

Investigations of Probabilistic Inference.

**University of Colorado.
Robert M. Hamm, 303-492-2936**

**Annual Interim Report for project on
THE USE OF PROTOCOL ANALYSIS AND PROCESS TRACING
TECHNIQUES TO INVESTIGATE PROBABILISTIC INFERENCE.**

**Contract MDS-903-86-K-0265
September 1, 1986 to September 1, 1988.**

**Army Research Institute
for the Behavioral and Social Sciences.**

Report for period September 1, 1986 to September 1, 1987.

Tech Report # 87-12

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION		1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION/AVAILABILITY OF REPORT	
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE			
4. PERFORMING ORGANIZATION REPORT NUMBER(S)		5. MONITORING ORGANIZATION REPORT NUMBER(S)	
6a. NAME OF PERFORMING ORGANIZATION Institute of Cognitive Science	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION	
6c. ADDRESS (City, State and ZIP Code) University of Colorado, Box 345 Boulder CO 80309-0345		7b. ADDRESS (City, State and ZIP Code)	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Army Research Institute	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER MDA-903-86-K-0265	
8c. ADDRESS (City, State and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600		10. SOURCE OF FUNDING NOS.	
		PROGRAM ELEMENT NO.	PROJECT NO.
11. TITLE (Include Security Classification) Investigations of Probabilistic Inference			
12. PERSONAL AUTHOR(S) Robert M. Hamm			
13a. TYPE OF REPORT Annual Interim	13b. TIME COVERED FROM 9/1/86 TO 9/1/87	14. DATE OF REPORT (Yr., Mo., Day) 9/18/87	15. PAGE COUNT 14
16. SUPPLEMENTARY NOTATION Contracting Officer's Representative was Michael Drillings			
17. COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Inference, probability, probabilistic inference, base rate fallacy.	
FIELD	GROUP		
19. ABSTRACT (Continue on reverse if necessary and identify by block number) Reports first year results of project "The Use of Protocol Analysis and Process Tracing Techniques to investigate Probabilistic Inference." Results include that the most recently presented information is given more attention in situations that require the integration of information about expectancies with uncertain information about what is happening at present. In addition, subjects have difficulty distinguishing two types of conditional probability information, the reliability p(evidence/hypothesis) and the probability that the hypothesis is true if the evidence is observed, p(hypothesis/evidence).			
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐		21. ABSTRACT SECURITY CLASSIFICATION	
22a. NAME OF RESPONSIBLE INDIVIDUAL		22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL

Table of Contents

1. Executive Summary.	1
2. Introduction.	3
3. Studies undertaken.	3
3.1. Questionnaire Study.	4
3.1.1. Basic Results paper.	4
3.1.2. Two Models paper.	5
3.1.3. Complementarity and Resuscitation Paper.	6
3.2. Follow-up Studies.	6
3.2.1. Verbal/Numerical Study.	7
3.2.2. Confusion Study.	7
3.3. Protocol Analysis and Process Tracing Studies.	7
3.3.1. Mathematical problems allow a protocol analysis shortcut.	8
3.3.2. Think Aloud Study 1.	8
3.3.3. Think Aloud Study 2 -- the Confusion Study.	9
3.3.4. Think Aloud Study 3 -- the Insurance Expert Study.	10
4. Potential applications of the research.	10
5. Bibliography.	11

Investigations of Probabilistic Inference.

1. Executive Summary.

Scientific Objectives. Taking correct action in conditions of uncertainty requires the ability to utilize two types of information: statistical base rate information about what is likely to be the case, based on the history of what usually happens, and imperfectly reliable information about what is now the case. Past work on probabilistic inference has demonstrated that when these two types of information conflict, novices tend to neglect the base rate information and to put unwarranted confidence in the information about the present situation, even though this information is unreliable. This research will describe the process by which novices, as well as experts in probabilistic inference and experts in the substance of the problem, combine the two types of information in making probabilistic inferences. The goal is to understand both novice and successful inference processes so that people can be taught strategies for correct inference.

Approach. The approach is to determine the strategies people use on probabilistic inference problems through various types of analysis, including process tracing and protocol analysis. Preliminary work analyzing novice answers has been completed. Work with experts, and protocol analysis with novices, is under way. In the preliminary study, novices were asked to estimate the probability that a hypothesis was true in three probabilistic inference word problems. In each problem, they answered before and after the presentation of each of three types of information -- base rate, evidence, and reliability of evidence.

Findings. A number of hypotheses were evaluated. The typical subject can be described as using strategies that depend on the kind of information that is available, rather than universally applied weighted averaging processes, or normative strategies. Many subjects respond with numbers that are available in the problem presentation. The more recent information has a greater impact. Comparison of production system simulations of the typical responses and the normative responses shows that the neglect of the base rate information is due in part to a misunderstanding of the reliability information. Specifically, subjects do not distinguish between two conditional probabilities, the probability that particular evidence would be seen if a hypothesis were true, and the probability that a particular hypothesis would be true if evidence were seen.

The confusion between reliability $p(E/H)$ and the conditional probability of the hypothesis given the evidence $p(H/E)$ is being investigated in a further study which presents $p(H/E)$ where $p(E/H)$ is usually given, to see whether subjects respond differently. The finding that subjects frequently respond using numbers available in the word problem is being followed up by presenting probabilistic information verbally, or requiring verbal rather than numerical responses, or both. The performance of experts in the field of insurance is being investigated in a think-aloud and process tracing procedure.

Potential applications. Many military operational contexts require the integration of information about expectancies (prior probabilities that a hypothesis will be true) with uncertain information about what is happening at present. If the statistical information is neglected, it could lead to an excessive amount of "false alarms". If, as demonstrated here, the most recent information is given more attention,

Investigations of Probabilistic Inference.

2. Introduction.

Probabilistic reasoning is a basis for action in a wide variety of vital contexts. A decision maker in a combat situation must interpret potentially unreliable intelligence information concerning enemy troop movements. An officer must draw conclusions concerning a new subordinate given stereotypical expectations based on the subordinate's ethnicity, race, or sex, and on impressions derived from brief interactions with the subordinate. It is important to understand how people make probabilistic inferences, what determines their accuracy or inaccuracy, and how their accuracy can be improved. Erroneous methods of interpretation of battlefield intelligence could unnecessarily decrease the probability of victory. Methods of evaluating subordinates that do not take account of both reasonable expectations based on the subordinate's group, and evidence about the individual, could lead to inefficient allocation of manpower resources, as well as resentment and low morale.

Past research has shown that people's use of probabilistic information in reasoning deviates from the uses that are prescribed by the normative methods of probabilistic inference. For example, neglect of base rate information has been shown in a number of probabilistic inference word problems (Bar-Hillel, 1980; Tversky and Kahneman, 1982), where the proper combination of base rate and case information is prescribed by Bayes' Theorem.

The primary goal of this project is to develop an understanding of the variety of strategies that people can use to make probabilistic inferences, so that we can know how people can do this reasoning most accurately. One approach to this goal is to study the performance of people who have been trained in the application of the normative techniques of probabilistic inference, i.e., the mathematics of probability. A distinct approach is to study the strategies used by people who are expert in the area of the word problem. This may reveal usable heuristics, or ways that people structure their information environments and their decisions that allow sensible inferences to be made without resort to the formal normative probabilistic inference procedures.

Discover of accurate heuristic strategies that leaders and decision makers could be taught to use as a mental habit, as part of their automatic interpretation of the world, could lead to accurate performance of probabilistic inference in uncertain situations without reliance on external computer aids, such as those which perform Bayes' Theorem calculations. These aids have had low acceptance in decision making contexts (cf Shortliffe, 1984), partly because of fear-based psychological barriers in potential users, partly because of the practical inconvenience of the requirement of accurately entering the full set of pertinent data in the system, and partly because of the potential for catastrophic results due to minor clerical errors (Hammond, 1981; see Hamm, in press). Methods of probabilistic inference that are well-founded, even if not perfectly accurate, and that can be integrated into decision making practice, may potentially be of great value.

3. Studies undertaken.

This paper reports on studies that have been completed and are in progress addressing the problem of describing people's strategies and processes of making probabilistic inferences. The first body of work to be reviewed is the Questionnaire Study, which presented incomplete probabilistic inference word problems to naive subjects. Its results are presented in a separate report (Hamm, 1987) and in two

situations, e.g., when presented with only the evidence and the base rate.

Three classes of hypothesis were proposed to explain how people answer these word problems and why the answers often neglect the base rate information -- variants of normative probabilistic reasoning, heuristic strategies, and non-normative information integration. Findings included:

1. Many subjects responded with numbers that are available in the problem presentation. Often the use of an available number is normatively correct, which implies that the novices have some understanding about appropriate reasoning in these situations. However, many of the subjects' wrong answers also used numbers available in the word problem. This implies that they may be adopting the simple strategy of answering with whatever numbers are available. It is thus possible that the subjects who answered correctly may have done so just by luck.
2. The more recent information has a greater impact. For example, when the subjects had all three pieces of information, if base rate information was presented most recently, more subjects used it as their answer than if it was presented first and the evidence and the reliability information followed it. This identifies another condition that influences the subjects' likelihood of using or neglecting the base rate information (see Bar-Hillel, 1980).
3. There is no universally applied weighted averaging scheme that accounts for the average response in all conditions. Rather, some form of "contingent strategies" theory is needed to account for the data. Contingent strategies means, broadly, that people will adopt different information processing strategies in different conditions (when given different combinations of information), rather than applying one strategy (weighted averages) in all conditions.
4. The neglect of the base rate information is due in part to a misunderstanding of the reliability information, specifically, a confusion between $p(H/E)$ and $p(E/H)$.

A production system model, that embodies a form of a contingent strategies theory, was produced. This model exactly predicted the most common response made by subjects in each of the possible situations (situations are defined by combinations of available information). This shed additional insight on the "neglect of base rate" -- no rule in the production systems model expressed a process that would be characterized as underweighting base rate. Rather, when reliability, evidence, and base rate information were all present, the rules took the reliability information $p(E/H)$ to be $p(H/E)$, which is the answer the problem asks for, and hence used it as the answer. Base rate was not used, but this was a reasonable response given the interpretation of $p(E/H)$ as $p(H/E)$, rather than the result of a mistaken judgment of the relative relevance of statistical (base rate) and case (evidence) information (Bar-Hillel, 1980).

As just described, this study has identified a major barrier to accurate probabilistic inference, which is that people do not know how to interpret the conditional probabilities in which the reliability information is often couched in probabilistic inference word problems. Training would presumably correct this problem. Even if this barrier were to be surmounted, we still lack knowledge of how to train people to best integrate the statistical and case information (but see Lichtenstein and McGregor, 1984). Yet no progress at all can be possible when subjects confuse $p(H/E)$ and $p(E/H)$.

3.1.2. Two Models paper.

The production system just described represents a model from the information processing or artificial intelligence school of modeling cognition. A distinct approach is to model subjects' behavior as involving intuitive judgment and choice processes. For example, subjects' responses could be produced by a two stage process,

1. an intuitive judgment of the probability of the hypothesis,

3.2.1. Verbal/Numerical Study.

In order to investigate the subjects' use of available numbers as answers, discovered in the Questionnaire Study, a further study is under way (with Donna Hughes) in which verbal phrase expressions of probability are used in addition to numerical expressions of probability. The presentation mode and response mode are varied independently, across subjects, producing four conditions: verbal phrase presentation and verbal phrase response, verbal presentation and numerical response, numerical presentation and verbal response, and numerical presentation and numerical response (as in the previous study). If using available responses is a general tendency, reflecting subjects' superficial processing of the problem, then it should occur in both the verbal-verbal and numerical-numerical conditions. But in the verbal-numerical and numerical-verbal conditions, this easy answer is not possible, and we will see how subjects respond. However, the use of available numbers may be due to subjects' discomfort with numerical probabilities, which may not extend to verbal phrase probabilities. If so, then there should be less use of available numbers in the verbal-verbal condition.

Beyth-Marom (1982) has investigated the use of verbal phrase probabilities on the part of professional economic forecasters, and found them to be very imprecise in comparison with numerical probabilities. She recommended that numerical probabilities be used. However, if it should be found that the tendency to use available probabilities as one's answer occurs more often with numerical probabilities than with verbal phrase probabilities, this would be a reason to qualify her advice (see also Zwick, 1987).

The issue of the confusion between $p(H/E)$ and $p(E/H)$ is also investigated with the four word problems in this study.

The study also investigates the accuracy of subjects inferences when given verbal and numerical presentation of information, and using verbal phrases and numbers for their responses. The findings will be relevant to evaluation and restructuring of many situations in which Army personnel currently communicate degrees of uncertainty using verbal phrases.

3.2.2. Confusion Study.

The hypothesis that subjects confuse $p(H/E)$ and $p(E/H)$ is investigated as part of Think Aloud Study 2, described below.

3.3. Protocol Analysis and Process Tracing Studies.

The analysis of the data of the Questionnaire Study showed that a "contingent strategies" theory is required to describe the subjects' behavior across situations. This conclusion was reached by a general argument: that there is no single process, describable by a single mathematical model of the dependency of the subjects' answers on the available information, that can explain the typical or mean answer in every situation (every possible combination of information). The production system simulation extended this conclusion by modeling the most common answers using simple rules or strategies -- different rules being applied in different situations. But both these analyses are based only on the subjects' answers, not on direct or indirect observation of the processes they use to produce the answers.

It is important to know the psychological processes that produce the different behavior in different situations, i.e., the processes that the production system is simulating with its simple rules. What is the subject thinking of? Is there evidence, other than the fit between data and theory, that subjects are using the type of strategies that the production system model assumes? Two methods are available for this:

counted the number of sentences in which the subject mentioned the base rate, after all the information had been presented. This was on the average 1.4 sentences (13% of sentences) for the Cab problem, 4.7 sentences (33%) for the Doctor problem, and 2.3 sentences (14%) for the Twins problem.

To determine whether there is a relation between the mentioning of base rate and its use in producing the answer, the correlation between the number and proportion of sentences mentioning base rate, and the absolute deviation of the subject's answer from the base rate, was calculated. Findings were: the more sentences the subject said, the lower the answer (closer to the base rate) ($r = -.60$, $p = .001$) and the closer the answer to the base rate ($r = -.55$, $p = .003$); the more sentences that mentioned the base rate, the lower the answer ($r = -.73$, $p = .000$) and the closer the answer to the base rate ($r = -.61$, $p = .001$); the higher the proportion of sentences mentioning the base rate, the lower the deviation of the answer from the base rate ($r = -.13$, $p = .280$) and the lower the answer ($r = -.22$, $p = .156$).

It seems that the primary relation is: the more the subject talked (and presumably, thought) about the problem, the lower the answer (and the closer to the base rate). The effect of the mentioning of the base rate, per se, is secondary, though talking about the base rate seems to bring the answers closer to the base rate. This analysis suggests that the joint use of information processing (protocol) analysis and input/output analysis may be fruitful (see Einhorn, Kleinmuntz, and Kleinmuntz, 1979).

A second question is whether the subjects consider the possibility that the hypothesis might be false (e.g., the Green cab be responsible for the accident), and further, the possibility that the evidence (cab called Blue by witness) might occur if the cab is Green. Only one subject mentioned this idea, on one problem. This points to a blind spot in naive subjects' considerations on probabilistic inference word problems. This is an opportunity for training, and a possible point of contrast between novice and expert behavior.

3.3.3. Think Aloud Study 2 – the Confusion Study.

A second study in the information processing tradition has been designed, with the data almost all collected and tapes transcribed. This study involved both verbal protocol analysis and a form of process tracing. It also contrasts the behavior of two groups of subjects: undergraduates and mathematics graduate students. The substantive focus of the study is the confusion between $p(H/E)$ and $p(E/H)$. It serves as a pilot study for the verbal protocol analysis and process tracing methods.

The subjects are given a questionnaire containing three problems. The order is counterbalanced across subjects. Analyses will be made both of the subjects's answers, and their information processes. The first test of the confusion hypothesis involves a contrast between the first and third problems, one of which contains $p(H/E)$ in the "reliability" paragraph, the other having $p(E/H)$. It will be seen whether subjects' numerical answers depend on this variable, or (as hypothesized) it is all $p(H/E)$ to them. Verbalizations will be analyzed to see whether the subjects' interpretations of the information reveal only the hypothesized confusion, or distinctions are made.

The second test of the confusion hypothesis involves a process tracing method, on the second problem. After the problem is described but before any information (evidence, reliability of evidence, or base rate) is given, the subject will be shown four paragraphs that have blanks in them, and told that the information for the blanks is available to them. The paragraphs convey the type of information (base rate, evidence, reliability $p(E/H)$, and $p(H/E)$) but do not specify it. The subjects are asked to think aloud while

5. Bibliography.

- Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. Acta Psychologica, 44, 211-233.
- Beyth-Marom, Ruth. (1982). How probable is probable? A numerical translation of verbal probability expressions. Journal of Forecasting, 1, 257-269.
- Einhorn, Hillel J., Kleinmuntz, Don N., and Kleinmuntz, Benjamin. (1979). Linear regression and process-tracing models of judgment. Psychological Review, 86, 465-485.
- Ericsson, K. Anders, and Simon, Herbert A. (1985). Protocol Analysis: Verbal Reports as Data. Cambridge, MA.: MIT Press.
- Hamm, Robert M. (1987). Diagnostic inference: People's use of information in incomplete Bayesian word problems. Publication 87-11, Institute of Cognitive Science, University of Colorado, Boulder.
- Hamm, Robert M. (in press). Clinical intuition and clinical analysis: Expertise and the cognitive continuum. In J. Dowie and A. Elstein (Eds.), Professional Judgment. Cambridge: Cambridge University Press.
- Hammond, K. R. (1981). Principles of organization in intuitive and analytical cognition. (Report #231). Boulder: University of Colorado, Center for Research on Judgment and Policy.
- Kahneman, D., and Tversky, A. (1982). Variants of uncertainty. In D. Kahneman, P. Slovic, and A. Tversky, (Eds.), Judgment under uncertainty: Heuristics and biases. Cambridge: Cambridge University Press, pp 509-520.
- Lichtenstein, S., and MacGregor, D. (1984). Structuring as an aid to performance in base-rate problems. Report #84-16, Decision Research, Eugene, Oregon.
- Payne, John. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. Organizational Behavior and Human Performance, 16, 366-387.
- Robinson, Lori B., and Hastie, Reid. (1985). Revision of beliefs when a hypothesis is eliminated from consideration. Journal of Experimental Psychology: Human Perception and Performance, 11, 443-456.
- Schum, D.A., and Martin, A. W. (1980). probabilistic opinion revision on the basis of evidence at trial: A Baconian or a Pascalian process? (Report No. 80-02). Houston, Texas: Rice University.
- Shafer, Glenn. (1976). A mathematical theory of evidence. Princeton, N.J.: Princeton University Press.
- Shortliffe, E. (1984). Coming to terms with the computer. In S. J. Reiser and M. Anbar (Eds.), The machine at the bedside: Strategies for using technology in patient care. Cambridge: Cambridge University Press, pp 235-240.