INSTITUTE OF SCIENCE AND TECHNOLOGY

The University of Michigan

Progress Report No. 3

15 April 1969 - 15 April 1970

Grant NGL 23-005-171

Ward Edwards

Institute of Science and Technology

Engineering Psychology Laboratory

The University of Michigan

Ann Arbor, Michigan

20 April 1970

Prepared for: National Aeronautics and Space Administration
Ames Research Center
Mail Stop 200-17
Moffett Field, California 94035

Attention: Dr. Edward Huff
Program and Research Office

I N S T I T U T E   O F   S C I E N C E   A N D   T E C H N O L O G Y
The University of Michigan

Progress Report No. 3

INTRODUCTION

We seem to be unable to write progress reports more often than once a year; fortunately, we do keep Ames Laboratory personnel well informed about progress by other means.

The basic scientific facts about this year's work are these: we are beginning at last to see in print the results of previous years' work; we are well on our way toward nailing down the misaggregation theory of human conservatism in probabilistic inference; we are now deeply involved in experiments on multi-level inference systems; and we have found it surprisingly difficult to make the move toward greater emphasis on utility and lesser emphasis on probability that we have been advertising for a year now as "imminent." It is, however, still imminent.

Administratively, the most significant events were receipt of $80,000 on 29 August 1969 to cover the period 1 July 1969 to 30 June 1970, and the receipt of $75,000 on 29 August 1969 as step-funding money to be held till the program terminates. As I understand the rules, that latter money is non-expendable so long as year-to-year funding continues (more precisely, it is spent but replenished from the year-to-year funding), and is used to taper the program off gradually (over a two-year period) when NASA decides not to continue it further.

Finally, a major event was the trip of Dr. Edward Huff, from Ames, and Edwards to MSC in February. That visit was originally intended to explore some applications of the Michigan findings to the launch phase, but its actual result was quite different. Both Dr. Huff and I got very excited about the application of multi-dimensional utility measurement techniques to the problems of selecting, scheduling, and rescheduling (after failure or malfunction) of experiments for AAP and subsequent manned space flights. A draft memorandum on this topic has already been prepared and is now being reviewed at Ames; hopefully, a NASA publication will eventually result. We anticipate that (assuming continued support) we will be able to reorient major portions of our program toward this problem, and thus be able both to make some directly useful inputs to MSC (especially in connection with experiment scheduling and rescheduling) and to get the realistic context that we have needed in which to get going on the abstract multidimensional utility problem.

In the summer of 1969 we added three new graduate student research assistants to the program. We also have enjoyed during the 1969-1970 academic year the services of Dr. Tom Wallsten, a new PhD from the University of Pennsylvania here for a year on a post-doctoral fellowship, at no cost to the grant.

Research in Progress

1. Utilities and personal probabilities as determiners of controller decisions. See publication No. 1.

2. The role of computer displays and controls in PIP. For a description of this project, see Progress Report No. 1, page 5. The results showed that the use of a computer display increased degree of conservatism. That is, subjects who were shown the probabilities of the hypotheses translated from their likelihood ratios for a single datum were more conservative than subjects who had no such display. This result is congruent with a previous finding that probability estimates are more conservative than odds estimates.

The response device also influenced the degree of conservatism. Subjects who recorded their likelihood ratio estimates on logarithmically spaced scales by positioning mechanical lever arms were less conservative than subjects who wrote their likelihood ratios on response forms.

The fact that the displays and controls do have systematic effects upon the performance of the PIP system is important. It suggests the need of more research designed to find out which kinds of apparatus provide the most appropriate interviewing techniques for the task of eliciting likelihood ratios in a PIP system.

3. The effect of the denominator on likelihood ratio estimation in PIP. For a description of this project see Progress Report No. 1, page 5. Data analyses are still continuing on this project. At present we have no conclusions to report.

4. _Estimating the value of information for decisions._ See Publications Section, No. 3.

5. _Action selection with continuous variables._ For a description of this project, see Progress Report No. 1. A draft report has been prepared by Dr. Wendt, and is now being reviewed in this laboratory.

6. _The effect of instructions on conservatism._ For a description of this project see Progress Report No. 1, page 8. Data analyses are now complete. An article describing this research is being written.

7. _Misaggregation vs. misperception._ For a detailed description of the development of this project see Progress Report No. 1, page 10. Data analyses on both the first and second experiments are now complete. The results of the first experiment provided evidence that misaggregation is a cause of conservatism. Posterior odds were very conservative when subjects estimated them directly. However, the corresponding posterior odds were nearly optimal when they were calculated by Bayes's theorem with likelihood ratios estimated by subjects. That is, subjects made veridical estimates of likelihood ratios but they were conservative in estimating posterior odds.

An important criticism of the first experiment was that the posterior odds were nearly always much larger than the corresponding likelihood ratios. For example, after several trials Bayesian posterior odds were on the order of 300,000:1 whereas likelihood ratios for a single datum never exceeded 75:1. Thus the results of the first experiment could be explained as an artifact of a response bias. That is, it may be that subjects are veridical within the range of small numbers but their estimates deteriorate outside of that range. In response to this criticism, the first experiment was repeated by using *d'* (the distance in standard deviation units between the means by two normally distributed populations) as an independent variable, and by manipulating the relative magnitudes of the Bayesian likelihood ratios and posterior odds.

The results are similar to those of the first experiment. The estimated posterior odds are, on the average, more conservative than the corresponding likelihood ratios. This result holds for all three levels of *d'* (1, 1.6, and 2.2) and all levels of cumulative diagnostic value. With respect to veridicality, the estimates of posterior odds were universally conservative. Average likelihood ratios were nearly veridical with the intermediate and high *d'* conditions; but they were about 50% more extreme than the corresponding Bayesian estimates in the condition of low diagnostic value (*d'*=1). In general, these results favor misaggregation as a determiner of conservatism and they argue against the response bias interpretation of the earlier results. Two papers will result from these experiments. One is in draft and being revised; the other hasn't been written yet.

8. *A bibliography of research on behavioral decision processes to 1968*. See Item 5 in the Publications Section.

9. *Nonstationary processes and conservative inference*. See Progress Report No. 2, section 9 for a detailed description of this experiment. It has now been accepted for publication in the *Journal of Experimental Psychology*.

10. Information purchase: the flat EV function. See Progress Report No. 2, page 7, for a detailed description of this experiment. We have spent considerable time during the last year making an unsuccessful attempt to extend the results of this experiment. We have searched for sensible laboratory simulations of real-world situations in which the function relating expected value to number of data purchased is relatively peaked around its maximum. So far we have been successful only when the optimal number of data is quite small. For example, there are realistic situations in which the purchase of a single datum is clearly superior to either the purchase of no data at all or to the purchase of more than a single datum. But we have not been successful in finding a realistic situation where the function is peaked and the optimum number of data should be purchased is moderate or large. These constraints appear to be incompatible.

11. Optional stopping and data diagnosticity. For a description of these two experiments and their results see page 8 of Progress Report No. 2. A report of these experiments has been submitted to the Journal of Experimental Psychology.

12. Conservative revisions about normally distributed populations. See page 9 of Progress Report No. 2 for a description of these experiments and the principal results. These experiments comprised DuCharme's PhD thesis and are now completed. He has prepared an article describing these

experiments which he has submitted to the Journal of Experimental Psychology. It has been accepted pending some specified revisions.

13. Conditional dependence. See Progress Report No. 2, page 11 for a description of this experiment and results. The data analysis of this experiment is now complete and a write-up of the results is in progress.

A second experiment on conditional dependence is in progress. In that experiment a subject observes a sequence of data sampled from one of two populations. A sequence contains two types of data-independent and conditionally dependent. Data are independent if the evaluation of the diagnostic impact of a datum does not depend upon which prior data have occurred. Data are conditionally dependent if the evaluation of a datum is dependent upon the previous data. The degree of dependence is manipulated as an independent variable.

Subjects in one condition estimate cumulative posterior odds after each datum in the sequence, whereas in the other condition subjects make estimates of likelihood ratios associated with each datum in the sequence. Thus, this is another replication of the PIP-POP design, but this time in a situation where data may be conditionally dependent as well as have probabilistic relations to the populations from which they are sampled. Results are just beginning to be collected for this experiment and as yet we have no good feeling for what they will look like.

14. Cascaded inference. An inference must be cascaded when a situation is hierarchically arranged so that a datum does not provide

direct evidence about an upper level hypothesis of interest. Rather, it provides information about a state at an intermediate level which, in turn, must be used as a datum for the higher level hypothesis. See Progress Report No. 2, page 11 for a description of the first experiment on cascaded inference. The major result was that subjects were excessive rather than conservative when revising odds estimates in a task requiring cascading.

A follow-up experiment was designed to test the generality of excessive cascaded inferences. It manipulated diagnosticity by using the following likelihood ratios which increase in diagnostic value: 5:4, 22:14, 23:7, and 4:1. These likelihood ratios were achieved in the tasks involving cascaded inference by combining component likelihood ratios of somewhat greater diagnostic value. The design included four control conditions of noncascaded inference in which the likelihood ratios were equal to each of the cumulative likelihood ratios of the cascaded task.

The results of the second experiment are now analyzed. The mean posterior odds estimated by subjects in the cascaded tasks were more excessive than the mean posterior odds estimated in the corresponding noncascaded tasks. Otherwise, the pattern of results in the two tasks was similar; as data increased in diagnostic value the subjects responded too slowly to the rate of increase. Thus, the primary difference between cascaded and noncascaded tasks is that subjects are more conservative in the noncascaded task.

Why? A third experiment asked this question and it seems to have answered it too. Recall that we are dealing with three levels--data at Level 1, states at Level 2, and hypotheses at Level 3. The results of the third experiment showed that the excessiveness is due to a tendency of subjects to ignore the implications of all except the most likely state at Level 2. They seem to make inferences in the following manner. They observe a datum at Level 1 and, as a result, infer which of the states at Level 2 is most likely to be true. They then decide what posterior odds at Level 3 would be correct if they knew for sure which state were true at Level 2 and, since they're only partially certain of the state at Level 2, they hedge their estimates at Level 3. But this hedging takes into account only the fact that they're not completely sure of which state at Level 2 is true; it fails to consider what would be true at Level 3 if an alternative state at Level 2 were true. The experiment manipulated the Bayesian posterior odds at Level 3 by varying probabilities of other than the most likely state at Level 2. Almost all subjects failed to react to this variation. This result not only went a long way toward explaining the excessive inferences made in cascaded tasks; it also suggested the design of a PIP-like system for cascaded inference. The next experiment tested that system.

A PIP system, like most other man-computer systems in the area of decision making, is based upon the assumption that performance will improve by breaking complicated tasks into smaller components, eliciting quantitative estimates about each component, and then combining those estimates by the appropriate arithmetic. This philosophy led to the following experimental design: subjects observed a lower level datum. In the intuitive condition, subjects directly revised odds estimates about the Level 3 hypotheses as a result of that datum. But in the computer-aided condition, subjects made step-wise inferences. They first assumed that one intermediate state was true and under that assumption revised estimates about the upper level hypotheses. They then assumed that an alternative intermediate state was true and under that assumption revised estimates about the upper level hypotheses. Finally, they estimated the probability that each intermediate level state was true on the basis of the information provided by the lower level datum. These various single-step inferences, either from Level 1 to Level 2 or from Level 2 to Level 3, were then combined by the appropriate arithmetic to yield resultant odds about the upper level hypotheses. As expected on the basis of the earlier experiments on cascaded inference, subjects were excessive when they were unaided in making two-step inferences. However, they were nearly optimal when the computer combined their several single-step inferences into a resultant two stage inference.

We are encouraged by this result. Previous research has shown that the use of a PIP system increases optimality by decreasing conservatism in single-stage inference. This experiment shows that use of a system created by the same philosophy increases optimality in a task calling for cascaded inferences by getting rid of the suboptimal excessive performance.

15. The speed vs. accuracy tradeoff in choice reaction time. For a full description, see Progress Report No. 2, item 15. Two papers reporting these experiments have been prepared; journal submission is imminent.

New Research

16. Reliability of data. As the basic algorithm within an inference system, Bayes's theorem prescribes how odds should be revised as a result of the occurrence of a datum. But what if the data are unreliable so that you do not know which one occurred? Then the problem is to make an inference about a hypothesis when you are uncertain about which datum has occurred. In this case not only are the data fallible with respect to implications about hypotheses, but also there is unreliable information about which datum occurred. You suspect that it was datum A, but it may have been datum B. How do you use this knowledge to revise your posterior probabilities about which hypothesis is true?

We represented this kind of data unreliability in an experiment in the following manner. The experimenter sampled data from a population. On the basis of the sampled data the subject revised his odds estimates about which population was being sampled. But sometimes, after observing a datum, the experimenter informed the subject that he might lie about it; that is, he might say that datum A occurred when it was really datum B. We were interested in the quality of performance as a function of the displayed probability that the experimenter would lie about the datum.

The results are currently being analyzed. One result is that subjects consistently revised their odds estimates more than the amount required by Bayes's theorem. They acted as if they were more sure than they should have been about which datum occurred. There are formal similarities between cascading and data unreliability. This excessiveness suggests psychological similarities as well. Data analysis continues on this experiment.

17. Group vs. individual performance in inference and decision making. Most psychological research in the area of statistical decision theory has focussed on subjects as individuals. In many applied settings, however, decisions are made by committees. Therefore, we conducted an experiment that compared the performance of individuals with the performance of groups.

Subjects performed three tasks--first as individuals, then in four-man groups, and then again individually. In the group situation unanimous decisions were required. There were two action-selection tasks, each with an unambiguous definition of greater risk. The choice dilemma task consisted of twelve life-situation problems. For each, the subject was to advise a hypothetical person on the lowest probability of success for which this person should select a risky alternative consisting of a greater and a lesser valued outcome in preference to an intermediate valued sure-thing outcome. The second action-selection condition was a gambling task in which subjects wagered their own money. The third task involved inference; subjects estimated 80 likelihood ratios, one for each of 80 data. Each datum was sampled from one or two normal populations with equal variances but different means.

In the action selection tasks, groups were on the average riskier than the average individual within that group on his pregroup individual session. The intercorrelations within individuals and within groups between different measures of risk in the two tasks were low. This is interpreted as evidence against a general risk-seeking or risk-averse trait.

A striking conformity effect occurred in the likelihood ratio estimation task. Subject's postgroup individual estimates resembled their group's estimates regardless of the group rule and regardless of the individual's pregroup estimates. 22 of 24 subjects conformed. Subjects were easily swayed from their initial likelihood ratio scale. However, for all individual and group sessions the correlations between the subjects' (or groups') estimates and the Bayesian values were very high. Thus subjects were consistent. They understood the nature of the task. But they did not know the appropriate numbers to use. This study shows that much research needs to be done in order to learn how to train subjects to express their inferences on a meaningful scale. This research comprised Goodman's PhD thesis and is now completed. The thesis is of publishable length and will be submitted to a journal soon.

18. Multidimensional utilities. Progress Report No. 2 promised a shift of emphasis in our research in the direction of utilities. The momentum of our research on inference, however, has made it rather difficult to change direction abruptly. Consequently, we have completed only one experiment on multidimensional utilities but more are planned for next year.

The one experiment on utilities was carried out in the school of Landscape Architecture. The task facing subjects, who were students trained in the area of design, was to make two decisions: which of two lots should be used for building a house and which of two designs should be used for a playground. Considerable background material was prepared for each of the two choices. The purpose of the experiment was to compare aided versus unaided procedures for eliciting utilities for each of the alternatives within each of the two problems. In the unaided condition, the subjects intuitively aggregated utilities across all dimensions without any computer assistance. In the aided condition the experimenter decomposed each alternative (e.g., lot) into elementary dimensions of worth, elicited judgments from subjects about utilities along each dimension, elicited judgments of weights for each dimension, and then combined the judgments.

The following procedure was followed. Each subject was first presented with considerable background material upon the problem. He then responded by using a light pen to interact with a cathode ray tube attached to a computer. He estimated utilities for each consequence on a zero to 100 scale. These were, of course, unaided aggregate estimates.

For the second stage, the experimenter listed each of the major dimensions of utility on the CRT display. The subject estimated the utility of each consequence, again on a zero to 100 scale. After estimating the utilities along each dimension, he judged the relative importance of the dimensions by spreading 100 points among all of the dimensions. The computer then calculated the aggregate utility for each consequence by taking a weighted linear average of the utilities across dimensions. At this time, however, this calculated aggregate utility was not displayed to the subject. At stage 3, the subject again estimated the aggregate utilities for each consequence. These aggregate estimates benefited from the experience of considering, dimension by dimension, each of the events in the decomposition stage of the experiment. Utilities estimated in stage 3 were, like those estimated in stage 1, intuitively aggregated without the aid of a computer, which served merely as a display and response device. At the end of stage 3 the computer displayed to the subject both his intuitively aggregated utilities and the computer aggregated utilities for the consequences. The subject was then invited, at stage 4 of the experiment, to modify either his decomposed, dimension-by-dimension utility estimates, or his intuitively aggregated utility estimates.

The results were surprising. Recall that many experiments in the area of inference have shown that a system discriminates more thoroughly between hypotheses when a computer is used to aggregate estimated likelihood ratios across data. Consequently, our expectation for this experiment was that the procedure of decomposition and computer aggregation would result in greater discrimination between the utilities of different events. This did not happen. Instead, there was close agreement between utilities derived from computer and intuitive aggregation, and the degree of agreement increased as the stages progressed.

The big surprise, of course, was that the two systems produced similar results. We suspect the reason resides in the type of arithmetic used for computer aggregation. The normative model for inference is Bayes's theorem, and it is a multiplicative model. That is, the posterior odds based upon an array of data is equal to the product of the likelihood ratios associated with each individual datum. The algorithm used by the computer for multidimensional utilities, on the other hand, is additive. The machine-aggregated utility is a weighted mean of the utility estimates for each dimension. Thus, for an inferential system the individual likelihood ratios are multiplied together. The product of these likelihood ratios, the output of the system, is driven further and further away from the general range of the individual estimates as data accumulate. Not so for utilities. With the weighted average model used in this experiment, the intuitively aggregated utility was of the same magnitude as the utilities estimated for each dimension. In fact, it was a measure of central tendency. Thus we expect that the result of this experiment is essentially correct; intuitively aggregated utilities will <u>not</u> show a systematic bias compared to machine-aggregated utilities in the way that intuitively aggregated posterior odds are systematically biased (conservative) when compared with machine-aggregated odds.

This experiment showed agreement between intuitive and computer aggregated utilities. The serious problem is that, when there was a difference between the two systems, we have no way of knowing which system gave the best output. In our last proposal, we outlined an experiment with known organizational utilities so that it would be possible to have an objective criterion against which to evaluate the results. For a variety of reasons we did not go in that direction in this experiment. However, the difficulty of not being able to determine optimum performance confirms our original expectation that organizational utilities are an important component of research in this area. They will certainly be used in our next experiment on multidimensional utilities.

19. Multiattribute utilities as tools for selection, scheduling, and rescheduling of experiments for manned space flights. This topic has so far been strictly conceptual, having produced only a memorandum from Edwards to Huff. That memorandum essentially makes three points: that selection, scheduling, and rescheduling are all parts of the same problem and all depend on highly subjective value judgments; that a technique known as weighted linear averages (see preceding item) permits an approach to disaggregating these value judgments; and that by so doing it permits the expertise of many different experts to be brought to bear, each working within the topic he knows best.

We hope to be able to report considerable progress, both conceptual and empirical, on this topic in the next progress report.

Publications

1. Edwards, W. Controller decisions in manned space flight. In *Applications of Research on Human Decision Making*, Proceedings of a symposium on Application of Research on Human Decision Making, 1968, Washington, D. C.: NASA Scientific and Technical Information Division, NASA-SP-209, 1970, 93-106. For a summary of the study, see Progress Report No. 1. The published version covers more data than were included in the original speech.

2. Edwards, W., Phillips, L. D., Hays, W. L., & Goodman, B. C. Probabilistic information processing systems: Design and evaluation. *IEEE Trans. Syst. Sci. Cybernetics*, 1968, 3, 248-265. A Probabilistic Information Processing System (PIP) uses men and machines in a novel way to perform diagnostic information processing. Men estimate likelihood ratios for each datum and each pair of hypotheses under consideration (or a sufficient subset of these pairs). A computer aggregates these estimates by means of Bayes's theorem of probability theory into a posterior distribution that reflects the impact of all available data on all hypotheses being considered. Such a system circumvents human conservatism in information processing, the inability of men to aggregate information in such a way as to modify their opinions as much as the available data justify. It also fragments the job of evaluating diagnostic information into small, separable tasks. The posterior distributions that are a PIP's output may be used as a guide to human decision making, or may

be combined with a payoff matrix in order to make decisions by means of the principle of maximizing expected value.

A large simulation-type experiment compared PIP with three other information processing systems in a simulated strategic war setting of the 1970's. The difference between PIP and its competitors was that in PIP the information was aggregated by computer, while in the other three systems, the operators aggregated the information in their heads. PIP processed the information dramatically more efficiently than did any competitor. Data that would lead PIP to give 99:1 odds in favor of a hypothesis led the next best system to give 4 1/2:1 odds.

An auxiliary experiment showed that if PIP operators are allowed to know the current state of the system's opinions about the hypotheses it is considering, they perform less effectively than if they do not have this information.

This paper reports work done before the NASA program began; only its preparation was supported by NASA.

3. Wendt, Dirk. Value of information for decisions. Journal of Mathematical Psychology, 1969, 6, 3, 430-443. Information that will reduce the risk of a decision may be costly in time, effort, or money. The maximum amount that should be invested in the information, its fair cost, depends upon prior probabilities of the hypotheses, payoffs, and the diagnosticity of the data source. These are the independent variables of this experiment. Ss estimated the fair costs by means of the Marschak bidding procedure. The Ss' bids changed in the direction appropriate

to each of the three independent variables, but not enough to be optimal.

4. Edwards, W. Man-machine systems for policy mediation and intellectual control. Talk given at the Fourth Annual NASA-University Conference on Manual Control, Ann Arbor, Michigan, March 22, 1968. No further publication of this speech is planned, though many of its ideas are embodied in a theoretical paper now at the gestation stage.

5. Edwards, W. *A bibliography of research on behavioral decision processes to 1968*. Human Performance Center Memorandum Report No. 7, January 1969.

6. Swensson, Richard G. *The elusive tradeoff: Speed versus accuracy in choice reaction tasks with continuous cost for time*. Human Performance Center Technical Report No. 13, University of Michigan, December 1968.

7. Chinnis, James O., & Peterson, C. R. Nonstationary processes and conservative inference. *Journal of Experimental Psychology*, in press. This experiment tested the hypothesis that people are conservative processors of fallible information because they treat stationary data generating processes as if they were nonstationary, i.e., subject to change. The task included a probabilistic relation between data and the process generating the data and also a possibility that changes could occur in that process. Performance on this task was compared with performance on a similar, but stationary task. The *Ss* behaved differently in the two situations,

appropriately assuming a non-zero probability of change only in the nonstationary task. In addition, the pattern of conservatism in the two tasks requires rejection of the hypothesis that conservatism is due to inappropriate assumptions of nonstationarity.

8. DuCharme, W. M. A response bias explanation of conservative human inference. Human Performance Center Technical Report No. 19, December 1969. Conservative human inference has been attributed to misperception or misaggregation of data, but it may be caused by response biases. In the present experiments subjects revised odds estimates about which of two normal distribution data generators were being sampled. An analysis of special sequences and a plot of revised odds against theoretical odds in Experiment I showed a bias in subjects' response functions. They revised odds optimally only over a range of ± 1.0 log odds. When the experimenter set different levels of prior odds, subjects shifted their response functions so that the optimal range centered around the set prior odds. A second experiment showed that the biased functions remained invariate over changes in data generator familiarity and diagnosticity. Subjects were biased over either cumulative evidence impact or the number system, but within their optimal range they neither misaggregated nor misperceived data.

9. Goodman, B. C. Risky decisions by individuals and groups. Unpublished PhD thesis, University of Michigan, 1970. This study investigates the shifts between individual and group performance in two action selection tasks (a choice dilemma task in which subjects

equate a risky option with a sure thing and a gambling task in which subjects wager their own money) and in one Bayesian diagnosis task (likelihood ratio estimation). 27 male subjects performed each task alone. Then 24 of these subjects were formed into 6 four-man leaderless groups and repeated each task. Three subjects, serving as individual controls, performed each task alone a second time. Finally, all 27 subjects repeated each task again alone.

The group decisions in the choice dilemma task reproduced previously found patterns of shifts (compared with mean pregroup performance) toward the risky option or toward the sure thing. In the gambling task groups tended to prefer higher variance gambles than did the average group member on his pregroup performance. A striking conformity effect occurred in the likelihood ratio estimation task: the estimates of 22 of the 24 test subjects more closely resembled their group's values than their own pregroup estimates. However, no conclusion can be drawn about whether groups or individuals make more extreme likelihood ratio estimates.

Both group and individual correlations between measures of performance in all three tasks were low. Thus proclivity for a risky option in the choice dilemma, preference for higher variance in gambling, and tendency to extreme likelihood ratio estimates seem to be unrelated.

10. Snapper, K. J., & Fryback, D. G. Inferences based on unreliable reports. Submitted to the Journal of Experimental Psychology. Inferences may be based on direct observation of events, or on reports from indirect sources about the occurrence of events. A direct observation

will be more diagnostic than a report if the source of the report is not completely reliable. Previous studies have investigated Ss' inferences based on either directly observed events or completely reliable reports. This study investigated Ss' inferences based on partially reliable reports. Ss responded to reduced report reliability by using a formally inappropriate rule which led to overestimation of the diagnostic impact of a report.

PERSONNEL

Name	Title	Fraction-of Full Time
Edwards, Ward	Research Psychologist	.45
Peterson, Cameron	Research Psychologist	.40
Wallsten, Thomas	Post-Doctoral Fellow	.75 (No cost to NASA)
Fryback, Dennis	Research Assistant	.25
Goodman, Barbara	Research Assistant	.50
Jensen, Floyd	Research Assistant	.25
Saltzman, Mark	Research Assistant	.25
Snapper, Kurt	Research Assistant	.50
Vernon, David	Research Assistant	.25
Wheeler, Gloria	Research Assistant	.25
Bender, Linda	Secretary	.50
Johnson, Annette	Secretary	.50
Hastings, William	Programmer	.30