Investigation and Interpretation of an Event with Multiple Causes

Stephen A. Ginpil
INVESTIGATION AND INTERPRETATION
OF AN EVENT WITH MULTIPLE CAUSES

STEPHEN A. GINPIL

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN PSYCHOLOGY

UNIVERSITY OF RHODE ISLAND
1982
Previous research suggested that subjects tend to investigate a potentially multiply caused event by initially exploring the most likely cause. Such search was said to be "truncated" in that it did not go on to explore additional causes if the involvement of the initially hypothesized cause was demonstrated. It was suggested that this search strategy led to subjects' encountering a narrow range of information which would not allow them to adequately understand such an event.

The present study allowed subjects to investigate three hypothesized causes of an event by sequentially selecting three separate pieces of information. Subjects could choose to investigate the same cause or a different one at each question selection. Perceived likelihood of all hypothesized causes was measured before and after each new piece of information was encountered. After all information was selected, additional measures were taken, two of which called for the subject to interpret the information received.

The information available was confirmatory with respect to all three causes, although the nature of the task only allowed subjects to be certain of the involvement of one. Depending on the three pieces of information selected, subjects' feedback ranged between solid proof of one cause's involvement with no information about others, to proof of one cause with information making other cause likely, but not certain.

Results indicated that there was a tendency for subjects to truncate search. In addition, most subjects, regardless of the scope of
information encountered, agreed with a statement which implied that only a single cause was operating. Subjects also felt that they completely understood why the hypothesized event occurred, even though it was impossible to have thoroughly explored all three causes.

These results are seen to be the result of an initial set of "sufficiency" assumptions. Specifically, it is suggested that subjects seek only to account for the event and encountered information, and to do so as parsimoniously as possible.

It is argued that the appropriateness of this strategy largely depends upon the cost of an incomplete understanding of the event.
# Table of Contents

A. Introduction .......................... page 1
B. Method ................................. page 17
   Subjects ................................ page 19
   Procedure ................................ page 25
   Design .................................. page 27
C. Results ................................. page 28
   Introduction ............................ page 41
   Confirmatory Search Patterns ......... page 57
   Truncated Search ........................ page 70
   Evaluation of Alternatives .......... page 74
   Opinion Variable ...................... page 79
   Recommend Variable ................... page 87
   Relationships among Variables ..... page 92
D. Discussion ............................. page 95
   Serial Search .......................... page 104
   Truncated Search ...................... page 107
   Scope of Search ........................ page 110
   Effects of Scope of Search .......... page 116
   Opinion Variable ...................... page 123
   Recommend Variable ................... page 128
E. References ............................. page 128
ACKNOWLEDGEMENTS

I would like to express my appreciation to Charles Collyer for his concerned, sympathetic, and thorough supervision. Thanks are also due to Jerry Cohen, Jan Kulberg, Wayne Velicer and George Willis for their comments and suggestions. I wish also to thank Chuck Folkers, Sandy Neill, John Norcross, and Joe Rossi for their help during the administration of this study, and Julie Lee for typing several pages at short notice.

The greatest debt of gratitude is owed to my wife, Lynnette, who provided me with assistance and encouragement at every stage in the production of this dissertation.
<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Frequencies of Various Question Selections by Booklet Form, Question Selection Stage, and Hypothesis Rank.</td>
<td>31</td>
</tr>
<tr>
<td>2. Estimated Odds in Favor of Choosing Questions A, B, &amp; C by Hypothesis Rank and Question Selection Stage.</td>
<td>34</td>
</tr>
<tr>
<td>3. Frequencies of Various Question Selections at Each Question Selection Stage by Booklet Form.</td>
<td>39</td>
</tr>
<tr>
<td>4. Frequencies of Hypotheses Rated Certain by Third Question Selected and Number of Hypothesis Explored for Subjects who Explored the Same Hypothesis at the Second and Third Stages.</td>
<td>44</td>
</tr>
<tr>
<td>5. Frequencies of Hypothesis Specific Question Selections for Some Hypotheses Containing Certain Ratings at Various Rating Stages.</td>
<td>46</td>
</tr>
<tr>
<td>6. Means, Standard Deviations and Frequencies of Ratings of Usefulness of Additional Hypothesis Specific Information.</td>
<td>48</td>
</tr>
<tr>
<td>7. Frequencies of Usefulness Variable Matching Responses for Subjects with a Single Certain Hypothesis by Booklet Form and Hypothesis.</td>
<td>52</td>
</tr>
<tr>
<td>8. Frequencies of Matching Usefulness Variable Responses by Number of Hypotheses Explored.</td>
<td>54</td>
</tr>
<tr>
<td>9. Frequencies of Matching Usefulness Variable Responses by Similarity of Last Two Hypotheses Explored.</td>
<td>54</td>
</tr>
<tr>
<td>10. Frequencies of Number of Hypotheses Rated Certain by Number of Hypotheses Explored.</td>
<td>56</td>
</tr>
<tr>
<td>11. Frequencies of Number of Hypotheses Rated Certain by Similarity of Last Two Questions Selected.</td>
<td>56</td>
</tr>
<tr>
<td>12. Frequencies of Various Combinations of Changes in Ratings of Alternatives by Booklet form and Rating Stage Transitions.</td>
<td>59</td>
</tr>
<tr>
<td>13. Frequencies of Number of Alternatives Rated Impossible by Booklet Form, Rating-Stage and Hypothesis Confirmed.</td>
<td>66</td>
</tr>
<tr>
<td>14. Estimated Odds in Favor of No Alternatives Rated Impossible by Rating Stage, and Hypothesis Confirmed.</td>
<td>66</td>
</tr>
<tr>
<td>15. Frequencies of Alternative Hypotheses Rated Impossible by Booklet Form and Number of Hypotheses Explored.</td>
<td>69</td>
</tr>
</tbody>
</table>
Table | Page
---|---
16. Frequencies of Opinion Variable Responses by Final Question Selected, Booklet Form, and Number of Hypotheses Explored. | 71
17. Frequencies of Recommend Variable Responses by Final Hypothesis Confirmed, Booklet Form, and Number of Hypotheses Explored. | 77
18. Frequencies of Opinion Variable Responses by Recommend and Match Variables and Number of Alternatives Crossed Out at the Final Stage. | 82
19. Frequencies of Recommend Variable Responses by Opinion Variable Responses and Number of Hypotheses Rated Certain at the Final Stage. | 85

Appendices | 133
---|---
A. Log-Linear Analysis
B. Booklet Form
Introduction

It has long been accepted that rationalistic methods of arriving at conclusions are easily distorted so as to support the prejudices of the individual. More recently, it has become clear that even empirical criteria do not guarantee accuracy. Data are often selectively perceived or recalled if they conform to the expectations of the subject. Thus, the more data the subject collects, the easier it becomes for him to discover "illusory correlations" i.e., to perceive nonexistent relationships in the data (Smedslund, 1963; Chapman & Chapman, 1967). In addition, even scrupulously recorded data are often accepted or rejected according to inconsistent ("double") standards (e.g., Crary, 1966; Markus, 1977; Shrauger & Lund, 1975). Results which are pleasing may be readily accepted, whereas discordant findings may be dismissed as being the work of extraneous variables. Apparently it is always possible to maintain a belief in a discredited hypothesis ceteris paribus (Lakatos, 1970). Psychological research demonstrates this trend in its reluctance to embrace experimental findings which fail to reject the null hypothesis (Greenwald, 1975), so much so that McGuire (1973) has commented that many experimenters act as though their job was that of "demonstrating that an obviously true hypothesis is correct" (p. 449).
These errors may be attributable to the fact that data collection is usually guided by a particular hypothesis (Elstein, Shulman, & Sprafka, 1978) and may involve a commitment to that hypothesis. Thus the basis of these distortions of the empirical method is the presence of pre-existing bias, and is therefore within the long tradition in psychology of ascribing error to motivational forces (e.g. Freud, 1917).

Current studies, however, have tended to liken human thought processes to those of the computer (Newell & Simon, 1972), and to favour mechanical, rather than clinical, explanations of error (e.g. Norman, 1979). While it is clear that in the applied setting, strict impartiality will seldom be found (e.g. Mitroff, 1974), the conclusion of recent cognitive theorists is that even in the absence of motivational bias, human thought does not necessarily follow the prescriptions of logic (e.g. Kunreuther & Slovic, 1978). It has been suggested that the origin of errors of reasoning lies in the fact that human problem solving makes use of generalized, "rule of thumb" strategies rather than precise algorithms (Tversky & Kahneman, 1973; Tversky & Kahneman, 1974). These simplified procedures, or "heuristics", although normally extremely efficient under certain conditions are said to result in significant inaccuracy.

Within the hypothesis-testing literature, one such heuristic
has been identified, usually referred to as "confirmatory bias" or "confirmatory hypothesis testing" (Doherty, Mynatt, Tweney, & Schiavo, 1979; Johnson-Liard & Wason, 1977; Mynatt, Doherty & Tweney, 1977; Snyder & Swann, 1978; Wason, 1968). Although the definition of this heuristic has varied somewhat, it is perhaps most clearly stated as a strategy of information search, whereby hypotheses are tested by observing whether or not events predicted by the hypothesis in question do, in fact, occur. This may be contrasted with a strategy of searching for events which are incompatible with the hypothesis. One consequence of confirmatory search is that only a limited amount of hypothesis-specific information will be collected. Also, implicit in this definition is that search terminates after some number of confirmations of the hypothesis have occurred, i.e., when it is considered to be "proved."

In almost every instance where confirmation bias has been reported, emphasis has been placed on its inappropriateness. For example, a specific weakness of this approach is that it may tend to be self-fulfilling when the information is gathered by interacting with another person. This is because the person being questioned is likely to alter his responses in line with the confirmatory expectations of the questioner (Snyder & Swann, 1978).

More generally, a confirmatory approach to hypothesis
testing has been objected to on logical grounds. Popper (1959) noted that although inconsistent results clearly indicate that a hypothesis is false, results which are in accord with expectations, do not offer similarly conclusive proof that the hypothesis is correct. It is argued that since any number of successful confirmations cannot unequivocally show a hypothesis to be true, experimenters should, instead, attempt to falsify their hypotheses.

This criticism of confirmatory approaches, however, has itself been criticized on philosophical grounds (e.g. Quine, 1953; Feyerabend, 1975). It is even less persuasive on a practical level where the aim is to be able to make reasonably likely, rather than certain, conclusions based on some limited amount of investigation. In a Bayesian sense, each successful prediction increases the probability that the hypothesis is true. Although one prediction may be of only trivial importance, successive predictions of events, each of which has a low baseline probability of occurring, may provide a very good approximation to certainty. In fact, the effectiveness of the falsification strategy has been shown to be inferior to naturally occurring confirmatory approaches (Mynatt, Doherty, & Tweney, 1978; Tweney, Doherty, Worner, Pliske & Mynatt, 1980). Thus, confirmatory hypothesis testing, like other heuristics, is probably a generally efficient strategy.
There are, however, certain situations in which an incorrect hypothesis might receive successive confirmations. For example, this would be likely in a limited testing situation, if the incorrect and true hypotheses both implied the same sort of things, as in the case of two diseases with similar symptomatology.

Even in the case of an unlimited search (i.e., when any number of predictions may be investigated), repeated misleading confirmations will still occur if the tested (incorrect) hypothesis is a subset of the true one. An example of this is provided by Wason (1968) who instructed subjects to test their hypotheses about a rule governing the generation of a sequence of numbers (e.g., 2-4-6). The correct rule, ("any three numbers in ascending order"), was very general; however most subjects selected for testing a hypothesis which was some subset of this (e.g., "three positive numbers with the second number being twice the first, and the third number three times the first"). It should be noted that every number sequence generated by such a subset hypothesis, will be confirmed (i.e., will receive feedback that it is consistent with the true rule).

A similar relationship often occurs in those situations where the event to be explained has a number of potential causes which are not incompatible with one another. If a number of causes were simultaneously operating, then an
hypothesis about the involvement of any one of them would again be a subset of the correct hypothesis, and thus subject to repeated confirmation.

The confirmation of subset hypotheses is not, in itself, inappropriate, since it implies only the possible involvement of the hypothesized factor. Error results only if the assumption is made that the tested hypothesis fully characterizes the true one. In behavioral terms, this corresponds to a termination of information search after one hypothesis has been extensively supported. It should be emphasized that this termination would result not from an inability to generate appropriate alternatives (e.g. Pitz, Sachs & Heerboth, 1980), but rather because available alternatives were ignored.

There are a number of studies in other fields that seem to suggest that subjects do not consider alternatives if the one under consideration seems tenable. This seems to be a factor in explaining the general tendency to be overconfident about the correctness of current knowledge (Koriat, Lichtenstein, & Fischhoff, 1980), and to see the occurrence of known past events as having been easily predicable (Slovic & Fischhoff, 1977).

It is also analogous to what has been termed the "win-stay, lose-shift" strategy in concept identification
studies (Bruner, Goodnow, & Austin, 1956; Levine, 1966; Wickens & Millward, 1971). In these studies, subjects retained a hypothesis under consideration as long as it led to correct predictions, but adopted a new one if it did not. Similarly, Gettys and Fisher (1979) found that the addition of a new hypothesis to a pre-existing set depended upon it being at least half as plausible as the most plausible hypothesis currently in the set. Presumably then, subjects with a highly plausible hypothesis would be less inclined to add new ones.

None of these studies are directly relevant to a criticism of confirmatory search, however, since they do not deal with the situation where the hypothesis under consideration is both subject to repeated confirmation (or otherwise made to appear certain), and yet potentially incorrect. One study that does meet these criteria is Wason's (1968) "2-4-6" task. Subject were told that they could test their hypotheses as long as they wished, but that they were not to announce their conclusion until they were certain that they were correct. The general finding was that this announcement (and thus certainty), followed from repeated confirmation of a single hypothesis. This then, suggests the weakness of confirmatory search in the detection of extremely general causes.

The other case in which an incomplete hypothesis will be
repeatedly confirmed, that of the multiply caused event, has not been well studied, but there are suggestions that the same premature elimination of other hypotheses also occurs. This issue was examined most directly by Shaklee and Fischhoff (1982). Subjects were told of a number of events, and given three causes of each event to consider. Two of each of the three causes were designated as "potential" causes, whereas the third was labelled a "known" cause (i.e., definitely involved). Subjects were instructed to choose from among the three questions (one of which provided information about each cause) that question that they would most like answered in trying to explain each event. Results indicated that most subjects (across all events) picked the "known" cause to inquire about significantly more often than any of the other "potential" causes. This was interpreted as an unwillingness to explore additional causes given the certainty of one—a process referred to in the paper as "truncated search."

These results, however, are open to an alternative interpretation. Presumably, subjects who truncated their search did so because they felt that the potential causes did not require investigation. If so, such a subject would have nothing to gain from asking any question, since those relating to other causes would be deemed irrelevant whereas information about the "known" cause would be seen as redundant. Thus, it could be argued that, if subjects were
forced to choose, all questions should have an equal chance of selection, since they would all be considered to be equally uninformative.

On the other hand, simply being told that a factor was a "known" cause may not have generated certainty about the involvement of that factor in the minds of the subjects. If this were the case, the decision to ask about the "known" cause may have been an attempt to personally validate this conclusion, and as such, would represent a propensity to clarify the role of the most likely hypothesis as a first step, a strategy referred to by Shaklee and Fischhoff as "serial search." From this point of view, the question of whether there would have been further exploration was not answered.

A pilot study was conducted by the author to further examine this issue. Subjects were presented with a scenario which contained information about a man getting to work each morning, and were asked to generate factors that might reasonably be expected to cause him to be late. Each factor was assigned a probability value according to its perceived likelihood of occurring. The scenario was constructed in such a way that one particular hypothesis was mentioned by almost all subjects. Subjects were then told that the man in the scenario actually was late and information was then provided in stages, each stage presenting evidence which further supported the involvement of the salient hypothesis but was
largely irrelevant to all others. Subjects assigned new probabilities to each hypothesis at every stage, and were also given the opportunity to add new hypotheses or eliminate any that, in the light of current information, no longer seemed worthy of further consideration. The final piece of information implied the certain involvement of the salient hypothesis. Since hypotheses were generally independent of one another (e.g. "alarm fails to go off," "car does not work properly"), no hypotheses should have been eliminated since they were all initially thought to be reasonably likely, and should have remained so. On the other hand, if a truncated search was being employed, the subject should have, by the final stage, lost interest in all but the then certain hypothesis. Results indicated that very few subjects retained all their initial hypotheses, and that 49.5% of the sample had no alternatives left after the final stage. This figure is a conservative one in that some of the additional hypotheses were very close in meaning to the targeted one, and thus at least partially supported by the sequential information. If such overlapping hypotheses were not counted, then the percentage of subjects having no alternatives rose to 57.8%. Furthermore, not all subjects perceived the targeted hypothesis as certain (i.e., assigned it a value of 1.00) at the final stage. If these subjects were eliminated from the sample then the figures would be 62.7% and 74.5% respectively for the two definitions. To some extent, the converse relationship was also observed,
i.e., subjects whose most probable hypothesis was relatively low were somewhat more likely to add new hypotheses to the set. This finding is similar to that reported by Gettys and Fisher (1979).

This investigation, however, involved a number of design weaknesses. For example, by controlling the information that the subject received, there was no opportunity for the subject to display a non-confirmatory search pattern. In other words, the subject was not free to choose information that might have falsified his hypothesis. It is also possible that this design gave rise to demand characteristics which suggested that alternative hypotheses should be ignored. There are also several problems which stem from the fact that subjects generated their own hypotheses. Those with larger numbers of initial hypotheses may have been more likely to eliminate some of them but less likely to eliminate all of them. Also there was no control over the specificity of hypotheses. Thus, one subject may have hypothesized "flat tire" as a cause of someone being late, whereas another may have simply been testing for "transportation difficulties." Bartlett (1958) suggests that extended problem solving activities are avoided by redefining open, general questions (e.g. "what is wrong with this patient?") to more specific formulations (e.g. "is anemia present?"). The latter form of the question provides a well-defined stopping point for the investigation, and
thus would be expected to lead to earlier termination of search.

Another difficulty is that neither the pilot study nor the work of Shaklee and Fischhoff gives any indication why alternative hypotheses are overlooked. Shaklee and Fischhoff state that the significance of truncated search is that the investigator who uses it "...risks overlooking other potent influences on the event of concern" and "...will judge cause-effect relationships on the basis of a narrow subset of the potentially relevant information" (p. 15). This seems to imply that it is the absence of complete information per se which leads to difficulties. However, limited information per se can only lead to difficulties if the investigator chooses to judge it as adequate. Conversely, even complete information may not lead to correct solutions if certain aspects of it are ignored. Clearly there must be some pre-existing reason why a truncated search is employed and this, rather than the resultant narrow scope of search, must be considered the more fundamental cause.

Two processes in particular suggest themselves as likely explanations. It is possible that subjects were making the assumption that only one cause could be operating. This would seem to explain the logical fallacy of "affirming the consequent" (e.g. Copi, 1978) in which a person given the premises "if P, then Q" and "Q, " incorrectly concludes "
This deduction would be correct only if the first premise was "if and only if P, then Q," i.e., there can be only one cause of Q. This assumption would also be consistent with the widespread and often mistaken tendency, in argument, for people to treat the other person's point of view as contrary, i.e., to assume that the simultaneous validity of both arguments is impossible.

This "prejudice that a phenomenon can not have more than one cause" was mentioned (as a criticism of Bacon) by John Stuart Mill (1843, 1974) and has in recent works gone under the labels of "misguided parsimony" (Kanouse, 1972) and the "hydraulic" model of causation (Nisbett & Ross, 1980). Nisbett & Ross suggest that studies on perceived intrinsic versus extrinsic motivation (e.g. Lepper, Greene, & Nisbett, 1973) demonstrate that subjects see these two types of causes as being mutually exclusive.

It has also been noted that there is a tendency to formulate theoretical hypotheses in overly simple terms, not taking into account, among other other things, the possibility that more than one cause may be operating (Petrinovich, 1979).

Stronger experimental support for the employment of a single cause assumption can be seen in the work of Elstein et al (1978). This study described the problem solving activities of physicians attempting to arrive at correct diagnoses. In
one case, the correct diagnosis involved two disease processes. Most doctors, in the course of their investigation, happened upon the symptoms of both illnesses. The major diagnostic error found, however, was termed "overinterpretation," defined as "treating noncontributory cues as relevant to a particular hypothesis" (p. 107). Thus, rather than hypothesize about the existence of a second disease process, a single disease was said to account for all the data collected. It could be argued that the phenomenon which led to "overinterpretation" errors at the integration stage, might also, in a differently structured task, cause a premature termination of information search.

Another way of explaining why subjects might choose to ignore additional causes would be to assume that they were adopting a "satisficing" mode of operation (Simon, 1961). Simon argued that people do not strive for optimal or maximizing solutions, but rather for ones which are "good enough." Simon (1981) suggests that subjects go through several "generate-test" cycles in which alternatives are generated and then tested against various satisficing requirements. Testing is said to terminate when an alternative is able to meet these requirements. Thus "...the expected length of search for an alternative meeting specified standards of acceptability, depends on how high the standards are set, but it depends hardly at all on the total size of the universe to be searched. The time required
for a search through a haystack for a needle sharp enough to sew with, depends on the density of sharp needles, but not on the total size of the stack" (Simon, 1981, p. 140).

Simon's conception of satisficing seems to differ from other types of heuristics in that it is assumed that this strategy is resorted to consciously because of the impossibility of finding an optimal solution.

Within the context of multiple causation, an optimal solution would clearly be one which accurately identifies all involved factors. A satisficing solution however, is likely to be that which, in itself, would be "good enough" to account for the resultant effect, i.e., a single sufficient factor.

In conclusion then, the literature suggests that not all causes of a multiply caused event will be explored, and it is reasonable to assume that some sort of single cause or sufficiency heuristic determines this tendency to truncate search. If so, the judged adequacy of causal information should also reflect this heuristic, and should not be strongly influenced by the amount of information received.

In view of this, the following study attempted to achieve four objectives. The first objective was to describe search tendencies, and to see whether the truncated search
phenomenon could be replicated. The second was to see how adequate subjects would judge the information acquired during the search to be. The third intention was to examine these adequacy judgements as a function of subjects' scope of search (i.e., how many hypotheses had been investigated). Lastly, since it was expected that subjects would pay inadequate attention to alternative hypotheses, it was hoped that some light could be shed on the reasons for this behavior.
Method

Subjects

A total of 355 undergraduates at the University of Rhode Island participated in the two parts of this study. All subjects volunteered so as to obtain extra credit in an introductory psychology course. The first portion of the study was based upon the responses of 147 subjects, while the second employed the responses of 190 subjects. Results from nine subjects in each section were eliminated due to inconsistencies in their responses which suggested that they had not properly understood the instructions. Both samples were predominantly female (100 out of 147, and 132 out of 190, respectively), which reflected the population of students in these classes (67 percent female). The average age of subjects was 18.62 years in the first part of the study and 19.16 years in the second.

Analyses of these figures revealed that there were no significant differences between the groups in terms of discard rate (chi-square = 0.28, df = 1, n.s.), or ratio of males to females (chi-square = 0.08, df = 1, n.s.). The difference in mean age between the samples was statistically significant, (t = 2.14, df = 335, p < .05), but of little practical importance. This is especially so in light of the fact that the older sample was recruited approximately three months later within the same academic year, and therefore probably stood in the
same relation, relative to the class as a whole, as did the earlier group of subjects.
Subsequent analyses have assumed that these two samples were essentially equivalent.
Procedure

A common procedure was followed in both conditions. Subjects were asked to select a single booklet from one of several stacks, each of which was said to contain a different type of form. A different identifying letter appeared on the cover sheet of each of the booklets in the different stacks. This was meant to emphasize the point later made in the instructions that many types of forms were being used which varied in the amount and sufficiency of information they contained. In fact, no such differences existed, but it was intended that subjects understand that the information provided to them was not necessarily wholly adequate.

Subjects were told that they would be presented with an event and three potential causes of that event, and that their task was to determine the involvement of those causes, as best as possible, within the limitations of the information provided by the booklet selected. To this end, subjects were told that they would be given an opportunity to select, from a variety of questions, those which they deemed most important to have answered. It was stated that the results of their inquiries would be highly individualistic depending, as they did, both on the type of booklet and the questions selected. It was explained, however, that this did not matter since the experimenter's major concern was to see what effect the
information selected would have on their estimation of the likelihood of involvement of a particular cause.

By way of illustration, subjects were asked to generate possible causes of an event, (a student failing to show up for his final examination), and practice was given in rating each cause according to how likely it seemed that it was involved in causing the event. New information was then introduced and it was clearly demonstrated that such information, depending upon its relationship to the hypothesis in question, may make the hypothesis appear to be more likely, less likely, or (if totally irrelevant) unchanged from its former rating. Subjects were told that the hypotheses that they had generated for this demonstration would be used in subsequent administrations of this task, and that those which they would themselves encounter had been suggested by previous subjects. The purpose of this deception was to emphasize that the task was not an experimenter designed "test," and thus make more believable the assertion that there was no one "right" answer for which they should be looking.

After the completion of the instructions, subjects began work on the task by following the directions specified in their booklet (a copy of which appears in the Appendix). All booklets described the same event: that of a foreign student in America who did poorly on a mid-term exam and was in
danger of failing the final, and having, as a result, to leave the university and return to his native country; and three possible causes of that event: lack of interest in the subject matter, inability to understand the English used in lectures, and poor health. These causes will be referred to as A, B and C respectively. It was stated that all other possible causes had been ruled out and that only these three needed to be considered.

Instructions then called for each hypothesis to be evaluated as to its perceived likelihood of involvement in causing the described event. All ratings were made on separate evaluation sheets. Each sheet contained three continuous lines, four inches in length, marked off at the extremes by the words "impossible" and "certain," with the word "neutral" appearing at the mid-point. Each line appeared next to a short description of one of the hypotheses, and subjects were asked to rate by making a mark on each line. In addition to these ratings, the evaluation sheet also called for subjects to indicate if they felt, in a practical sense and regardless of the rating given, that they had already made up their mind, one way or another about the hypothesis' involvement.

Three questions were then presented, each obviously pertaining directly to one of the three possible explanations. The subjects were asked to choose that
question deemed most helpful in determining why the described event occurred. After receiving the answer to the selected question, subjects were again asked to evaluate each of the listed hypotheses.

This question-evaluation cycle was repeated until a total of three questions were selected and four evaluations were made, (each on a separate evaluation sheet). No matter which question was selected at a given stage, all subjects were offered the same choice of new questions at the subsequent stage. Also, no matter which choice was made, the answer to the question always provided confirmation for the relevant hypothesis. The strength of the confirming information was meant to increase at each stage, with the final piece of information making the hypothesis appear virtually certain.

Subjects were given control over question selection by means of a 41 page "scrambled" booklet containing the instructions, questions and confirmatory answers, as well as some misleading information. Differences in question selection were accommodated by referring each question selected to a different page in the book for further information. All pages were randomly distributed throughout the booklet. Thus it was unlikely that subjects, casually flipping to the referred page, would have been able to find answers to specific questions which were not selected. Moreover, the misleading pieces of information (printed on pages ostensibly not intended to be seen) were stated in
negative terms so as to prevent subjects who nevertheless did see them from assuming that they would necessarily have all their questions answered in an affirmative manner.

After the completion of the final evaluation, subjects were directed to answer four further items in the booklet. The first asked subjects to rate the potential usefulness of information which might derive from a series of additional questions, although no actual information was made available. The questions included one relevant to each of the hypotheses as well as three arbitrary questions. This has subsequently been referred to as the Usefulness variable.

A second item (termed Opinion) then asked subjects to indicate their degree of agreement with a conclusion stated in the form "If T had not occurred, the event would not have occurred," where T was the hypothesis presumably made certain by the third piece of information. The exact wording of this question was somewhat different in the two parts of the study. In the first part, the sentence used was "If T had not occurred, the student would not have done poorly on the mid-term exam." In the second part of the study, the last part of the sentence was changed to "would have done well on the mid-term exam."

The third item (referred to as the Recommend variable)
consisted of two questions, one to be answered by those who felt that they had acquired a complete understanding as to why the event occurred, and one by those who did not feel that they had such an understanding. Subjects were instructed to read both questions and choose the one which best described their opinion. The former group was then asked to suggest actions which the student described in the scenario should take so as to guarantee that he would not fail his final exam. The latter group was asked to indicate what additional information they would have required for a complete understanding of the event.

The final item consisted of two questions (both of which had to be answered) which directly asked the subject to indicate what effect knowing of the definite involvement of the "T" cause would have on the likelihood of occurrence of the other two potential causes, in the case of some other foreign student.
Design

The major restrictions observed in the design of the booklet information involved the attempted conditional independence of potential causes (or at least the absence of any negative association such that the occurrence of one would have made the involvement of another less likely), as well as the specificity of new information made available. Thus the answers to questions selected on each hypothesis were always confirmatory of only that specific hypothesis and largely irrelevant to the others.

Although the same basic information was used in both parts of the study, differences existed in the description of the severity of the event. In the first part, the student was simply described as having done "poorly" on the mid-term exam, and in danger of failing the final (and thus having to leave the university). The description was thus "categorical" in nature, telling subjects only that his performance had been unsatisfactory, but giving no indication of his standing within that category.

In the booklet used in the second part of the study, the student was now described as having done "extremely poorly" (with the word "extremely" printed in block letters), and a sentence was added which stated that his grade had been "the lowest in the class." All subsequent references to the
student's performance within the booklet were altered to remind subjects of the extent of the academic difficulties, (e.g. references to the student doing "poorly" were changed to "so poorly" etc.). With the exception of these changes, and the previously mentioned wording alteration in one other question, the booklets used in the two parts of the study were identical.

For ease of reference, the booklets used in the first and second part of the study will be referred to as the categorical and quantitative forms respectively.
Results

Results will be reported in six sections, each of which addresses one aspect of the analysis. The first examined whether search was confirmatory and whether this resulted in an (at least initial) narrowing of the investigation (i.e., serial search). The second section reported whether the results of the first section persisted after certainty had been achieved (i.e., truncated search).

A truncated search implies not only that alternatives are ignored, but also that the scope of search has been restricted. Sections three through five looked at three other measures of subjects' treatment of alternatives (Ratings, Opinion and Recommend responses), and included an examination of the effect of scope of search on each measure. The final section investigated the relationship among measures.
Confirmatory Search Patterns:

I: Association between Rank and Question Selection

Confirmatory search should have tended to focus information search on one target hypothesis. A subject should therefore have attempted to confirm his hypothesis by choosing the question relevant to it. Since the information received was supportive in all cases, the same hypothesis should have been retained and this should have led to the choosing of a question pertaining to the same hypothesis at subsequent selections.

It was found that 25.2% and 27.4% of the categorical and quantitative samples respectively behaved in this manner (i.e., chose all three questions relevant to one hypothesis). Since the difference between the groups was not significant (Pearson Chi-square = .21, df = 1, n.s.), the samples were combined giving a total rate of 26.4%. In all, 27 search patterns were possible, and this total rate, which represented only three of those patterns (AAA, BBB, CCC) was significantly higher than the 11.1% that would have been expected if subjects had chosen questions randomly at each stage (Pearson Chi-square = 79.85, df = 1, p < .001).

The degree to which subjects engaged in this type of restricted search, although directly related to the practical issue of access to information about other possible causes,
was only in part a measure of the tendency to seek confirmation. Since no formal means were taken to equate questions with regard to the perceived amount and helpfulness of information that they provided, the selection of one question over another must be presumed to have been affected not only by the subject's strategy but also by the relative attractiveness of the questions available. In fact, it would have been possible to have produced restricted search in all subjects simply by making the questions relating to some specific hypothesis relatively more interesting than any of the other questions at each stage.

A more appropriate test of confirmatory search would be to compare the conditional probabilities of subjects choosing a particular question given their choice of hypotheses. The expectation would be that a subject entertaining a single hypothesis $X$ should be more likely to select the question relating to $X$ than those who were not considering this hypothesis.

The hypothesis under consideration was inferred from the subject's ratings. The distance of the subject's marks on each line was measured in units of one sixteenth of an inch, giving a scale with at least ordinal properties, ranging from 0 to 64. By dealing only with those cases in which there were no ties in the ratings at any given stage, it was possible to rank hypotheses in terms of their relative
likelihood of having been under active consideration. It was further assumed that, for the vast majority of these subjects, only a single hypothesis was being investigated at any one time. This followed not only from the lack of tied ratings, but also from findings to this effect by Tweney Doherty, Worner, Pliske & Mynatt (1980).

The following analyses made use of all subjects (whose ratings were not tied) no matter how extreme their ratings. The assumption made here was that a subject who rated a hypothesis extremely highly, and also indicated that he had already decided on that hypothesis' involvement, would nevertheless still have a high probability of choosing the question relating to that hypothesis. This issue is discussed in further detail in the next section.

The relationship between a hypothesis’s ranking and question selection was analyzed separately for each hypothesis, although the data for all hypotheses appear in Table 1. An examination of the first two columns of Table 1 gives the results for the A hypothesis (student had no interest in the subject matter). The probability of choosing question A, as opposed to either questions B or C (denoted as "other"), was always highest when hypothesis A was ranked highest. This result held across all rating stages and booklet forms (i.e., categorical or quantitative). Moreover, in all cases but one (the third stage in the quantitative form), the
Table 1

Frequencies of Various Question Selections by Booklet Form, Question Selection Stage, and Hypothesis Rank

<table>
<thead>
<tr>
<th>Form Stage Rank</th>
<th>A</th>
<th>Other</th>
<th>B</th>
<th>Other</th>
<th>C</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form Stage Rank</td>
<td>A</td>
<td>Other</td>
<td>B</td>
<td>Other</td>
<td>C</td>
<td>Other</td>
</tr>
<tr>
<td>1 21 10 9 20 72 10 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 2 21 24 1 23 28 25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 16 42 1 5 15 43</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category 1 16 11 40 10 7 38</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 5 22 38 19 5 33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 4 64 5 10 2 37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 19 4 39 42 10 9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 2 12 20 5 29 14 43</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 21 47 0 8 3 44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantitative 1 12 10 27 94 13 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 23 25 1 30 40 41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 28 62 0 8 16 46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 15 12 65 6 18 39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 4 38 38 30 1 44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 6 80 7 9 1 52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 16 8 57 56 11 15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 2 17 30 6 26 7 77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 38 54 4 14 7 46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
probability of choosing question A was also higher when the hypothesis was ranked second than when it was seen as being least likely.

The data in Table 1 (Form type by hypothesis Rank by Question selected by selection Stage), was subjected to log-linear analysis by program BMDP4F (Dixon, 1981). The best fit to the data was given by the model: FRS, RQS. (likelihood ratio Chi-square=4.99, p=.8356, Pearson Chi-square=4.97, p=.8368, df=9).

The FRS interaction results from the fact that the Question selected was treated as a response variable. All log-linear analyses will follow the practice of reporting only higher order terms, using the first letter of the variable to represent it. Thus QSR represents a second order interaction among the variables Question selected, Stage and Rank, and includes all lower order terms QS, QR, RS, Q. See the section on log-linear analysis in the Appendix for a more detailed explanation of this technique.

The most interesting interpretation of this model is that the relationship between hypothesis rank and question selected, although significant, differed in its level of association across the ranking stages. In other words, the rank:question relationship at the first stage was not necessarily equal to the relationship between these
variables at the second or third stages. The type of form used seemed not to have been important.

The estimated odds of choosing the A question, calculated from the above model, are shown in the first section of Table 2. It is clear from these results that the second order interaction is due to the fact that, at the first stage, the odds in favour of choosing the A question when A was ranked highest are relatively close to those for when A was ranked second. In other words, although being ranked highest always leads to the greatest probability of selecting the hypothesis-specific question at all stages, this tendency, for the A data, is somewhat less at the first stage.

There is some evidence that the stage one A question was somewhat ambiguous, and this may have contributed to the above finding.

Examination of the data for the B and C hypotheses in Tables 1 (poor English comprehension, illness), revealed the same pattern of ordered conditional probabilities already noted for the A data. The only exceptions to this finding with regard to the B hypothesis (second and third columns) occurred in the reversal of the second and third rankings in the first stage of the categorical form and the third stage of the quantitative form. For the C data, the only reversal occurred between the second and third rankings at the third
Table 2
Estimated Odds in Favor of Choosing Questions A, B, and C by Hypothesis Rank and Question Selection Stage

<table>
<thead>
<tr>
<th>Rank</th>
<th>Stage</th>
<th>A Question Selected</th>
<th>B Question Selected</th>
<th>C Question Selected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>1.16 : 1</td>
<td>1.35 : 1</td>
<td>2.92 : 1</td>
<td>.28 : 1</td>
</tr>
<tr>
<td>2</td>
<td>.90 : 1</td>
<td>.15 : 1</td>
<td>.58 : 1</td>
<td>.06 : 1</td>
</tr>
<tr>
<td>3</td>
<td>.42 : 1</td>
<td>.07 : 1</td>
<td>.58 : 1</td>
<td>.03 : 1</td>
</tr>
<tr>
<td></td>
<td>4.19 : 1</td>
<td>.33 : 1</td>
<td>.88 : 1</td>
<td>.96 : 1</td>
</tr>
<tr>
<td></td>
<td>.39 : 1</td>
<td>.03 : 1</td>
<td>.08 : 1</td>
<td></td>
</tr>
</tbody>
</table>
stage in the quantitative form. When subjected to separate log-linear analyses (Stage by Rank by Form by Question selected), it was found that the model: PRS, QS, QR fit both sets of data. The likelihood ratio Chi-square and Pearson Chi-square, for the B data were 12.82 (p=.4621, df=13), and 13.84 (p=.3849) respectively. The same statistics for the C hypothesis data were 19.84 (p=.0992) and 19.39 (p=.1115, df=13).

As in the previous case, a significant rank by question interaction was discovered, unaffected by form type. In these instances, however, no differences in this relationship were found across the stages. The QS terms in the B and C models reflect the fact that the relative attractiveness of the questions relating to B and C questions respectively were not constant across all stages.

The estimated odds in favour of choosing the hypothesis-specific question, based on the appropriate models for the B and C data, are shown in the second and third sections of Table 2 respectively. These results show the same ordered relationship between odds and rank at each stage. In both the B and C data, the difference between the odds of the first and second ranked hypotheses was always greater than the corresponding difference between the second and third ranked hypotheses.
Further evidence for confirmatory search was found by examining the relationship of hypothesis rank to rated usefulness of additionally offered information. It will be recalled that after the completion of the final rating, subjects were asked to rate six additional questions from 1 (totally useless) to 5 (essential) according to how necessary it would be to know the answer to each so as to understand why the event described in the scenario occurred. One question related to each of the three hypotheses, while the other three questions offered information that was largely irrelevant to any of them.

In the first analysis, the rated usefulness of the question about the A hypothesis was the dependent variable in a two-way ANOVA (form type by rank). There were three levels of the rank dimension, corresponding to whether the A hypothesis was ranked highest, second, or third at the final rating stage. Subjects with tied ratings for any pair of hypotheses were excluded from the analysis.

The results indicated that there were significant differences in perceived usefulness of the A information ($F(2,267) = 6.52, p < .01$). The form effect and form by rank interaction were not statistically significant ($F(1,267) = .29, n.s. and F(2,267) = .27, n.s. respectively$).

Newman–Keuls follow-up tests were performed on the rank
dimension (collapsed over form). Results indicated that subjects who had rated the A hypothesis most highly at the final stage felt that further information about that hypothesis would be more useful than those subjects who had rated the A hypothesis second or third (p<.01). There were no significant differences between the second and third ranked groups.

These analyses were repeated for the B and C relevant information. In both cases, the results were similar to those just reported; a significant rank effect, with no significant effects due to form or the interaction form by rank. The B data values were F(2,267)=4.61, p<.01, F(1,267)=1.03, n.s. and F(2,267)=.03, n.s. respectively. Because the homogeneity of variance assumption was not met with the C data, a critical value with 1 and (n-1) degrees of freedom was used for the F statistics (Winer, 1971, p. 206). The corresponding values for the C data were (using the harmonic n) F(1,39)=30.77, p<.01, F(1,39)=.16, n.s. and F(1,39)=1.09, n.s.

In both the B and C data the first rank was significantly different from either the second or third (p<.01) but the Newman-Keuls tests on the C data differed in showing a significant difference between the second and third ranked groups (p<.05). The B data however, like that relating to the A information, showed no such differences and there was
in fact a (nonsignificant) reversal in the means between these two groups (i.e., the second ranked groups rated the information to be less useful than did the third group). It would appear that the major distinction with regard to perceived usefulness of additional hypothesis-specific information was between those who ranked that hypothesis most highly and those who did not.

II: Serial search
Subjects demonstrated a tendency to choose questions relating to their highest ranked hypotheses. It was expected that, since all questions gave confirmatory answers, this should have led to a tendency to stay with the same hypothesis across question selections.

A four-way log-linear analysis (Initial question by Second question by Third question by booklet Form) resulted in the following model: IST, P, (likelihood ratio
Chi-square=22.79, \( p = .5897 \), Pearson
Chi-square=20.36, \( p = .7280, \text{df.}=25 \) The data appear in Table 3. Each of the three questions selected had some relation to the others but not all 2-way relationships were equal at all levels of the third variable. As before, no differences were noted between booklet forms.

In general, the results indicated that subjects who selected a particular question at one stage were significantly more
Table 3
Frequencies of Various Question Selections at Each Question Selection Stage by Booklet Form (N=337)

<table>
<thead>
<tr>
<th>Third Selection</th>
<th>Second Selection</th>
<th>Initial Selection</th>
<th>Form</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>17</td>
<td>14</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>9</td>
<td>18</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>26</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>B</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>17</td>
<td>26</td>
</tr>
<tr>
<td>B</td>
<td>B</td>
<td>14</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>13</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>B</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>A</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>B</td>
<td>A</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>B</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>6</td>
<td>16</td>
</tr>
</tbody>
</table>
likely to choose the question relating to the same hypothesis at the subsequent stage. The three way interaction term in the model stems from the fact that, not only is the probability of choosing a question at the third stage higher for those who chose it at the second stage, but it is higher yet for those who also chose it at the first stage. The implications of this finding are discussed further in a later section.

These data then, argue for a tendency for subjects to confine their information search to a single hypothesis, that which is considered most likely. This tendency appeared to operate to the same extent in both booklet form types.
The previous section presented evidence that subjects tended to confine their information search to a single hypothesis, given the confirmation of that hypothesis. If, however, the subject became convinced of that hypothesis' involvement, there would be no disadvantage to dropping that line of questioning and switching to investigate the role of some other potential cause. Although the information provided was structured so that certainty of involvement was not generally apparent until the final piece of information, some subjects did indicate that they were certain before that time, and their subsequent question selections provided a means to examine this issue.

Before examining this problem directly, it should be noted that there has already been some evidence, presented in the previous section, suggesting that subjects do not proceed to investigate other hypotheses once they become certain of one. This conclusion can be inferred from two sets of evidence.

First, there is the fact that no inter-stage differences were noted (at least for the B and C data) with regard to the relationship between hypothesis ranking and question selection. This would not have been expected given that the proportion of "certain" hypotheses clearly increased from
stage to stage (due to the accumulation of information as well as increased strength of confirmation at each stage). Even the A data, which did show inter-stage differences, provided little support. Since "certain" hypotheses were very likely to have been ranked highest, they should have had the effect of reducing the difference in conditional probabilities between the first and second ranked hypotheses. In the A data however, the inter-stage difference lies in the fact that the stage one results (with fewest "certain" hypotheses) showed smaller differences between the first and second ranks when compared with the results at the second and third stages.

The second piece of indirect evidence can be seen in the second order interaction noted in the relationship between questions selected. It was found that the probability of choosing some specific question at the third (final) selection was higher for those who had chosen that question at the second stage, and among those subjects, higher still for those who had also chosen it at the first selection. Thus a subject whose first two questions were BB had a better chance of choosing B again than if his first two questions had been AB or CB. These results were again not what would have been expected if "certain" hypotheses were abandoned, since it is likely that the BB group, for example, had a higher proportion of subjects who were certain about B than did the AB or CB groups. To test this,
certainty was defined as being rated at least 61 (out of 64) in addition to being (separately) marked as being beyond practical doubt. The data which appear in Table 4, were analyzed as a three way contingency table (Certainty by Hypothesis by Number explored). Since all subjects in this analysis chose the same hypothesis to explore at the second and third selections (denoted by the Hypothesis variable), the total number of hypotheses explored was either one or two depending upon the first selection. Treating Certainty as a response variable, the model NH,NC,HC was found to fit the data, (likelihood ratio Chi-square=1.96, p=.3749, Chi-square=1.61, p=.446, df=2). Aside from the fact that the B hypothesis produced more certainty than did the A hypothesis, the model indicates that there was a significantly higher probability of certainty among subjects who had only explored a single hypothesis.

A direct test of the effects of certainty was made by comparing the conditional probabilities of choosing a hypothesis-specific question for hypotheses deemed certain with the corresponding probabilities for those that were not. The definition of certainty was as above, with the added restriction that neither of the subject's (same stage) alternative hypotheses met any of the requirements of certainty. In order to avoid extremely unreliable estimates, hypotheses were examined only if at least ten subjects found it to be certain as judged by these
Table 4
Frequencies of Hypotheses Rated Certain by Third Question Selected and Number of Hypotheses Explored for Subjects Who Explored the Same Hypothesis at the Second and Third Stages (N=180)

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Number Explored</th>
<th>Certain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>23</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>
standards. Only one hypothesis met these criteria at each stage.

These are presented, collapsed over form type, in Table 5. Comparison figures were provided by subjects, all of whom had distinct (i.e., not tied) ratings, ranked highest to lowest, with no hypotheses meeting any of the criteria of certainty.

As can be seen, the probability of a subject choosing a question relating to a hypothesis considered certain, tended to be about the same (stage 2), or higher (stages 1 and 3) than if the hypothesis was rated highest, but not certain.

Thus, subjects, when certain of a hypothesis' involvement, did not shift away from that hypothesis in order to explore alternative hypotheses.

This same issue was also addressed by comparing the mean rated usefulness of additional information for subjects who had indicated, at the final stage, that they were certain of a hypothesis' involvement with those who were not. All definitions and restrictions were as in the previous analysis.

Three two-way ANOVAs (form type by rank) were conducted, one for each of the hypothesis-specific pieces of information.
Table 5

Frequencies of Hypothesis Specific Question Selections for Some\textsuperscript{a} Hypotheses Containing Certain Ratings at Various Rating Stages

<table>
<thead>
<tr>
<th>Rating\textsuperscript{b}</th>
<th>Hypothesis Specific</th>
<th>NOT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1 B Hypothesis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Certain</td>
<td>5 (.38)</td>
<td>8 (.62)</td>
</tr>
<tr>
<td>Highest</td>
<td>35 (.20)</td>
<td>138 (.80)</td>
</tr>
<tr>
<td>Second</td>
<td>2 (.05)</td>
<td>40 (.95)</td>
</tr>
<tr>
<td>Third</td>
<td>1 (.09)</td>
<td>10 (.91)</td>
</tr>
<tr>
<td>Stage 2 C Hypothesis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Certain</td>
<td>4 (.19)</td>
<td>17 (.81)</td>
</tr>
<tr>
<td>Highest</td>
<td>12 (.20)</td>
<td>47 (.80)</td>
</tr>
<tr>
<td>Second</td>
<td>5 (.08)</td>
<td>58 (.92)</td>
</tr>
<tr>
<td>Third</td>
<td>2 (.03)</td>
<td>61 (.97)</td>
</tr>
<tr>
<td>Stage 3 B Hypothesis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Certain</td>
<td>42 (.59)</td>
<td>29 (.41)</td>
</tr>
<tr>
<td>Highest</td>
<td>28 (.44)</td>
<td>36 (.56)</td>
</tr>
<tr>
<td>Second</td>
<td>8 (.19)</td>
<td>34 (.81)</td>
</tr>
<tr>
<td>Third</td>
<td>2 (.15)</td>
<td>11 (.85)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Hypotheses were included only if at least 10 subjects rated it certain.

\textsuperscript{b}Rank among remaining noncertain hypotheses

\textsuperscript{c}Row probabilities appear in brackets
The rank dimension thus consisted of four levels, those that had a single "certain" hypothesis, and three groups without any "certain" hypotheses, distinguished by the relative standing of their hypotheses at the final stage. The form and Rank by Form interaction values for the A data were $F(1, 71)=1.95,\text{n.s.}$ and $F(3, 71)=.54,\text{n.s.}$ respectively. The corresponding figures for the B data were $F(1, 146)=.01,\text{n.s.}$ and $F(3, 146)=.29,\text{n.s.}$ In addition, there were no significant differences among ranks with regard to the rated usefulness of the A and B data ($F(3, 71)=.99,\text{n.s.}$ and $F(3, 146)=.68,\text{n.s}$ respectively). Moreover, the trend of the means of the "non-certain" groups are not consistent with previous findings. The second ranked group in the A data has a higher mean than that of the highest ranked, and in the B data, the order of all three means is exactly opposite to expectation (see Table 6).

These findings, however, may not be as inconsistent as they first appear. First, the means of the "non-certain" groups are very close to one another, and it is not at all clear that this trend would be maintained or prove to be statistically significant if a larger sample were available.

Second, these "non-certain" comparison groups may consist of somewhat atypical subjects, since they are composed of that 15% of the total sample that had not decided on the certainty of any of their hypotheses after receiving three
Table 6
Means, Standard Deviations and Frequencies of Ratings\textsuperscript{a} of Usefulness of Additional Hypothesis Specific Information

<table>
<thead>
<tr>
<th>Information Type</th>
<th>Booklet Form</th>
<th>Hypothesis Rank</th>
<th>Mean</th>
<th>S.D.</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categorical</td>
<td>Certain</td>
<td>2.60</td>
<td>1.43</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>First</td>
<td>2.40</td>
<td>1.17</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Second</td>
<td>2.17</td>
<td>0.75</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Third</td>
<td>2.50</td>
<td>1.29</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Certain</td>
<td>3.29</td>
<td>0.69</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Quantitative</td>
<td>First</td>
<td>2.62</td>
<td>1.04</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Second</td>
<td>2.88</td>
<td>1.46</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Third</td>
<td>2.36</td>
<td>0.92</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>Certain</td>
<td>2.98</td>
<td>1.08</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>Categorical</td>
<td>First</td>
<td>2.83</td>
<td>1.17</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Second</td>
<td>2.71</td>
<td>0.76</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Third</td>
<td>3.00</td>
<td>1.29</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Certain</td>
<td>3.11</td>
<td>1.00</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>Quantitative</td>
<td>First</td>
<td>2.60</td>
<td>0.97</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Second</td>
<td>2.92</td>
<td>0.76</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Third</td>
<td>2.78</td>
<td>0.67</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Certain</td>
<td>4.47</td>
<td>0.52</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Categorical</td>
<td>First</td>
<td>4.25</td>
<td>0.50</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Second</td>
<td>3.71</td>
<td>0.76</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Third</td>
<td>3.89</td>
<td>0.78</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Certain</td>
<td>4.92</td>
<td>0.28</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Quantitative</td>
<td>First</td>
<td>4.67</td>
<td>0.50</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Second</td>
<td>4.00</td>
<td>0.63</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Third</td>
<td>3.58</td>
<td>0.79</td>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a}Rating ranged from 1 (useless) to 5 (essential).
pieces of confirmatory information. As such, their responses may not necessarily be representative.

Most important, however, is the fact that in both the A and B data the "certain" group means were higher than those of any other. In addition, the differences between the "certain" groups and the others were always larger than any intra ("non-certain") group differences.

The data relating to the C information again displayed the highest mean usefulness ratings for the "certain" group, and this time significant differences between the groups were found \((F(3,72)=12.59, p<.01)\) although the form and interaction effects were not significant \((F(1,72)=2.11, \text{n.s.} \text{ and } F(3,72)=1.74, \text{n.s} \text{ respectively})\). Newman-Keuls tests indicated that the "certain" and highest ranked "non-certain" groups were indistinguishable, but that both were significantly different from the second and third ranked groups \((p<.01)\). These latter two groups were also found to be indistinguishable.

Once again, there is no indication that deciding on the certainty of a hypothesis' involvement in any way reduces the interest in acquiring further information about that hypothesis.

The interval-level properties of the usefulness ratings also
allow an opportunity to adjust for differences in the quality of information offered. Thus, it is possible to discover what percentage of subjects with "certain" hypotheses at the final rating would be most interested in further information about that hypothesis, given the availability of other pieces of information of equal inherent attractiveness.

Subjects were first divided according to final question selection and then random deletion was performed on the larger groups until all three groups were of equal size. These groups were then combined, and their responses were used to estimate the means and standard deviations of each of the three hypothesis-specific usefulness questions. These figures were, in turn, used to produce standard scores for all subjects used in the analysis.

Subjects were included in this analysis if they had decided that only one of their hypotheses was certain at the final rating. The definition of certainty used here was the presence of both an indication of practical certainty and a rating of at least 48. Within the sample as a whole, 244 subjects had at least one hypothesis that met this definition at the final rating and 218 subjects had only one such hypothesis. The additional criterion for inclusion in this analysis was that neither of the subjects' alternative hypotheses met either of the criteria for certainty.
The response variable, Match, referred to whether the highest standardized usefulness score was that belonging to the certain hypothesis. Form and Hypothesis marked certain were analyzed as explanatory variables. The resulting model: PH,M was found to have a likelihood ratio Chi-square of 8.23 (df=5, p=.1438) and a Pearson Chi-square of 7.91 (p=.1613). The data appear in Table 7.

Thus the Match variable was unaffected by form or hypothesis type, and in general 57.6% of subjects were found to rate information relating to their certain hypothesis to be more useful than that relating to either of their non-certain hypotheses. This figure is well in excess of the 33.3% that would have resulted from random selection among three choices. Rationally, it might be expected that subjects with only a single certain hypothesis and two alternatives which were far from certain, would be most likely to expand their information search. Prior results however indicate that just the opposite is true, and thus it is likely that this 57.6% figure represents, for the three choice case, an upper limit on the tendency to ignore information relating to alternatives, at least when information about those alternatives is explicit and readily available.

If the inclination to narrow information search is strongest when there is only one highly rated or certain hypothesis, it should be expected that the tendency to rate more highly
Table 7

Frequencies of Usefulness Variable Matching Responses for Subjects With a Single Certain Hypothesis by Booklet Form, and Hypothesis (N=177)

<table>
<thead>
<tr>
<th>Form</th>
<th>Match</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categorical</td>
<td>No</td>
<td>12</td>
<td>21</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>14</td>
<td>19</td>
<td>10</td>
</tr>
<tr>
<td>Quantitative</td>
<td>No</td>
<td>5</td>
<td>30</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>15</td>
<td>34</td>
<td>10</td>
</tr>
</tbody>
</table>
information relevant to the highest hypothesis should vary as a function of the number of hypotheses explored. Those who investigated only a single hypothesis, for example, should be more likely to have rated that hypothesis as certain and less likely to have so rated alternatives.

To test this, all subjects, except those with tied ratings, were classified according to whether they had explored one, two, or all three hypotheses. A subject was considered to have chosen matching information if the most highly rated standardized usefulness score was that relating to the highest rated hypothesis. The results, shown in Table 8, indicated that although all groups had above chance levels of matching responses, there was a tendency for matching to increase when fewer hypotheses were explored. This trend just misses statistical significance (Pearson Chi-square = 5.495, df=2, p = .06). If the first question selected was ignored however, and subjects classified only as to whether the same hypothesis was explored at both the second and third selections, statistical significance was achieved. The data, shown in Table 9, have a Chi-square value of 6.148 (df=1, p = .013).

A corollary of these results is that subjects with no certain responses at the final rating should have had a different distribution of search patterns from those who did, a result already shown in the special case of data in
Table 8
Frequencies of Matching Usefulness Variable Responses
by Number of Hypotheses Explored

<table>
<thead>
<tr>
<th>Match</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>33</td>
<td>78</td>
<td>35</td>
<td>146</td>
</tr>
<tr>
<td>Yes</td>
<td>55</td>
<td>84</td>
<td>27</td>
<td>166</td>
</tr>
<tr>
<td>TOTAL</td>
<td>88</td>
<td>162</td>
<td>62</td>
<td>312</td>
</tr>
</tbody>
</table>

Table 9
Frequencies of Matching Usefulness Variable Responses
by Similarity of Last Two Hypotheses Explored

<table>
<thead>
<tr>
<th>Match</th>
<th>Same</th>
<th>Different</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>72</td>
<td>74</td>
<td>146</td>
</tr>
<tr>
<td>Yes</td>
<td>105</td>
<td>61</td>
<td>166</td>
</tr>
<tr>
<td>TOTAL</td>
<td>177</td>
<td>135</td>
<td>312</td>
</tr>
</tbody>
</table>
Table 4. The data, shown in Table 10, are highly significant (Pearson Chi-square=63.06, df=6, p<.001). They indicate that those with no certain ratings were more likely to have investigated all three hypotheses and less likely to have investigated one or two, relative to those who ended up with a single certain rating. In fact, the largest group of those who investigated all hypotheses (43.8%) had no certain ratings.

Alternatively, if only the last two questions selected were used to classify subjects (see Table 11) an equally clear pattern emerges. Investigation of the same hypothesis at both the second and third selections was associated with the certain rating of a single hypothesis. Investigation of different hypotheses led to greater proportions of subjects with multiple certain ratings and with no certain ratings. In spite of this, a multiple certain rating was always a less common response than one that indicated no certainty.

In summary then, the tendency to investigate a single hypothesis is not apparently affected by becoming certain of the involvement of that hypothesis. Even after certainty is arrived at, subjects still have a tendency to seek further information about it rather than exploring information relevant to alternative hypotheses.
Table 10

Frequencies of Number of Hypotheses Rated Certain by Number of Hypotheses Explored

<table>
<thead>
<tr>
<th>Number of Certain Hypotheses</th>
<th>Number of Hypotheses Explored&lt;sup&gt;a&lt;/sup&gt;</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>16 (.18) 42 (.25) 35 (.44)</td>
<td>93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>71 (.80) 120 (.71) 27 (.34)</td>
<td>218</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2 (.02)  6 (.04) 10 (.12)</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0 (.00)  0 (.00)  8 (.10)</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>89  168  80  337</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 11

Frequencies of Number of Hypotheses Rated Certain by Similarity of Last Two Questions Selected

<table>
<thead>
<tr>
<th>Number of Certain Hypotheses</th>
<th>Last Two Questions&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Same</th>
<th>Different</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>28 (.16)</td>
<td>65 (.42)</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>150 (.83)</td>
<td>68 (.43)</td>
<td>218</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2 (.01)</td>
<td>16 (.10)</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0 (.00)</td>
<td>8 (.05)</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>180</td>
<td>157</td>
<td>337</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Column probabilities (rounded to two decimal places)
I: Ratings

Another means of investigating a subject's tendency to be content with incomplete evidence, involves examining changes in ratings, from one stage to another, for hypotheses other than the one confirmed by the intervening question information.

A piece of evidence was considered to have had a confirmatory effect if it produced an increase in the rating of the targeted hypothesis of at least four units from one rating to another, or if the resultant rating was at least sixty-two units. An alternative was considered to have increased or decreased if the change in ratings between stages was at least four units higher or lower respectively. If there were less than four units change in either direction, the rating of that hypothesis was considered to have stayed the same. Analysis was restricted to those cases in which the information had a confirmatory effect (as defined above) and all alternative hypotheses were originally rated between four and sixty units.

A four way log-linear analysis was conducted (Change by Form type by Stage by confirmed Hypotheses). Since each subject had two alternative hypotheses to consider at each stage, the change variable was divided into six categories.
corresponding to all possible combinations; both increase, one increase:one decrease, one same:one decrease, both same, one same:one decrease, both decrease. The data appear in Table 12. Since the major purpose of this analysis was to investigate the effect of Stage, Hypothesis confirmed and Form type on the ratings of alternative hypotheses (Change), these former variables were specified as explanatory and thus the FSH interaction term was included in all models. A constant of .5 was added to each cell in order to eliminate the increase:increase zero submarginal. Using this approach, the model FSR, C, with 85 degrees of freedom, was found to provide an acceptable fit to the data. The likelihood ratio Chi-square and Pearson Chi-square were 89.28 (p=0.3542) and 85.75 (p=0.4568) respectively.

This outcome means that there was a distinct tendency, unaffected by form, confirmed hypothesis, or stage, to rate alternatives lower. A more precise description of this tendency can be found by comparing the estimated odds. If the probability of an alternative being rated higher or lower, given change, were the same, the odds of a same:decrease response would be even with respect to same:increase. In fact the odds in favour of the same:decrease pattern were 3.02:1. Similarly the odds in favour of a decrease:decrease response as compared with an increase in both alternatives, were 8.68:1.
### Table 12

**Frequencies of Various Combinations of Changes in Ratings of Alternatives**

**by Booklet Form and Rating Stage Transitions**

<table>
<thead>
<tr>
<th>Form</th>
<th>Stage</th>
<th>Increase:</th>
<th>Increase:</th>
<th>Same:</th>
<th>Same:</th>
<th>Same:</th>
<th>Decrease:</th>
<th>Decrease:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Increase</td>
<td>Decrease</td>
<td>Increase</td>
<td>Decrease</td>
<td>Increase</td>
<td>Decrease</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>A Confirmed, Alternatives are B and C</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Categorical</td>
<td>1-2</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2-3</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3-4</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>7</td>
<td>9</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Quantitative</td>
<td>1-2</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2-3</td>
<td>1</td>
<td>6</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3-4</td>
<td>2</td>
<td>9</td>
<td>5</td>
<td>11</td>
<td>18</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td><strong>B Confirmed, Alternatives are A and C</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Categorical</td>
<td>1-2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>6</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2-3</td>
<td>2</td>
<td>10</td>
<td>6</td>
<td>15</td>
<td>23</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3-4</td>
<td>0</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>9</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Quantitative</td>
<td>1-2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2-3</td>
<td>3</td>
<td>8</td>
<td>1</td>
<td>23</td>
<td>27</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3-4</td>
<td>1</td>
<td>6</td>
<td>7</td>
<td>17</td>
<td>12</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td><strong>C Confirmed, Alternatives are A and B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Categorical</td>
<td>1-2</td>
<td>1</td>
<td>9</td>
<td>5</td>
<td>9</td>
<td>14</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2-3</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3-4</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Quantitative</td>
<td>1-2</td>
<td>2</td>
<td>12</td>
<td>3</td>
<td>10</td>
<td>17</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2-3</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3-4</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>
It could be argued that the reason subjects lowered the ratings of their alternative hypotheses was because the hypotheses were actually incompatible. Thus, it may have been the case that, in spite of the experimenter's efforts to the contrary, subjects deemed these particular causes to be mutually exclusive or at least negatively associated with one another.

The conditional independence of hypotheses was directly examined. Each subject was, at the end of the task, asked to indicate, in the case of some other hypothetical foreign student, what effect the known certainty of hypothesis T would have on the likelihood of occurrence of the two alternative hypotheses. The T hypothesis was that one confirmed by the final piece of information. Thus, each subject answered two questions. In addition, there were three classes of subjects depending on their final question selection.

Even though these two conditional independence questions were answered after completion of the ratings, and may have been influenced by them, they can also be interpreted as causal variables affecting the direction of alternative hypotheses ratings.

Subjects were asked to indicate whether the alternative hypothesis in question would become more likely, be
unchanged, become less likely or be rendered impossible given the certainty of the hypothesis confirmed by the final question selection. Owing to the fact that the "impossible" response was rarely indicated (a point which will be taken up further in a later section), such responses as did occur, were combined with the "less likely" category.

It would be necessary to find that a large number of subjects chose the "less likely" response if the results of the previous section were to be explained in terms of perceived incompatibility of specific hypotheses. In fact, the modal category in all cases was that of "no change" (i.e., independence), and in all but two situations (the effects on B given A, and the effects on C given B) deviations from independence were biased in favour of the "more likely" response. There were no differences in response patterns between form types. The Chi-square values for the Form by Independence response tables, all of which had two degrees of freedom, were .16, 2.3, 1.1, 1.5, .37, and .68 for the effects on B given A, C given A, A given B, C given B, A given C, and B given C respectively.

The relationship between these conditional independence questions and actual change in ratings was examined for differences which occurred as a result of the final piece of information; the one which was intended to make the confirmed hypothesis appear certain. Definitions of rating
change and restrictions on subject selection were as stated in the previous section. Thus, for example, one table consisted of a cross-tabulation of subjects' estimation of the effects on B given that A was certain against those subjects' actual changes in ratings for the B hypothesis which occurred when A was confirmed at the final stage. Each of the six tables had four degrees of freedom. In order to control the cumulative Type I error rate, all analyses were conducted at the .01 level of significance. None of the Chi-square values were statistically significant. The values were 3.88, 6.36, 11.28, 6.67, 11.26, and 3.85 respectively for the tables: B given A, C given A, A given B, C given B, A given C and B given C.

Two tables, A given B, and A given C would have been significant at the .05 level. If subjects' notions of conditional dependence were influencing their ratings, it would have been expected that those who thought, for example, that A was less likely given B, would have been more inclined to actually rate A lower when B was confirmed. This pattern would have produced a positive valued Gamma statistic (Goodman & Kruskal, 1954, 1963). In the case of these nearly significant tables however, the Gamma values were negative, although not significantly so. The Gamma values and their corresponding "t" scores were -.183 and -1.211 for the A given B table, and -.384 and -1.338 for the A given C table. In all six tables, however, the majority of
subjects who had decreased the rating of their alternative hypotheses, had indicated in the conditional independence questions, that these hypotheses should have been unaffected or more highly rated.

It should also be noted that the lack of relationship between these two measures was not necessarily influenced by the amount of increase in the confirmed hypothesis. The conditional independence question assumed that the confirmed hypothesis had become certain, and for a large number of subjects the final piece of information had just this effect, and thus the measures should be comparable. For those subjects, whose confirmed hypotheses increased in perceived likelihood but did not attain certainty, the condition should not be fundamentally different. For example, a subject who indicated that hypothesis $K$ would become "less likely" given the certainty of hypothesis $T$, should, if a continuous scale were available, make a response somewhere between "no change" and "less likely" if hypothesis $T$ had increased in likelihood, but not to the point of certainty. On a discrete scale, this would manifest itself in an increase in the number of subjects expecting "no change." Since this "regression" type effect should operate to the same extent if the expected effect on hypothesis $K$ were toward "more likely," the overall cell proportions would be unaffected.

The main point though, is that any redistribution that
increased the number of "no change" responses undermines even further the explanatory power of perceived conditional dependence with regard to decreases in the ratings of alternative hypotheses.

It would still be possible to argue that the decrease in the ratings of alternative hypotheses was an artifact of the experimental task deriving from the nature of the information made available at each stage. Thus, even though a subject may have felt that hypotheses T and K were independent of one another, it would nevertheless still have been possible to have received information that simultaneously made T more likely and K less so.

It is unlikely that this occurred however, not only because great care was taken to avoid information with multiple implications, but also because of the generality of the effect. In each of the six possible situations a decrease in the rating of the alternative hypothesis was the most common response. An earlier analysis in this section also noted that there were no differences in rating patterns between stages. Therefore, it is extremely unlikely that every piece of information at every stage was interpreted as being relevant to the alternative hypotheses and that this information was consistently interpreted as making the alternatives less likely.
II: Elimination of Alternatives
The previous section demonstrated that subjects tended to lower their ratings of alternative hypotheses. It would be of more practical interest to discover whether these hypotheses were, in addition, deemed too unlikely to be worthy of further consideration.

A hypothesis was assumed to have been ruled out of consideration if the subject both indicated a lack of practical interest in it and gave it a rating of less than seventeen units. This was meant to be a very broad definition that would be quite sure of including all subjects who had lost interest in the hypothesis. Analysis was conducted using only those subjects who had not ruled out any of the three hypotheses at the initial evaluation. A further restriction was that the information made available was shown to have had a confirmatory effect on the target hypothesis (i.e., caused a rating of at least sixty-two units or an increase of at least four units).

The results of the 4-way analysis (Form by Hypothesis confirmed by Stage by Number rated impossible) appear in Table 13. Since only 24 subjects in this analysis had eliminated both of their alternatives, these responses were collapsed with those who had eliminated only one. Treating the number of eliminated hypotheses as a response variable, the model: PH, SN, NH was found to provide an adequate fit,
Table 13
Frequencies of Number of Alternatives Rated Impossible by Booklet Form, Rating Stage and Hypothesis Confirmed

<table>
<thead>
<tr>
<th>Hypothesis Form</th>
<th>Numbera</th>
<th>Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Categorical</td>
<td>0</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Quantitative</td>
<td>0</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Categorical</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Quantitative</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Categorical</td>
<td>0</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Quantitative</td>
<td>0</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>

a Subjects with two alternatives rated impossible were included with those who had only eliminated one.

Table 14
Estimated Odds in Favor of No Alternatives Rated Impossible by Rating Stage and Hypothesis Confirmed

<table>
<thead>
<tr>
<th>Stage</th>
<th>Hypothesis</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>25.0 : 1</td>
<td>8.2 : 1</td>
<td>5.8 : 1</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>10.4 : 1</td>
<td>3.4 : 1</td>
<td>2.4 : 1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5.6 : 1</td>
<td>1.8 : 1</td>
<td>1.3 : 1</td>
<td></td>
</tr>
</tbody>
</table>
The results, aside from indicating that subjects who selected the final A question were less likely to eliminate alternatives, showed that there was a tendency to increase the proportion of eliminated hypotheses at the later stages.

The estimated odds against the elimination of any hypotheses appear in Table 14. In spite of the just-mentioned result, the odds were always at least 1.3 to 1 against any hypotheses being eliminated. Even with this very inclusive definition, there was never any stage, no matter which hypothesis was confirmed, at which the majority of subjects eliminated even one of their hypotheses.

The relationship between the number of hypotheses investigated and the number of alternatives eliminated at the final stage was also examined. It was expected that a subject who had only investigated one hypothesis (i.e., chose the question pertaining to the same hypothesis at each stage), would be more likely to eliminate alternatives when compared with subjects who had investigated two hypotheses or those who had had information on all three. All definitions and restrictions were as in the previous paragraph.

Log-linear analysis of the data (Number of hypotheses
investigated by Alternatives eliminated by Form) appears in Table 15. This analysis treated Alternatives eliminated as a response variable and added a constant of .5 to each cell to get rid of the zero sub-marginal at A=3, N=3. The model, NF, NA (likelihood ratio Chi-square=8.89, p=.1796, Chi-square=8.6, p=.1972, df=6) indicated that these variables were indeed significantly related, independent of form type. In general, the fewer the number of hypotheses investigated, the more likely it was for alternatives to be eliminated. In spite of this, 61.5% of subjects who investigated only one hypothesis still had not eliminated any alternatives, and only 20% of this group had eliminated both.

Thus, although subjects tended to lower the ratings of alternative hypotheses, it was unusual for any of these hypotheses to be deliberately abandoned, and more rare still for both alternatives to be eliminated.
Table 15
Frequencies of Alternative Hypotheses Rated Impossible by Booklet Form and Number of Hypotheses Explored

<table>
<thead>
<tr>
<th>Form</th>
<th>Alternatives</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categorical 0</td>
<td>17</td>
<td>37</td>
<td>28</td>
<td>82</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>18</td>
<td>9</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Quantitative 0</td>
<td>23</td>
<td>60</td>
<td>35</td>
<td>118</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>20</td>
<td>2</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>65</td>
<td>140</td>
<td>74</td>
<td>279</td>
<td></td>
</tr>
</tbody>
</table>
Opinion variable:

The next analysis examined subjects' agreement with a statement implying that only the cause confirmed by the subject's final question selection was involved in causing the event described in the booklet scenario. This variable was investigated as a function of form type, number of hypotheses explored, and final question selected. Three levels of the second explanatory variable were used, corresponding to whether the subject had, after making all three question selections, chosen questions relating to one, two, or all three hypotheses. The response variable was originally constructed with five categories: "almost certainly true," "probably true," "unknown," "probably false," and "almost certainly false." There were, however, only thirteen responses in the last two categories (12 and 1 responses respectively) and so, they were combined with the "unknown" responses whenever they appeared.

The data appear in Table 16. The fitted model FNQ, NO, QO had 26 degrees of freedom with a Likelihood ratio Chi-square of 27.16 ($p=.4008$) and a Pearson Chi-square of 23.68 ($p=.5943$). These results may be interpreted as stating that there were no form effects but that the number of hypotheses explored and final question selected (which determined the content of the statement) each had an independent effect.
Table 16
Frequencies of Opinion Variable Responses by Final Question
Selected, Booklet Form, and Number of Hypotheses Explored

(N=336)

<table>
<thead>
<tr>
<th>Question Form</th>
<th>Opinion</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Certain</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Categorical</td>
<td>Probable</td>
<td>10</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>1</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>A</td>
<td>Certain</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Quantitative</td>
<td>Probable</td>
<td>11</td>
<td>22</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>1</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Certain</td>
<td>11</td>
<td>18</td>
<td>2</td>
</tr>
<tr>
<td>Categorical</td>
<td>Probable</td>
<td>2</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>Certain</td>
<td>11</td>
<td>27</td>
<td>1</td>
</tr>
<tr>
<td>Quantitative</td>
<td>Probable</td>
<td>10</td>
<td>23</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Certain</td>
<td>2</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Categorical</td>
<td>Probable</td>
<td>3</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>Certain</td>
<td>9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Quantitative</td>
<td>Probable</td>
<td>5</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

\(^a\) "Unknown" responses include "probably false" and "almost certainly false" responses.
The QO effect was due to the fact that the B hypothesis had more "almost certainly true" responses and fewer unknown" responses relative to the A hypothesis.

The lack of form effect, although in keeping with previous findings, is especially noteworthy here in that the wording of the statements under the two conditions reflected not only a difference in the severity of the event described but also implied greater effects for the implicated cause. Thus, the categorical booklets contained statements in the form "If X had not occurred, the student would not have done poorly on the mid-term exam" where X was the cause implicated by the final piece of information. The quantitative booklets on the other hand, used the form "If X had not occurred, the student would have done well on the mid-term exam."

The most important finding here is that the number of hypotheses explored affected subjects' responses. Specifically, those who explored all three hypotheses were significantly more likely to have answered "unknown" and less likely to have answered "almost certainly true" as compared with those who only investigated a single hypothesis.

The category that included subjects who had had exposure to confirmation of two hypotheses produced no distinct results
on the Opinion response variable. On closer inspection of the original data, it was found that the subjects first question selection had very little effect. Thus, with regard to affecting the response variable, there were only really two distinct groups, defined by whether the second and final question selections related to the same hypothesis. The groups which explored two hypotheses seemed to be a mixture of two type of subjects.

It should be noted however that, in spite of the effects of the scope of search, even those subjects who had investigated all three hypotheses only had 30% of their responses falling into the (collapsed) "unknown" category. In other words, 70% of this subgroup felt that other causes were almost certainly or probably not operating.

In summary, most subjects were willing to agree with a statement that implied that only a single cause was operating. This tendency varied, to some extent, as a function of the number of hypotheses explored. Even so, all subjects had at least a 70% agreement rate regardless of prior search scope.
Recommend Variable:

Subjects were asked whether they felt that the information that they had collected allowed them a sufficient understanding of the event to confidently recommend corrective action or that additional information was deemed necessary. Subjects answered different sections of this question depending upon which of the above two situations they felt best described their feelings. Within each subsection, subjects were then asked to either list corrective action or indicate what additional information was required.

These responses were classified as belonging to one of eight categories. Three categories were used to represent those responses that related to only one of the three hypothesized causes. An additional three categories were used to describe a reference to any one of three possible pairs of causes. The seventh category was reserved for subjects whose responses implicated all three causes. The eighth category was used for responses which were too general or ambiguous to be clearly identified with one or more of the hypothesized causes.

The reliability of these classifications was checked by having an independent rater classify the responses of 75 subjects which had been randomly selected from the sample.
Agreement was found in 67 of the 75 cases (89.3 percent). Given the fact that chance agreement was only expected to be 1 in 8, these results were taken as evidence of substantial agreement. Subsequent analyses were based on the writer's classifications alone.

A total of 39 subjects (11.6 percent of the total sample) were excluded from analysis because of the ambiguity of their responses. With regard to the remaining sample, 254 subjects (85.2 percent) indicated that they felt that they fully understood why the event occurred. This group consisted of 187 (62.8 percent) who recommended action relating to a single hypothesis, 48 (16.1 percent) whose recommendations related to two hypotheses, and 19 (6.4 percent) whose recommendations implicated all three hypotheses. For the sake of further analyses, these latter two categories were collapsed into a "multiple action" group.

Of the 44 subjects (14.8 percent) who felt that further information was required, 24 (8.1 percent) wanted information relating to one hypothesis, 11 (3.7 percent) wanted information on two hypotheses, and 9 (3.0 percent) requested further details relating to all three hypotheses. These three categories were all collapsed into one "need information" group.
Three explanatory variables were included in the log-linear analysis; Number of hypotheses explored, Form type, and Hypothesis confirmed as a result of final question selection. The wording of the response variable was the same for all subjects within form type. Those receiving the categorical booklet however, were asked to recommend corrective action if the information that they had selected had enabled them to fully understand why the student in the scenario "did poorly" on his mid-term exam. The phrase in quotations was "in the quantitative booklet, changed to "received the lowest grade in the class."

The data appear in Table 17. The fitted model NPH,RH,NR had 26 degrees of freedom and a Likelihood Ratio Chi-square of 28.10 (p=.3537) and a Pearson Chi-square of 25.47 (p=.4927). Once again there were no effects of form type but final question selection and number of hypotheses explored did affect the response variable.

Subjects who had confined their information search to exploration of a single hypothesis were more likely to recommend intervention related to a single hypothesis and less likely to recommend multiple interventions when compared with subjects who had explored all three hypotheses.

The effect of final question selection was somewhat complex.
Table 17
Frequencies of Recommend Variable Responses by Final Hypothesis Confirmed, Booklet Form, and Number of Hypotheses Explored (N=298)

<table>
<thead>
<tr>
<th>Hypothesis Form</th>
<th>Recommend</th>
<th>Number Explored</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Categorical</td>
<td>One Action</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Multiple Actions</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Need Information</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>Quantitative One Action</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Multiple Actions</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Need Information</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>One Action</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Multiple Actions</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Need Information</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>Quantitative One Action</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>Multiple Actions</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Need Information</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>One Action</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Multiple Actions</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Need Information</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>Quantitative One Action</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Multiple Actions</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Need Information</td>
<td>1</td>
</tr>
</tbody>
</table>
The B question group had significantly more single cause action recommendations than did the A group, but significantly fewer multiple cause recommendations than did the C group. The A group had significantly more requests for additional information than did the C group.

The effect of the first question selected seemed stronger in this analysis than in the Opinion variable results. For example, if T were the hypothesis confirmed by the final question selection, and A any other hypothesis, then a ATT group tended to request additional information or make multiple action recommendations more often than a TTT group.
Relationships Among Variables:

This final section is concerned with the relationship among variables which appear to measure aspects of the subjects' lack of interest in the possible involvement of additional causes.

The first analysis considers four such variables. The first variable, Crossout, refers to the number of hypotheses rated below 17 units and marked as being too unlikely to be worthy of further consideration. Subjects who had so rated any of their hypotheses at the initial evaluation were eliminated from this analysis. Remaining subjects were dichotomized according to whether they had eliminated at least one hypothesis in this way.

The second variable, Match, measured whether the highest standardized usefulness rating given to additional information belonged to that hypothesis rated highest at the final evaluation. Subjects were removed from this analysis if their highest rated hypothesis was either tied with any other or if it was not the same hypothesis confirmed by the final question selected. This latter restriction was necessary to coordinate these responses with those of the Opinion variable (discussed in a later paragraph), the wording of which was determined by the final question selected.
The Match variable is very close to the measure of truncated search used by Shaklee & Fischhoff (1982). The major distinction between the two measures is that, in the Shaklee and Fischhoff study, the subjects being offered the additional information were explicitly told that one of the causes was "known" to be operating. With regard to the Match variable, however, subjects were not excluded from this analysis if their highest rated hypothesis was not rated certain, and thus these subjects may not be strictly said to have "known" of that hypothesis' involvement.

The justification for the inclusion of these subjects is twofold. First, the final piece of information received was actually proof of the definite involvement of that cause, and thus, in this sense, was equivalent to telling subjects that it was a "known" cause. Subjective certainty on the part of the subject may be too strict a definition in that it is likely that at least some of the subjects in the Shaklee and Fischhoff study were also uncertain, in spite of what they had been told. The comparative degree of certainty is not precisely known. On the one hand, subjects in the present study may have been more prone to uncertainty as a result of having been exposed to information that confirmed alternative hypotheses. On the other hand, it is possible that the feedback from a selected question may have been more convincing than simply being told to assume certainty.
The issue of certainty may not however be of great importance even if differences between the samples were found. Previous analyses have shown that "certain" hypotheses lead to very much the same sort of search behavior as do those that are ranked highest, but not certain. Thus there is no evidence to suggest that the validity of the Match variable would depend on the subjective certainty of the leading hypothesis.

The third variable, Opinion, measures agreement with a statement which could only be true if there were no causes operating other than the one confirmed by the final question selected. Subjects' responses were dichotomized according to whether they expressed agreement with the statement ("almost certainly true" or "probably true").

The final variable, Recommend, measures subjects' willingness to recommend intervention, based on what should rationally have been perceived as inadequate information. Three categories were used according to whether subjects had recommended intervention directed towards a single cause, two or three causes, or felt that further information was required. Subjects who recommended action or requested information that could not be clearly related to at least one of the potential causes were excluded from analysis.

The crosstabulation appears in Table 18. The fitted model
Table 18
 Frequencies of Opinion Variable Responses by Recommend and Match Variables and Number of Alternatives Crossed Out at the Final Stage (N=222)

<table>
<thead>
<tr>
<th>Match</th>
<th>Recommend</th>
<th>Cross Out</th>
<th>Agree</th>
<th>Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>None</td>
<td></td>
<td>39</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Some</td>
<td></td>
<td>23</td>
<td>0</td>
</tr>
<tr>
<td>One Action</td>
<td>None</td>
<td></td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Some</td>
<td></td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Multiple Action</td>
<td>None</td>
<td></td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Some</td>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Need Information</td>
<td>None</td>
<td></td>
<td>59</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Some</td>
<td></td>
<td>27</td>
<td>2</td>
</tr>
<tr>
<td>Yes</td>
<td>None</td>
<td></td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Some</td>
<td></td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>One Action</td>
<td>None</td>
<td></td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Some</td>
<td></td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Multiple Action</td>
<td>None</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Some</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


OR, C had 17 degrees of freedom with a Likelihood Ratio Chi-square of 18.85 (p = .3374) and a Pearson Chi-square of 14.50 (p = .6312). In spite of the fact that each of these four measures varies as a function of the number of hypotheses explored, the model indicates that only the Opinion and Recommend variables are significantly associated with one another. Specifically, subjects who agreed with the stated opinion were more likely to recommend intervention related to a single cause and less likely to recommend multiple interventions compared with those subjects who did not agree. There was no difference between the groups in the proportion of subjects who did not feel that they had enough information to recommend action.

In this last analysis, the final rating was classified according to the number of hypotheses rated Certain (i.e., marked as being beyond practical doubt and at least 48 units) in order to examine its relationship to the Opinion and Recommend variables (as previously defined).

As in the previous analysis, data were eliminated if the Recommend response was not clearly related to one or more of the hypothesized causes or if the highest rated hypothesis at the final stage was not that confirmed by feedback from the third question selected. In addition, subjects were eliminated from analysis if one or more of the hypotheses rated certain, were also rated certain initially (i.e., before
any information was seen).

The results which appear in Table 19, are fit by the model RO,RC, (likelihood ratio Chi-square=3.09, p=.7980, Pearson Chi-square=2.99, p=.8099, df=6). The RO interaction indicates, as in the previous analysis, the association between agreement with the stated opinion and the recommendation of a single intervention, and between an "unknown" response and the recommendation of multiple corrective actions. The RC interaction is due to the tendency for those who have a single certain hypothesis to recommend a single action, whereas those with more than one hypothesis rated certain tended to recommend multiple corrective actions.

The lack of association between the number of hypotheses rated certain and the Opinion variable is somewhat surprising. The RO interaction suggests, at least among those few subjects in this analysis that gave an "unknown" response, that their lack of agreement was due to concern for the involvement of alternatives. Then, those with multiple certain hypotheses would be expected to produce a much lower proportion of agreement responses.

The reason this did not occur is that a majority of subjects with multiple certain responses followed an unusual pattern of exploring all three hypotheses but only rating the first two explored as being certain. Thus, these subjects were
Table 19
Frequencies of Recommend Variable Responses by Opinion Variable Responses and Number of Hypotheses Rated Certain at the Final Stage (N=221)

<table>
<thead>
<tr>
<th>Certain</th>
<th>Opinion</th>
<th>Recommend</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>One Action</td>
<td>Multiple Actions</td>
<td>Need Information</td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>Agree</td>
<td>26</td>
<td>9</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>One</td>
<td>Agree</td>
<td>116</td>
<td>14</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>4</td>
<td>8</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Two</td>
<td>Agree</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

The "Two" category includes responses of subjects who rated all three hypotheses certain.
largely excluded from this analysis. An examination of the excluded subjects with multiple certain hypotheses showed that only 8 of the 15 had given an agreement response to the Opinion variable.

One possibility then, (discussed further in a later section) is that it required both an uncertainty about the discussed cause as well as certainty about the involvement of both alternatives to produce a reasonable probability that a subject would not agree with the stated opinion.

Results then, have shown that subjects tended to display both serial and truncated search. Ratings of alternative hypotheses decreased at each stage, but were, nevertheless, seldom ruled out of consideration. In spite of this, most subjects were willing to agree with a statement that implied that only a single cause was operating. In addition, most subjects indicated that they had encountered enough information to fully understand why the event occurred, and then went on to suggest that intervention relating to a single cause would be sufficient to insure that the event did not reoccur. The responses to the Recommend variable showed some relationship to the Opinion responses and to the number of hypotheses deemed certain. All variables were related to the scope of the subject's search, but the size of this relationship was fairly small in all cases.
Discussion

Serial Search:

The results of this study confirm the "serial" nature of search, defined by Shaklee and Fischhoff as the initial clarification of one cause before consideration of any others. In that study, it was found that subjects tended to choose to investigate that hypothesis initially made to appear most plausible. The implication is that, since the results of such an investigation, in the case of a multiply caused event, will likely yield information compatible with the hypothesis in question, that hypothesis should continue, at subsequent stages, to be seen as being more plausible than any of the other alternatives. As such, it should continue to be investigated, producing a series of questions, all relating to the same hypothesis.

This study was able to demonstrate both this relationship between the relative rank of a hypothesis and question selection, and the similarity of sequential questions. With regard to the first relationship, subjects were most likely to investigate a hypothesis if it was rated more highly than any of the others, and there was also a tendency, after the completion of question selection, to rate additional information relating to such a hypothesis as being most
useful.

There was also evidence that this process tended to maintain the relative superiority of the investigated hypothesis, both by increasing its own rating, and by leading to a decrease in the ratings of alternatives. Stability of ranking, being dependent upon receiving compatible feedback, would thus not necessarily be present in the case where events had a single cause, since there would then be less chance of randomly selecting a correct hypothesis. As such, investigations would not necessarily have a serial character, unless some sort of bias in the interpretation of results was also acting.

The fact that serial search tendencies were found in this study is both a replication of the Shaklee and Fischhoff finding, and an extension of these results to a true sequence of investigations. The original demonstration of serial search showed only that subjects preferred a hypothesis related question, or a set of such questions at a single selection. Strictly speaking, a search could not be said to be serial unless as in the present study this tendency persisted through feedback and subsequent selection stages.

Evidence presented here showed that the question selected at the first stage was related to the second question selected
and it, in turn, was related to the final selection. It might however, still be argued that if the similarity between successive questions decreased over time, relatedness might eventually disappear over a long sequence of investigations. It was not possible to directly test the validity of this idea by examining the stationarity of the inter-stage transition matrix because of the changing relative attractiveness of questions at each stage. It was possible, however, to examine the stability of the underlying rank by selection interaction, and this relationship was found to be equal at all stages, at least with regard to the B and C hypotheses' data. The A data's second order interaction was found to derive from a weaker effect (lower relative effect of being ranked highest) at the first stage. These particular results then, seem to argue for a strengthening of the relationship at the later stages. Thus, for the sample as a whole, the tendency to choose, at a given selection, the question related to the highest ranking hypothesis, remained at least as strong at the later stages.

If looked at as a within subjects measure however, there is some suggestion that the tendency to limit search to a single hypothesis, may in some cases, increase with each selection. For example, it was shown that not only was the selection of a question at the third stage most likely for those who had also selected it at the second stage, but it
was more likely still for those who had, in addition, selected it at the first stage. Thus search tendencies were dependent upon more than just the immediate past selection. Similarly, it was shown that subjects whose second and third selections related to the same hypothesis, were more likely to rate the (standardized) usefulness of further information as being highest for that hypothesis, as compared with those who had chosen to investigate that hypothesis at the third stage only. It would appear then, that the more subjects chose to investigate a particular hypothesis, the more likely they were to continue to investigate it in the future.

As a final point, it should be noted that the rank-question selection relationship can be seen as a result of subjects following a confirmatory search strategy. Viewed in this way, the most highly rated hypothesis is the one under most active consideration (perhaps the only one), and the questions selected are those which promise to provide evidence supporting this hypothesis. A subject wishing to clarify the situation by use of disconfirming strategies would instead have been expected to have selected questions relating to the lower ranked hypotheses. If the answers received from those questions were not compatible, it would have allowed an opportunity to eliminate the weaker explanations to which they related. Alternatively, if the results were confirmatory, this might have been seen as a
challenge to the major hypothesis, at least among those subjects who were unwilling to consider the possibility of several simultaneously operating causes.

Thus, results indicate that confirmatory search patterns were employed and that this led to a tendency to investigate the same hypothesis at consecutive question selections.
Truncated Search:

The significance of serial search rests upon the demonstration of "truncation", defined as the termination of search after the role of one cause has been suitable clarified. Thus, a subject displaying a serial, truncated search, would tend to investigate only one cause and then stop the investigation, leaving other potential causes unexplored.

Truncation, in this study operationalized as the persistence of the rank-selection relationship after the involvement of a hypothesis was judged to be certain, was indeed found to occur. This was determined both by analyzing the pattern of questions selected by those who became certain before seeing all three pieces of information, as well as by examining the rated usefulness of additional information for those who were certain after receiving feedback from all of their questions. In each analysis, subjects who were certain of a hypothesis' involvement seemed as interested in further information about it as did those who had rated that hypothesis highest but not certain. Since certainty will most likely result from information received from the preceding investigation, this finding is equivalent to saying that certainty does not interrupt the tendency to search in a serial manner.
Defining truncation as the continuation of serial search requires some comment. A more intuitive definition might involve a measurement of the subject's inclination to terminate all further search, but all such measures would be liable to overestimate disinterest in alternatives by including responses which reflect a more general apathy about the task as a whole. The definition used here, and in the Shaklee and Fischhoff study, is a much more strict criterion in that it assumes that subjects, who might under natural circumstances have stopped investigating, would, nevertheless, still avoid exploring alternatives if forced to continue their collection of information.

This amounts to saying that subjects, in general, prefer to examine redundant or unnecessary information rather than widening the scope of their search. The validity of this measure then, rests upon the assumption that the information chosen, was in fact, not needed. Shaklee and Fischhoff simply told their subjects that one of the hypotheses was "known" to be involved, but it is not clear whether such a manipulation alone would have been enough to remove all doubt about its role. The present study, in relying upon the subjects' own ratings was less likely to overestimate subjective certainty.

Two other controls were added to this study. First, subjects were asked to estimated the effect that the certain
involvement of one hypothesis would have had on others. The fact that only a minority of subjects saw certainty of one hypothesis as making alternatives less likely, reduces the risk that the specific hypotheses used in this study were perceived as being mutually exclusive or negatively related. In other words, subjects did not fail to investigate alternatives because they were seen as being incompatible with the hypothesis deemed certain.

If hypotheses were seen as being perfectly independent of one another, the probability of an alternative hypothesis operating, given the certain involvement of some other hypothesis, would simply be the baseline probability of that alternative. Another possibility then, is that subjects ignored alternatives because their baseline probabilities were considered to be too low to be worthy of serious consideration. Data from the final evaluation however, renders such an explanation untenable. Almost all subjects had at least one alternative that had not been ruled out of consideration at the final stage, and a large majority had indicated that both alternatives were still being considered. It is not possible to judge whether these final ratings tended to be below their perceived baseline probabilities, since no unconditional estimates were taken. It is clear however, that these alternatives were, for the most part, still considered to be reasonably likely.
Scope of Search:

The basic assumptions of the Shaklee and Fischhoff article are that a truncated search will leave the investigator with an incomplete account of a multiply caused event, and that it is this limited exposure to information that will lead to difficulties in dealing with this event.

Both of these assumptions are however open to question. With regard to the first point, it is important to note that, although subjects display a tendency to limit their search to an investigation of a single hypothesis, this is not the same thing as assuming that only one type of confirmatory information will be encountered. It is clear that the extent to which investigators would be drawn to explore other hypotheses should, at a minimum, vary as a function of the number of alternatives, their relative plausibility, and the accessibility and richness of information about them. In the case of this study, although subjects always had a higher probability of choosing a particular question if they had previously chosen it, that probability was nevertheless, often quite low. Moreover, after attempting to control for question differences, it was found that only 57.6% of subjects with one certain hypothesis indicated that further information about that hypothesis would be more useful than information related to either of the two non-certain alternatives. At that rate, only 19.1% of subjects would
have been expected to have chosen all three questions relating to the same hypothesis, assuming that all alternatives had been equated for perceived attractiveness.

It is difficult to estimate how search would actually proceed under naturalistic conditions where questions are unlikely to be equated and where selections are made from a total set of questions containing an unequal number relating to each hypothesis. It is possible that, since all sources simultaneously vie for attention, rather than, as in this study, being limited to nonoverlapping choice subsets, there may be an increased chance that several types of information will be seen.

Still, there are several reasons to believe that naturalistic search may be somewhat more narrow than demonstrated in this study. For example, it is often the case that one hypothesis is, in fact, on the whole, much easier to investigate than others. In addition, most casual investigations may not delineate alternatives as explicitly as those in this study, nor is it certain that investigators would be able to formulate questions to test these alternatives as readily as they might devise ways to examine the leading hypothesis.

The scope of a search then is likely to vary from one situation to another. To the extent that the search is a
casual and brief investigation where confirming information relating to the major hypothesis is easily available, it is very probable that alternatives would not be explored. However, in the case of a more consequential investigation where a number of reasonably plausible alternatives were explicitly acknowledged, it seems much more likely that a variety of information would come to light in the course of the investigation.
Effects of Scope of Search:

The previous section suggested that the tendency to engage in a truncated search, is only one of many factors which determine the scope of search. The scope in turn, been shown to affect a number of measures which reflect the subjects' tendency to ignore alternative hypotheses. The connection between truncation and the ignoring of alternatives is thus indirect. This is a likely explanation for the finding that the "match" variable, a measure of truncation independent of actual feedback, was unrelated to either the "crossout", "opinion" or "recommend" variables.

The purpose of this section however, is to argue that even the more direct scope of search variable, plays only a limited role in determining how subjects treat alternatives.

Those subjects who limited their search to an investigation of only one hypothesis, tended to rule out of consideration more alternatives at the final evaluation than did those who had investigated two or three hypotheses. The former group was also relatively more likely to strongly agree with the statement in the "opinion" variable, and to feel confident that a single action would guarantee that the student, described in the scenario, would not fail his final exam.

In spite of these findings, the same general pattern of
responses were found for all groups, regardless of the scope of their search. It was found that even among those who had seen only one kind of information, 61.5% of subjects still had not ruled out any of their hypotheses at the final stage, and only 20% had eliminated both. Although these figures do indicate a relative disinterest in alternatives, it must still be acknowledged that the actual proportion of subjects expressing no further interest, even by this very liberal definition, is quite small. A tendency was noted for the number of ruled out hypotheses to increase at each stage, presumably as a result of the increasing confirmation of the investigated hypothesis. Since the final confirmation was probably as strong as one could normally expect to encounter, it is unlikely that the proportion of eliminated alternatives would rise much beyond those levels seen at the final evaluation.

It was also found that the group that had investigated a single hypothesis tended to have a somewhat greater proportion of subjects who strongly agreed with the stated conclusion, and fewer that were unsure or disagreed. In spite of this, agreement was still the norm for all groups. Among those who had seen all three types of information, "probably true" was the modal response, leaving only a minority of subjects unsure or in disagreement.

With regard to the "recommend" variable, the vast majority of
subjects in all groups seemed to feel that they had a complete understanding of all causal forces and were thus in a position to recommend corrective action. A greater scope of search did not, even in a relative sense, diminish this overall confidence, but rather seemed to encourage the recommendation of multiple, as opposed to single, interventions. Since the majority of such multiple recommendations made reference to only two causes, this typically amounted to ignoring one potential cause.

It should also be noted that, given the nature of the information made available, subjects often seemed willing to recommend corrective action relating to a cause which was not decisively implicated.

In summary then, encountering information suggestive of the existence of more than one cause did have some effect on behavior. In spite of this, the average subject, although hesitant to directly rule out hypotheses, was willing to act as though such alternatives were of no consequence. This was demonstrated not only by their agreement with a stated conclusion, but also in regard to their own decisions about corrective actions.

The possibility exists that the apparent robustness of these findings, in the face of varying degrees of diverse information, is simply an artifact of the manner in which
the information made available increased in confirmatory strength at each stage. Thus it could be argued that prior information was simply overwhelmed by the more compelling evidence encountered at later stages.

To the extent that investigators would, under natural conditions, simply stop their information search after becoming certain of a cause, and to the extent that that certainty would result from a particularly strong piece of confirmatory evidence, it is likely that the final information considered would be, as in this study, the most persuasive.

More importantly, although the strength of confirmations increased at each stage, the earlier pieces of information, (especially those that the second stage) were far from trivial. There is evidence that these earlier confirmations, in some cases, actually overwhelmed later information. It was earlier reported that those subjects who had investigated a wider range of hypotheses were less likely to have rated any of their hypotheses as being certain,(judged by very inclusive criteria). The final piece of evidence was however worded in such a way as to remove all practical doubt as to the involvement of that hypothesis. Thus, it can only be assumed that those subjects who did not rate the hypothesis considered last to be certain, were not fully processing that information, and
that receiving prior information relating to other hypotheses, increased the likelihood that this would occur.

Another point should be made about the scope of search variable, and that is that it measures not only the type of information seen, but also all those factors acting on individuals which predispose them to adopt their particular pattern of search. Thus, it can not be assumed that the type of person who chose to exclusively investigate only one hypothesis, would have reacted in the same way to other information, as did those who actively sought it out.

This does not necessarily weaken the validity of this measure since it is likely that, under natural conditions, information will be available only to those who actively pursue it. Even if this were not the case however, there would still be little reason to expect that those subjects now receiving the information, would be more affected by it than those who sought it independently. On the contrary, it is more likely that those those who actively chose to investigate a particular hypothesis would be more willing to accept their findings. Thus, the limited effects of scope of search might have been even less substantial if information had been passively received.

In conclusion then, the scope of search seemed to have had only minor effects on Opinion and Recommend variable
responses. It would appear that encountered information, influenced indirectly by some single-cause or sufficiency heuristic (discussed further in later sections), is then evaluated by this heuristic. Apparently the decisions produced by this heuristic were only slightly dependent upon the scope of the considered information.
Opinion Variable:

Subjects' ratings seemed to indicate that even those who had investigated only a single hypothesis, nevertheless still kept open the possibility that alternatives were operating. This apparent concern for alternatives did not, however, carry over to the more applied Opinion and Recommend variables.

The Opinion variable can be likened either to a casually offered assessment of a situation, having no particular implications, or instead as a definite proposal raised for discussion. Viewed from the latter perspective, it is possible to imagine a diagnostic committee in which some member gives his opinion as to what is causing the event in question in just such a manner. If general agreement could be obtained, it is reasonable to assume that discussion might then end, and an official recommendation (i.e. on behalf of the committee) be made, based on an assumption of a single cause.

Insofar as this variable does mirror behavior in such a situation, it is very likely that group consensus would be reached. Results here indicated that even subjects with exposure to a wide range of confirmatory evidence (including those who objectively rate multiple causes as being certainly involved) were still likely to respond with some
degree of agreement, a response which implied the absence of additional causes. In addition, although the B hypothesis (perceived initially as being the most plausible) led to a greater proportion of subjects who strongly agreed with the statement, general agreement was still the norm regardless of the hypothesis explored at the final stage.

The only situation in which this is not the rule occurs when subjects are both less than certain about the putative single cause and, in addition, are very confident about the involvement of two alternative causes. When this situation does not occur, only a small minority of subjects withhold agreement, presumably out of a relatively rare concern for the involvement of other causes. A wide scope of search although limited in its effect, is nevertheless associated with an increase in both types of subjects.

If all three possible causes are denoted $X$, $Y$ and $Z$, then a statement in the form "If not $X$, then no effect" implies both "$X$ is a cause" and "$Y$ and $Z$ are not". One possibility then, is that subjects needed to have doubts about both assertions in order to be reasonable likely to disagree. This conclusion is open to question however, in that the rating responses of the group which met these criteria were most atypical and this may also be true of their judgement processes. In addition, the wide scope of search associated with this group may, with respect to the
Opinion variable, be less causal, and more a response itself to some set of initial assumptions.

Because the form of the Opinion variable was determined by the final question selected, most subjects felt at least that this particular hypothesis was more likely than not, and so probably tended to agree with the "X is a cause" implication. Another possibility then, is that the belief in the involvement of X, in some way, overrides doubts about the non-involvement of Y and Z. Thus, it may be the case that, that when subjects both agree and disagree in part, that the agreement judgements get more heavily weighted. This type of process is similar to that reported by Gollob, Ross, and Abelson (1973) who found that confirming instances generally have more impact on inductive conclusions than do disconfirming instances. In this respect, the stated opinion is analogous to a hypothesis, and the subject's belief in the involvement of X is a piece of confirmatory data which is given more weight than the incompatible (disconfirming) beliefs about the involvement of Y and Z.
Recommend Variable:

This variable required the subject to first judge the adequacy of the information that they had encountered, and then decide either on appropriate action or specify further information needed.

Rationally speaking, the very best one could have done was to encounter information which definitely implicated once cause and gave less conclusive support to the two others. Therefore, an appropriate response should have been to indicate that there was not enough information on which to base action. This however, was a most unusual response. Most subjects instead indicated that they had enough information to completely understand why the event occurred. If this pattern of confidence had been observed only in that group of subjects who had explored all three hypotheses, it may have been possible to interpret this result as indicating a low threshold of practical certainty; that subjects were simply willing to act on less than definitive evidence. As will be discussed later, this may well be true, but it is not adequate to explain why the proportion of similarly confident subjects is as least as high within the groups that had explored only one or two hypotheses. Nor does a low threshold theory account for the fact that the vast majority of recommendations were for less than three actions. This issue is discussed further in a later section.
Given the decision to proceed with recommendations, a wider scope of search did, as mentioned in a previous section, increase the probability that multiple actions would be advocated. Within the group that had explored two hypotheses however, 86.2% of such recommendations made reference to only two actions, and even within the group that had explored all three hypotheses 59.4% of multiple actions were limited to implicating two causes. Thus, these recommendations must still be judged inadequate by a simple rational standard. Similarly, although investigation of the strongest hypothesis (B) led to more single action recommendations, even the weakest hypothesis (A) led to very few requests for further information.

Returning now to the low threshold possibility, it would appear that subjects often showed a willingness to act on hypotheses not rated as being certain. For example, among subjects who had not rated any of their hypotheses as being certain, 58% nevertheless felt confident enough to recommend a single action, and a further 23% recommended multiple actions. It should be noted that this latter figure is higher than that for subjects with a single certain hypothesis (only 13% of whom recommended multiple actions). This can be explained by the fact that a relatively large proportion of subjects with no certain responses, nevertheless had rated more than one hypothesis as being likely (i.e. above 32 units). This may be taken as additional
evidence that the threshold for inclusion in action is often less than the absence of all practical doubt.
Single Cause Heuristic:

Many of the results of this study are consistent with the idea that subjects started their investigations with an implicit assumption that the event must have had only a single cause. Under such a "single-cause" heuristic, the decrease in the ratings of alternatives and the larger number of such alternatives ruled out of consideration at later stages, could be seen as the reciprocal relationship expected for ratings of mutually exclusive events. In addition, if there is only one possible cause, there would be no need to continue search after one seemed certain (truncated search). The relative absence of more than one hypothesis rated certain, the general agreement with the Opinion variable, and the confidence of complete understanding displayed on the Recommend variable are also compatible with this assumption. It would also explain why a wide scope of search increased the proportion of subjects who did not rate any of their hypotheses as certain. From this point of view, a series of data confirming different hypotheses should have been expected to produce confusion about which one of the causes was "really" involved.

The most convincing justification for the operation of a "single-cause" heuristic comes indirectly from the work of Kelley (1972). Kelley argued that observers employed a "multiple sufficient scheme" when it was believed that any
one of several causes alone could have produced an effect. Thus, in the case of two potential causes X and Y, the presence of either one was assumed to account for the effect. The implication of this was, since the causes were not mutually exclusive, that if the effect occurred and X was known to have been present, there should be ambiguity about the role of Y since, in this situation, either its presence or absence was compatible with the presence of the effect.

A number of studies however, found that subjects often considered the presence of Y given X to be not merely ambiguous or uncertain, but distinctly unlikely (Kun, 1977; Kun and Weiner, 1973; Shultz, Butkowsky, Pearce, and Shanfield, 1975; Smith, 1975). Uncertainty about the presence of a second cause was however reported by Shultz and Butkowsky (1977). That study found that approximately 50% of subjects judged a second cause to be present, and 50% judged it to be absent given prior knowledge of the presence of another sufficient cause.

If X and Y are independent, then the probability of Y given X should be equal to the baseline or unconditional probability of Y. Kelley’s “multiple sufficient scheme” seemed to amount to nothing more than that, and therefore the degree of uncertainty experienced should be determined by the baseline probability of Y. In spite of this, none of
the above mentioned studies controlled for this probability.

One study which did however, was Kun and Murray (1980). Those authors found that the conditional probability of $Y$ given $X$ was, on average, below the baseline probability of $Y$. To explain this finding, Kun and Murray referred to Kelley's discussion of "graded effects" in which he noted that "if the attributor distinguishes degrees of a given effect, he is likely to assume that it increases with the cumulative strength of relevant causes" (p.158). On the basis of this, Kun and Murray argued that subjects are sensitive to quantitative or qualitative variations in the effect under consideration, and that they make an initial estimation as to whether one or more causes are likely to be operating. This was termed the "variant effect scheme".

Kun and Weiner (1973) had previously found support for Kelley's assertion that a multiple sufficient scheme was employed to account for common events, but unusual or extreme events were likely to elicit a "multiple necessary scheme" (i.e. an assumption that both $X$ and $Y$ would be needed to cause the effect). Presumably then, in extreme events, the subject judges that he is observing more than one cause "worth" of effect, and therefore implicates both $X$ and $Y$. Conversely, in most common events, subjects judge that they are observing only a single cause "worth" of effect and, as such, he can be quite sure that $Y$ is not operating, given
the presence of X.

In this study, the median rating of alternative hypotheses at the final evaluation was 17 units in the group which had explored only one hypothesis. This figure, is halfway between "impossible" and "neutral" and can be taken to mean than subjects were somewhat sure that alternatives were not involved. A strict interpretation of the variant effect scheme however, should have implied that subjects would be more than somewhat sure of non-involvement (even though none of the studies cited above reported certainty of non-involvement). Perhaps this can be explained in terms of the difficulties in precisely determining the number of causes operating. Also, there may be some hesitancy on the part of subjects to rate causes as being either absolutely certain or impossible. In addition, the nature of this study, which called for multiple reassessments of the same cause makes the operation of "anchoring" effects quite likely (Tversky and Kahneman, 1974). Since the initial evaluation of an alternative starts off high, it is possible that the final rating is biased in an upward direction.

Similarly, the rated independence of causes noted earlier is not incompatible with a variant effect scheme in that it requires, not that events be mutually exclusive, but only that they be treated as such, given the judgement that only one effect "worth" of effect has been observed.
Other objections to this interpretation, however, are less easily dismissed. First, few hypotheses were eliminated at the final evaluation. The measure used did not call for a belief in the strict impossibility of the hypothesis, but only that, in a practical sense, the hypothesis was no longer worthy of further consideration. It is true that this reluctance to dismiss alternatives was not always reflected in other measures, but the literature cited in support of the variant effect scheme has usually relied on similar sorts of ratings as the major dependent variable.

Second, although it might be unreasonable to expect that ratings would sum precisely to 64 units at each stage, nevertheless it should follow that no more than one of the three (presumably exhaustive and mutually exclusive) causes could simultaneously be rated above "neutral", assuming that such a rating was taken to mean that there was a .5 probability of involvement. In fact, when all three hypotheses were explored, 36.3% of subjects rated two of their hypotheses over 32 units, and 48.8% rated all three hypotheses as such.

Similarly, it is difficult to explain why a wide scope of search should often have resulted in the recommendation of multiple actions if only one was assumed to be operating.

Perhaps the strongest objection to the variant effect scheme
is the absence of "form" effects found in this study. If subjects discriminate among effect sizes, then they ought to have reacted to the categorical effect as a common event requiring only a single cause explanation. The quantitative effect however, should have been seen as an unusual event, and thereby evoked a multiple necessary scheme. It is conceivable that the absence of differences between the two forms was due to the fact that subjects included in their understanding of "did poorly" that case where the student received the "lowest grade in the class". But, since the opposite merging is not logically consistent, (i.e. "the lowest grade in the class" can not logically accommodate all instances of "did poorly") it should have followed that both forms would have evoked multiple necessary schemes.

One possible explanation comes from Kun and Weiner (1973) who found that subjects were somewhat more likely to attribute a failure to the action of a single cause, even in the case of an unusual or extreme failure. Even so, in that study the ratings of alternatives in the extreme failure scenario were higher than those in the common failure scenario. Moreover, since the subjects for this study were drawn from an introductory psychology course consisting of approximately 500 students, it seems hard to believe that they could have looked upon receiving the "lowest grade in the class" as being anything but an extreme failure.
Sufficiency Heuristic:

Another way of accounting for the results of this study would be to assume, not that subjects necessarily expected a single cause solution, but rather that they were willing to accept an incomplete solution as sufficient. Any sufficiency heuristic consistent with obtained results would have to consist of at least two sub-strategies. It would have to seek to account for both the event and other information encountered in the exploration of that event, and it would have to try to do so as parsimoniously as possible.

The first point implies that in general, no attempt would be made to account for information not encountered, nor would the implications of such information be considered (i.e., "out of sight, out of mind"). Such a strategy would explain a preference for narrow search since, initially, although any hypothesis would be sufficient to account for the event, later only the investigation of the previously explored hypothesis could also promise to account for the just encountered information. Thus, a subject who explored only a single cause would still be in a position to sufficiently explain, not only the event, but also the three pieces of information encountered.

Similarly, it would explain the actions of the large proportion of subjects who indicated that they completely
understood why the event occurred. Viewed from this perspective, such an assertion would mean only that the subject felt completely confident that he could account for the information he had seen.

The relative lack of hypotheses ruled out of consideration is also expected since it requires only that subjects not be willing to make judgements in the absence of encountered information. In other words, an actual decision that a particular cause was not involved would have necessitated the encountering of information incompatible with the existence of that cause. On the other hand, even though not explicitly ruled out of consideration, a cause might well still be ignored because a sufficient explanation was possible without it.

The question of why the ratings of non-confirmed hypotheses tended to decrease, is not directly addressed by the operation of a sufficiency heuristic. Presumably, and in contrast to the findings of Kun and Murray (1980), this decrease represents a return to baseline estimates for that hypothesis. If a particular hypothesis had, in the past, been the object of confirmatory information, its unconditional probability estimate would be expected to increase. Even so, this higher figure may still have been less than the estimate of involvement when none of the three causes had yet been rendered certain. This relationship
would be particularly likely if attempts were made to dismiss prior confirmatory evidence, a strategy discussed further later in this section. In other words, it is likely that the baseline probability of an alternative, even one which had received some measure of prior confirmation, was lower than the probability of its involvement initially.

The exclusive concern with encountered information has also been found in work on judgement of covariation among events (e.g. Jenkins and Ward, 1965; Smedslund, 1963; Ward and Jenkins, 1965). In these studies, subjects were presented with a series of 2 x 2 contingency tables, which described each of the two variables as being either present or absent. A common failing was the almost exclusive reliance upon the size of the "present-present" cell relative to the entire sample, in determining covariation, thus ignoring the implications of events that did not occur. This is also similar to the finding in the concept utilization studies that people prefer to use positive instances of concepts over negative ones (e.g. Hovland and Weiss, 1953).

It is also possible that this strategy accounts for one early demonstration of the "representative" heuristic termed the "law of small numbers" (Tversky and Kahneman, 1971). This phrase referred to the inordinate confidence that researchers often had in data that was based on a very limited sample. An alternative conceptualization of this
phenomenon might be that investigators strive only to explain what they have encountered (i.e. a small sample), and therefore , their confidence would depend only on the consistency of obtained results and their ability to explain these results.

The second component of a sufficiency heuristic, the seeking of parsimony, is useful in explaining a number of results. For example, the majority of subjects who encounter information confirming two causes, nevertheless do not recommend action relating to both of them. Also it was unlikely that a subject who had investigated all three cause would recommend three interventions. Thus, it was possible for subjects to ignore even information encountered, and this suggests that subjects try to arrive at a satisfactory account of the problem with the fewest number of explanatory factors. Thus, a wide scope of search makes it more difficult to maintain a simple, single cause, explanation, and this is why multiple actions are often recommended (e.g. when all three hypotheses have been explored). On the other hand, since it was relatively easy to dismiss the first, rather weak, confirmation, relatively few of these multiple recommendations implicated all three causes.

An agreement response on the Opinion variable was also more likely if the subject has attempted to devalue information
which confirmed alternatives. Even so, the notion of disproportionate weighting discussed in an earlier section, is still required to explain agreement in cases where alternatives have not been adequately dismissed.

Encountered information may be dismissed in at least three different ways. One strategy is what Elstein et al. (1978) termed "overinterpretation" (originally mentioned as support for the use of a single cause heuristic). This amounts to assuming that information of diverse origins has been brought about by the same cause. This "collapsing" of information would be relatively easy to accomplish if it was possible to posit a common cause post hoc, or if the pieces of information were general events which commonly occur as a result of a wide variety of causes. This study, however, defined all causes a priori. In addition, the format of question selection emphasized the unique relevance of each piece of information to a particular hypothesis, and the subjects' understanding of this is reflected in the differential impact of information on confirmed and alternative hypotheses. Thus, it is unlikely that many subjects were able to dismiss information in this manner.

A second approach would be to attribute encountered information to some cause which had no relationship to the effect in question. For example, subjects who were informed that the student did not think that the material covered in
his course would be of any use to him in the future, may have been able to assume that this was due the nature of pre-established career plans, rather than because he was bored. Since having non-related career plans is not generally considered a cause of failure, this information could be ignored in accounting for the failure.

This approach runs up against some of the same objections mentioned previously with regard to "overinterpretation". Still, instructions for the task included the fact that not all information was necessarily relevant to all hypotheses. This could have also been interpreted to mean that some information would not be related to any of the listed hypotheses.

The third strategy would be to redefine the sufficiency of a hypothesized cause. For example, a subject may be willing to admit that lack of interest in a course could account for failure in that course, and therefore, the certain discovery that the student had no interest would be taken as an explanation of the effect. In the case where the failure has already been imputed to a lack of language skills however, the certain discovery of lack of interest may no longer necessarily imply to the subject that it was also causally involved, because, it could be argued, that this condition does not inevitably result in failure. Thus, a subject when faced with definite evidence for more than one cause, may
succeed in eliminating one or more by adopting a more stringent definition.

It is possible that this strategy accounts for the fact that a wide scope of search often results in none of the hypotheses being rated certain at the final evaluation. Subjects may have been convinced by both the second and third pieces of information, but after adopting the stricter definition of causality, both hypotheses remain likely, but neither now seems certain.

As a final point, it should be noted that the lack of form effect is not incompatible with the use of a sufficiency heuristic. In contrast to a single cause heuristic which would determine the number of causes which are most likely to be operating, a sufficiency heuristic would be concerned only with the minimal number of causes that could be operating. Although the event described in the quantitative form was more likely than that described in the categorical form to have been multiply caused, both events had the same minimal causal estimate. In other words, it was possible that even the quantitative event had only one cause.
Closing Remarks:

The tacit assumption up to this point has been that an optimal approach to investigating an event entails a thorough examination of all potential causes. This however, makes no allowances for the specific goals of the investigator, or the practical costs involved in such thoroughness.

There are obviously many reasons for conducting an investigation, and not all necessarily require such a complete understanding. For example, if the goal was to be able to recreate an event, (as opposed to having the power to prevent its reoccurrence), it would be enough to identify a single, (manipulatable) sufficient cause.

Similarly, there might be quite specific interest in the involvement of only certain causes. Thus, if in an investigation into an unusual accumulation of wealth, it was found that this was due, at least in part, to a misappropriation of funds, it is unlikely that there would then be any interest in determining the additional contributions of hard work, business acumen or luck.

But even in those cases where an investigation of all causes would be desirable, it is not likely that it would be practically attainable. First, it is not clear that it would
even be possible to draw up a complete list of potential causes. Assuming however, that an acceptably complete list could be produced, the difficulties associated with investigating each cause might well be overwhelming. In all probability, some sub-optimal form of investigation would have to suffice, determined by such factors as the number of potential causes, their likelihood of involvement, their estimated contribution to the effect, and the ease with which they could be explored or manipulated.

Perhaps the most important factor in determining the sufficiency of an incomplete solution is the relative costs of being in error. If errors can be corrected by subsequently taking into account previously overlooked factors, it would usually be wasteful to examine every possibility initially. Moreover, if it was discovered that the original understanding of the event was inadequate, an examination of the failed intervention may very well provide clues as to the precise nature of the overlooked cause(s), thus adding in another way, to the efficiency of this approach.

Thus, a determination of the extent to which an investigation should be exhaustive should be achieved by balancing the costs of such an investigation against those incurred as a result of acting on incomplete information.
The behaviour of most subjects in this study would, if applied to typical problems of this sort, probably result in reasonably efficient solutions. Within the context of this particular task, however, such behavior seemed needlessly reckless.

This is so for two reasons. First, although there were practical limitations which prevented a comprehensive investigation, there was never any emphasis placed upon arriving at a solution. In fact, the instructions were quite explicit in saying that some subjects would not have an opportunity to gain a complete understanding of the event, the implication being that the inability to reach a decision might be an appropriate state. In spite of this, it was quite unusual for a subject to indicate either that he did not know if the Opinion statement was true, or that more information was required before actions could be confidently recommended.

The second point is that it was clear from the nature of the task, that an inadequate intervention would have unacceptable consequences. The scenario description stated, not only that the student was in danger of failing his course, but that such a failure could lead to his being dismissed from the university and having to return to his native country. Given the severity of these consequences, and the fact that the failure could not be subsequently
corrected, it was essential that no causes be ignored.

It may have been that subjects were uncomfortable with suspending judgement. If this were the case, an acceptable solution in the presence of inadequate information, would have been to take a cautious approach by recommending interventions relating to all potential causes. Although there was some tendency to recommend that a cause be acted upon even if it were not deemed certain, nevertheless, it was seldom recommended that that all potential causes be acted upon.

Another possibility is that subjects saw their actions as only an initial step; a way of beginning early action on those causes known to be operating, to be followed (were the investigation allowed to continue) by an investigation of causes whose involvement had not previously been determined. Although this may have been true of some subjects, it will not do as a general explanation since it is clear from the results of the Match variable that, after receiving all three pieces of information, there was disproportionately little interest in unexplored causes.

In conclusion then, acting upon less than complete information, although perhaps often the most reasonable approach, seems less defensible when there is an option to postpone judgement and where the consequences of an
unsuccessful intervention would be irreparable. It is the presence of both of these conditions in this study that marks the behaviour of subjects as inappropriate. Apparently the tendency to act upon incomplete information is strong even when the usual advantages of this approach are not at a premium.
References


Lakatos, L. Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge*. London: Cambridge


Log-linear analysis is a versatile method of simultaneously dealing with several dimensions of categorical data. It assumes, much like the ANOVA procedure that all departures from a "grand mean" (in this case, a mean frequency) are due to the actions of certain "effects".

For the sake of illustration, consider a two by two crosstabulation of Eye Colour (brown, blue) and Sex (male, female) for a sample of 100. In the absence of any effects, there would be a uniform or "mean" frequency (M) of 25 for all cells. Departures from this figure of 25 could occur for three reasons. First, the numbers of males and females might not be equal. Second, there might not be equal number of brown and blue eyed people. Third, there might be some interaction between the two variables. These effects are symbolized (using the first letter of the variable) as S, E, and SE respectively. Thus a complete ("saturated") description as to why cell frequencies differ from 25 requires taking into account the S, E and SE terms. Taking the log of frequencies allows an additive "model" to be stated as $M + S + E + SE$.

In practical terms, a single term, such as E, implies that the marginal frequencies for that variable will not be equal. A first order interaction term such as SE implies
that the "odds-ratio" is not equal to 1, or in this case, that the proportion of blue-eyed males is not the same as the proportion of blue-eyed females. Similarly a second order interaction implies that the first order interaction between two of the variables is not equal at all levels of the third.

Each model has associated with it a set of expected cell frequencies. The expected values for the saturated model are the same as the observed frequencies, but models with less terms have expected values that differ from those observed. In an attempt to see whether a table can be described more parsimoniously, terms may be removed and a "goodness of fit" statistic can be calculated to test the similarity of observed and expected values. This is usually done by using the Pearson Chi-square or Likelihood Ratio Chi-square, both of which are asymptotically distributed as chi-square.

It can now be seen that the normal Pearson Chi-square test of independence is just a special case of this technique. In terms of the example, it would be equivalent to testing the fit of the model $M + S + E$. If there is no interaction, or if this effect is not very large, a reasonably good fit will be obtained, signified by a statistic value which does not exceed its critical value.

Theoretically, any term or set of terms may be eliminated to
test for the fit of the remaining effects. In practice however, only hierarchical models are considered. This means that, a term may not be eliminated if a higher order term containing that effect remains in the model. For example, the term $E$ could not be removed from the model if the $SE$ term remained. For this reason, it is sufficient to specify only the highest order terms which contain a particular effect. Thus $ABC$ is understood to stand for a model which contains the terms $M,A,B,C,AB,AC,BC$, and $ABC$.

If the intent of the analysis is to describe the effects of some set of "explanatory" variables on one or more "response" variables, then all terms which describe the relationships among the explanatory variables are included in all models tested. For example, in a four way table $A$ by $B$ by $C$ by $D$, with $D$ as a response variable, all models would include the $ABC$ term. The marginals of the explanatory variables are, in this case, considered to be fixed.

When dealing with a dichotomous response variable it is common to specify models in terms of the log of ratios, rather than the log of cell frequencies. For example, if the response variable was voting behavior with 30 people voting and 10 not voting, the "logit" or "log-odds" of the voting marginal would be the log of 3 (30 divided by 10). Estimated odds for any marginals or submarginals can be calculated based upon the fitted model, and correspond to
the ratio of expected values that would have been obtained had a non-logit analysis been conducted.

Within the appropriate constraints then, the initial objective of a log linear analysis is to find the most parsimonious model that provides an acceptable fit to the data. A concise description of this technique can be found in Fienberg(1980) and a more elementary introduction is available in Knoke & Burke(1980).
II: Booklet Form

The following material was presented to subjects in the form of a 41 page "scrambled" booklet. The wording on these pages represents the "categorical" form. Differences between this and the "quantitative" form are detailed in the Design section.

Nineteen pages of redundant, misleading, or instructional material has not been included. For example, the Recommend variables and Usefulness / were each repeated three times but differed only in their subsequent page references. Misleading information included not only answers to various questions, but also the pages containing the series of questions from which those answers could be chosen. Instructional pages were largely reminders about previous instructions such as how to rate the hypotheses and how to find information in the booklet.

The four rating sheets have been included at the end of the booklet.
Consider the following situation.

A university student from another country came to America to study. In his home country, he had already taken a few courses, but these had all been in subjects that had particularly interested him.

As a condition of being allowed to study in the United States, the student had to demonstrate that he was able to read and understand university level books written in English (although it was not necessary for him to demonstrate that he could understand spoken English).

The student was not eligible for any financial aid but did manage to get a part-time job which earned him enough money to pay for his tuition and living expenses.

In one course he has done poorly on the mid-term exam, and if he does not do better on the final exam he will fail this course.

A failure in any course would be very serious and could lead to the student having to leave the university and return to his native country.

Your task is to decide, as best as possible, why the student is doing poorly so that appropriate advice can be given to him as to how to improve his grades on the final.

GO TO PAGE 4.
The following possible causes (hypotheses) were the most frequently mentioned by previous people participating in this study.

ASSUME THAT ALL OTHER POSSIBLE EXPLANATIONS HAVE BEEN RULED OUT and that only the following hypotheses need be considered in trying to figure out why the student is doing poorly in this course.

1) The student had no interest in the course. He found it boring and, as a result, avoided studying it.

2) The student, although able to read English, had trouble with listening comprehension. Because of this, he couldn't understand what was said in lectures.

3) The student's job was too demanding for him. As a result, he became too exhausted and sick to study properly.

GO TO PAGE 5
The student works 20 hours per week at a job which involves quite a lot of heavy physical labor.

Now that you know this information, rate EACH of the potential causes according to how likely it seems to you that it was involved in causing the student to do poorly on the mid-term exam.

Use the Rating Sheet; Page 2 for this.

Remember to make only one mark on each line.

If you feel that any of the potential causes is so likely (even if not certain) that you no longer have any serious doubt about it, circle the letter of that cause.

If you feel that any potential cause is so unlikely (even if not impossible) that you no longer consider it to be a serious possibility, cross out (x) the letter of that cause.

Go to page 38
Second piece of information

The student's boss has often noticed that he looked tired and ill. Because of this, the student has frequently been sent home early to rest.

-----------------------------------------------------------------

Now that you know this information, rate EACH of the potential causes according to how likely it seems to you that it was involved in causing the student to do poorly on the mid-term exam.

USE THE RATING SHEET; Page 3 for this.

Remember to make only one mark on each line.

If you feel that any of the potential causes is so likely (even if not certain) that you no longer have any serious doubt about it, circle the letter of that cause.

If you feel that any potential cause is so unlikely (even if not impossible) that you no longer consider it to be a serious possibility, cross out (x) the letter of that cause.

GO TO PAGE 23
The student had great difficulty holding a conversation in English, largely due to the fact that he could not understand very much of what was being said to him. In lectures, he understood almost nothing of what the professor said.

Now that you know this information, rate each of the potential causes according to how likely it seems to you that it was involved in causing the student to do poorly on the mid-term exam. USE THE RATING SHEET; Page 4 for this. Remember to make only one mark on each line.

If you feel that any of the potential causes is so likely (even if not certain) that you no longer have any serious doubt about it, circle the letter of that cause.

If you feel that any potential cause is so unlikely (even if not impossible) that you no longer consider it to be a serious possibility, cross out (x) the letter of that cause.
If you were told for certain that some OTHER student has not been interested in the material covered in the course, what effect would this information have on the likelihood that this person's English was not good enough to understand what was said in the lectures?

(Circle One)

1. It would make it more likely.
2. It wouldn't have any effect; the two are unrelated
3. It would make it less likely.
4. It would make it impossible.

What effect would this information have on the likelihood that the student's job was making him too tired to study properly?

(Circle One)

1. It would make it more likely.
2. It wouldn't have any effect; the two are unrelated
3. It would make it less likely.
4. It would make it impossible.

END OF TASK
HAND IN YOUR PAPER
If you were told for certain that some OTHER student had trouble understanding the English used during lectures, what effect would this information have on the likelihood that he was not interested in the course material (i.e. and would not have been interested in it even if it was taught in his native language)?

Circle one

1. It would make it more likely.
2. It wouldn't have any effect; the two are unrelated
3. It would make it less likely.
4. It would make it impossible.

What effect would this information have on the likelihood that the student's job was making him too tired to study properly?

Circle one

1. It would make it more likely.
2. It wouldn’t have any effect; the two are unrelated
3. It would make it less likely.
4. It would make it impossible.

END OF TASK

HAND IN YOUR PAPER
If you were told for certain that some OTHER student has been too tired to study properly, what effect would this information have on the likelihood that he wasn't interested in learning the course material (even if he had been feeling well)?

1. It would make it more likely.
2. It wouldn't have any effect; the two are unrelated
3. It would make it less likely.
4. It would make it impossible.

What effect would this information have on the likelihood that his English was not good enough to understand the lectures?

1. It would make it more likely.
2. It wouldn't have any effect; the two are unrelated
3. It would make it less likely.
4. It would make it impossible.
Taking into account all three pieces of information that you have now received, how necessary would it be to know the answer to each of the following questions in order to understand why the student did poorly on the mid-term exam?

Rate each of the questions from 1 to 5 as follows:
1) totally useless
2) not very useful
3) somewhat useful
4) very useful
5) essential

(put the number in the space at the end of each question)

a) How often has the student read the university newspaper during lectures in this course?:--------

b) How many other members of his family are also studying in America at other universities?:-------

c) How entertaining does the student find it to watch American television (English speaking programs)?:-------

d) On average, how many times per week does the student buy his lunch at the cafeteria?:-------

e) On average, how many hours sleep does the student get each night?:-------

f) On average, how many letters a week does he receive in the mail?:-------
The student often felt tired, and has had headaches almost every day during the semester. He has, on several occasions, gone to doctors complaining that these symptoms made it impossible for him to concentrate properly on his schoolwork.

Now that you know this information rate each of the potential causes according to how likely it seems to you that it was involved in causing the student to do poorly on the mid-term exam.

USE THE RATING SHEET; Page 4 for this.

Remember to make only one mark on each line.

If you feel that any of the potential causes is so likely (even if not certain) that you no longer have any serious doubt about it, circle the letter of that cause.

If you feel that any potential cause is so unlikely (even if not impossible) that you no longer consider it to be a serious possibility, cross out (x) the latter of that cause.

GO TO PAGE 19
What is your opinion of this conclusion?

If the student had had good English listening comprehension skills, he would not have done poorly on the mid-term exam.

Circle one

1. almost certainly true
2. probably true
3. unknown
4. probably false
5. almost certainly false
Final Question Selection

Choose one question from those below that you now feel would be most useful to have answered.

a) Did the student spend less time studying the material from this course than other courses?

b) Did the student have difficulty holding a conversation in English?

c) Did the student have any health problems during the semester commonly associated with fatigue?

Write the letter of the question you wish to choose on the RATING SHEET; Page 4.

If you wish to choose question A, GO TO PAGE 37

If you wish to choose question B, GO TO PAGE 11

If you wish to choose question C, GO TO PAGE 21
The student did significantly better on the part of the exam that tested his knowledge of the textbook than he did on the part which examined material given in class. This pattern also occurred in another course which was divided into text and lecture based sections.

Now that you know this information rate each of the potential causes according to how likely it seems to you that it was involved in causing the student to do poorly on the mid-term exam.

USE THE RATING SHEET: Page 3 for this.

Remember to make only one mark on each line.

If you feel that any of the potential causes is so likely (even if not certain) that you no longer have any serious doubt about it, circle the letter of that cause.

If you feel that any potential cause is so unlikely (even if not impossible) that you no longer consider it to be a serious possibility, cross out (x) the letter of that cause.
What is your opinion of this conclusion?

If the student had been interested in the course, he would not have done poorly on the mid-term.

Circle one
1. almost certainly true
2. probably true
3. unknown
4. probably false
5. almost certainly false
First piece of information

The student is very quiet when in the company of American students but is more talkative when with other foreign students from his own country.

Now that you know this information rate each of the potential causes according to how likely it seems to you that it was involved in causing the student to do poorly on the mid-term exam.

USE THE RATING SHEET; Page 2 for this.
Remember to make only one mark on each line.
If you feel that any of the potential causes is so likely (even if not certain) that you no longer have any serious doubt about it, circle the letter of that cause.
If you feel that any potential cause is so unlikely (even if not impossible) that you no longer consider it to be a serious possibility, cross out (x) the letter of that cause.

GO TO PAGE 38
In order to investigate why the student did poorly on the mid-term exam, choose one question from those below that you feel would be most useful to have answered.

a) Was the course compulsory? Did the student have to take it as a requirement of his degree?

b) Was the student quiet and not very talkative when in the company of American students?

c) Did the student's part-time job involve long hours or strenuous work?

Write the letter of the question you wish to choose on the RATING SHEET; Page 2.

If you wish to choose question A, GO TO PAGE 33

If you wish to choose question B, GO TO PAGE 27

If you wish to choose question C, GO TO PAGE 8
What is your opinion of this conclusion?

If the student had not been so tired, he would not have done poorly on the mid-term exam.

Circle one
1. almost certainly true
2. probably true
3. unknown
4. probably false
5. almost certainly false
The course was compulsory. It had to be taken as a requirement of the student's degree.

Now that you know this information, rate EACH of the potential causes according to how likely it seems to you that it was involved in causing the student to do poorly on the mid-term exam.

USE THE RATING SHEET; Page 2 for this.

Remember to make only one mark on each line.

If you feel that any of the potential causes is so likely (even if not certain) that you no longer have any serious doubt about it, CIRCLE the letter of that cause.

If you feel that any potential cause is so unlikely (even if not impossible) that you no longer consider it to be a serious possibility, CROSS OUT (x) the letter of that cause.

GO TO PAGE 38
The student feels that the subject matter of this course is unlikely to be of use to him in the future.

Now that you know this information, rate EACH of the potential causes according to how likely it seems to you that it was involved in causing the student to do poorly on the mid-term exam.

USE THE RATING SHEET; Page 3 for this.

Remember to make only one mark on each line.

If you feel that any of the potential causes is so likely (even if not certain) that you no longer have any serious doubt about it, circle the letter of that cause.

If you feel that any potential cause is so unlikely (even if not impossible) that you no longer consider it to be a serious possibility, cross out (X) the letter of that cause.

GO TO PAGE 23
Final piece of information

The student has only put in one half the time studying this course that he has on the others. He often says that he finds the subject matter of the course so boring that when he sits down to study, he always puts the work of other courses first.

Now that you know this information, rate EACH of the potential causes according to how likely it seems to you that it was involved in causing the student to do poorly on the mid-term exam.

USE THE RATING SHEET; Page 6 for this.

Remember to make only one mark on each line.

If you feel that any of the potential causes is so likely (even if not certain) that you no longer have any serious doubt about it, circle the letter of that cause.

If you feel that any potential cause is so unlikely (even if not impossible) that you no longer consider it to be a serious possibility, cross out (x) the letter of that cause.

GO TO PAGE 31
Second Question Selection

Choose one question from those below that you now feel would be most useful to have answered.

a) Did the student feel that the information taught in this course would probably not be useful to him in the future?

b) Did the student do better on the part of the exam based on information from the text than he did on material testing comprehension of what was said in the lectures?

c) Did the student’s boss at work ever notice that he looked tired?

Write the letter of the question you wish to choose on the RATING SHEET: Page 3.
If you wish to choose question A, GO TO PAGE 35
If you wish to choose question B, GO TO PAGE 24
If you wish to choose question C, GO TO PAGE 9
If you feel that the booklet that you selected gave you enough information to fully understand why the student did poorly on the mid-term exam, what could you suggest to help the student? In other words, what should the student do to make sure that he doesn't fail the final exam?

If you feel that the booklet that you selected did NOT give you a chance to fully understand why the student did poorly on the mid-term exam, what information do you feel would have been required for such an understanding?
AGE: __

SEX: MALE FEMALE (circle one)

HYPOTHESES

A. -CIRCLE if have no serious doubt
   -CROSS OUT if not serious possibility
   NO INTEREST IN SUBJECT /
   DID NOT WORK AS HARD
   AS COULD HAVE

   LIKELIHOOD RATINGS (make one mark on each line between brackets)
   (---------) IMPOSSIBLE NEUTRAL CERTAIN

B. -CIRCLE if have no serious doubt
   -CROSS OUT if not serious possibility
   ENGLISH COMPREHENSION POOR/COULDN'T UNDERSTAND LECTURES

   IMPOSSIBLE NEUTRAL CERTAIN

C. -CIRCLE if have no serious doubt
   -CROSS OUT if not serious possibility
   TOO TIRED/COULDN'T STUDY PROPERLY

   IMPOSSIBLE NEUTRAL CERTAIN

AFTER FINISHING THIS PAGE, RETURN TO BOOKLET
FIRST QUESTION SELECTED: ___

<table>
<thead>
<tr>
<th>HYPOTHESES</th>
<th>LIKELIHOOD RATINGS (make one mark on each line between brackets)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. CIRCLE if have no serious doubt -CROSS OUT if not serious possibility NO INTEREST IN SUBJECT/DID NOT WORK AS HARD AS COULD HAVE</td>
<td>IMPOSSIBLE NEUTRAL CERTAIN</td>
</tr>
<tr>
<td>B. CIRCLE if have no serious doubt -CROSS OUT if not serious possibility ENGLISH COMPREHENSION POOR/COULD’NT UNDERSTAND LECTURES</td>
<td>IMPOSSIBLE NEUTRAL CERTAIN</td>
</tr>
<tr>
<td>C. CIRCLE if have no serious doubt -CROSS OUT if not serious possibility TOO TIRED/COULD’NT STUDY PROPERLY</td>
<td>IMPOSSIBLE NEUTRAL CERTAIN</td>
</tr>
</tbody>
</table>

AFTER FINISHING THIS PAGE, RETURN TO BOOKLET
SECOND QUESTION SELECTED: 

<table>
<thead>
<tr>
<th>HYPOTHESES</th>
<th>LIKELIHOOD RATINGS (make one mark on each line between brackets)</th>
</tr>
</thead>
</table>
| A. -CIRCLE if have  
no serious doubt  
-CROSS OUT if not  
serious possibility  
NO INTEREST IN SUBJECT/  
DID NOT WORK AS HARD  
AS COULD HAVE  
(IMPOSSIBLE | NEUTRAL | CERTAIN)  
B. -CIRCLE if have  
no serious doubt  
-CROSS OUT if not  
serious possibility  
ENGLISH COMPREHENSION  
POOR/COULDN'T UNDERSTAND LECTURES  
(IMPOSSIBLE | NEUTRAL | CERTAIN)  
C. -CIRCLE if have  
no serious doubt  
-CROSS OUT if not  
serious possibility  
TOO TIRED/COULDN'T STUDY PROPERLY  
(IMPOSSIBLE | NEUTRAL | CERTAIN)  

AFTER FINISHING THIS PAGE, RETURN TO BOOKLET
### FINAL QUESTION SELECTED: ___

#### HYPOTHESES

| A. -CIRCLE if have no serious doubt -CROSS OUT if not serious possibility |
| NO INTEREST IN SUBJECT/ DID NOT WORK AS HARD AS COULD HAVE | LIKELIHOOD RATINGS (make one mark on each line between brackets) |
| 📅 | ⏰ | 🕒 |
| IMPOSSIBLE | NEUTRAL | CERTAIN |

| B. -CIRCLE if have no serious doubt -CROSS OUT if not serious possibility |
| ENGLISH COMPREHENSION POOR/CAN'T UNDERSTAND LECTURES | |
| 📅 | ⏰ | 🕒 |
| IMPOSSIBLE | NEUTRAL | CERTAIN |

| C. -CIRCLE if have no serious doubt -CROSS OUT if not serious possibility |
| TOO TIRED/CAN'T STUDY PROPERLY | |
| 📅 | ⏰ | 🕒 |
| IMPOSSIBLE | NEUTRAL | CERTAIN |

### AFTER FINISHING THIS PAGE, RETURN TO BOOKLET