ARI Research Note 90-73

Investigations of Probabilistic Inference

Robert M. Hamm

for

Contracting Officer's Representative
Michael Drillings

Office of Basic Research
Michael Kaplan, Director

July 1990

United States Army
Research Institute for the Behavioral and Social Sciences

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

EDGAR M. JOHNSON
Technical Director

JON W. BLADES
COL, IN
Commanding

Research accomplished under contract
for the Department of the Army

Institute of Cognitive Science, University of Colorado

Technical review by

Michael Drillings

Accesion For	
NTIS CRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

DTIC COPY INSPECTED

NOTICES

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS --
2a. SECURITY CLASSIFICATION AUTHORITY --	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE --	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) --	5. MONITORING ORGANIZATION REPORT NUMBER(S) ARI Research Note 90-73
6a. NAME OF PERFORMING ORGANIZATION Institute of Cognitive Science 6b. OFFICE SYMBOL (If applicable) --	7a. NAME OF MONITORING ORGANIZATION U.S. Army Research Institute
6c. ADDRESS (City, State, and ZIP Code) University of Colorado, Box 345 Boulder, CO 80309-0345	7b. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences 8b. OFFICE SYMBOL (If applicable) PERI-BR	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER MDA903-86-K-0265
8c. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61102B PROJECT NO. 74F TASK NO. N/A WORK UNIT ACCESSION NO. N/A

11. TITLE (Include Security Classification)

Investigations of Probabilistic Inference

12. PERSONAL AUTHOR(S)

Hamm, Robert M.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Interim	FROM 86/09 TO 87/09	1990, July	16

16. SUPPLEMENTARY NOTATION

Contracting Officer's Representative, Michael Drillings

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Inference; Probabilistic inference; Probability; Base rate fallacy

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." Research indicates 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 Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Michael Drillings	22b. TELEPHONE (Include Area Code) (202) 274-8722 22c. OFFICE SYMBOL PERI-BR

DD Form 1473, JUN 86 Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

INVESTIGATIONS OF PROBABILISTIC INFERENCE

CONTENTS

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 underway. 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 univerally 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 that 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 number of "false alarms." If, as demonstrated here, the most recent information is given more attention, then the flow of information in operational situations should be designed so that base rate information is presented concurrently with or after the current information, so that it is not neglected. If novices have difficulty distinguishing the two types of conditional probability information, then training should be designed to overcome this difficulty.

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 that 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 additional papers summarized below.

The results of the Questionnaire Study raised several questions that are addressed in subsequent studies currently being conducted. The Verbal/Numerical Study is designed to test whether subjects will rely extensively on probabilities that are presented in the word problem, when the probabilities are verbal phrases rather than numbers. The Confusion Study (Think Aloud Study 2) addresses the issue of confusion between the reliability of evidence, p(evidence/hypothesis) or p(E/H), and the conditional probability that the hypothesis is true given the evidence, p(hypothesis/evidence) or p(H/E).

Protocol analysis and process tracing are powerful methods for discovering how people represent their problem situations and use strategies for solving the problems. Think Aloud Study 1 used novice subjects to test the feasibility of discovering whether people paid attention to probabilistic information, and what strategies they used to answer the questions. Think Aloud Study 2 used novice subjects and subjects with experience with probability mathematics (mathematics graduate students) to develop the analysis of think aloud protocols and use it to investigate the issue of the confusion of p(H/E) with p(E/H). It also involved aspects of process tracing methodology. Think Aloud Study 3 will combine process tracing methodology and protocol analysis and will contrast the strategies of novice subjects, mathematics experts, and people expert in the substantive area of the word problem.

3.1. Questionnaire Study.

A questionnaire study in which 265 undergraduate students answered three probabilistic inference problems has been completed. Three papers are based on the data from this study.

3.1.1. Basic Results paper.

Extensive analyses of the results of the Questionnaire Study are reported in the paper "Diagnostic Inference: People's Use of Information in Incomplete Bayesian Word Problems" (Hamm, 1987). The word problems were the Cab problem, used before by Tversky and Kahneman (1982) and others, and two variants, the Doctor problem and the Twins problem. The Cab problem tells subjects that

> "A cab was involved in a hit and run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:
>
> (a) 85% of the cabs in the city are Green and 15% are Blue.
> (b) A witness identified the cab as Blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colors 80% of the time and failed 20% of the time.
>
> What is the probability that the cab involved in the accident was Blue rather than Green?" (Tversky and Kahneman, 1982, pp 156-157).

The procedure in this study differed from this example, however, in that subjects were required to respond with their probability that the named hypothesis is true 4 times during the problem, after the basic situation is described and again after each piece of key information is presented. The three pieces of key information are the evidence (e.g., in the Cab problem, that the witness reported a Blue cab), the reliability of the evidence (that the witness is right 80% of the time) and the base rate (that 15% of the cabs in the city are Blue). The three pieces of information were presented in each possible order, to different subjects. This allows us to study how subjects make probabilistic inferences in a number of 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,

2. a probabilistic choice process which selects one of the available numbers as the answer, as a function of how near it is to the intuitive judgment.

A paper in preparation compares these two models in terms of their assumptions and the ease with which they account for 5 aspects of the data from the Questionnaire Study. It suggests methods for combining the advantages of the two approaches.

3.1.3. Complementarity and Resuscitation Paper.

A paper in preparation (with Jose A. Lucero) addresses two phenomena distinct from the issue of combining statistical and case-based evidence. These are the subjects' understanding of the complement of a probability, and the occurrence of "resuscitations", i.e., judging the probability of a hypothesis to be 0 at one stage, and then to be a non-zero number at a later stage.

Complementarity. The questionnaire asked subjects not only for the probability that a particular hypothesis was true (e.g., that the cab involved in the accident was Blue) but also for the probability of the complementary hypothesis (that the cab was Green). Given the word problem's definition of these events as mutually exclusive and exhaustive, the correct answer to the second question is the probabilistic complement of the first, i.e., p(Green) = 1 - p(Blue). It was found that a high proportion of subjects gave complementary answers. Evidence was sought for subjects' use of variant conceptions of subjective probability, such as one proposed by Schafer (1976) in which someone with little evidence might have very low subjective probability for a hypothesis and for its complement (see Kahneman and Tversky, 1982), so that the probabilities would add up to much less than one. If this theory is correct, then as the subjects get more information, the sum of their probabilities for the mutually exclusive and exhaustive events should approach 1.0. This pattern occurred very rarely among the subjects whose answers were noncomplementary.

Resuscitations. If a Bayesian probability estimator is receiving a stream of information pertinent to one's estimate of the probability of a hypothesis, and if the probability ever hits 0 or 1.0, there is no way that it can return to an intermediate value. Subjects' probabilities for a hypothesis have been observed to be "resuscitated" after hitting 0, and to return from 1.0 (Schum and Martin, 1980; Robinson and Hastie, 1985). Such behavior was observed in this study, as well, though it was infrequent.

The import of our analysis of complementarity and resuscitations is that these are rules of probability that most naive subjects follow. This finding contradicts some pessimistic conclusions reached on the basis of previous research, concerning people's general inability to do any type of probabilistic reasoning. However, the occasional occurrence of noncomplementary estimates and of resuscitations should alert us to the possibility that people use numerical probabilities to mean something other than what a strict interpretation of the numbers would imply (Kahneman and Tversky, 1982).

3.2. Follow-up Studies.

Two studies have been designed to explore findings from the Questionnaire Study. The first deals with the use of available numbers, and the second deals with the confusion between p(H/E) and p(E/H).

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 forcasters, 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:

process tracing and verbal protocol anslysis. Both methods are used in the approach that Simon (1976) characterizes as "information processing", aimed at characterizing cognitive behavior in terms of the sequence of operations undertaken by a mind metaphorized as a machine that processes symbolic information, establishes goals, and uses control strategies to manipulate information in short term and long term memory. Verbal protocol analysis is a method by which a theory of how the mind's information processing capabilities are applied to the problem is constructed A theory is produced that is consistent with the verbal trace of the problem solver, assuming that the contents of the short term memory are verbalized (Ericsson and Simon, 1985). Process tracing attempts to build the same sort of model using behavioral evidence, usually pertaining to the order in which external information is searched (e.g., Payne, 1976).

Both methods are being used in this project, as appropriate.

3.3.1. Mathematical problems allow a protocol analysis shortcut.

In addition to the production system described above, another use of the information processing metaphor has been possible which avoids the use of verbal protocol analysis and process tracing as data collection methods. This takes advantage of a special characteristic of mathematical problems: there are only certain routes by which one can arrive at given mathematical answers. Thus, if we look at the mathematical answers that the subjects produce, we can know (except for ambiguities) what mathematical strategies they used to produce those answers. This analysis assumes, of course, that subjects are indeed directly manipulating the mathematical symbols (rather than making judgments, rounding, using available numbers, or using conventional numbers).

Hamm (1987) applied such an analysis to the answers of the subjects of the Questionnaire study. One hundred and one possible mathematical operations were computed and compared with the subjects' answers in each situation. It was found that very few of the subjects' answers could be unambiguously interpreted as having been produced by the application of mathematical operations to the numbers given in the problem.

This procedure could be used to assess people's approaches to any problem where there is the option of using mathematical operations in its solution. It could reveal the sources of wrong answers -- Are people trying to use a mathematical operation that just happens to be wrong, or are they not even trying mathematical operations at all, but just responding with available numbers or with guesses?

3.3.2. Think Aloud Study 1.

A pilot study (with Edson Sellers) was done to determine the feasibility of coding the transcripts of subjects' verbalizations while solving probabilistic inference word problems. Since this will not be written up elsewhere it is described here. Ten student subjects thought aloud while solving two or three word problems: variants of the Cab problem, the Twins problem, and/or the Doctor problem (see Hamm, 1987). Their answers at each juncture in the problem were transcribed, unitized into sentences, and coded with respect to whether they mentioned the base rate, the probability of the complementary hypothesis p(~H), the reliability p(E/H), the likelihood of seeing the evidence if the complementary hypothesis were true p(E/~H), and others. The identification of these concepts offered few problems.

One explanation of the typical response on probabilistic inference word problems of this type is that the subject is ignoring the base rate. To see whether the verbal protocol data is consistent with this, we 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 mentioning 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 (-.13, $p = .280$) and the lower the answer (-.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 involves 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, 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 if 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 they evaluate the types of information and specify the order in which they want to see the information. This provides evidence concerning subjects' ability to distinguish the concepts of p(H/E) and p(E/H) when shown the two at the same time.

The coding schemes developed for the analysis of verbal protocols in this study, and the process tracing method, will be used in the third Think Aloud Study, with substantive experts.

3.3.4. Think Aloud Study 3 – the Insurance Expert Study.

Problems have been prepared for use by insurance experts. Four problems have been prepared, dealing with the possibility that an automobile insurance client is accident prone, the possibility that a client involved in an automobile accident was drunk, the possibility that a health insurance client has AIDS, and the possibility that a new client had hidden a history of diabetes. These have been constructed so that a sequence of information, in typical form, is received by the main actor in the story. Experts have been consulted to assure that the probabilistic information in the stories is plausible.

These problems will be presented in a manner that involves both the collection of verbal protocols, and the offering of choices so that processing may be traced. Subjects will vary both on whether they have substantive experience with the content of the problems, and on whether they have training in the mathematics of probabilistic inference.

This study represents the culmination of the project, contrasting expert and novice with respect to the types of information processing that they use. The procedures and analyses depend on the results of the previous studies.

4. Potential applications of the research.

The studies being conducted as part of this project have the potential for improving probabilistic inference, and the decision making which depends on it, in two ways. First, the studies already completed have already identified common strategies for making probabilistic inferences. Knowing the strategies used, which often produce wrong answers, helps us know where to start in corrective training. Second, description of the strategies used by real world experts in making probabilistic inferences, coupled with the evaluation of their performance, will help us discover strategies that work. Since these strategies are already in use by experts, we know that they have high acceptance and that people can learn to use them. Still at issue is the question of how much better experts perform than novices.

5. References

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.

Tversky, A., and Kahneman, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic, and A. Tversky, (Eds.), Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press, pp 153-160.

Zwick, Rami (1987). Combining stochastic uncertainty and linguistic inexactness: Theory and experimental evaluation. PhD Dissertation, Department of Psychology, University of North Carolina, Chapel Hill.