9</td>
<td>60.0</td>
<td>50</td>
<td>18.5</td>
<td>10-95</td>
</tr>
<tr>
<td>6</td>
<td>Probable</td>
<td>61.5</td>
<td>60.0</td>
<td>60</td>
<td>18.0</td>
<td>10-99</td>
</tr>
<tr>
<td>7</td>
<td>Fairly Likely</td>
<td>54.1</td>
<td>63.0</td>
<td>60</td>
<td>21.3</td>
<td>2-90</td>
</tr>
<tr>
<td>8</td>
<td>Possible</td>
<td>50.6</td>
<td>50.0</td>
<td>50</td>
<td>16.9</td>
<td>4-80</td>
</tr>
<tr>
<td>9</td>
<td>Fair Chance</td>
<td>48.9</td>
<td>50.0</td>
<td>50</td>
<td>20.7</td>
<td>1-100</td>
</tr>
<tr>
<td>10</td>
<td>Unlikely</td>
<td>22.9</td>
<td>20.0</td>
<td>10</td>
<td>15.5</td>
<td>0-70</td>
</tr>
<tr>
<td>11</td>
<td>Fairly Unlikely</td>
<td>21.3</td>
<td>20.0</td>
<td>10-20</td>
<td>14.9</td>
<td>0-65</td>
</tr>
<tr>
<td>12</td>
<td>Improbable</td>
<td>16.3</td>
<td>10.0</td>
<td>10</td>
<td>15.3</td>
<td>0-70</td>
</tr>
<tr>
<td>13</td>
<td>Very Unlikely</td>
<td>14.9</td>
<td>10.0</td>
<td>20</td>
<td>12.5</td>
<td>0-60</td>
</tr>
<tr>
<td>14</td>
<td>Quite Unlikely</td>
<td>14.4</td>
<td>10.0</td>
<td>10</td>
<td>12.6</td>
<td>0-50</td>
</tr>
<tr>
<td>15</td>
<td>Highly Improbable</td>
<td>12.6</td>
<td>10.0</td>
<td>10</td>
<td>17.7</td>
<td>0-90</td>
</tr>
</tbody>
</table>
However, error criteria of number of correct estimates, smallest absolute error, and smallest signed error (mode, median and mean, respectively) each imply a three-interval scale with intervals of 0-40, 40-70, and 70-100.

The mean numerical assignments to the six-mirror-image pairs of probability phrases indicate an underlying asymmetry in the subject's use of the probability scale. When attached to "likely," adverbs were ordered as follows: "very" > "quite" > no adverb > "fairly." However, when attached to "unlikely," adverbs were ordered, "quite" > "very" > "fairly" > no adverb. When attached either to "probable" or to "improbable," adverbs were ordered the same way, that is, "highly" > no adverb. This symmetric set of phrase pairs did not lead to a symmetric response set, apparently due to scale compression in the lower half of the probability scale. The mean range from "fairly likely" to "very likely" is 27.9 whereas the mean range from "unlikely" to "quite unlikely" which includes "fairly likely" and "very unlikely" is 8.7; the mean range from "probable" to "highly probable" is 21.5, whereas the mean range from "improbable" to "highly improbable" is 3.7. Differences in the ordering of adverbs attached to "unlikely" when compared to "likely" are confounded with the compression in the lower half of the probability scale. A decrease in the range of numbers into which a set of phrases is mapped encoded would increase the chances for a reversal in the ordering of phrases.

A hierarchical cluster analysis was used to further identify the interrelationships between phrases. The product moment correlation was used as a measure of association in an unweighted analysis. The association value within any set is the average correlation computed from the original correlation matrix. This analysis has the advantage of showing a continuum of clusters as it looks progressively (in steps) for the most compact and isolable clusters, then for the next most compact and so on, ending with the whole set. The results are displayed as a hierarchical line network or dendrogram, where the length of a line segment joining a pair of phrases reflects the level of association between the phrases (Figure 1). There are three clusters with a relatively high degree of intra-cluster association: the first consisting of the phrases ranked 1 to 4, the second of the phrases ranked 5 to 8, and the third of the phrases ranked 9 to 12 and 14. The remaining two phrases, "very unlikely" and "highly improbable," ranked 13 and 15 are not highly correlated with one another; they "cluster" together only at a relatively low level of association, and they join the third of the

---

Figure 1. Dendrogram representing results of unweighted pair-group cluster analysis
three principle word clusters at an even lower level. This is perhaps
due to the relatively low inter-subject variability on these two phrases.
The degree of fit between the derived cluster structure and the original
data matrix was assessed using the cophenetic correlation between the
original measures of association and the predicted measures of association
based on the dendrogram. This value was .73, indicating an acceptable
fit. These results parallel those obtained using a Scheffe criterion
and, considering the relatively low number of subjects, the differences
are surprisingly minor.

The percent of agreement in encoding a phrase across sentences was
used as a measure of within-subject consistency (Table 3). Forty-eight
percent of the phrases were encoded into exactly the same number for at
least two of the three sentences and on eight percent of the phrases
into the same number for all three sentences. Relaxing the criterion
of agreement to include encoding into numbers within ±5 of each other,
subject encodings agreed on at least two of the three sentences for
74% of the phrases and on 27% of the phrases for all three sentences.
Thus, subjects on the average encoded at least eleven phrases on two
of the three sentences, including four phrases on all three sentences,
within ±5 of each other. The high percentage of agreements was uni-
formly distributed across subjects: the standard deviation of the two
distributions of pair agreements over subjects was 2.29 and 2.00 for
exact and ±5 agreements, respectively, and 1.22 and 2.20 for the dis-
tribution of exact and ±5 agreements on all three sentences.

Correlation of subjects' responses by phrase with age, sex, and
number of years of school showed that all were non-significant. Al-
though inter-subject differences account for more than five percent of
the total variance, none of the variables examined in this study were
related to these differences.

The fifteen probability phrases used in the present experiment
were among those used in two earlier studies. Although the nu-
merical values obtained for specific phrases differ in all three sets
of results, the differences in the numerical encoding are less than 15
for any phrase (Appendix). Three of the phrases ("likely," "probable"
and "improbable") were used by Cohen, Dearnley and Hansel and the
size of the differences from the present results and from previous
studies is in the same range, less than 15 for any phrase. The pattern
or rank ordering of phrases from the 1967 and 1970 studies is also
similar to the present data (Kendall coefficient of concordance =
.992, p < .002).

19 Levine & Eldridge, 1970.

- 10 -
Table 3

MEAN PERCENT OF ENCODING AGREEMENTS ACROSS SENTENCE CONTEXTS

<table>
<thead>
<tr>
<th>Rank</th>
<th>Any 2 Sentences</th>
<th>All 3 Sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Type of Agreement</td>
<td>+5</td>
</tr>
<tr>
<td>1</td>
<td>Highly Probable</td>
<td>57</td>
</tr>
<tr>
<td>2</td>
<td>Very Probable</td>
<td>43</td>
</tr>
<tr>
<td>3</td>
<td>Very Likely</td>
<td>46</td>
</tr>
<tr>
<td>4</td>
<td>Quite Likely</td>
<td>39</td>
</tr>
<tr>
<td>5</td>
<td>Likely</td>
<td>33</td>
</tr>
<tr>
<td>6</td>
<td>Probable</td>
<td>40</td>
</tr>
<tr>
<td>7</td>
<td>Fairly Likely</td>
<td>36</td>
</tr>
<tr>
<td>8</td>
<td>Possible</td>
<td>64</td>
</tr>
<tr>
<td>9</td>
<td>Fair Chance</td>
<td>43</td>
</tr>
<tr>
<td>10</td>
<td>Unlikely</td>
<td>43</td>
</tr>
<tr>
<td>11</td>
<td>Fairly Unlikely</td>
<td>53</td>
</tr>
<tr>
<td>12</td>
<td>Improbable</td>
<td>50</td>
</tr>
<tr>
<td>13</td>
<td>Very Unlikely</td>
<td>57</td>
</tr>
<tr>
<td>14</td>
<td>Quite Unlikely</td>
<td>61</td>
</tr>
<tr>
<td>15</td>
<td>Highly Improbable</td>
<td>50</td>
</tr>
</tbody>
</table>

Mean | 48 | 74 | 8  | 27 |
DISCUSSION

The scale used by individuals in this study for numerically mapping uncertainty phrases was relatively stable but differed, sometimes radically, between individuals. For example, the phrase "fair chance" was encoded into exactly the same number for two out of three sentences by over 40% of the subjects, but the numbers ranged from 1 to 100. The data indicated that the mapping was not influenced by sentence context or by the individual's group membership. Further, there was no finding of a significant correlation between numerical assignments and the age, the sex, or the educational level of the individual. Thus, although an individual's probability scale may be stable, neither the weather forecaster's nor the intelligence assessor's use of a probability phrase would be likely to agree with a user's numerical interpretation of the phrase or with each other's. This suggests that an individual's encoding of probability phrases might be used as an indicator of other characteristics such as risk-taking.

If a probability scale were labeled in accord with a typical subject's numerical encoding, the resulting scale, although veridical to the subject's reported impressions, might look so peculiar as to confuse or mislead both subjects and experimenters (Figure 1). The scale would be asymmetric to reflect the asymmetry between mirror-image pairs and compressed for values below 0.5. It would also be a discrete scale with perhaps three intervals in addition to the anchor points of "impossible" and "certain" events. These three intervals would be shifted toward the upper half of the scale to reflect the lack of differentiation in the lower half. Thus, phrases attached to points on the probability scale to facilitate an individual's understanding of the scale may in fact be confusing to him. This conclusion is also suggested by the fact that individuals often claim that numerically reported subjective probabilities do not fit their verbal conceptualizations.

The use of a standardized lexicon with a small number of expressions or the direct use of numerical scales to communicate degrees of uncertainty should be investigated. Numerical encoding provides an index of the success of communicating degrees of uncertainty using qualitative expressions. The present results and those of earlier studies 23, 24, 25 26

---

23 Levine & Eldridge, 1970.
24 Rigby, L. V., and Swain, A. D. In-flight target reporting—How much is "a bunch"? Human Factors, 1971, 13, 177-182.
suggest that the variability within and between individuals in interpreting qualitative expressions is so high as to often result in a significant degree of misunderstanding. Further, the increased use of the tools of probability and decision theory anticipated with the development of ARTADS (Army Tactical Data Systems), particularly TOS (Tactical Operations System), will require numerical values. This strongly suggests that numerical scales rather than a lexicon of qualitative phrases are the most promising method for improving communication of the degree of uncertainty in intelligence data and products.

CONCLUSIONS

The conclusions of principle interest in identifying sources of individual differences in the numerical encoding of probability phrases are:

1. Encoding of probability phrases into numerical equivalents was not influenced by sentence context.

2. Encoding of probability phrases into numerical equivalents was similar across military personnel and college students and was not correlated with age, sex, or education beyond high school.

3. Individuals were relatively consistent in their encoding of a given probability phrase, but are likely to differ from other individuals.

4. Individuals' numerical encodings indicated the use of an underlying asymmetric probability scale comprised of a small number of intervals.


Letter, MAS (Army) (69) 559, from NATO Assistant Chief of Staff for Intelligence to Military Agency for Standardization, OTAN/NATO, Avenue Brussels/Zaventem B-1110, Brussels 39, Belgium; dated 20 February 1970, Subject: Proposed Agenda Item for Next Meeting of the Intelligence Procedures Interservice Working Party (NC).


Samet, M. G. Subjective interpretation of the source reliability and information accuracy rating scales, ARI Technical Paper. 1073 (in press).


Rigby, L. V., and Swain, A. D. In-flight target reporting--How much is "a bunch"? *Human Factors*, 1971, 13, 177-182.

APPENDIX

STATISTICS ON PROBABILITY PHRASES FROM PREVIOUS EXPERIMENTS


<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>1</td>
<td>Highly Probable</td>
<td>187</td>
<td>89</td>
<td>90</td>
</tr>
<tr>
<td>2</td>
<td>Very Probable</td>
<td>187</td>
<td>87</td>
<td>89</td>
</tr>
<tr>
<td>3</td>
<td>Very Likely</td>
<td>185</td>
<td>87</td>
<td>90</td>
</tr>
<tr>
<td>4</td>
<td>Quite Likely</td>
<td>188</td>
<td>79</td>
<td>80</td>
</tr>
<tr>
<td>5</td>
<td>Likely</td>
<td>188</td>
<td>72</td>
<td>75</td>
</tr>
<tr>
<td>6</td>
<td>Probable</td>
<td>188</td>
<td>71</td>
<td>75</td>
</tr>
<tr>
<td>7</td>
<td>Fairly Likely</td>
<td>188</td>
<td>66</td>
<td>70</td>
</tr>
<tr>
<td>8</td>
<td>Possible</td>
<td>178</td>
<td>37</td>
<td>49</td>
</tr>
<tr>
<td>9</td>
<td>Fair Chance</td>
<td>188</td>
<td>51</td>
<td>50</td>
</tr>
<tr>
<td>10</td>
<td>Unlikely</td>
<td>178</td>
<td>18</td>
<td>16</td>
</tr>
<tr>
<td>11</td>
<td>Fairly Unlikely</td>
<td>187</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>12</td>
<td>Improbable</td>
<td>187</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>13</td>
<td>Very Unlikely</td>
<td>186</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>14</td>
<td>Quite Unlikely</td>
<td>187</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>15</td>
<td>Highly Improbable</td>
<td>181</td>
<td>6</td>
<td>5</td>
</tr>
</tbody>
</table>

---

aRank obtained in the present study.
bN = 20.