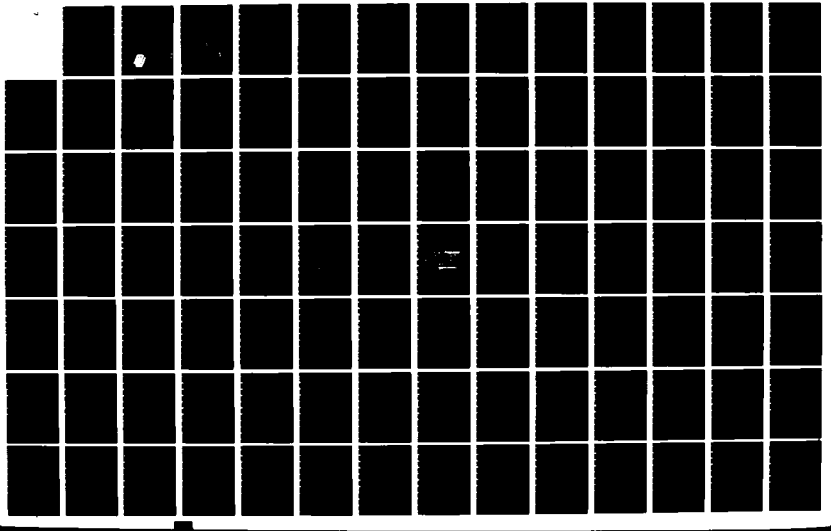
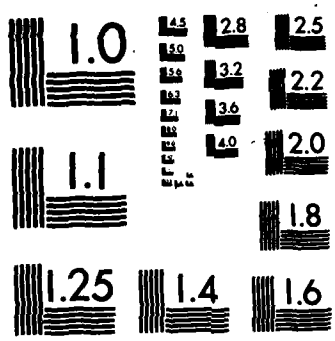
AD-A137 962 RESEARCH AND THEORY ON PREDECISION PROCESSES(U) 1/2
OKLAHOMA UNIV NORMAN DECISION PROCESSES LAB C F GETTYS
30 NOV 83 TR-11-30-83 N00014-80-C-0639

UNCLASSIFIED F/G 5/10 NL





(12)

RESEARCH AND THEORY ON PREDECISION PROCESSES

CHARLES F. GETTYS

TR 11-30-83 NOVEMBER 1983

DECISION PROCESSES LABORATORY
UNIVERSITY OF OKLAHOMA

OFFICE OF NAVAL RESEARCH
CONTRACT NUMBER N00014-80-C-0639
WORK UNIT NUMBER NR197-066

RESEARCH AND THEORY ON PREDECISION PROCESSES

CHARLES F. GETTYS

TR 11-30-83 NOVEMBER 1983

PREPARED FOR

OFFICE OF NAVAL RESEARCH
ENGINEERING PSYCHOLOGY PROGRAMS
CONTRACT NUMBER N00014-80-C-0639
WORK UNIT NUMBER NR 197-066

DECISION PROCESSES LABORATORY
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF OKLAHOMA
NORMAN, OK 73019
(405) 325-4511

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: TR-11-30-83	2. GOVT ACCESSION NO.: AD-A137962
3. RECIPIENT'S CATALOG NUMBER:	
4. TITLE (and Subtitle): RESEARCH AND THEORY ON PREDECISION PROCESSES	5. TYPE OF REPORT & PERIOD COVERED: FINAL
	6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): Charles Gettys	8. CONTRACT OR GRANT NUMBER(s): N00014-80-C-0639
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Decision Processes Laboratory, Department of Psychology, University of Oklahoma, Norman, OK 73019 (405) 325-4511	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR 198-066
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Engineering Psychology Programs, 800 N. Quincy St. Arlington, VA 22217	12. REPORT DATE: 30 November 1983
	13. NUMBER OF PAGES: 97
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release: distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Outcome generation	Causality
Decision theory	Act generation
Mental models	Hypothesis generation
Causal schemata	Predecision processes

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This monograph discusses six years of research and theory building at the Decision Processes Laboratory concerned with predecision processes, the cognitive processes that occur prior to making the actual decision. These processes include problem detection, the process by which the decision maker decides that a problem exists; act generation, the process of creating candidate acts that might solve the problem, hypothesis generation where →

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Block 20 (continued).

various states of the world are identified that might affect the outcomes of various actions; and outcome generation, a process where the possible results or outcomes of actions are generated.

There are nine substantive chapters in the monograph. The first five chapters are concerned with modeling the various predecision processes and describe the empirical research that addresses these models. Chapter 6 is devoted to research on various topics such as schemata, causal explanation, small group research, individual differences, and expertise in various predecision processes. Chapter 7 discusses recommendations for improving predecision performance, including specific attempts to aid the decision maker, and Chapter 8 presents, in summary form, the major conclusions of this program of research. In Chapter 9, general suggestions are made for further research in the area. Also included are titles and abstracts for all technical reports produced in both contracts.

TABLE OF CONTENTS

i. ACKNOWLEDGMENTS

This is the final report of a three-year program of research entitled "The predecision processes of act and outcome generation" sponsored by the Office of Naval Research, Engineering Psychology Programs. The contract between the Office of Naval Research and the University of Oklahoma was N00014-80-C-0639, work unit number NR197-066, Charles F. Gettys, principal investigator. The chapter on problem detection was jointly authored by Charles Gettys, Frank Durso, and Tom Dayton.

This report also discusses ideas and data from a previous ONR contract concerned with a closely related topic, that of hypothesis generation. This contract between the Office of Naval Research and the University of Oklahoma was N00014-77-C-0615, work unit number NR197-040, and was also sponsored by the Engineering Psychology programs of ONR. Several chapters of this report are edited sections from the final report of the hypothesis generation project.

We wish to thank Martin A. Tolcott who, until his recent promotion, was director of the Engineering Psychology Programs. His role as a facilitator of our research was invaluable.

In addition, we would like to thank Bill Vaughan, the new leader of the Engineering Psychology group, and Gerald Maleki and John O'hare of ONR who facilitated our research in various ways. The scientific direction that we received from these individuals was invaluable, and it would be difficult to imagine a more cooperative atmosphere in which to do research.

Various individuals contributed to this program of research at the University of Oklahoma. Becky Pliske, now at the Army Research Institute, was a post-doctoral research associate for the final two years of the present contract. She contributed both ideas and organizational skill to the project, and almost all of the act and outcome generation research profited from her touch.

Much of the research to be reviewed here was conducted by Stanley D. Fisher, Thomas Mehle, Carol Manning, Suzanne Baca, Nancy Shelton, Jeff Casey, and Peter Engelmann. In addition, the chapter on problem detection was co-authored by Tom Dayton, who contributed many theoretical ideas, and a considerable editorial talent to its creation. As graduate students at the University of Oklahoma, they spent countless hours developing theory, and conducting and implementing this research. Any successes that have been achieved are largely due to their inspired and painstaking efforts.

Undergraduate researchers who worked on these projects were Jeff Casey, Dee Ann Fisher (nee' Patterson), Jim Laughlin, Frank Savala, Steve Holmstrom, Paul Foster, Jason Beckstead, Chip Minty, Michelle Kelley, and Julie Thompson. Their efforts, as paid and unpaid volunteers, are especially appreciated. We would also like to thank Dale Umbach and Bradford Crane for their help in mathematical proofs, and Chuck Rice and Jim White of the University Computer Services for software and consulting services, respectively.

Tom Harrison and the staff of the Office of Grants and Contracts provided editorial and secretarial assistance for which we are most

grateful. Glenda Smith, Janie Mooney and the Psychology Department office staff made substantial secretarial contributions, and the late Dick Rubrecht cheerfully repaired our computers "one more time". Kamran Sadeghi furnished electronics help after Dick's death.

We wish to thank Alan Nicewander who collaborated with us on one of our studies, and critically read several of our manuscripts. We also want to acknowledge the help of other members of the faculty of the psychology department. Jack Kanak, Rick Acosta, Steve Smith, Vesta Gettys, Frank Durso, Rich Reardon, Larry Toothacker, and Joe Rodgers who read and commented on several of our manuscripts, and freely contributed their ideas. These individuals also served as consultants and made valuable contributions of ideas from their specialties.

My interest in this topic can be traced to conversations between Mel Moy of the Navy Personnel Research and Development center and myself on the topics of threat detection and crisis prevention. These discussions lead to the conclusion that it would be profitable to model predecision processes.

Charles Gettys
December, 1983

CHAPTER 1. INTRODUCTION

This is the final report for the project "The predecision processes of act and outcome generation" sponsored by the Engineering Psychology Programs, Office of Naval Research. The project began August 15, 1980 and ended September 30, 1983. The goal of this project was to develop theory and to do research on act and outcome generation processes. The strategy employed in this project was to blend concepts drawn from three areas: decision analysis, behavioral decision theory, and cognitive psychology. As part of this project, 18 experiments were conducted, and 9 technical reports were issued concerning the processes of act and outcome generation.

This is not a typical final report. Rather than write a brief overview of the experiments conducted in the present contract, we have chosen to present the gist of our thinking on predecision processes in monograph form. In doing this, we review research and theory developed in our previous hypothesis generation contract, research and theory from the present contract, and speculative theory that we have recently developed on problem detection.

The actual order of development of the three major theories presented here, problem detection, hypothesis generation, and act and outcome generation, was hypothesis generation (1978-1980), act and outcome generation (1980-1983), and problem detection (1983). Our theories for the various predecision processes are in the order in which we believe they come into play in problem structuring, not in the order that we developed them, although such a description might better display the development of our thinking. For example, our ideas on problem detection profited by six years of research on related topics. The problem detection theorizing also is so recent that it has not been the subject of empirical work; we will undoubtedly refine our theory when data is collected. We have also made slight changes to the description of our hypothesis generation model and research to reflect our current thinking on this topic.

The discussion that follows is organized according to topics and does not attempt to explain experimental procedures and results in detail. To attempt this task would result in several hundred more pages of text that would be largely redundant with our previous technical reports. Instead, as various topics are discussed, reference is made to previous technical reports which contain these details, or to reports which contain relevant references to the general literature. So that interested readers can obtain more information, these technical reports are cited using numerals (ie. 1, 5, 9). The titles and abstracts for these technical reports are presented in section 11. The technical reports numbered 1-9 are from the hypothesis generation contract, and those numbered 10-18 are from the act and generation contract.

There are nine substantive chapters in the monograph. The first five chapters are concerned with modeling the various predecision processes and describe the empirical research that addresses these models. Chapter 6 is devoted to research on various topics such as schemata, causal explanation, small group research, individual differences, and expertise in various predecision processes. Chapter 7 discusses recommendations for improving predecision performance, including specific attempts to aid the decision maker, and chapter 8 presents, in summary form, the major conclusions of this program of research. In chapter 9, general suggestions are made for further research in the area of predecision processes.

CHAPTER 2. WHAT ARE PREDECISION PROCESSES?

In general terms, predecision processes are the cognitive processes that occur prior to the final evaluation that leads to a decision. Predecision processes may include the recognition that a problem exists that may require a decision and further action, problem definition and analysis, the generation of possible actions that might possibly solve the problem, the generation of possible states of the world that may affect the outcomes of possible actions, and the generation of outcomes themselves. In the sections that follow, comments are made as to why predecision processes have received relatively little attention, the concept of an ill-defined problem is discussed, and each of these predecision processes are further defined in the context of ill-defined problems.

The traditional focus of decision theory. Decision theory has traditionally focused on the act of deciding itself. Most decision theory inquiries start with fully structured problems, problems where the possible actions, the states of the world that determine the outcomes, and the outcomes are all specified. The techniques of decision theory are applied to the evaluation of outcomes, or the choice of action. However, the structure of the problem is usually a given. This emphasis is probably a historical accident due to the origin of modern decision theory in economics (Von Neumann & Morgenstern, 1947) at a time when psychology had little to offer to the understanding of how decision problems are structured.

One notable exception to this general picture is found in decision analysis. In decision analysis, attempts are made to capture the structure of the decision problem by eliciting this structure from the decision maker (Raiffa, 1968) by using various elicitation techniques. This technology, however, has been created by decision analysts interacting with their clients and adopting techniques that seem to be effective, but there has been almost no research that attempts to understand the cognitive mechanisms used by decision makers when they structure decision problems. Raiffa (1968), for example, stated that explaining how humans develop problem structure was a problem he wanted to "duck".

The importance of understanding predecision processes. The importance of understanding predecision processes should be obvious. As the structure of the decision problem is the model that the decision maker uses in making the decision, the adequacy and completeness of the model determines the quality of the decision to a large degree. These remarks apply with even more force to intuitive decision making. Decision analysis is a collection of informal elicitation techniques which have been adopted because they seem to tease the structure of the decision problem from the client. It seems reasonable to assume that the decision problem structure of the intuitive decision maker is less complete and less adequate than that of a client aided by a decision analyst.

Without understanding the extent to which decision makers can create a problem structure that is isomorphic with reality, the concern with adopting the optimal decision seems to be somewhat pointless. It is a well-known principle of decision theory that optimality is always defined in terms of a model. Therefore, a decision can be optimal in terms of the model, but non-optimal in terms of the actual situation because of a lack

of isomorphism between the model and the situation being modeled. The implications of the results to be presented here are that there is reason to be concerned about how completely a decision maker can structure a decision problem. This concern, if it is valid, suggests that spending a great deal of additional effort studying how to optimize decisions which may be based on incomplete models may be less profitable than spending a comparable effort to understand the extent to which humans can produce good problem structure.

Therefore, there is the distinct possibility that the cart has been put before the horse in the development of decision theory. A more rational approach might be to first study the extent to which decision makers can produce problem structures that are isomorphic with reality. Then, if it can be shown that such isomorphism exists, develop optimization techniques that work with that structure. The purpose of the projects described in this report was to start the study of predecision processes and to determine the extent of the isomorphism between decision models and reality.

Ill-defined problems. Problem structuring is most important in the class of problems that are termed "ill-defined" (Taylor, 1974). These problems are typically non-routine problems for which no "standard operating procedure" exists, and for this reason are often the most challenging and difficult problems that the decision maker has to face. Ill-defined problems are problems which must be formulated in a fruitful manner by creating structure where little or no structure existed previously. Ill-defined problems may be ill-defined because the decision maker's present state is poorly understood, the goal state is poorly understood, or the transformations necessary to move the decision maker from the present state to the goal state are poorly understood. For example, a task force commander may experience a surprise attack in force. Following this attack, the present state may be poorly understood until the commander has damage reports and time to take stock of losses, the goal state may be poorly defined because the original goal may no longer be reachable, and the transformations necessary to reach the original goal or any alternate goal may also be poorly defined due to a lack of information about the commander's present resources.

The extent to which the problem is ill-defined is a major determinant of its difficulty, since the most difficult problems are often those in which the present states, goal states, and transformations are all poorly defined. In these situations, the decision maker must first define the missing parts of the problem structure before the decision can be made.

This is not to say that well-defined problems are necessarily easy. Chess (de Groot, 1965), for example, is played by a rigid set of rules, the beginning state and the goal state are explicitly defined, and the molecular operations to achieve the goal are exactly specified. Despite this structure, it is a difficult and fascinating game because the combinatorial possibilities of the moves are so high.

Brief descriptions of various predecision processes.

In the following paragraphs, brief descriptions of various predecision processes are presented. The processes described below are one possible

categorization of important predecision processes, other categorizations are possible, and no claim is made that these categories are exhaustive. It is also important to note that although these processes are presented sequentially, the decision maker does not necessarily proceed through these processes in a step-wise manner. Rather, it seems much more probable that the decision maker changes from one predecision process to another at will. Thus problem definition, act, hypothesis, and outcome generation may occur repeatedly while thinking about the a problem, as the decision maker discovers new dimensions and ramifications to the problem.

Problem detection. Problem detection is the process that alerts the decision maker to the need to make a decision. Without this process, decisions would not get made because the the decision maker would never realize the necessity of stopping planned activities, and charting a new course of action (Corbin, 1980).

Problem detection, although it creates the opportunity for a new decision, has its roots in earlier decisions and plans for action, and it is in the context of these plans that problems are detected. Problem detection occurs because the decision maker realizes that previous actions are not likely to result in the desired goal. As will be discussed extensively in the chapter devoted to problem detection, we believe that problems are detected by a comparison of the flow of events from the decision maker's environment with the expected course of events-- an **act/event scenario** which is created by the decision maker at the time of taking action. Events which are **anomalous**, or unexpected in terms of the decision maker's scenario may be the stimuli for problem detection.

Problem analysis and definition. Once the decision maker has detected a problem, the nature of the problem can be identified, and this process often suggests a possible remedy. If such a remedy is not obvious, the definition of the problem may be improved by further analysis. Problem definition in ill-defined problems may involve the identification of goals, which may be multiple and conflicting, the identification of problem constraints, and the identification of control variables, or "operators" which may suggest ways to solve the problem (Newell & Simon, 1972). We believe that these problem characteristics are organized by the decision maker into what may be termed a "mental model" (cf. Gentner and Stevens, 1983). The mental model is the decision maker's representation of the decision situation, the mental structure of the problem. It is based in part on causal schemata (Tversky & Kahneman, 1980) which specify the causal relationships between actions which manipulate of the control variables of the problem and possible outcomes of these actions.

Once the problem is defined, the stage is set for act, hypothesis, and outcome generation. However, we imagine that the process of analysis and definition continues throughout the time spent working on a problem. If, for example, undesirable consequences of an act are discovered, this may stimulate further problem analysis.

Although problem analysis and definition is perhaps the most important of the predecision processes, we have devoted only one study to it explicitly, although a number of our studies are indirectly relevant to it. Problem analysis and definition is not treated separately in this paper because it was not a major topic in our projects, However, most of our

research, in a real sense, was concerned with this topic.

Hypothesis generation. Hypothesis generation is important in two contexts. In inductive inference tasks it is the process that generates alternate explanations for data. It is also an important process in outcome generation. In outcome generation, a decision maker should be aware of possible states of the world that may influence the outcomes that result from a particular action, and often may have to generate these hypothetical future states of the world in ill-defined problems. Hypothesis generation probably is also important in problem analysis and definition where the decision maker attempts to generate explanations for anomalous events. The hypothesis generation process in all of these situations probably involves similar mechanisms, however, our research has been almost entirely concerned with generating explanations for data.

Act and outcome generation. These two processes are treated together, as we theorize that outcomes are generated by tracing the possible consequences of actions. We believe that the decision maker generates actions by using a mental model which is created during problem analysis and definition. This mental model may include actions that have been used to address similar problems in the past. In the case of problems that are ill-defined in respect to possible actions, the decision maker may choose to generate additional actions to supplement those that are immediately suggested by similarities between the present problem and other problems that the decision maker has previously solved. How this process may occur is the subject of several studies in the the present contract, and further discussion of act and outcome generation is deferred until the chapter devoted to this topic.

CHAPTER 3. A MODEL OF THE PROBLEM DETECTION PROCESS

"The best laid schemes o' mice and men gang aft a-gley"
-Robert Burns (1759-1796)

Problem detection is the least understood of the decision processes, yet it is one of the most important, as it triggers or initiates the remaining processes. Decision theorists (e. g. Corbin, 1980) have noted its importance but have not proposed models of it, nor systematically studied it. However, several recent theoretical developments in cognitive psychology and behavioral decision theory have made it possible to create models of the problem detection process. These developments include the work on causal scenarios and schemata (Tversky & Kahneman, 1980), the proposal of a simulation heuristic (Kahneman and Tversky, 1971), causality and cues to causality (Einhorn & Hogarth, 1981), and scripts and plans (Schank and Ableson, 1977). These new ideas facilitate the development of a model and taxonomy of problem detection.

In this chapter we develop a working model of the mental processes involved in problem detection--one that is yet to be refined and revised by research. We then describe the problem detection taxonomy we are developing by examining our model in relation to the problem detection task environment.

A problem detection model.

Problem detection as a cyclic process. Problem detection cannot be understood in isolation from the other decision processes. Decision making is cyclic. After detecting a problem the decision maker decides which steps to take to correct the problem, and after taking these actions reenters the problem detection phase in anticipation of the next problem. Therefore, problem detection can be viewed as the first step in solving the next problem, but the cognitive information used to detect the new problem is derived from earlier decisions to take a particular action. The precursors to problem detection are previous decisions and actions, and any model of problem detection must start with these precursors as an input. The first task, therefore, in creating a problem detection model is to specify the precursors to the problem detection process.

Precursors to problem detection: the plan, and act/event scenarios. Because decision making is a goal-oriented process (Newell & Simon, 1972), a precursor to problem detection is the plan (Schank & Ableson, 1977) that the decision maker creates to reach that goal. This plan is based on world knowledge such as cause and effect relationships (Einhorn and Hogarth, 1981), and consists of generic goal-directed actions generated by the decision maker (10, 12) together with the general effects or outcomes these actions should produce. The plan is a causal schemata (Tversky & Kahneman, 1980) that specifies how, in a general way, the decision maker expects to achieve the desired goal.

Next we assume the general plan is fleshed out. The decision maker uses the general plan to create a detailed act/event scenario by simulating in imagination a series of actions and events (outcomes) that lead to the goal (Kahneman and Tversky, 1981). These scenarios are relatively precise recipes for reaching the goal--they include specification of the actions

to be taken and their consequences. The act/event scenario is the decision maker's detailed plan for action. Its existence can be established and its contents measured in ways we will describe later.

Components of the problem detection process.

The component mental processes of problem detection are summarized and incorporated into the model presented in the remainder of this section. This model is a first approximation to the one we intend to develop. Four components appear logically necessary from an analysis of the problem detection task. Briefly:

1. Anomaly detection: matching the act/event scenario. The initial component is a matching process wherein discrepancies between the act/event scenario and the acts and events that actually occur are detected. The outputs of this process are acts and events that are anomalous, that is, are not anticipated in the act/event scenario.

2. Assessing causal relationships. This is a process in which the causal relationships, if any, between anomalous events and the act/event scenario are established.

3. Assessing relevancy/importance of events. Here the problem detector decides if the causally-related anomalous event is important enough to add to the scenario to produce a revised act/event scenario.

4. A goal assessment process. The revised act/event scenario is examined to see if it still leads to the goal by simulating the effects of the anomalous event. If the decision maker concludes that chances of reaching the goal are about the same as in the original scenario, the anomalous event is dismissed. If, however, the chances of reaching the goal are reduced, a problem has been detected and a new action is required.

Developing the problem detection model.

Two things will be accomplished in this section. First, notation for describing act/event scenarios will be developed. Second, the problem detection model will be developed using this notation.

Notation for describing act/event scenarios. We next must develop notation for two important problem detection situations. The first situation is one in which environmental uncertainty is low and the effects of actions quite predictable. For this reason, the decision maker's scenario consists mostly of actions. In this low environmental uncertainty case the act/event scenario can be represented:

Act/event scenario = {a, b, c, ... goal},

where a, b, c ... is a sequence of actions leading to the goal. An example of this type of scenario is starting a car with a dead battery. Act a might be moving the car with a good battery next to the problem car, act b might be opening the hoods of both cars, etc.

However, when environmental uncertainty is high, the representation of the act/event scenario is somewhat more complicated, because both

actions and the uncertain events resulting from these actions should be represented. Imagine that a doctor is using penicillin to treat a patient with a bad case of pneumonia. The action would be to administer penicillin to the patient and the doctor's expectation is that the patient will improve. Possible events include the patient's improvement, lack of improvement, and an allergic reaction to penicillin. In this case the act/event scenario might be as follows:

Act/event scenario = { a->e(a), b->e(b),... goal},

where **a** is the act of administering penicillin, the symbol "->" denotes that an event is caused by an act, **e(a)** is the expected event (the patient improves), **b** is the act of keeping the patient in bed for several days, **e(b)** is the further reduction of the lung infection, etc. Thus the scenario might be represented in the decision maker's mind as, "When I give the patient penicillin the pneumonia will be cured. Then by keeping him in bed for several days I can completely cure him."

Arranging the components into a model. Although a task analysis of problem detection suggests the above component processes are necessarily involved, the exact number of components that the model should have, and the order in which they are performed, is uncertain without research. Perhaps some components could usefully be combined with others to simplify the model, and the order changed. The model, shown in figure 1 below in schematic form, is a reasonable first approximation of the problem detection process to use until more evidence becomes available. The remainder of this section elucidates this working model.

A tentative model of problem detection

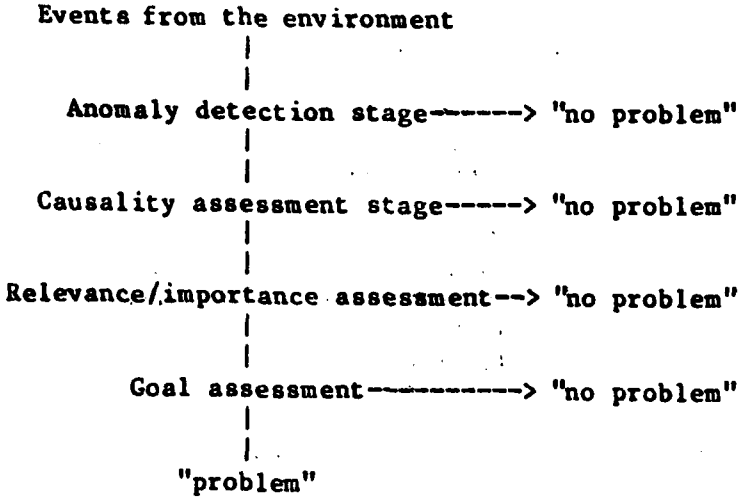


Figure 1. The tentative model, with its stages and exit points. Processing resumes at the anomaly detection stage if an event is classified as "no problem".

A tentative arrangement of the stages in the model is shown. Events from the environment are either passed from stage to stage for further processing, or classified as "no problem" in which case processing returns to the anomaly detection stage. Problem detection occurs if an event passes through all four stages. If a consistent event is in the scenario, processing does not go beyond the anomaly detection stage. Inconsistent events produce problem detection errors if they exit prior to the last stage. We now turn to detailed discussion of what takes place in each stage.

Anomaly detection. Assume, as described earlier, that a goal has been set, a plan constructed, an act/event scenario generated from the plan, and the scenario put into effect by taking the first action in the sequence. The initial anomaly detection stage consists of step by step matching between the nominal course of events--the act/event scenario--and the actions and events that actually occur. If the observed act or event matches the appropriate act or event in the scenario, the event is classified as confirming--as events are following the expected course--and the matching process continues. If, on the other hand, an actual event or act does not match, further processing is required to analyze the anomaly. For example, the detection of an anomalous event might occur when the e(a) event from the scenario is compared to the actual event E'(A) as follows:

Penicillin act/event scenario = {a->e(a), b->e(b),... goal},

Actual actions/events = A, E'(A).....

anomaly detection:

where the actual course of events is indicated by capital letters. An anomaly would be detected when the doctor compared the event expected in the act/event scenario, e(a), with the event that actually occurred, E'(A), which is the unanticipated, anomalous event of the patient developing a rash. The event in question--the rash--may not have been part of the act/event scenario, and hence may involve a problem. On the other hand the rash may not be due to the penicillin but to some other cause. If the anomaly is detected, processing is transferred to the causality assessment stage. However, if an event is classified as confirming, monitoring continues (see figure 3.1).

Assessing causality. Once an event is classified as anomalous rather than confirming, its causal relationship to the act/event scenario is examined. In terms of the previous example, the doctor would search for a causal relationship between penicillin administration and the rash. One cue to possible causality (Einhorn and Hogarth, 1981) is the temporal proximity of penicillin administration and the subsequent rash. Furthermore, the doctor's understanding of drugs may or may not include rash as a symptom of drug reaction. The doctor may decide there was a causal relation or there was not.

Causality determination makes heavy use of the decision maker's knowledge of cause and effect relationships. The interpretation of events and the expected consequences of acts all involve the problem detector's

interpretations of causality, and these interpretations are made using causal knowledge. As Einhorn and Hogarth (1981) have noted, causes are seen as differences in a causal field, and users with different causal fields may attribute different causes to an event. The patterns of causality are complex. Several events may be sufficient to cause change, in which case there are multiple possible causes. Other events may be necessary but not sufficient, in which case all of the events must be present to cause change. Events may also be seen as indirect causative agents. For example, we might believe that the patient contracted pneumonia because he worked too hard and became exhausted.

Assessment of relevance and importance. If the doctor concludes there is no causal connection between penicillin and the rash, then we assume the event is dismissed as irrelevant and the doctor returns to the first stage of monitoring and anomaly detection. Note that if the the doctor was wrong, and the rash was a symptom of an acute reaction to the penicillin, he or she has failed to detect a problem.

If, on the other hand, a causal connection between the penicillin and the rash is found, the doctor must assess its relevance and importance. Is the rash a minor result of the penicillin, or does it represent an allergic reaction which would be exacerbated by continuing treatment? If the doctor concludes the latter then **the act/event scenario is revised** to incorporate this new conclusion.

Goal assessment by simulating the effects of the anomalous event. The next step is to determine if the revised scenario is still consistent with the goal of recovery. The consequences of an allergic reaction are simulated to determine the likelihood of reaching the goal, given the revisions of the scenario. The doctor may conclude the goal of recovery can still be reached despite the risk, and continue to administer penicillin. However, the doctor may conclude that the goal can not be reached by this route because the patient is too allergic to penicillin to warrant its continued use. Having detected the problem the doctor must abandon or modify the act/event scenario and resolve a new decision problem--how to treat a patient who has pneumonia and is allergic to penicillin.

Types of problem detection failures. According to our model, successful problem detection occurs when a decision maker classifies an event as anomalous, determines that it is causally related to the scenario, revises the scenario appropriately, and by means of a simulation using the new scenario determines it will be difficult or impossible to reach the goal.

Our model also predicts several ways in which a decision maker can fail to detect a problem. First, an anomaly may go undetected. Our past research on hypothesis generation (9) suggests many anomalous events are not represented in act/event scenarios and hence may not be recognized as significant. For example, the doctor may not view the skin rash as significant because his scenario does not contain rash as a possible consequence of administering penicillin.

Second, even if the event is classified as anomalous, its causal connection to the scenario may go undiscovered and the event dismissed as

unimportant, leading to another type of failed problem detection.

Third, an event may be classified as anomalous and causally related, but the decision maker may not believe it relevant or important enough to warrant revising the scenario, and so may make an error of another type.

Fourth, even though the problem detector has revised the scenario, the simulation process, where the effects of the change in the scenario are assessed, may be faulty. An erroneous inference may be made that the anomalous event is consistent with reaching the goal, again failing to detect the problem.

A problem detection taxonomy.

A problem detection taxonomy can be produced by simultaneously considering the types of environmental events that can produce problems, with the types of errors the problem detector can make. This taxonomy should be of great value in research, and should also be of value independently of our theory, as a way of analyzing problem detection performance. Before the taxonomy is described we need to introduce additional notation and define the types of environmental events that can produce problems.

Modeling the environment. In the examples of problem creation and failed problem detection to be presented later we will be discussing acts that may produce a series of events. These sets of acts and events that would lead to the goal are represented as act/event scenarios:

Act/event scenario = {a->{e1, e2, e3}, b, c, ...goal},

where a, b, and c represent acts by the decision maker, the set {e1, e2, e3} symbolizes events 1, 2, and 3, and the arrow -> indicates the events are consequences of act a.

The act/event scenario distinguishes between actions and events because this is important in our problem detection model. However, an act is a special kind of event, an event created or produced by the problem detector. In terms of the causal relationships in the environment, the distinction between acts and events may not be necessary, but it is important when talking about our problem detection model.

A further distinction can be made between situations where the actions and events must occur in a certain sequence, and situations where the sequence is not important. For example, some of the actions in starting an aircraft engine should be performed in a logical order, while others can be performed any time before the starting switch is thrown.

Environmental conditions that can produce a problem. The minimum act/event scenario consists of those actions and events that are necessary to reach the goal. If acts are considered to be special types of events, there are only three environmental situations that can create a potential problem. These are 1) a necessary event that is omitted from the scenario (an omission), 2) an event is added to the scenario (a commission) that diverts the scenario, and 3) a sequencing error where events occur in the wrong order. In all three situations the goal cannot be reached because the

necessary chain of cause and effect happenings leading to the goal has been broken. The third case can be thought of as a simultaneous omission and commission error, in which an event is omitted from the causally correct position in the scenario and inserted (committed) at an incorrect location. For example, striking a match before closing the cover is a minor sequencing error.

Producing a problem detection taxonomy by combining a problem detection model with environmental conditions. A problem detection taxonomy can be produced by combining the problem detection model with the three environmental conditions described above. As there are four stages to the model and three environmental conditions, the factoral combination produced has 12 cells (four stages by 3 environmental conditions). For the sake of expositional simplicity, this taxonomy will not display other important variables, such as the difference between acts and events.

Table 3.1: An abbreviated problem detection taxonomy

Problem detection model stages:

	A. Anomaly assessment	B. Causality assessment	C. Relevance/ importance assessment	D. Goal assessment
Omitted Act or Absent Event (CASE 1)	A. --> A B. ----------> B C. ----------> C D. ----------> D			
Committed Act or Intruding Event (CASE 2)	A. --> A B. ----------> B C. ----------> C D. ----------> D			
Incorrect sequences of Acts or Events (CASE 3)	A. --> A B. ----------> B C. ----------> C D. ----------> D			
	(CASE 4)	(CASE 5)	(CASE 6)	(CASE 7)

Note: Case numbers are arbitrary. Cases were selected from a much larger set to illustrate each of the three rows and each of the four columns of the taxonomy. Arrows terminate at the model stages where problem detection fails.

Table 3.1 is the abbreviated version of the problem detection taxonomy. Shown across the top of the table are the four model stages:

anomaly detection, causality assessment, relevance/ importance assessment, and goal assessment, where the effects of the anomalous event are simulated. The major headings in the rows are the three environmental conditions that can produce problems: omissions, commissions, and incorrect sequences. The directed lines (arrows) show the places where problem detection can fail at the various stages of the model. The notation A, B, C, D refers to the depth of processing an event receives before it is correctly or incorrectly classified. The Case notation (e.g., Case 1) refers to examples of these problem detection errors, which will be presented after the table as illustrations of its contents.

Examples of the taxonomy gained from accidents.

One useful way of illustrating this taxonomy is to discuss cases in which problem detection failed. Perhaps the best documented sources of failed problem detection are accounts of accidents. Although we are not primarily interested in accidents per se, by studying these accounts we can further refine our problem detection model and taxonomy. First we give examples of each of the major types of environmental conditions:

Case 1. Errors of Omission. Omission errors occur when a necessary act or event is omitted from the scenario. Omission errors are frequent, and many of these problems could have been detected in time to recover from the error. For example, a plane flew 250 miles off course in the North Atlantic because the operator of the inertial navigational system forgot to enter a coordinate into the navigational computer. In a second case, a corporate jet crashed on take-off because the pilot forgot to release the parking brake.

Case 1. Omission error:

Act scenario = { a, b, c, d.... goal}
actions = { A, C, D....

error—action B was omitted.

Case 2: Errors of commission. Commission errors are also frequent. They occur when an act or event not present in a scenario occurs and causes a deviation from the ideal path to the goal. For example, a man was cleaning his clothes with naphtha (a flammable solvent) and lit a cigarette in the midst of the fumes. Other commission errors occur when an incorrect action is substituted for a similar correct action. During World War II, there was a rash of training accidents involving lowering the landing gear when the plane was on the ground preparing for take-off. The pilot pulled the lever raising the landing gear rather than pulling a second, adjacent lever that lowered the flaps.

Case 2. Commission error:

```
act scenario = {a, b, c, d....., goal}
actions      = {A, B, Z, C....
                    ^
                    |
          error—action Z was performed
```

Case 3: A sequencing error. The sequencing error is interesting. In this case the decision maker performs the correct actions but in the wrong order. An incorrect sequence for lighting a gas stove is to turn on the gas and then strike a match.

Case 3. Sequencing error:

```
Act scenario = { a, b, c, d.... goal}
actions      = { A, C, B, D....
                    ^
                    |
        error—action C was performed out of sequence.
```

Examples of failures at the various model stages.

The following four examples are cases where problem detection failed due to errors in one of the four model stages. These examples are also illustrations of various environmental cases:

Case 4: Failures of the anomaly detection stage. This example occurred when an expected event did not occur. Before the recent crash of a jet into a bridge when taking off from National Airport in Washington, the jet was "de-iced" (Act A), and the pilot expected e1 (removal of ice from wings). The pilot probably did not confirm that the ice was removed (E1), and the plane crashed on take-off (Act B) because the pilot did not detect the anomalous omission.

Case 4. Failure to detect an anomalous event:

```
Act/event scenario = { a->{ e1, e2 }, b, .... goal}
Actual act/events  = { A->{     E2 }, B, .....
                                      ^
                                      |
     error—act B was performed when event E1 was
absent (i.e., act A did not achieve the desired effect).
```

Case 5: Failures of the causality assessment stage. During an overhaul of an aircraft carrier a welder used a torch to cut the bolts fastening part of the aircraft launcher to the deck. The bolts, red hot from his torch, fell into a ventilation duct (an anomalous event). Unknown to him, the duct had been filled with trash due to a missing cover, and the

bolts started a fire. Evidently he did not see a causal relation between hot bolts in a metal duct and fire, since metal does not burn.

Case 5. Failure of the causality assessment stage:

Act/event scenario = {a->{e1, e2}..... goal}
Actual act/events = {A->{E1, E13, E2}....
^
|
Event E13 was unanticipated, misclassified as causally unrelated and consistent with goal.

Case 6: Failure of the relevance/importance stage. A pilot had problems with his fuel gauges and had them repaired. Resuming his flight, he ignored a sudden drop in the gauges and ran out of fuel a few minutes later. He apparently detected the anomaly and classified it as causally significant, but the existence of an alternative explanation (the gauges are malfunctioning again) caused him to classify it as irrelevant.

Case 6. Failure of the relevance/importance stage:

Act/event scenario = {a->{e1, e2}..... goal}
Actual act/events = {A->{E1, E13, E2}....
^
|
Event E13 was unanticipated, classified as causally related but misclassified as irrelevant.

Case 7. Failures at the goal assessment stage. Captain Robert F. Scott of the British Navy encountered a series of setbacks in his race for the South Pole. His comparison of his act/event scenario to his actual journey probably yielded many anomalous events such as distance covered, physical condition of the party, and remaining food. He must have realized the causal significance of these anomalous events and revised his act/event scenario accordingly, yet he persisted. One reason for his persistence may have been errors in his goal assessment using the revised scenario. He may have overestimated the probability of reaching his goal and returning. Alternatively his assessment may have been accurate, but he was determined to continue whatever the odds.

Extended examples of problem detection failures.

It should be apparent by now that problem detection can only be understood by considering both the problem detection model and the problem environment simultaneously. We include three extended examples that serve two purposes. First they make the point that one problem detection error can lead to another, and second they illustrate psychological processes not yet discussed.

The first mishap occurred during a flight when icing conditions were encountered. Airspeed in a modern jet is sensed via a pitot tube which leads to a pressure transducer. Pressure is translated into airspeed in the

cockpit display. The pitot tube de-icer was not turned on (an omission, failure of anomaly detection, point A in table), and the pitot tube became partially blocked by ice. The pilot, seeing the airspeed dropping and believing the jet was about to stall, went into a dive (a commission, incorrect problem detection at point D in table). The engineer, alarmed by airspeeds that were approaching the design limits of the aircraft, made a successful problem detection and corrected the pilot's error of omission by turning the de-icer on. This incident is of particular interest because the pilot's actions would have been appropriate if the indicated airspeed had been accurate, but because of the false alarm over the earlier omission, the wrong action was taken.

A second example illustrating the importance of having a correct scenario is found in the following mishap during a night landing. A pilot who had never landed at Cairo airport made a careful study of the route book for runway 05 and reassured himself that the approach to the runway was flat. Nearing the field, he turned onto the wrong runway, runway 34 (a commission error, anomaly detection failure, point A in table). To his surprise, he immediately lost sight of the airport lights. He continued his approach despite being unable to see the runway lights (absence of an event, causality assessment failure, point B in table), and flew his aircraft into the top of a high sand dune which had been obscuring the lights. The accident was attributed to the pilot's preparation to land on runway 05 (involving a flat approach), his mistaken use of runway 34 (for which a flat approach was inappropriate), and his failure to pull up when loosing sight of the runway lights. Evidently the pilot used the wrong scenario for runway 05 and attributed his failure to see the lights to a cloud.

A third mishap illustrates how events can be misinterpreted as consistent with an event scenario when in fact they suggest a problem. An airliner was flying from Tripoli to a small town, Kano, to the south. As Kano had no radio beacon, the navigator used dead reckoning based on magnetic and gyro compasses and star sightings. An initial error of commission (anomaly detection failure, point A in table) occurred when the navigator set 60 degrees magnetic declination into the magnetic compass rather than the correct 6 degrees. At this point the pilot detected the 54 degree discrepancy between the magnetic and gyro compasses (correct problem detection), and asked the navigator to take a star sighting to decide which of the two compasses was correct. The navigator distrusted the gyro compass and used a dead-reckoning plot based on the magnetic compass to select stars for the sighting. Then he failed to detect a problem upon making the sighting, when he misidentified the stars in the narrow field of view of the astro compass and concluded his sightings were in agreement with the magnetic compass (a case not discussed in the taxonomy, resulting from perceptual confusions).

Finally, when the dead-reckoning and astro navigation suggested they should be near Kano, the pilot made a problem detection error when the expected thunderstorms were not seen (omission, failure of relevance/importance stage, point C in table). Thirty minutes later the engineer made the correct problem detection by identifying the error in setting the compass, but the plane ran out of fuel and made a crash landing in the desert after flying almost to the East coast of Africa, 54 degrees off course.

Correct problem detection. In discussing cases where problem detection fails we have focused on what can go wrong with the processes, but it is important to emphasize that problems are successfully detected most of the time. Failures, when they do occur, provide interesting insights into the problem detection process in much the same way that perceptual illusions illustrate perceptual processes. However, failures may not occur frequently enough to form the sole basis for a study of problem detection. Much can be learned by using response time to study successful problem detection. For these reasons it is very useful to study latencies as well as errors in the problem detection research.

CHAPTER 4. A HYPOTHESIS GENERATION MODEL AND RELATED RESEARCH

The hypothesis generation task.

An earlier contract was devoted to the study of hypothesis generation, i.e., the process by which the decision maker generates the relevant states of the world. In terms of problem structuring, the decision maker should be able to generate the possible states of the world that may affect the outcomes of any acts that are taken. For some problems this task may be easy. The decision maker may generate hypothesized states of the world related to a problem which has been experienced before. In these situations possible hypotheses may be readily retrieved from memory because they are few in number and routine in nature. Another important class of problems exists where hypothesis generation is a crucial component of problem structuring. Examples of tasks which require hypothesis generation include medical diagnosis, automotive and electronic trouble shooting, and the scientific process itself. Tasks in this category are particularly difficult to solve when the number of possible hypotheses is large and the decision maker cannot rely on past experience to narrow the field to several obvious hypotheses.

It is particularly important that the decision maker include the actual state of the world in the problem structure because any subsequent decision that fails to consider that state of the world may be wrong. For example, if your auto mechanic fails to entertain the hypothesis that a dirty carburetor is responsible for your car's bad performance, you may pay for a series of adjustments or part replacements that do nothing to correct the problem. Similarly, if your doctor fails to consider the disease that you actually have, the whole treatment regime may be inappropriate, or even dangerous to your health. Therefore, one important part of the hypothesis generation task is the inclusion of the true state of the world in the set of possible hypotheses. It is important that the set of hypotheses generated by the decision maker should be as complete as possible. Ideally, the set should be exhaustive; however, a practical decision maker usually neglects improbable hypotheses because these states of the world appear so unlikely that they can safely be neglected.

The hypothesis set that the decision maker creates should contain plausible hypotheses. The construct of "plausibility" includes the notion that for a hypothesis to be included in the set of hypotheses it should be sufficiently probable to be worth further analysis. This does not necessarily involve an assessment process as detailed and thorough as is typically implied by the term "probability assessment." All that is logically necessary at the early stages of problem structuring is that the decision maker make a rough "go/no go" decision in regard to each hypothesis. Hypotheses that pass this crude plausibility test may be more carefully assessed in later stages of decision analysis.

While it is likely that plausibility assessment and probability assessment share common elements, there are a few clear differences. The first major difference is in the nature of the task requirements. In a probability assessment task, assessments are usually made about the relative likelihood of a set of specified hypotheses known to the decision

maker. In a hypothesis generation task, hypotheses are evaluated with respect to whether or not they should be considered further. This evaluation is complicated by the fact that the evaluation should be relative to both previously-specified hypotheses that the decision maker may have and unspecified hypotheses that are yet to be generated by the decision maker. These task differences suggest that calling the process of deciding if a hypothesis should be included in the set of hypotheses "probability assessment" may be premature and misleading because of the task differences between the two processes. We do not know at this time if the same psychological processes are used in both types of assessment, although it seems quite certain that both processes share common elements.

Hypothesis generation tasks also have the characteristic that generated hypotheses should be consistent with any available information. This information may be specific data or knowledge about the task. Obviously, hypotheses that are inconsistent with the available evidence should not be considered. Information provided by data and the task has a second important role, since it serves as a basis for the memory search processes described in the next section. Although the emphasis will be on memory search processes, the importance of the data as constraints to the logical possibility of hypotheses should be kept in mind.

The hypothesis generation process can operate in a number of different ways depending on the task requirements. For example, during a "brain-storming" session, decision makers may be asked to generate any hypotheses that come to mind irrespective of their plausibility or implausibility. In another situation, the decision maker's task may be to generate all hypotheses that are logically consistent with the data, even though some of the hypotheses are unlikely. In a third situation, the decision maker's task may be to generate a set of plausible hypotheses and to be concerned with whether or not each hypothesis in that set is sufficiently plausible to be included as a candidate for subsequent decision analysis.

Overview of the hypothesis generation model.

The hypothesis generation model that has been developed has three components or subprocesses. The first subprocess is an executive process. The executive subprocess controls hypotheses generation according to the demands of the task. It initiates memory searches and controls plausibility assessment. The memory search subprocess is responsible for both retrieving hypotheses from memory, and for furnishing information necessary for plausibility assessment. The third subprocess is that of plausibility assessment. In this subprocess hypotheses may be checked to see they are logically consistent with the data. More sophisticated plausibility judgments may also be made. The plausibility assessment subprocess decides if a hypothesis is sufficiently plausible to warrant further processing. Figure 4.1 shows this model in summary form. In the three sections that follow, each of the subprocesses and their experimental results are discussed.

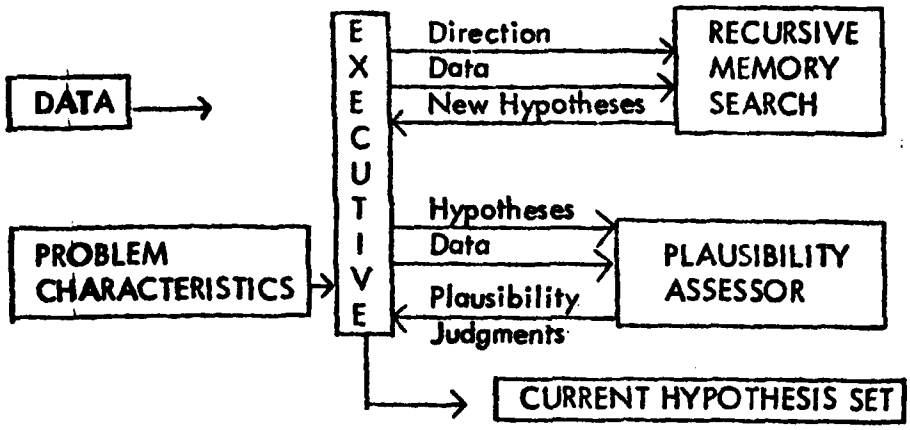


Figure 4.1. Major subsystems in hypothesis generation model.

Hypothesis retrieval from memory.

When the hypothesis generation process begins, the decision maker has an empty hypothesis set which must be populated. A reasonable goal is to develop a set of hypotheses that is as complete as possible. To accomplish this end, hypotheses must be retrieved from memory. The model assumes that available data and other task information are used to search memory. Memory is assumed to be organized in a semantic net (1, 3). Searches are made for each datum. If a hypothesis consistent with the available data is encountered in this search process, then it is tagged in memory to reflect this encounter. When a hypothesis accumulates a critical number of tags, the executive notes this fact, and the hypothesis is retrieved from memory for further processing. A detailed discussion of the memory search subprocess has been provided (1), but some of the results obtained during an evaluation of the model are of greater interest.

The first point of interest is whether or not the search and retrieval process produces candidate hypotheses which are logically consistent with all data. An analysis of the hypothesis generation task suggests that this should be a minimum requirement of any hypothesis included in the final hypothesis set. When does consistency checking occur? Does the memory search subprocess necessarily produce hypotheses that are logically consistent with all data or is consistency checking performed after retrieval from memory? Perhaps a hypothesis must be tagged by all data before it is retrieved by the executive. One assumption of this version of the model is that a hypothesis would not receive a tag from a datum if it is inconsistent with that datum. In a second version of the model it might be assumed that any hypothesis encountered in the memory search may be retrieved for further processing. Under this assumption, retrieval could follow from a single tag.

The "one-tag" version and the "all-tag" version are limiting cases of the tagging model. A task analysis suggested that it was unlikely that the "one-tag" version would be correct. If a hypothesis suggested by any of the data is retrieved for further processing, then using the "one-tag" version, the decision maker would have to process a large number of hypotheses most of which would be inconsistent with one or more data. If, however, all hypotheses suggested by the data had to be tagged by all data, then the decision maker would retrieve very few hypotheses, and would probably fail to retrieve many relevant hypotheses. It seems reasonable to assume that the decision maker should choose a strategy that lies somewhere between these two extremes.

The tagging model was designed so that the criterion number of tags was a free parameter, and this model was used as a measurement tool to address this issue. A study (1) was conducted where decision makers retrieved hypotheses from either a set of six data, or subsets of these data which consisted of three data, or only one datum.

The criterion number of tags for retrieval to occur was estimated, and was found to be between two and three tags for these data. Subsequently, we have shown that this conclusion does not depend on the assumptions of the tagging model; other similar models would yield the same conclusions.

The major implication of this result is that hypotheses are retrieved from memory using two or three data as retrieval cues. Therefore, retrieved hypotheses are at least partially consistent with the available data. These results also suggest that the memory search process may produce hypotheses that will be discarded in subsequent assessment because they are not logically consistent with the rest of the data.

Recently, Thompson (1983) performed an extensive modeling effort in this area. His results caused him the reject models similar to our "one-tag" and "all-tag" models as we did. The model he favored, called the "Activation Mean" model is a Thurstonian Model where the subject is assumed to retrieve a hypothesis in response to multiple data if the mean activation from the multiple tags exceeds a criterion. His Activation Mean model seems to fit the data well, but unfortunately he did not provide any relative comparisons between his model and ours. As both his model and ours seem to be excellent fits for their respective data sets, the choice between models will have to await further research.

A second point of interest deals with the efficiency of the hypothesis retrieval process. In order to study this process, the retrieval performance of the subjects was compared to a "minimally-adequate hypothesis set" developed by the experimenters. This minimally adequate hypothesis set consisted of the three most-plausible hypotheses which the experimenters felt should be included in an "adequate" set of hypotheses generated by the subjects. The set for each problem was chosen conservatively and many other plausible hypotheses were excluded. Only 19.9% of the subjects were able to retrieve these three hypotheses. We also explored the effect of relaxing the definition of adequate performance. We found that 50% of the subjects were able to retrieve two out of three of the "minimally adequate" hypotheses, while 92% of the subjects were able to

retrieve one of the three. This result was our first indication that the hypothesis generation process was less than adequate, and it has been replicated many times using more objective criteria of performance. Similar results are discussed in chapters 5, 6 and 7 of this monograph. The results discussed here are important because they suggest that the memory search process is involved in the deficiencies in hypothesis generation reported throughout the research on hypothesis generation.

Checking hypotheses for logical consistency.

Results from the tagging study (1) of the memory search model suggest that the decision maker will often retrieve a hypothesis from memory using several data. This newly retrieved hypothesis may or may not be consistent with all of the remaining data that were not used in its retrieval. A consistency checking process may exist in which the decision maker checks the newly-retrieved hypothesis for logical consistency with any remaining data. Such a process should be relatively fast, as compared to hypothesis retrieval. Using the hypothesis as a retrieval cue, the decision maker should perform a high-speed memory scan to examine whether the hypothesis is consistent with the remaining data. For reasons of efficiency, the consistency checking process should be self-terminating, ie. the consistency checking should stop if a datum is encountered which is inconsistent with the newly-retrieved hypothesis. If a hypothesis passes this consistency check, then it is logically consistent with all of the data, and it has met the minimum plausibility requirements. Plausibility assessment may stop at this point, or it may continue, depending upon the demands of the task. Figure 4.2 shows the relationship between hypothesis retrieval and consistency checking in our model.

A series of experiments (3) was conducted to investigate the nature of consistency checking. In the first experiment, we asked whether or not consistency checking exists. Subsequent experiments were conducted to examine the speed of consistency checking relative to hypothesis retrieval, and whether or not consistency checking is a self-terminating process.

The first experiment was an attempt to demonstrate that consistency checking exists. An instructional manipulation was used in which subjects were instructed to either respond with the first hypothesis that occurred to them, irrespective of its consistency, or were instructed only to respond with a consistent hypothesis. Hypothesis generation problems containing various numbers of data were used. We predicted an interaction between the time necessary to generate a hypothesis in the two conditions and the number of data in the problem. While large differences were observed between the two conditions, the interaction was not significant. We believe that the inconclusive results of this experiment were due to the subjects' inability or unwillingness to respond with the first hypothesis that occurred to them even though they were instructed to do so.

In a study which was performed after the original technical report (3), the question of the existence of consistency checking was investigated again. In this study a somewhat different approach was used. Subjects were asked to generate consistent hypotheses in response to data. Immediately after they generated a hypothesis, they were shown a list of inconsistent hypotheses that had been generated by another group of subjects. Subjects scanned the list of inconsistent hypotheses, and

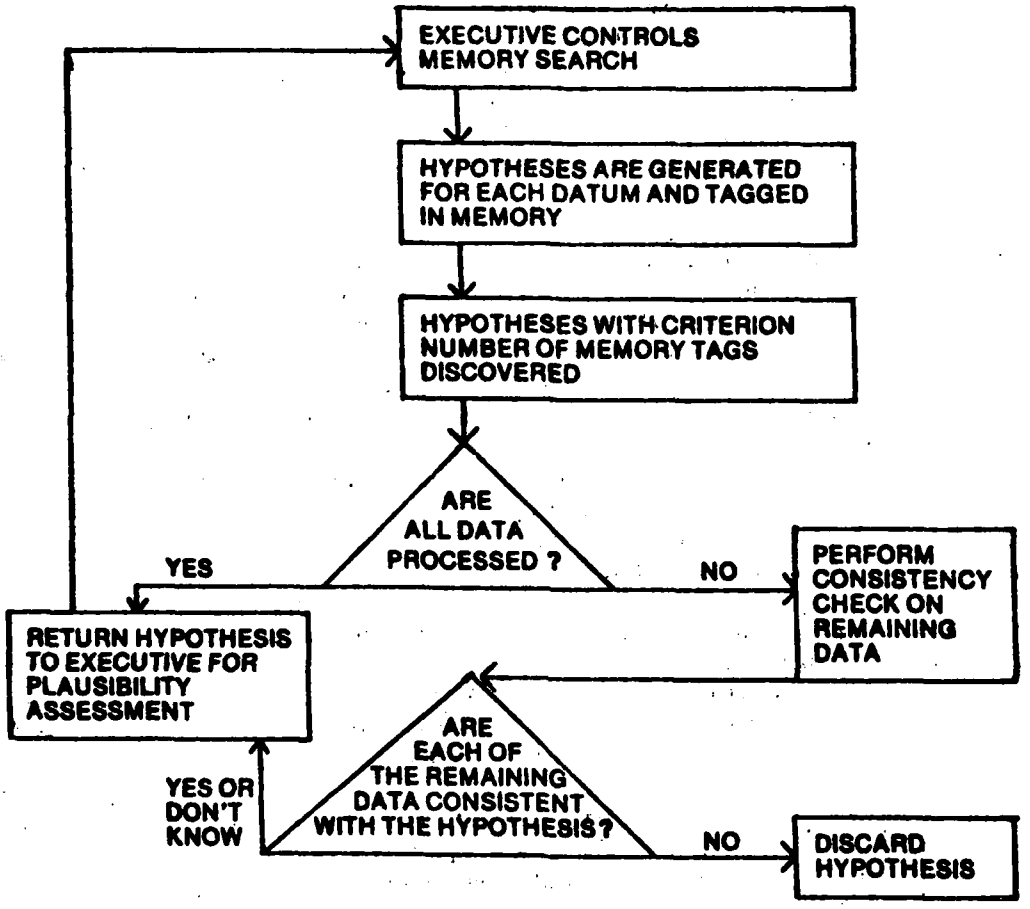


Figure 4.2. The hypothesis retrieval and consistency checking processes.

identified any that had "crossed their minds" during hypothesis generation.

It was estimated that subjects retrieved an average of 1.83 inconsistent hypotheses before they retrieved their first consistent hypothesis. This experiment contained a manipulation to control for the obvious demand characteristics. Subjects may have picked hypotheses from the list to please the experimenters. It is unlikely that these results could be explained in that way. It was concluded that subjects do check newly-retrieved hypotheses for consistency, and that inconsistent hypotheses are discarded at this time. These results also add support to the conclusion that memory is searched using only part of the available data. The memory search result implies that inconsistent hypotheses are retrieved from memory, and this consistency checking experiment demonstrated that inconsistent hypotheses are retrieved from memory and are then discarded.

The next experiment in this series (3) addressed our prediction that consistency checking is a more rapid process than hypothesis retrieval. Two experimental conditions were compared. Subjects in condition one generated hypotheses in response to varying amounts of data. Subjects in condition two were given the hypotheses that the first group had generated, and were asked to check them for consistency using the same data. Using a Sternberg memory search procedure (3), the time to process each additional datum was estimated. Subjects who generated hypotheses took 1.8 seconds per datum, while consistency checking subjects were able to process each datum in .7 second, i.e. between two and three times faster than hypothesis generation subjects.

The final experiment in this series examined the self-termination prediction. Subjects were provided with a hypothesis and were asked to check three-data problems for consistency with respect to that hypothesis. The position of a disconfirming datum in the data set was varied for problems where the hypothesis was inconsistent with the data. Subjects responded faster when the disconfirming datum was earlier in the sequence of data than when it was later. This result is consistent with a self-terminating process. The results of the experiments investigating the existence of consistency checking suggest that subjects retrieve hypotheses which are found to be inconsistent with a set of data. We believe that consistency checking occurs in the hypothesis generation process and that subjects tend to retrieve hypotheses in response to only part of the available data. Thus, the results support the predictions of the partial-retrieval consistency checking model of hypothesis generation rather than the alternate retrieval model which assumes that subjects retrieve consistent hypotheses using all data as retrieval cues.

The results of experiment two of this series demonstrated that less time is needed to process an additional datum during consistency checking than during hypothesis retrieval. These results are consistent with the predictions based upon the search properties of hypothesis retrieval versus the verification properties of consistency checking. Experiment three of this series provided evidence that consistency checking is a self-terminating process.

These results are important for an understanding of the hypothesis generation process. They more clearly define the role of memory in

hypothesis generation, and the processing of hypotheses subsequent to retrieval from memory. These results, when combined with our other research, are consistent with the following model of hypothesis generation: Hypotheses are retrieved from memory using several data. If the data are numerous, then retrieval is based upon only a part of the available data. Upon retrieval, hypotheses are checked for logical consistency with any remaining data using a high-speed semantic verification process. If a logical inconsistency is found between a hypothesis and a datum then processing stops, and the hypothesis is labeled as inconsistent. If, however, the hypothesis survives the consistency checking process, then further processing can occur depending on the task demands. The consistency checking process is faster than the retrieval process because retrieval involves a search for hypotheses that are suggested by several data, whereas consistency checking involves verifying semantic relationships among a hypothesis and data that are already active in memory.

Hypotheses that survive the consistency checking process have met the minimal task requirement for hypothesis generation, that of logical consistency with the data. They are not necessarily plausible hypotheses; plausibility can be established by further processing if the task requires this type of assessment.

Our use of the term "consistency checking" has been solely confined to high-speed semantic verification. We do not intend to imply that other processes which might be called "consistency checking" do not exist. Thus, a scientist may spend months determining if a hypothesis is consistent with data. This is not the process studied here, and this distinction becomes clearer if a scientist's work is termed "hypothesis assessment." We have studied the early phases of the hypothesis generation process, and we believe that in the first few seconds of hypothesis generation a hypothesis is retrieved from memory using part of the data and then checked for consistency with the remainder of the data.

Plausibility assessment of generated hypotheses.

After a hypothesis is retrieved from memory and checked for logical consistency, further processing may occur to determine if the hypothesis is sufficiently plausible to be included in the set of hypotheses that the decision maker is entertaining. Secondly, the decision maker must decide if more hypotheses should be included in the set of hypotheses, or if the set is complete enough to be satisfactory. Once the set is sufficiently populated with hypotheses, attention can be turned to other aspects of problem structuring. This task analysis suggests that the decision maker should have some sensitivity to the plausibility of both individual hypotheses and the collection of hypotheses called the hypothesis set.

As discussed previously, the task of estimating the plausibility of hypotheses is somewhat different than a probability or odds estimation task. The task of the decision maker in hypothesis generation is to populate an empty hypothesis set; whereas, in probability or odds estimation the task is to estimate the relative likelihood of an existing set of specified hypotheses. The probability estimator, for example, need only be concerned with the relative likelihoods of a set of enumerated hypotheses. The hypothesis generator, on the other hand, must judge a

specified hypothesis that has just been retrieved from memory against a diffuse unspecified set of hypotheses that potentially might be included in the hypothesis set. Before the plausibility of a hypothesis can be established, it must be compared to other alternative hypotheses which may or may not be available in memory. Thus, plausibility assessment would seem to be much more formidable than probability or odds estimation, and one might naturally expect that subjects' plausibility assessments will be found less accurate. This kind of judgment is analogous to the difference between absolute and relative judgments in perception where it is commonly known that relative judgments are easier to make than absolute judgments. The plausibility assessor may be making a judgment about a hypothesis in the absence of other hypotheses. As the hypothesis set becomes more populated, plausibility and probability assessment become more similar in nature, and for fully-populated sets the tasks become identical. The same argument holds for judgments of collections of hypotheses where the task is to generate a set of hypotheses which is as complete as possible. An optimal decision maker should continue to generate hypotheses until they believe that the collection of specified hypotheses equals the set of all possible hypotheses (neglecting "cost of thinking issues"). Figure 4.3 shows our model of the hypothesis assessment process.

The first research concerned with hypothesis assessment was an early study done by Gettys and Fisher (cited in 7) which was not a formal part of the contract on hypothesis generation. This study was devoted to the executive control of the hypothesis generation process, and it investigated the rules for deciding if a particular hypothesis or hypothesis set is plausible. Of particular interest in this study was the relationship between these rules and the memory search process. It was found that additional hypotheses were most often generated when data were presented which disconfirmed the set of currently-held hypotheses. The data were examined to see if a fixed criterion of plausibility was used to admit a newly-generated hypothesis to the current set of hypotheses. No evidence for such a fixed plausibility threshold was found. Instead, subjects seemed to be admitting hypotheses into the set only if they were close competitors with the most plausible hypotheses that had already been generated. This behavior was characterized as a search for "leading contenders" rather than a search for an exhaustive set of hypotheses.

The first study in this contract examined the question of whether or not subjects could evaluate the plausibility of hypotheses. Of interest were the plausibility estimates subjects made concerning sets of hypotheses differing with respect to plausibility or completeness. Subjects were given sets of hypotheses which varied in plausibility, and were asked to judge both the plausibility of each hypothesis individually and the collection of hypotheses. The judgments included estimates of the plausibilities of both specified hypotheses and the diffuse set of unspecified hypotheses. These judgments were evaluated by comparing them to a probabilistic model developed for this purpose.

The task which was modeled was that of generating possible academic majors for an hypothetical student at the University of Oklahoma. The hypotheses to be generated were based on the courses the student had taken. The enrollment records for all students currently enrolled in the University were used to determine the probabilistic relationships between majors and courses. A total of 166,858 enrollment records were tabulated to

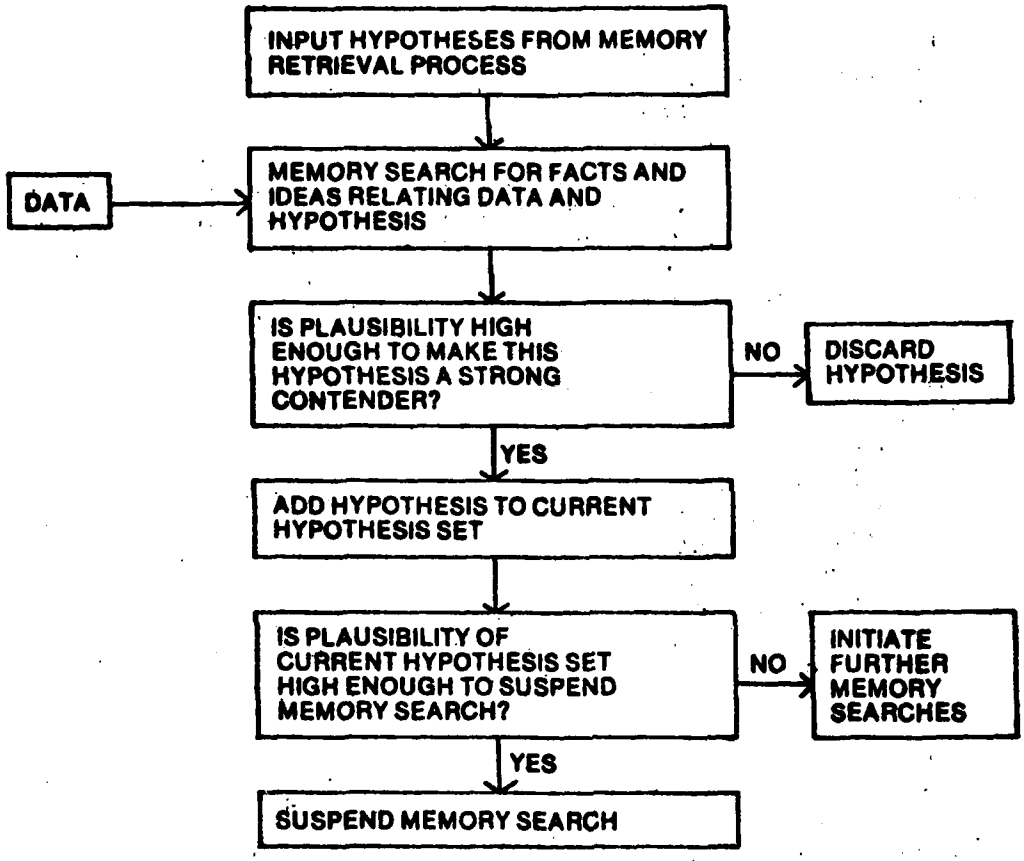


Figure 4.3. A model for the plausibility assessment process.

obtain the posterior probabilities of various majors given selected courses. These veridical values were compared to subjects' estimates to address the accuracy of calibration. This task had the necessary characteristic that the veridical relationships between majors and courses were known, and the task also had the property that most student subjects understood it intuitively. However, it should be noted that many of the relationships between courses and majors are complicated. Students enroll in a program of study for many complex reasons, including personal preference, advice from other students and advisors, and College and University requirements.

In the first experiment (1), subjects estimated the plausibility of three specified hypotheses and a diffuse catch-all hypothesis of all other hypotheses. They also estimated the plausibility of the specified collection of hypotheses versus the catch-all set. Two major results were obtained. First, as might be expected from the task analysis, plausibility estimates were quite variable, and were only weakly related to the veridical probabilities. Second, the overwhelming majority of these estimates were excessive in respect to the veridical probabilities. Both results were quite reliable, and have since been replicated in several situations (2,7).

It occurred to us that the explanation for this excessive certainty might be that the decision maker must populate the complementary set of unspecified hypotheses before the specified hypotheses (or sets of specified hypotheses) can be assessed accurately. We also had reason to believe that the retrieval of hypotheses from memory was impoverished. If this were the case, then attempts by the decision maker to populate the unspecified set of hypotheses would be only partially successful. Consequently, when plausibility estimates were made, the unspecified set of hypotheses was incomplete, and hence its plausibility was under-estimated. If the plausibility of the unspecified set was under-estimated, then the plausibility of the specified set was necessarily over-estimated.

The next study (2) was a test of this explanation. There were three groups of subjects in this study. One group was essentially a replication of one of the conditions of the previous study. Subjects estimated the plausibility of sets of specified hypotheses and the unspecified catch-all hypotheses much as before. In the other two groups, however, manipulations were introduced which were designed to increase the availability of hypotheses in the catch-all set. In one condition, subjects were encouraged to explicitly populate the catch-all set. This manipulation was chosen because it was believed that asking the subjects to make a formal search of memory for hypotheses would increase the number of "unspecified hypotheses" available in memory. The second manipulation consisted of showing the subjects exemplar hypotheses from an experimenter-generated catch-all set. This manipulation should also increase the availability of hypotheses in the catch-all set.

Both conditions which were designed to increase the availability of hypotheses in the catch-all set produced estimates that were less excessive. Therefore, we concluded that at least part of the excessiveness in plausibility assessment was due to the limited availability of hypotheses in the catch-all set.

Our studies up to this time had used only sets of hypotheses supplied by the experimenter. We were forced to used experimenter-supplied sets because of limitations in the software which determined the probabilistic relationships between courses and majors. We developed an algorithm which would efficiently process the 166,858 enrollment records for all courses and all majors. Then we were able to run a new study which both replicated the previous studies using experimenter-supplied hypotheses, and also also allowed us to study plausibility estimates for subject-generated hypotheses. Therefore, one comparison in this study was between experimenter-supplied and subject-generated hypotheses.

Previous studies employed a response mode which was a variant of the odds estimation technique. A direct probability estimation response mode was compared to the odds response mode. The motivation for this manipulation was to make sure that the excessiveness in plausibility estimates was not due to the response mode.

The results replicated our previous research and reinforced our conclusions. Plausibility estimates were excessive for both experimenter-supplied and subject-generated hypotheses. We had predicted that this would be the case because subjects should have difficulty populating the unspecified set of hypotheses in either condition. Somewhat to our surprise, however, subjects who generated their own hypotheses were significantly more excessive than subjects who worked with experimenter-supplied hypotheses. One possible explanation for this effect is that subjects who generated their own hypotheses nearly exhausted their set of plausible hypotheses in populating the specified set, and consequently did a poorer job of populating the unspecified set.

In both response mode conditions excessive estimates were found, although the subjects in the direct probability estimation condition were somewhat less excessive than subjects in the odds estimation condition. (This study was not issued as a technical report because it was a follow-up study for the availability study (2), but was included in the journal version of the availability study.)

There is a robust and important conclusion that can be drawn from the last three studies. We believe that plausibility estimates of hypotheses are excessive, and that this behavior can be traced to deficiencies of the hypothesis retrieval process.

Protocol analysis of hypothesis generation.

Mehle, in a doctoral dissertation (7), took a rather different approach to the hypothesis generation problem. Using a modification of Simon's protocol analysis technique, the hypothesis generation performance of expert and non-expert auto mechanics was studied in an automotive trouble-shooting task. This study used markedly different research strategies than the other studies in this contract, and it independently confirmed several of the observations that were made using more traditional techniques.

Subjects in the protocol analysis task were either undergraduates who professed some knowledge of cars, or expert auto mechanics from the University motor pool. Subjects were given a written description of a

malfunctioning automobile, and were asked to "think out loud" while generating hypotheses about the cause of the malfunction. Examination of the protocols revealed evidence for consistency checking. Hypotheses were generated, and then subsequently ruled out as inconsistent with the data.

In addition to the protocol analysis, both the number of hypotheses that the subjects generated were analyzed, and the plausibility estimates for collection of hypotheses that the subjects generated were analyzed. Experts and non-experts generated approximately the same number of hypotheses; the mean number of hypotheses generated per problem was 3.43 and 3.36 for the non-experts and experts, respectively. These means can be compared to the number of hypotheses that were logically possible for the problems. Information provided by the subjects was used to make this estimate in the absence of a completely authoritative source for this information. The hypothesis set for each subject was pooled with that of the other subjects by taking the union of all hypothesis sets. Illogical hypotheses were discarded from this pool (an average of .1 hypotheses per subject per problem). The number of hypotheses in the pooled set is actually a lower-bound estimate of the number of logically-possible hypotheses. The obtained pooled sets contained an average of 17.8 hypotheses per problem. By applying a mathematical model to this situation, Mehle was able to estimate the the number of hypotheses that were logically possible was 21.5 hypotheses in the average problem. Thus the average subject was generating approximately 19% of the logically possible hypotheses per problem. It was impossible to determine if the hypotheses generated by the subjects were implausible or plausible, but subjects' hypothesis sets certainly lacked the desirable characteristic of completeness.

The plausibility estimates of the sets of hypotheses generated by the subjects were also examined. There were no veridical probabilities for this task, but it was possible to exploit the fact that the sum of the probabilities of an exhaustive set should be one. The hypothesis generators in this experiment generated incomplete, impoverished sets of hypotheses. If all subjects' probability estimates are assumed to be true and if these estimates are assigned to the hypotheses in the pooled set, then a probability measure of 5.04 must be assigned to the more complete set of hypotheses developed by pooling. This measure would have been 1.00 had the whole group of subjects been veridical estimators. Thus, subjects were clearly excessive in this task. This result generalizes our earlier conclusions considerably, as it shows similar behavior in a task that was quite different from the "majors from classes task."

In summary, the protocol analysis study, even though it used different measurement techniques, reached much the same conclusions as other research. The data suggested that subjects were impoverished hypothesis generators whose plausibility estimates were excessive.

CHAPTER 5. ACT AND OUTCOME GENERATION

A model for act and outcome generation.

We developed a tentative model for act and outcome generation during the course of the project. This model served as an informal guide to some of our studies on act and outcome generation, and as such it is worth presenting here.

Assume that the decision maker has engaged in problem analysis and definition where the present state and the goal state are defined. At this point the decision maker knows where he or she is (the present state) and where he or she wants to go (the goal state). The problem may be recognized as familiar, and a number of possible actions may be readily available in memory. For example, suppose that your car refuses to start on a cold winter morning. This situation may bring to mind other similar or nearly identical problem situations that you have experienced in the past. On these occasions, calling a cab or trying to start a second car may have been solutions to the problem. Your ability to generalize from the present situation to past situations probably is based on a generalization gradient; that is, the problem situation itself reminds you of other similar situations. If this is the case, one might expect that the similarity of the present problem situation to earlier situations stored in memory will be a major determinant of the probability that these earlier solutions will be recalled. However, suppose that such a generalization process fails, and actions suggested by past experience either are not retrieved from memory, or if retrieved, are found to be infeasible for one reason or another.

If the actions that readily come to mind are infeasible, or nonexistent, then a deeper analysis of the problem is required. Actions can be suggested by an analysis of the problem space, and our model uses notions of a "means-ends" analysis borrowed from Newell & Simon (1972). The present state and the goal state differ on one or more dimensions in the problem space. We assume, following Newell & Simon, that the decision maker concentrates on those dimensions where the present state and the goal state are noticeably discrepant. Once these dimensions have been identified, we assume that the decision maker searches memory for operators that will reduce the difference between the present state and the goal state along that dimension. Thus if the present state is your home, and the goal state is your office, one strategy that you might employ is to discover operators that might reduce the distance between you and your goal. Thus, you might imagine taking a bus to within several miles of your office, and then calling your office and asking a colleague to pick you up at this intermediate point.

Subgoals can be defined. For example, you might define a subgoal of starting your car because if you could start the car quickly then you could get to work on time. Perhaps the engine was flooded by your unsuccessful attempts to start it, and a short wait would result in successful starting. If your home is at the top of a big hill and if your car has a manual transmission, you might contemplate rolling your car down the hill to start it.

The operators that the act generator uses are based on causal

knowledge (Einhorn & Hogarth, 1981); the act generator knows that certain variables cause change in the problem dimensions. For example, the car may not start because the battery will not turn the cold engine over fast enough. However, rolling a car down a hill will produce enough engine speed to start it. Thus, the latter type of act generation in our model is based on first identifying problem dimensions, and then finding causal operators to reduce the difference on that dimension between the present state and the goal state.

Notice that this example also illustrates the importance of hypothesis generation in interpreting the car's symptoms. The car may not have started for a number of reasons, such as a low battery, a frozen gas line, summer weight oil, etc. If the act generator makes an incorrect diagnosis the car may arrive at the bottom of the hill without starting, thus compounding the problem. Thus, problem analysis involves hypothesis generation followed by inferences; it is this higher-order structure that is analyzed to find the proper operators.

We assume that once an operator is found, the act generator may simulate the effect of taking that action in imagination, as discussed extensively in chapter 3. This "walk-through" process involves the construction of a scenario where the act generator imagines performing that operation, observing its effects, perhaps taking another action, etc., until the goal is reached. We assume that during this simulation process alternative outcomes are generated as the decision maker comes to places in the scenario where several outcomes seem plausible. Some outcomes are generated when the act generator reaches a chance fork in the scenario. The scenario has a number of plausible paths leading from the chance fork, and the act generator can not be sure of following the path leading to the goal. Sometimes the paths are associated with alternate states of the world that can be identified. As these alternative states of the world are created by the hypothesis generation process, hypothesis generation is also involved in outcome generation.

In review, our model proposes two processes for act generation. Some acts are generated by searching memory for similar instances of the problem. If this initial search process is not completely successful, a deeper analysis of the problem is made in an attempt to find operators that reduce differences in dimensions where the goal state differs most from the present state. Inference processes are important in choosing these operators. Once a potential operator is identified, its effects are simulated by creating a goal-directed scenario. Outcome generation occurs when chance forks are discovered in the scenario.

Measuring act and outcome generation performance.

Developing satisfactory ways of measuring act and outcome generation performance is an extremely challenging problem, a problem to which we devoted approximately 18 months of our contract. The difficulty of measuring performance is due to three sources. First, it is necessary to characterize the structure of decision problem to distinguish acts that are are minor variations of other acts from acts that represent a new and creative solution to the problem. Second, there are, in theory, an infinite number of acts that could conceivably be performed in most real-world problems. Third, even if the first two problems could be solved, it is

still highly desirable to characterize the subject's performance in terms of quality as well as quantity.

Our approach to solving the first two problems involved creating a hierarchical structure for the decision problem. First, we established empirically that a seemingly infinite set of actions can be organized into a hierarchical structure which represents generic solutions to the problem, major variations of these generic ideas, and minor variations of these major variations. This structure, which we call a tree, is finite in size, although surprisingly large, and can be managed. A hierarchical tree structure is also necessary for distinguishing minor variations of ideas from major variations. Second, we investigated several ways of creating a hierarchical tree, including having the experimenter create the tree, as well as multidimensional scaling and cluster analysis. Similar structures resulted using from both techniques.

What is good act generation performance? Criteria for good act generation should include both completeness and quality. The goal of the act generation process is to create the complete set of actions that should be evaluated in the decision process. When a decision maker evaluates various action alternatives, it is highly desirable that these alternatives include the highest utility actions possible. Obviously, a good act generator should not be expected to generate all possible actions that might solve the problem, these are virtually infinite. Instead, it seems more reasonable to define good performance as the generation of instances of most or all of the high-utility generic solutions to the problem. Thus completeness is defined in terms of capturing the high utility portion of the structure of a decision problem. not in terms of generating all conceivable solutions to the problem.

Developing a performance score for act generation. The hierarchical tree described above provides the structure that forms the basis for evaluating the quality of act generation performance. It is formed by pooling the responses of a large number of subjects and then using informal or formal techniques (to be described later) to recover the structure of decision problem. This structure is a lower-bound estimate, because if the ideas of other subjects were added to the structure it would expand somewhat. Once a satisfactory structure is created, the next step is to evaluate it using a utility estimation technique. Once these utility estimates are associated with the structure, high-quality ideas that are distinct from other ideas and minor variations of ideas can be identified.

The performance score itself is calculated from the utilities associated with the hierarchical tree. There are several scores which give complimentary information. If one is interested in the **breadth or completeness** of act generation performance, the analysis is confined to the generic ideas, or "limbs" of the tree. If one is more interested in the extent to which the act generator can create good actions, irrespective of whether they are related or unrelated to other actions, a "depth" score can be computed that includes variations of generic ideas, the branches of the tree. In both cases the performance score is computed in a similar manner. The actions of the decision maker are ordered in decreasing utility, and a cumulative function is calculated by taking the utility of the best action generated, adding it to the utility of the second best action, etc. This cumulative performance function obtained from a act generator is compared

with the equivalent function calculated from the hierarchical tree. Suppose, for example, that one wished to evaluate the act generation performance of a subject for the limbs of the tree. A performance score for the generic ideas generated by this subject would be calculated based on the actions this individual generated, using the limb structure provided by the hierarchical tree as a basis for classification. Then a similar function would be calculated based on all limbs of the tree, including those limbs generated by the subject, and those not generated. A comparison of these two functions would yield the desired measure of performance. If the two functions are identical, then the subject generated all the generic ideas in the tree, and performed optimally. Performance is suboptimal to the extent that the subject's function is lower than the function calculated on the entire structure. The depth score is calculated by by a nearly identical technique; the analysis is performed using branches as the basis for the calculations, rather than the limbs.

It is important to understand what these performance scores mean and what they do not mean. Both scores have the property that they measure the extent to which the act generator created the complete set of high utility actions. The "limbs" score reflects the breadth of performance and the "limbs and branches" score reflects the depth of performance. **Both scores capture performance in decision-relevant terms; they are a quantitative measure of the "goodness" of the act generation process as seen by the decision maker who will evaluate and choose among these actions.** However, these cumulative utility functions, because of our utility measurement techniques, are not an "extensive measurement" (cf. Roberts, 1979). This technical property means that the cumulative utility functions do not necessarily capture the utilities of **the collection** of actions, although they probably approximate the utility of the collection quite well. However as performance scores they have a the desired properties, with a minimum value of zero and maximum value dictated by the structure of the hierarchical tree.

The first act generation project. Our first study (10), for which we developed the performance score described above, used two ill-defined problems. The first problem involved generating actions directed at solving the difficult parking problem at the University of Oklahoma. Students played the role of student representatives to the parking committee, and were asked to record any action that came to mind. Students were encouraged to record any idea, good or bad, because we did not want to censor their ideas and they had an unlimited time to work on the problem. The second problem involved finding living arrangements for an impecunious Canadian friend who was broke, had exhausted his credit, his relative's ability to lend money, and was prevented by the U.S. Emigration service from working.

Following the initial study, two additional studies were performed to get utility estimates for the hierarchical tree constructed by the experimenters. Shown in figures 5.1 and 5.2 are the performance scores for both the "limbs" and "limbs and branches" analyses, which reflect the breadth and depth of subjects performance on the Parking problem. The results of the Living problem were very similar, and for that reason are not presented here. The function labeled "total" is the function based on the the limbs or branches of the hierarchical tree. The performance of the average subject is the function labeled "mean". As can be seen from these figures, the typical subject did not generate an adequate act set. However,

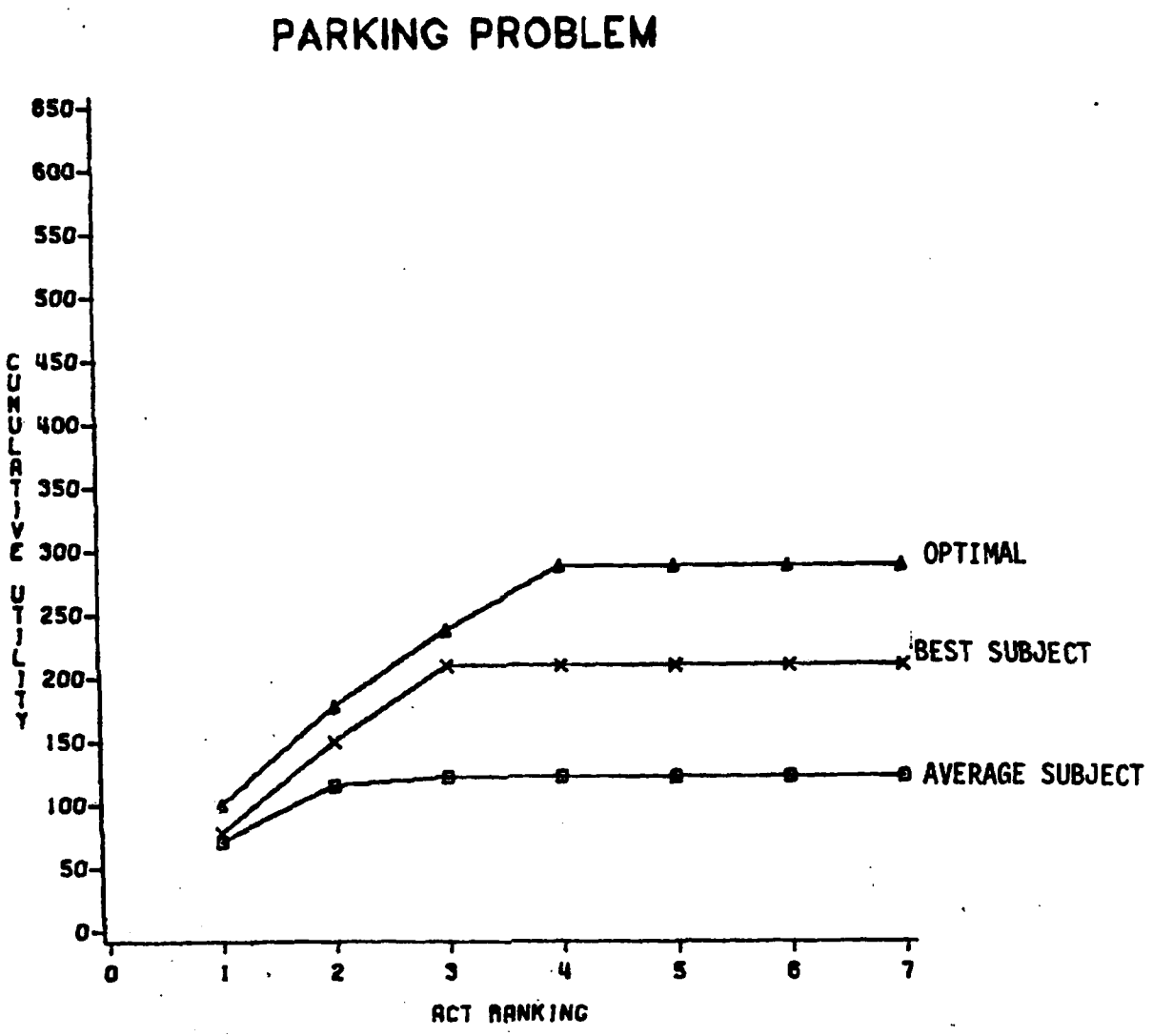


Figure 5.1. Performance scores from the "Limbs Only" analysis showing the "breadth" of performance.

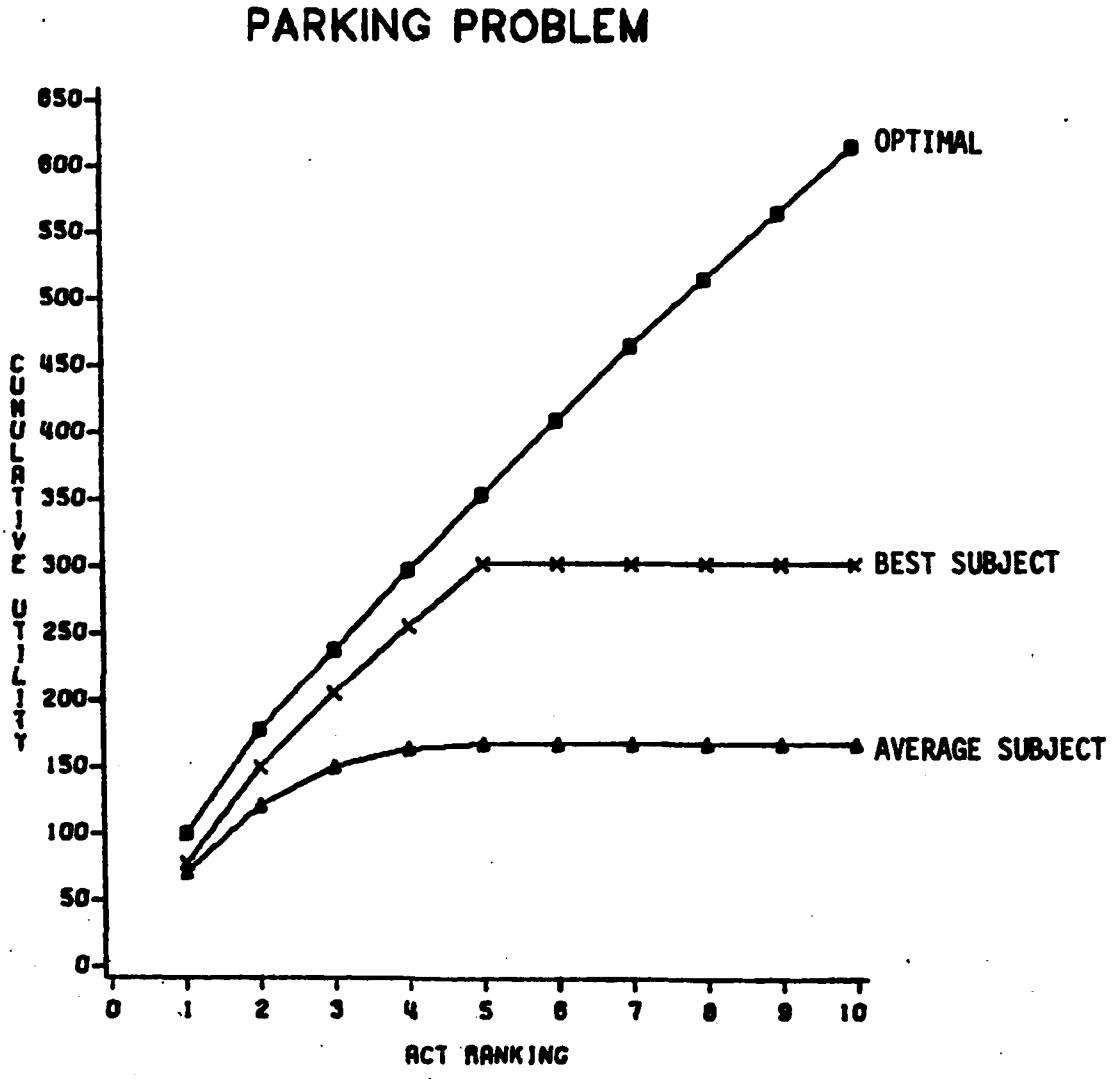


Figure 5.2. Performance scores from the "Limbs and Branches" analysis showing "depth" of performance.

the best subject in the group did a credible job (the function marked "best"). The closeness of the "best" subject function to the lower-bound estimate of optimal performance (the "total" function) is important because it indicates that the optimal measure of performance is reasonable, and reflects the average subject's failings accurately.

A more detailed look at the subject's performance can be obtained by examining the best 10 branches from the hierarchical structure. Shown in table 5.1 are the actions corresponding to these branches, and the probability of generating these branches. Two out three subjects generated the action "build a high-rise parking structure", an idea that was getting considerable publicity in the student newspaper at that time. However, the remaining high-utility actions are rarely generated. The typical subject generated 11.2 actions. Of these actions, 4.4 actions were of positive utility, and of these remaining actions about 2.5 had utility values higher than 45 points. (We chose 45 points as being near the middle of the range of utility scores, but similar conclusions would be reached with other arbitrarily chosen points.) Thus, the typical subject has between two and three ideas which are good enough to be serious candidates for implementation. However, there are 25 actions that had utilities greater than 45 points in the hierarchical structure. This demonstrates that the structure produced by the typical act generator is impoverished in this problem.

Table 5.1

The probability of generating the best 10 limbs and branches in the Parking problem tree

Utility	Probability of generating act	Actions
100	.07	Improve the trolley and CART systems.
77.5	.63	Build a high-rise parking lot.
60	.07	Put more small car parking spaces in student lots.
60	.03	Have employees park at Lloyd Noble, use trolley.
57.5	.03	Build lots around Norman, shuttle in students.
55	.10	Reduce the price of parking stickers for carpools.
55	.07	Advertise advantages of riding bike, motorcycle.
51	.10	Have more selection of afternoon, evening courses.
50	.03	Use some OU service vehicle parking for students.
50	.10	Use extra areas around fraternities for overflow.
616 sum of utiles		

We believe that this result is quite important as it is the first attempt to measure the impact of act generation performance on decision making, and the results suggest that act generation performance will have a big impact on the quality of decision making. Consider the quality of decision making with a problem structure of two or three good actions generated by the typical subject as compared with the more complete

hierarchical tree structure consisting of 25 good actions. The decision maker's menu of actions, if responses from a typical subject were to be used as the source of this menu, would be impoverished. Although a decision analysis on this impoverished set of actions probably would result in the implementation of one or more actions that would improve the situation somewhat, consider how much better off the decision maker would be if the menu had 25 actions from which to choose. For this reason, these results suggest that the quality of act generation will have a profound impact on the quality of decision making in ill-defined problems.

Follow-up studies on act generation. Because of the potential importance of our first studies, we did a second series of studies (12) using the parking problem which replicated and extended the results of the first studies. The second series of studies also addressing possible criticisms with improved methodology. In our first study we asked subjects to respond with everything that came to mind, and about 20% of the action generated were impractical or fanciful. Perhaps subjects would do better if they were attempting to generate quality actions, or perhaps they simply were not sufficiently motivated to do well. In these follow-up studies, we employed substantial monetary incentives. One group was given "quality" instructions, where they could earn up to $1.00 for each action they generated. A second group was given "quantity" instructions with a $.50 incentive for each action generated. This incentive was adjusted to be about equal to that of the "quality" group. A third control group was not given a incentive, but were just instructed to avoid minor variations of the same action, or frivolous actions.

Other methodological improvements included using a hierarchical structure based on multidimensional scaling and cluster analysis (12, to be described below), and improved utility measurement procedures which captured the utilities of both students and subject-matter experts.

The results of this second series of studies essentially replicated the findings of the first series. Shown in figure 5.3 are the performance scores for the three groups as compared to the lower-bound estimate of optimal performance using the utility values obtained from experts (the functions based on student utility values were virtually identical, and so are not shown). Notice that the subjects who were given substantial incentives for quality actions scored about the same as the control subjects. The group that was paid for quantity did somewhat better, but all three groups performed at about the same level as subjects in the previous series of experiments, replicating our previous results. Subjects in the new series generated about two to three good actions, out of at least 25 good actions.

Another result from this series of studies of great interest and importance is the estimates of the numbers of "ungenerated" actions at the end of their sessions. We asked subjects to estimate the number of remaining actions yet to be generated and the number of "good" solutions yet to be generated. A few subjects realized that the possibilities are almost infinite, but most subjects readily supplied a numerical estimate. The median estimates of the three groups were in the neighborhood of 4.5 to 6.0 for the first question, and 2.5 to 3.0 for the second question. Similar estimates were obtained in another study (17), were subjects estimated that there were between 4 and 5 "reasonable" actions that were yet ungenerated.

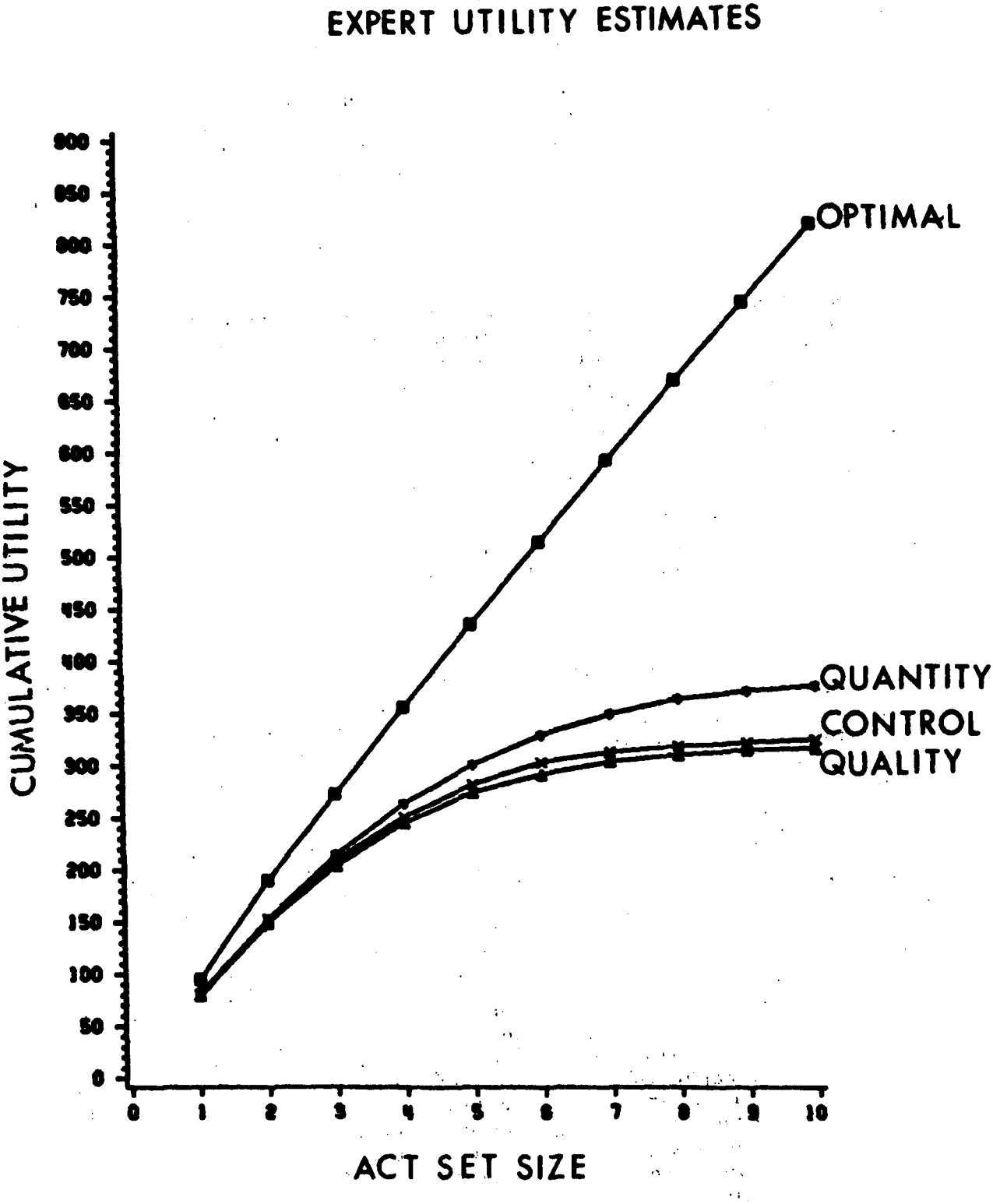


Figure 5.3. "Limbs and Branches" performance scores for the "Quality" and "Quantity" incentive conditions, compared to a Control condition and the lower-bound estimate of optimal performance.

Thus, the typical subject generates two to three good ideas, and thinks that about three to five other good ideas exist that they couldn't generate in the Parking problem, when in fact there are more than 25 good actions that could be suggested. This result is similar to that obtained in our hypothesis generation research, subjects apparently do not realize that their performance is impoverished, apparently because the same memory process that fail in the generation of actions are also involved in the estimation of the number of "ungenerated" actions. These results obviously are important. They suggest that impoverished act generation is a "silent" disease, similar to high blood pressure, of which the patient is not aware. This topic is discussed at greater length in chapter 8, where its implications are explored in more detail.

Describing the subjective representation of the decision space. To aid our studies of act generation performance, we also studied the subject's representation of the problem space using multidimensional scaling and cluster analysis. This research (12) was used both as a means of deriving a more objective hierarchical tree structure for use in our performance measures, and as a technique for better understanding typical subject's perception of the Parking problem. Do subjects see the Parking problem as a member of the general class of shortage problems, or as a unique problem with few relationships to other problems? If the Parking problem is seen as an instance of a more general shortage problem, then subjects should structure its space around generic strategies such as "increase the supply", "reduce the demand", or "use available resources more effectively".

Subjects made similarity judgments on 43 actions taken from the parking problem and these judgments were analyzed using nonmetric multidimensional scaling using the ALSCAL program (Shepard, 1974). Three factors were extracted. A second experiment was conducted to interpret these dimensions using multiple correlation techniques. Subjects rated each of the 43 actions in terms of the extent to which they represented generic strategies, specific strategies, personal goals (ie. being able to find a parking spot), and evaluative dimensions (ie. feasibility, cost, and political dimensions). It was found that the first dimension was best described as "involving alternative forms of transportation", the second apparently was best described as "change the current parking priorities", and the third dimension appeared to be related to "building new facilities".

Apparently the subjects did not structure the Parking problem as a generic shortage problem, as they were not inclined to describe the acts in these terms. Instead, they seemed to organize their thinking around specific strategies for solving the problem such as those listed in the preceding paragraph. These are concrete examples of the more general generic strategies. For example, "building new facilities" is a specific instance of the generic strategy "increase supply". Apparently the subjects prefer to think of the problem in concrete terms. A representation of the subject's space in these more concrete terms is shown in figure 5.4. Shown is a representation of the 43 acts arranged in the three-dimensional space described by these three dimensions.

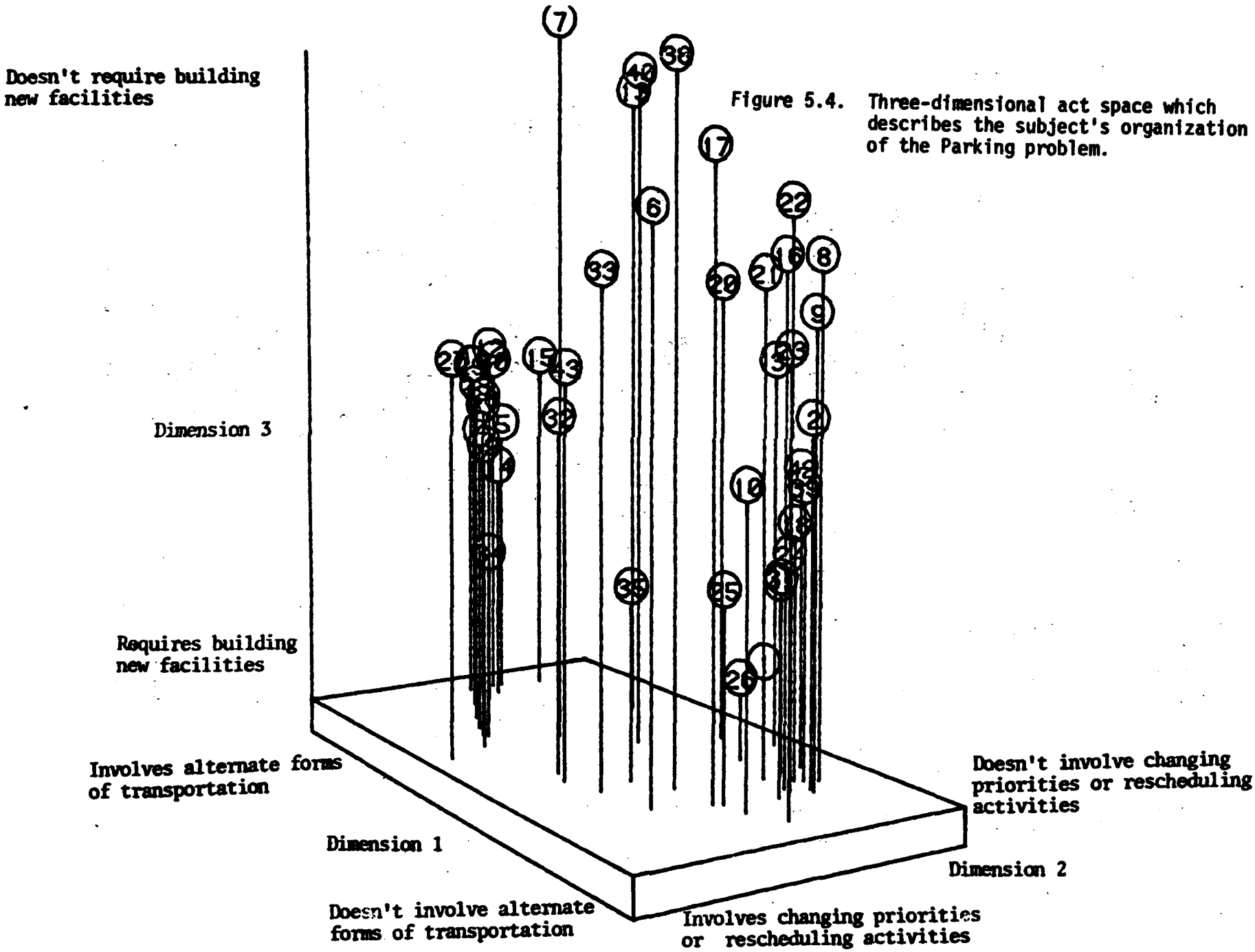


Figure 5.4. Three-dimensional act space which describes the subject's organization of the Parking problem.

Several types of hierarchical cluster analysis were performed on the 43 acts using the TAXON procedure in the NT-SYS statistical package (Rohf, Kishpaugh & Kirk, 1979). The Average and Complete Linkage methods gave similar results, and the Average method was chosen as the best structure because its clusters made more sense in several instances. A modified version of this cluster analysis was created for use in subsequent studies because it was necessary to incorporate additional acts that were generated after the cluster analysis was performed. This modified structure is shown in figure 5.5.

This structure was was used in the incentive study (12) and all subsequent studies using the Parking problem. Its major advantage is that it is empirically defined from the subject's responses, rather than being based on the idiosyncratic analysis of an experimenter. However, the results obtained do not appear to be influenced much by the technique used to define the structure, as will be discussed below.

Factors that might affect the generality of the act generation studies. The results of our act generation studies seemed to be so important that we performed a number of analyses to check the generality of these results. First, as mentioned above and reported in 10, we calculated the performance scores using both the tree structure created by cluster analysis and a tree structure created by one of the experimenters. Very similar results were found, suggesting that the result of impoverished act generation was not due to the particular structure used in the analysis.

Second, very similar results were found when subjects were given payoffs for quality of actions, quantity of actions (12), or simply told to respond with anything that came to mind (10). Thus the impoverished performance does not appear to be due to a lack of motivation, or nuances in the wording of the instructions.

Third, calculating the performance scores with utilities supplied by the type of subjects who were used as act generators or subject-matter experts had only a small effect on the performance scores, but the impression of impoverished performance remained using either source of utilities.

Fourth, in calculations done for the publication version of 10, we calculated the performance score measures repeatedly using the utilities supplied by each of the individual utility estimation subjects rather than using a median utility measure calculated for the group of utility estimators. While there was considerable variability produced by this manipulation, the same general conclusions of impoverished performance emerged.

Fifth, we obtained approximately the same effects in both the Parking and the living problems. Pitz, Sachs, & Heerboth (1980) report similar results in their study comparing various elicitation techniques.

Why is it that none of these manipulations cause a big enough change to substantially alter the conclusions? We suspect that when you start with performance such as that shown in table 5.1, where subjects were so unlikely to generate the best ideas, nuances such as instructions, the structure, and utilities used have little effect on the overall conclusion,

Figure 5.5. The decision tree used to classify acts developed from cluster analysis.

1.0 Create more parking spaces by building new facilities.
- 1.1 Build a highrise parking structure.
- 1.2 Build underground parking.
- 1.3 Build new parking lots on university land.
- 1.4 Expand existing surface lots.
- 1.5 Tear down old buildings to create space to build parking spaces.
- 1.6 Buy land to build additional parking lots.
- 1.7 Take the parking problem into account in future planning of expansion of the university.
- 1.8 Build additional remote lots and run buses to campus.

2.0 Obtain more parking spaces without actually building new facilities.
- 2.1 Use space more effectively (e.g., decrease the width of spaces).
- 2.2 Request the use of areas around campus (e.g., church lots) for additional parking.
- 2.3 Use city streets near campus for university parking.

3.0 Alternative forms of transportation--Group.
- 3.1 Encourage people to carpool.
- 3.2 Force certain people (e.g., commuters, faculty) to carpool.
- 3.3 Encourage people to use the C.A.R.T. system on campus.
- 3.4 Improve the C.A.R.T. system on campus.
- 3.5 Expand the C.A.R.T. system to include other areas of Norman.
- 3.6 Work with the near-by communities to form a mass transit system.

4.0 Alternative forms of transportation—Individual.
- 4.1 Encourage use of bicycles and/or motorcycles.
- 4.2 Make individual transportation safer.
- 4.3 Encourage other forms of individual transportation (e.g.,walking).

5.0 Change current university policies regarding parking.
- 5.1 Eliminate parking priorities.
- 5.2 Allow students to park in restricted areas (e.g., faculty/staff lots. during certain hours (e.g., after 6:00 p.m.).
- 5.3 Set time restrictions (e.g., 2 hour parking) on more lots.
- 5.4 Enforce existing parking regulations more strictly.
- 5.5 Make certain people (e.g., commuters) park in certain places.
- 5.6 Limit the number of cars on campus by not letting certain people (e.g., freshmen) have cars on campus.
- 5.7 Distribute a limited number of parking stickers.
- 5.8 Assign a specific space for each driver.
- 5.9 Outlaw cars on campus for everyone.
- 5.10 Allow certain people (e.g., those who have even number license plates) to only park on certain days of the week.
- 5.11 Increase the price of parking stickers.

6.0 Reduce the number of people who need to park.
- 6.1 Offer more correspondence courses.
- 6.2 Establish branch campuses of the university.
- 6.3 Reschedule activities and/or classes to change demand.
- 6.4 Provide housing or improve existing housing so people can walk.
- 6.5 Reduce the student population. For example, limit enrollment.
- 6.6 Have someone drop students & faculty/staff off and pick them up.

7.0 Indirect strategies for solving the problem.
- 7.1 Appeals to good judgment.
- 7.2 Ways to make money to solve the problem.
- 7.3 Suggestions for ways to come up with solutions.

8.0 "Flaky" acts (e.g., issue everyone a set of wings).

that performance is impoverished. Similar performance has been observed in two other studies yet to be described.

In all our research, we have found only one exception to this general conclusion. This exception is in one sense the exception that proves the rule. As will be discussed extensively in chapter 6, subjects who score exceptionally high in measures of divergent thinking do not show impoverished performance. Instead, they show exceptionally good performance, with the best of these subjects performing slightly below the lower-bound estimate of optimal performance. We believe the discovery of the importance of divergent thinking, and the fact that a few of these subjects approximate our estimate of lower-bound optimal performance, lends support to our characterization of **unselected** subjects as impoverished. If no subjects approached our estimate of optimal performance, it could be said that this estimate is too high.

CHAPTER 6. RECURRENT TOPICS IN HYPOTHESIS, ACT, AND OUTCOME GENERATION

Performance in small groups.

Group hypothesis generation. One strategy that has frequently been used to improve problem solving performance is to work in small groups rather than as individuals. The mounting evidence that individual hypothesis generators produced impoverished hypothesis sets suggested that it might be profitable to investigate group hypothesis generation to determine the improvement that working in a group affords.

In a study using the "Majors from Classes" task (8), subjects either generated majors from classes as individuals, or as a member of an interacting group of four subjects. The pooling technique was used again, but in this case the veridical posteripr probabilities of majors given classes were available, and were used rather than a count of logically-possible hypotheses. Thus the posterior probability of hypothesis sets generated by either individuals or small groups could be calculated.

The mean probability of the hypothesis set for individuals was .335 while interacting groups of four had a mean probability of .427. The means reported are the probabilities that the hypothesis sets contained the "true" hypothesis. Thus, as one might expect, group performance is superior to individual performance. However, both individuals and small groups were impoverished hypothesis generators. Although subjects in this task were told to neglect very unlikely ($p<.02$) hypotheses, and so could not be expected to have hypothesis sets with a probability of 1.00, the probability of the hypothesis set of an optimal subject would have been 0.906. There is ground for much improvement in these performances.

These results suggested a general way of examining at least two factors which affect group performance. One factor is the potential increase in information that the group provides. The adage, "Two heads are better than one," has validity in this sense. As group size increases, the amount of new information added by each new member should become less, but the total information possessed by the group should increase. The pooling process described earlier is one way to measure the information possessed by the group, and it provides a natural metric for expressing how the amount of task-relevant information increases as group-size increases.

The second major factor in interacting groups is the social interaction which occurs. Under certain conditions, social interaction may be facilitive, but it is usually found to inhibit group performance (8). When the performance of individuals, synthetic groups, and interacting groups are compared, it is possible to partition performance into an informational component and a social component. In the present experiment, the information that could be gained from pooling the information of four individuals is estimated to be a .205 increment in hypothesis set plausibility (.540-.335 =.205). Social interaction, however, caused a decrement in performance of .113, as calculated from differences in performance of the interacting and synthetic groups (.427-.540= -.113). The actual gain in performance of an interacting group over an individual is 0.092, and this difference results from the additive combination of informational and social factors.

Group act generation. The results of our study on group hypothesis generation left one interesting question unanswered. Why do people work together in groups when synthetic groups are far more efficient? One obvious answer is that interacting groups may fulfill other important motivational, social, or coordinating functions. However, there also may be an informational effect that makes it profitable to work together in groups. This effect, which we formally call the information exchange component, but informally call the "ping-pong" effect, occurs when the ideas of another person, when combined with your own, produce a synergistic product that is greater than the sum of its parts. This effect has also been called "piggybacking" (Day, 1980) and "hitch-hiking" (Stein, 1975), but its existence had not been empirically demonstrated.

However, it is also possible that group interaction has a negative net effect. This might occur if an individual spent time elaborating the ideas of others rather than thinking independently about the problem. Thus, two individual, working interactively, might chase each other's ideas, proposing minor elaborations, and achieve less to show for their time than if they had been working independently.

The problem in studying the "ping-pong" effect is the same as in the previous study; it is very difficult to disentangle social factors from informational factors. We used the same approach of devising experimental manipulations in our study on this topic (12). In addition to using interacting groups and individuals, we introduced a third information exchange (IE) condition which simulated the face-to-face interchange of ideas in an interacting group without allowing social interaction. In the third condition, the ideas generated by each subject were transmitted immediately via a computer to the other subject in a different room. Both subjects thought that the ideas received from the other subject were generated by the computer which they believed was running an experimental AI-based program. They were told the program was attempting to generate ideas related to theirs in an effort to help them. Subjects therefore believed that they were interacting only with a computer, when in fact, they were seeing the ideas of the other subject as they were generated.

We employed the Parking problem, and computed the performance score for individuals, interacting groups of size two, synthetic groups of size two, and IE groups of size two. Both the IE and the synthetic groups were found to be superior to the interacting group. However, there was no significant difference between the IE and the synthetic group on this summary measure of performance. The interacting group was not found to be significantly better than a single person working alone.

Using a simple additive model to partition performance, we estimated that the IE group showed a 6.9% improvement over a synthetic group of the same size. However, the IE group generated 41% more actions that the synthetic group, but only 15% of this 41% gain were unique actions. Thus about two thirds of the gain in the number of acts generated represented minor elaborations of the ideas of the other member of the IE group, and these elaborations did not yield higher utility actions.

In summary, the information exchange that interacting groups can exploit does seem to have a small positive effect. This gain, however, is swamped by other negative social effects that occur in a interacting group.

Furthermore, the actions that are generated by information exchange tend to be minor elaborations of the actions that have already been generated. It appears that the the suggestion of Osborn (1957) that "the average person can think up twice as many ideas when working with a group than when working alone" (pp. 228-229) is wildly optimistic. In fact, our interacting groups performed 41% less effectively than IE groups of the same size and no better than individuals working alone. Our research shows that one popular justification for working in interacting groups, that of information exchange, does not result in sufficiently large differences to justify working in interacting groups with the types of problems we studied. However, we found that the use of synthetic groups is very effective in improving act generation performance. The "Delphi technique" (Lindstone & Turoff, 1975) is one such method of exploiting the gains in information in a group, while reducing negative social factors.

Schemata, frames, and inferences in hypothesis, act and outcome generation.

Schemata in hypothesis generation. One informal observation that we made in several hypothesis generation studies was that our subjects appeared to be blind to certain classes of hypotheses. When asked to generate hypotheses, subjects sometimes generated hypotheses that seemed to be based on an implicit interpretation of the data. Other subjects seemed to adopt different interpretations of the data, and to generate a correspondingly different set of hypotheses. This observation suggests that sometimes interpretations of the data influence the memory retrieval process, thus biasing the subjects toward one type of hypothesis and against another type. This general phenomenon has received some attention in cognitive psychology. The organization of data into a meaningful pattern by making inferences about their meaning is termed a **schema** in cognitive literature.

When the hypothesis generator is attempting to add hypotheses to a set of hypotheses that have already been suggested, schemata might be expected to play an important role. This situation may occur when the hypothesis generator "inherits" a decision problem. As scientists we are constantly faced with inherited hypotheses which may bias our interpretation of the data and our generation of new hypotheses. Often "inherited" hypotheses suggest particular interpretations of the data which might seem forced in the absence of these hypotheses. In our natural desire to obtain closure, we may accept certain interpretations which relate data to hypotheses. These interpretations may come to represent the data and may even be encoded in memory in lieu of the data. When we attempt to generate new hypotheses, the schema that organized the data may be used instead of the data in searching memory. To the extent that this happens, the hypothesis generation process may be biased.

A study (6) was performed to investigate these ideas and to propose a partial cure for any such tendencies on the part of the hypothesis generator. In this study subjects were given several ambiguous data which could be interpreted by using several schemata. All subjects were encouraged to generate as many hypotheses consistent with the data as possible. The existence of an "inherited" hypothesis was simulated in some conditions by giving the subject one of several hypotheses to evaluate. These hypotheses were good exemplars of several different schemata that could be used to explain the data. The problems involved generating

possible hypotheses about an unknown geographical area known as "X". For example, subjects in one problem were told that one hypothesis that was consistent with area "X" was a bakery. Available data were that 1) Most people spend only a short time in area X, 2) Area X contains unusual smells, and 3) Area X is only open during business hours. Subjects who "inherited" the "bakery" hypothesis were more likely to generate hypotheses such as "restaurant," "fruit stand," or "flower shop." Other subjects were given this same problem but "inherited" the hypothesis "dump" rather than "bakery". These subjects were more likely to generate different hypotheses such as "chemical plant," "sewer treatment plant," or "public restroom." The two schemata that these two hypotheses suggest are "pleasant" and "unpleasant" areas, respectively.

As might be expected, subjects used some schema more often than others. Subjects in the "no hypothesis" condition were more than twice as likely to generate hypotheses consistent with the popular schema than the rare schema. If the hypothesis provided to the subjects suggested a schema that was popular, then there was relatively little change in hypothesis generation performance as compared to the "no hypothesis" subjects. If, however, the schema suggested by the hypothesis was rare, and hence less likely to occur to the subjects spontaneously, then there was a dramatic increase in the number of hypotheses generated that were consistent with that schema. There was also a corresponding decrease in hypotheses generated that were consistent with the popular schema. These results are evidence for the biasing effects of schemata.

We also explored a simple technique for reducing this bias. A second study was run using much the same procedure as the first, except that the subjects who "inherited" hypotheses were asked to generate a hypothesis which was consistent with the data "for another reason." For the subjects who successfully generated such a hypothesis, the bias was practically eliminated.

Frames or perspectives in act generation. Historians delight in explaining how battles are lost because the commander on the loosing side had a limited perspective of the situation. Recently, President Galtieri of Argentina invaded the Faulkland Islands, and received a humiliating rebuff from the British. Galtieri, the historians claim, did not properly anticipate the British reaction to an invasion of their territory. It is easy to make these analyses in hindsight (Fischhoff, 1975), but harder to show these effects under controlled laboratory conditions.

One project (14) was explicitly concerned with the effect of the decision maker's frame (Tversky & Kahneman, 1980) or perspective on act generation. In this series of two experiments, subjects were asked to play the role of either the French government, guerrillas who had invaded the French Embassy in a hypothetical South American country and captured French hostages, or the French hostages themselves. All subjects attempted to generate the actions that the French government would take to gain the hostages release, estimated the likelihood of the various French actions, and the French government's preference for the various actions.

Our predominant impression was that there was little or no effect attributable to the decision maker's perspective. There are two alternative explanations for the lack of a perspective effect. One, of course, has to

do with the inadequacies of a laboratory simulation to fully capture the nuances of a real situation. It is entirely possible that perspective effects might be observed in a real-world setting, or with a different problem. Alternatively, it may be that perspective only has an effect in hindsight in explanations of historians (Fischhoff, 1975).

Causal explanation in outcome generation. Our final project in this general area was an examination of the role of casual explanation in outcome generation (15). Consider the decision maker who is attempting to generate actions to reach some goal outcome. As was discussed previously in Chapters 3 & 5, we assume that the decision maker constructs a scenario leading to the goal. In this series of studies, we examined the effect that the construction of this goal-directed scenario has on the generation of other outcomes that do not lead to the goal. This topic is of considerable interest because act generation is a goal-directed activity. When a decision maker constructs a scenario leading to the goal, attention is initially focused on the chain of plausible actions and outcomes that lead to the goal. The construction of the scenario requires that certain differences in a causal field (Einhorn & Hogarth, 1982) be created. Therefore, it is possible that the creation of this initial goal-directed scenario makes it more difficult to construct alternate scenarios leading to other outcomes involving other causal factors.

For example, consider the entrepreneur who has invented a new "widget" in hopes of becoming wealthy. This individual may construct a scenario that involves forming a company to manufacture and market widgets. Widgets catch on, and soon every household has one, and the inventor retires to a life of wealth and leisure. As a matter of fact, most such ventures fail, usually because the entrepreneur fails to anticipate all the alternate outcomes, which represent the pitfalls in the plan. Often, the inventor is the only person who really needed a widget, or a competitor with more capital steals the essence of the idea, or the new firm is so undercaptialized that it fails before a market for widgets can be created. Therefore, the question of interest is whether the creation of an initial scenario makes it more difficult to create other alternate scenarios leading to other outcomes.

Our approach to studying this problem involved having subjects construct an initial scenario leading to one of several specified outcomes. Subjects were provided with a case history involving a young man who assumes a small-town Ford dealership upon the death of his father. Subjects were asked to write a plausible and convincing scenario leading to one of four outcomes which we provided. These outcomes involved either the success or failure of the dealership due to either the personality of the young man, or the economy. After the subjects had created the designated scenario, they engaged in a variety of activities that we hoped would capture any changes that creating the initial scenario might produce. Specifically, they were asked 1) to make their own judgment as to the probable outcome for the business, 2) to identify factors that would be important to its success or failure, 3) to generate at least five alternate scenarios about the business, 4) to rate these scenarios in terms of likelihood, 5) rate the importance of experimenter-supplied factors that might influence success or failure, and finally 6) to make predictions as to how these factors would turn out in the future.

The results suggested that after subjects have created a scenario

leading to a specified outcome, the factors that they use as causal factors in subsequent scenarios are biased in that they tend to focus on the same factors that were used in the original explanation. The important causal factors tend to remain the salient explanations in subsequent scenarios. For example, if their initial scenario was created to explain why the young man's personality lead to success of the car dealership, subsequent scenarios tended to focus on his personality traits, and tended to ignore economic factors.

Even though the content of subsequent scenarios was biased by the initial scenario, the number of success and failure scenarios and the likelihood estimates for the scenarios remained approximately equal in the various groups. This result is similar to that obtained by Pennington (1981) in hindsight and foresight judgments.

However, even after being forced to generate a number of alternate scenarios, a debiasing technique used successfully by Slovic and Fischhoff (1977), our subjects showed differences in their importance weightings of various causal factors, and in their predictions as to how these factors would turn out in the future. These results are quite interesting because they demonstrate how a single causal explanation can actually change what causative factors the subject believes are important, and their predictions about the future.

These results are also important because they demonstrate that the very act of creating a scenario causes the outcome generator to organize the world in a certain way, and this organization persists in the creation of other scenarios that might spring from that act.

The second study in this project examined two possible cognitive mechanisms that might account for the results of the first study. One such mechanism is selective encoding of information. Possibly subjects who were asked to explain why economic factors would lead to the failure of the dealership only recalled the information consistent with an explanation involving an economic failure.

Alternately, it may be that the initial inferences that the subjects make from the case history to support the first scenario are remembered and the subject does not explore other, alternate inferences that could be made. For example if they interpret the young man's poor academic performance and frequent changes in major as evidence that he is a "goof-off" to support a scenario of personal failure, other alternate inferences that could be made from the same data may be neglected in the generation of scenarios.

Our technique for distinguishing between these two alternate explanations involved using a recall test to determine the extent of selective encoding, and an inference procedure to determine if the groups were making different inferences from the case history. Subjects constructed an initial scenario as in the previous experiment, then they took the recall test and made inferences.

The data from the recall test showed that the subjects had a good recollection of the case history, but there was little evidence in favor of the selective encoding explanation. However, once the information in the

case history is used to make inferences supporting the initial scenario, these inferences persist, and the subjects have difficulty reinterpreting the information in an unbiased manner.

The results from this project have important implications for decision analysis. In decision analysis, the client is asked to identify the possible outcomes that might result from an action. This process is by necessity a serial process. Our results suggest that the process of creating the initial outcome for such an analysis may cause certain inferences to be made and causes certain causal factors to be seen as relevant. Subsequent outcomes tend to be generated using these inferences and these causal factors, even though the decision maker would be better off to explore other possible inferences that could be made in the decision situation.

Individual differences in predecision processes.

Individual differences in hypothesis generation. We noticed pronounced individual differences in hypothesis-generation ability among our subjects. Some subjects generated more than twice as many hypotheses as a typical subject, and although the typical subject generated impoverished hypothesis sets, there was an occasional exception to this rule. For practical reasons it might be useful to have a simple means of estimating the hypothesis generation ability of an individual, and the cognitive differences between good and poor hypothesis generators might be enlightening.

Our first study on this topic (5) was fairly traditional. First, we developed criterion measures of hypothesis generation performance. One criterion task was an abstract photo-reconnaissance task where the decision maker was given a simplified copy of a map from the U. S. Census tract. An unknown area was marked on the map, and the subjects' task was to generate as many hypotheses as possible about the identity of this unknown area using the map and several additional items of information. The criterion hypothesis generation score which was finally developed depended on both the quantity and quality of the hypotheses that the subject generated. Our choice of predictor variables was guided by several considerations. First, the divergent thinking involved in hypothesis generation seemed to be similar to the divergent thinking used in some creative activities. We surveyed this literature and identified several tests that were designed to measure divergent thinking and creativity. These tests were the Alternate Uses test, the Remote Associations test, and a subtest of the AC test of Creative Ability which we called "Possible Reasons". Second, other tests were included to measure such factors as inductive reasoning, and the ability to use the information provided by the tasks.

Alternate Uses was found to be by far the best predictor of hypothesis generation performance ($r=.27$), but none of the predictors accounted for much of the variance in this ability.

In the second study of this series (5), we took steps to increase the reliability of the criterion measure of hypothesis generation. The Alternate Uses test was retained, and the other tests of creative problem solving were dropped. Tests of general academic achievement (the ACT), and intellectual ability (the Information scale of the WAIS) were added to the battery of predictors. Several different versions of Alternate Uses were

also developed to measure possible cognitive skills that might be involved in hypothesis generation.

Our modifications of the Alternate Uses test were based on the following argument. The Alternate Uses test involves generating alternate uses for common household items, such as a coat hanger. Subjects are instructed to generate as many possible uses for a coat hanger as possible. Many of the possible uses for a coat hanger involve using a different schema than "a device for storing clothing in a closet". A coat hanger has many attributes which can be exploited in various ways. It is metal, it conducts electricity, it is ductile, it is long and thin, it is fairly rigid, it doesn't burn at household temperatures, etc. The implicit properties of this object could be used as retrieval cues to search memory. Various combinations of these attributes suggest different schemata such as "a device to open a car door" (long, thin, rigid, and ductile), or "marshmallow roaster" (long, thin, rigid and fire resistant). Therefore, a subject who performs well at this task might first analyze an object to determine implicit dimensions or attributes and then use various combinations of these dimensions as retrieval cues for alternate uses. Performance in the Alternate Uses task and in hypothesis generation might have two components, the retrieval of the implicit dimensions and the use of this implicit information to retrieve uses or hypotheses, depending on the task.

We modified the Alternative Uses test to create two new versions of the test to use in addition to the original version. One of the new versions measured the subjects' ability to retrieve the attributes of the household objects that might be useful retrieval cues, and a second version measured the subjects' ability to generate uses when these attributes or dimensions were explicitly provided by the experimenter.

There were several interesting results from this experiment. First, as has been found in every study dealing with this topic, hypothesis generation of the average subject was impoverished. The mean hypothesis generation score for subjects was about 3 "good" hypotheses per problem, while the lower-bound estimate of the maximum number of logically possible hypotheses was approximately 26 "good" hypotheses and 43 "fair" hypotheses per problem. Second, the correlation between the Alternate Uses test and the criterion measure of hypothesis generation was .51, a considerable gain in predictive power over the previous experiment. This correlation could undoubtedly be increased by item-selection and other methods of test refinement. Such further development could perhaps convert the alternate uses test from a research tool to a useful predictor of hypothesis generation performance. Third, achievement and general intelligence were shown to be only weakly related to hypothesis generation performance.

Both of the proposed components of hypothesis generation performance were shown to be important. The "retrieval of implicit attributes" component and the "retrieval of hypotheses from attributes" component were significantly related to hypothesis generation performance. An analysis of variance was performed on these data which showed that these two components are additive, uncorrelated factors. Subjects who scored below the median on both components generated, on the average, 2.15 "good" hypotheses per problem while subjects who scored above the median on both of these components generated, on the average, 3.6 "good" hypotheses per

problem, 67% better. This study, therefore, has identified two cognitive skills that appear to be important in hypothesis generation.

Individual differences in act generation. Large individual differences were also found in act generation performance. In fact, they were so large as to be the bane of our existence, and much of our effort was devoted to developing experimental procedures and manipulations that were robust enough to survive the error variance that these individual differences created. In our many act generation studies, for example, the worst subject typically generated two or three actions, while the best subject typically generated more than 30.

However, one project (17), designed to explore expert act generation, is the best example of the extreme impact that individual differences can have on act generation performance. Our initial goal was to examine expert act generation performance. However, as will be explained below, we ran into extreme individual differences in the course of the project, so extreme that we could really say little about expertise in act generation, but considerable about the effects of individual differences. For this reason, we discuss this project under the heading of individual differences, even though we have a section on expertise that immediately follows this section.

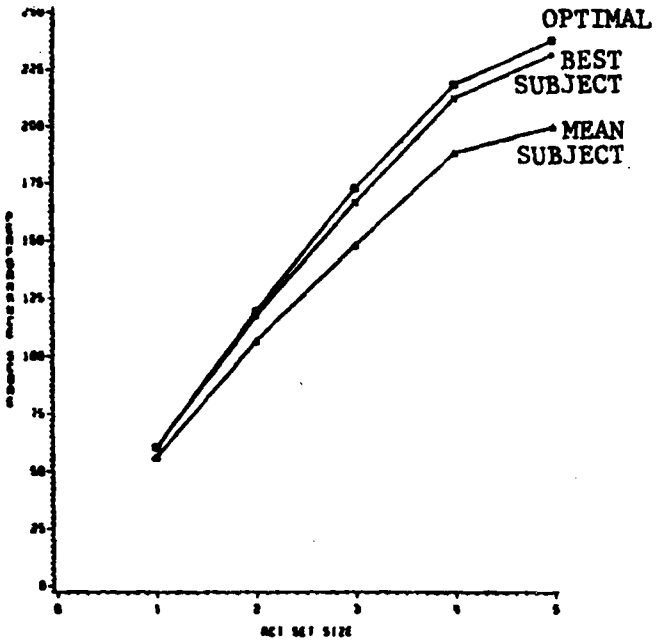
Our original goal was to study expertise in act generation using subject-matter experts. The experts and task that we chose for this purpose were graduate students at the University of Oklahoma, and the task involved generating all possible actions to improve the recruitment and retention of high-quality and motivated graduate students into the experimental psychology program. Graduate students generated actions in several sessions, and spent approximately five hours each in the experiment generating actions and making utility estimates over a period of at least a week.

The results were a surprise in the light of our earlier investigations on expertise in hypothesis generation. In the hypothesis generation studies we had used tasks that both experts and non-experts could perform, and we found that both experts and non-experts displayed similar impoverished hypothesis generation. (See the next section for a more complete discussion of these results.) The graduate student experts, however, did remarkably well, so well in fact that their act generation performance left little to be desired! Graduate students typically generated three to four times as many actions as the typical undergraduate, and these actions were of higher quality. Figure 6.1 shows the performance of the graduate students in the upper two panels, and typical undergraduate performance in the lower two panels. Shown are the "limbs" and "limbs and branches" cumulative performance scores described in chapter V. Although there are a number of differences between the upper two panels and the lower panels such as the nature of the problem, an informal comparison reveals the large difference in performance.

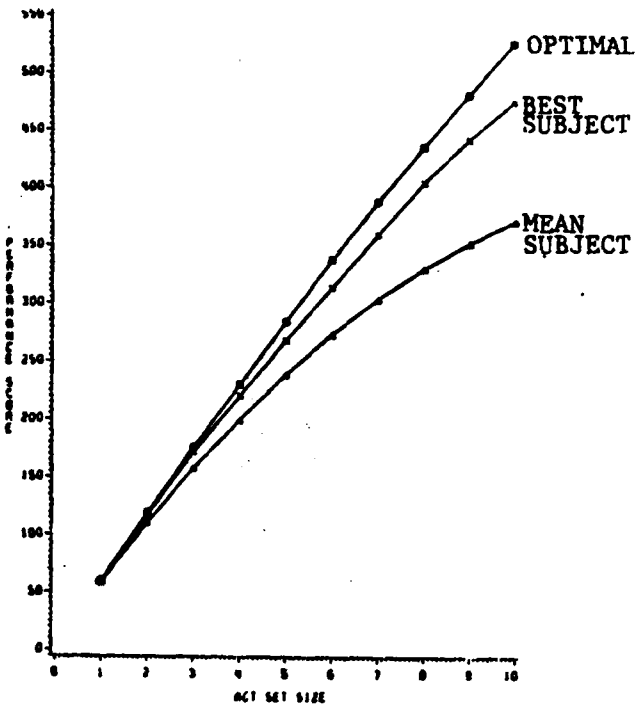
We were suspicious of an expertise interpretation of this result because our previous research on hypothesis generation suggested that both experts and non-experts have similar cognitive deficiencies and because act generation and hypothesis generation are so similar. Furthermore, we had not used a non-expert group for purposes of comparison because the task

GRADUATE DEPARTMENT PROBLEM

PERFORMANCE SCORE FUNCTIONS FOR LIMBS ANALYSIS

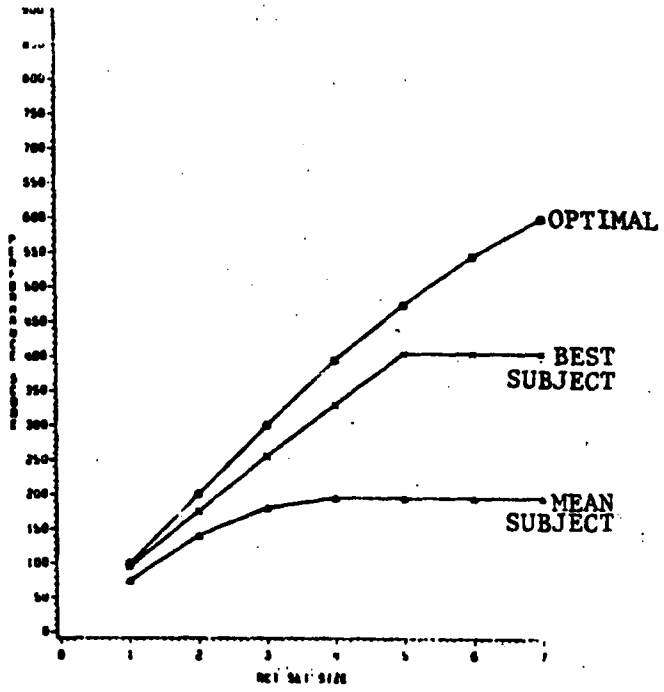


PERFORMANCE SCORE FUNCTIONS FOR LIMBS AND BRANCHES ANALYSIS



LIVING PROBLEM

PERFORMANCE SCORE FUNCTIONS FOR LIMBS ANALYSIS



PERFORMANCE SCORE FUNCTIONS FOR LIMBS AND BRANCHES ANALYSIS

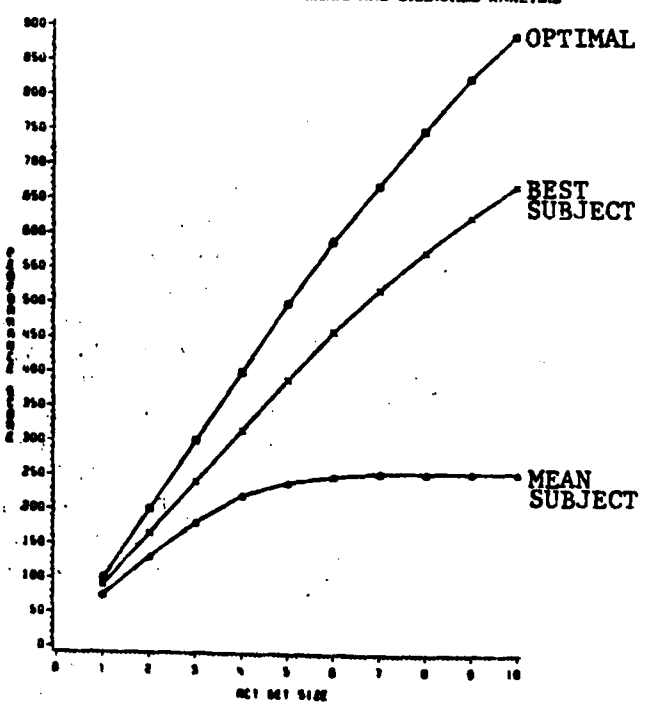


Figure 6.1. COMPARISON OF PERFORMANCE SCORE FUNCTIONS FOR EXPERTS ON THE "GRADUATE DEPARTMENT PROBLEM" AND NON-EXPERTS ON THE "LIVING PROBLEM" (FROM PREVIOUS RESEARCH). THE GRAPHS ON THE LEFT SHOW FUNCTIONS FOR LIMBS GENERATED AND THE GRAPHS ON THE RIGHT SHOW FUNCTIONS FOR LIMBS AND BRANCHES GENERATED.

chosen seemed to require more subject-matter expertise than any undergraduate possesses.

Consequently, we decided to perform a second study as check on the first study. We reasoned that if the effect in the first study was due to expertise, and if we used a task where our graduate students were not expert, we should see little difference between them and undergraduates. The task we chose was the "Living" problem used previously, and the graduate students used in the previous study and a group of urdergraduates performed this task under the conditions that we had used in our earlier experiments.

We were also interested in exploring an alternate explanation for these results. Graduate students are at the tip of the selection pyramid as compared to undergraduates. A graduate student has survived four more selection processes than the typical undergraduate. Graduate students have graduated from college, self-selected themselves in terms of applying to graduate school, been selected into graduate school, and all but one of the graduate student subjects had been admitted into candidacy for the Ph.D. program. One of the informal criteria used in the last three of these selections is creativity and divergent thinking ability. Therefore, we were interested in whether the good performance of the graduate students could be due superior divergent thinking ability, and administered the "Alternate Uses" test described above to both groups of subjects.

The results supported the divergent thinking explanation. First, graduate students scored nearly twice as high as undergraduates on the "Alternate Uses" test, and act generation performance correlated .43 (limb and branch scores) and .49 (limb scores) with this test. Second, there was almost no overlap in the two groups. The average graduate student generated 5.0 limbs (out of six), while the average undergraduate generated only 3.2. Only four undergraduates exceeded the performance of the worst graduate student on this measure. Branch performance was similar.

Figure 6.2 shows the performance scores of the graduate and the undergraduate students. Notice that the graduate students perform in a highly similar manner on a problem on which they are expert (figure 6.1) and on a problem on which they are not expert. The undergraduates show typical impoverished performance.

We also investigated other alternative explanations for this effect. Perhaps, for example, the graduate students were more expert on the Living Problem because of their greater age and experience. If this is the case, then age should be correlated with performance on this problem. Using a large sample of subjects from a previous experiment, we found that the correlation of age with performance on the Living Problem was -.05.

It appears that selecting subjects in terms of divergent thinking ability, which we inadvertently did in this project, has a profound effect on their performance. This was the first occasion in which we used subjects of the highest intellectual abilities, and the gain in performance is the largest we observed in any of our studies.

This project both validates and qualifies our earlier conclusions. In terms of validation, the good performance of the graduate students in

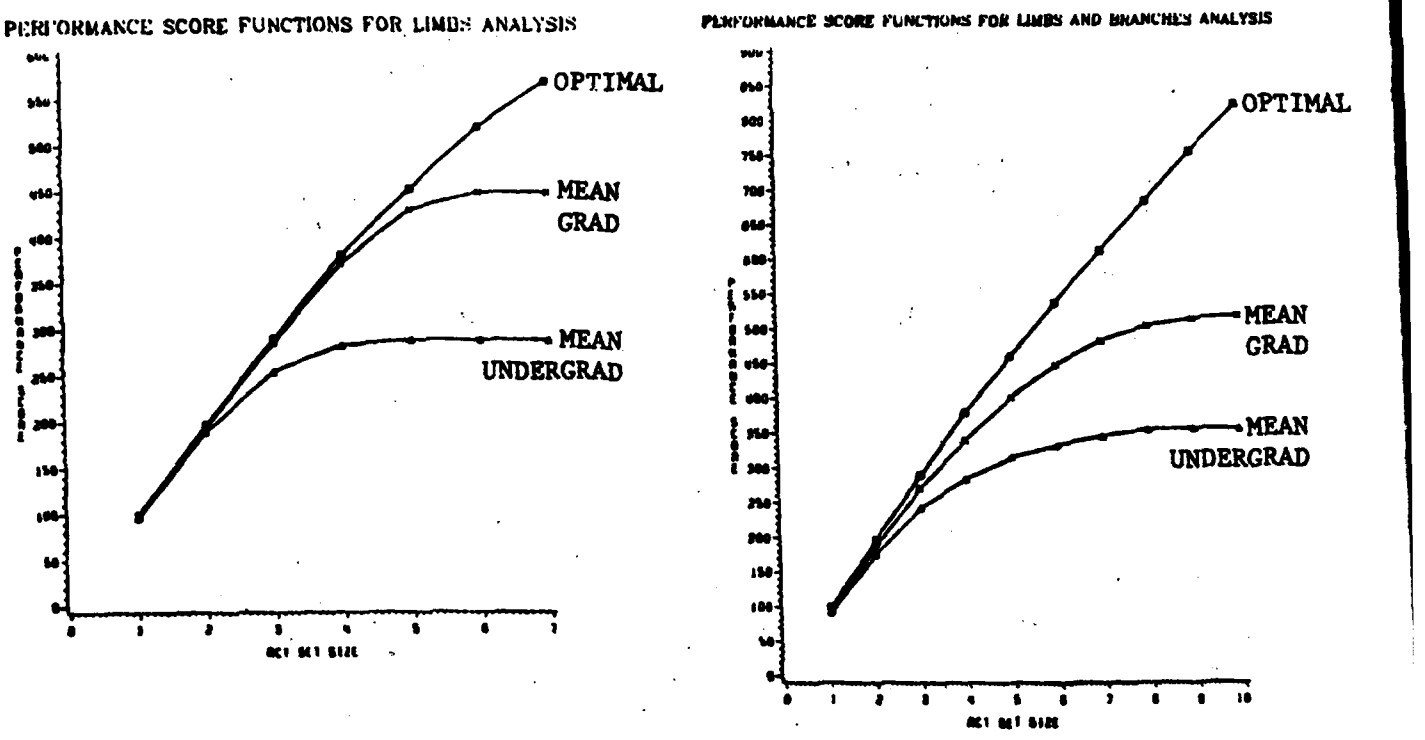


Figure 6.2. PERFORMANCE SCORE FUNCTIONS FOR LIMBS GENERATED AND FOR LIMBS AND BRANCHES GENERATED ON THE "LIVING PROBLEM" BY GRADUATE AND UNDERGRADUATE SUBJECTS (EXPERIMENT 2).

general, and the performance of the best graduate student in particular, suggests that our lower-bound estimate of performance is reasonable and that this estimate is not artifactually too high. In the absence of this result our critics could easily make this claim, even though they would not care to so if they inspected the raw data.

We should qualify our conclusion that act generators, like hypothesis generators, are impoverished. The exceptional individual, individuals in perhaps the upper few percent of the population, do not show impoverished behavior. They do remarkably well. However, they are not the typical individual. Our college student subjects are somewhat superior to but closer to the general population. Their typical behavior we believe is impoverished.

Generalizing to expert populations.

Expert hypothesis generation. Most of our studies employed populations of college students, and the generality of results obtained with college students has been questioned. We deliberately included groups of expert subjects in two studies (4, 7) as a check on the generality of our results obtained with college students. We were interested in determining if experts also generated impoverished hypothesis sets and made excessive plausibility estimates. Our purpose was not to show that expertise has no influence on hypothesis generation. In fact, the hypothesis generation tasks used were carefully chosen so that they could be performed by both college students and expert subjects. Other tasks, requiring the specialized knowledge of an expert, could not be performed by college students, and so were not considered as candidate tasks for these experiments.

Our initial bias was that expert subjects would show considerably different performance than non-experts. Much to our surprise, the experts we studied were quite similar to non-experts in the two performances in which we had the most interest. In the protocol analysis study (7), expert mechanics generated almost exactly the same number of hypotheses as non-experts, and both groups generated impoverished hypothesis sets. The quality of hypothesis sets generated by the experts could not be compared to that of non-experts due to task limitations, but both groups displayed similar excessive plausibility estimates.

Another study (4) was performed which involved expert curriculum advisor subjects. This study will be described in more detail in the next chapter, but the same general conclusions can be reached from this study. The results suggest that observed deficiencies in hypothesis generation can be generalized to experts. We do not claim that expertise is unimportant in hypothesis generation. We do believe, however, that even experts will generate impoverished hypothesis sets and will evaluate these sets as being more exhaustive than they really are.

It is important to realize that both of the expert populations studied above had occupations where there is little selection in terms of divergent thinking ability. It does not seem likely that selection as an auto mechanic or as a curriculum advisor has much to do with divergent thinking ability. This may be the reason why we found little difference between expert and non-expert performance. It is also possible, although we have no

evidence whatsoever for it, that much of the so-called "expertise" effect is really an "intellectual abilities" effect, as most experts are selected for training on the basis of their intellectual abilities. These questions await further research.

CHAPTER 7. IMPROVING PREDECISION PERFORMANCE

The primary goal of our research was not to find ways of improving predecision performance. However, two projects were devoted specifically to this topic, and this chapter mentions a number of other studies relevant to this topic.

An artificial memory aid for hypothesis generation.

Our research suggests that many of the deficiencies in hypothesis generation can be traced to difficulties in the hypothesis retrieval process from memory. The aiding study (4) employed an artificial memory to aid hypothesis retrieval. Hypotheses retrieved from the artificial memory were displayed to the subjects, and they could add these hypotheses to the set of hypotheses that they had generated if they wished. The artificial memory supplemented those hypotheses that the subjects were able to retrieve from memory, and exploited the differences between retrieval and recognition in memory. The basic philosophy behind the aid is that subjects may not be able to retrieve a plausible hypothesis from memory, but may be able to recognize that it is a plausible hypothesis. Thus the aid is designed to supplement the memory retrieval process.

We do not intend that this aid be implemented in its present form. The purpose of the investigation was simply to see if such an aid was feasible in certain limited situations.

The artificial memory. Hypothesis generators have used artificial memories of various sorts to aid hypothesis generation. The reference books of a doctor, or the maintenance manuals of a mechanic or an electronics technician are examples of artificial memory aids. These aids are primarily useful in routine situations were common problems are to be solved. They do not usually suggest hypotheses for rare complexes of symptoms or data. Nevertheless, these artificial memories are so useful that they are often consulted, and we often deplore their lack in problem-solving situations. Generally, the information contained in these reference books comes from an authoritative source, and this information is so difficult to collect and collate that it usually exists only for commonly encountered situations.

The problem of constructing an aid to hypothesis retrieval for situations that lack authoritative reference materials is interesting. Consulting an expert would be a possible solution, but we suspect that even experts retrieve incomplete hypothesis sets. Several experts might jointly create a more complete hypothesis set if their hypotheses were pooled; this is one reason why doctors often use consultants when making difficult diagnoses. One effective way to achieve more complete hypothesis sets is to pool the hypothesis sets of individuals, as was done in the group research (8).

A difficult problem still remains. The task of creating a pooled hypothesis set for every possible combination of data or symptoms is difficult or impossible for diagnostic situations where many data are possible. For example, if there are N data possible, and if a simplifying assumption is made that these data are not mutually exclusive, then the possible number of data complexes is 2^N-1, potentially a large set. Therefore it is impossible in many situations to convene a panel of

experts, and ask them to evaluate every possible data complex that might occur; there simple may be too many complexes. Perhaps the answer is to use expert judgment to construct an artificial associative memory, and then interrogate this memory to find hypotheses that are logically consistent with any complex of symptoms or data.

We constructed such an artificial memory. First we asked subjects to generate as many hypotheses as possible for each datum. These hypotheses were pooled across the subjects to create a more-complete hypothesis set than any individual could generate. This set, with comparable sets for all other possible data, was stored in a computer. Thus each datum had many plausible hypotheses associated with it in the computer memory. In use this memory was queried. The tagging model (1) developed for modeling human hypothesis retrieval was used to retrieve hypotheses suggested by a complex of data. Hypotheses were tagged in the artificial memory for each datum in the complex, and those hypotheses that received more than a criterion number of tags were retrieved from the artificial memory and displayed to the hypothesis generator.

An evaluation of the artificial memory. A study (4) was performed to evaluate the extent to which this artificial memory aided hypothesis generation. Subjects were given either one or three courses that a student had taken and were asked to generate as many plausible hypotheses as possible. When the subjects finished hypothesis generation, they either started the next problem, or they were shown the results of the search of the artificial memory. This display consisted of a list of hypotheses that had been retrieved from the artificial memory, and the subjects were allowed to add any hypothesis from this list to their hypothesis sets. There were two groups of subjects. One group were junior or senior students at the University of Oklahoma. The other group was more expert. This group consisted of Curriculum Advisors who were employed by the university to give students advice on course offering and schedule planning. These individuals are experts in the sense that they are intimately familiar with the typical courses of study for each major.

Performance was measured by calculating the posterior probability of the sets of hypotheses that the subjects generated in the aided and unaided conditions. This probability is the probability that the set of generated hypotheses contains the "true" hypothesis. Subjects were told to ignore implausible hypotheses ($P < .02$), and, for this reason, an optimal hypothesis generator should have had a hypothesis set that had a probability 0.906 for the average problem when implausible ($P<.02$) hypotheses are excluded from the calculation.

The unaided performance of both groups was impoverished. Non-experts had mean hypothesis set probabilities of .477, while experts had mean probabilities of .506. This difference is statistically reliable, but experts performed similarly to non-experts, in that both groups generated impoverished hypothesis sets. These number are directly interpretable. It will be recalled that these probabilities are the probability that the true hypothesis is contained in the set of generated hypotheses. An optimal hypothesis generator who generated all hypotheses whose posterior probability was greater than .02 would have a hypothesis set probability of 0.906. Therefore, the hypothesis sets of both experts and non-experts only contained the correct hypothesis about half the time.

Both groups increased the plausibility of their hypothesis sets when they used the aid. The non-expert's aided hypothesis sets had a mean probability of .57, while the experts mean probability was .603. The difference between groups was not reliable, but both groups were aided significantly by the aid. The experts showed an improvement .133, while the non-experts showed an improvement of .185 over their unaided performance. The aid, therefore, provides a noticeable, but not dramatic, gain in performance.

Perhaps the most interesting result comes from an examination of those hypotheses generated by the subjects, and not suggested by the aid. The posterior probabilities of these hypotheses totaled less than .01. In other words, the aid generated nearly all of the hypotheses that subjects were capable of generating, and had it been used as the sole source of hypotheses it would have been better than an unaided subject, and equal to an aided subject using the aid. The concept of using an artificial memory to aid hypothesis generation was shown to be viable for those situations where it seems worthwhile to construct such an aid.

As the artificially memory used was one that modeled unaided human performance, this result is similar to the "bootstrapping" result reported by Dawes & Corrigan (1974) in that a model which captures the behavior of a decision maker can sometimes exceed the performance of the unaided individual. If the artificial memory were to be optimized for aiding purposes, the aid would probably perform noticeably better.

Other possibilities for improving hypothesis generation performance.

Some of the results obtained incidentally during our study of the hypothesis generation process might also be usefully employed to improve hypothesis generation. These results will only be mentioned briefly here because they have already been discussed previously in chapters 4 and 6.

Group hypothesis generation. Our study of group hypothesis generation strongly suggests that using a group of several hypothesis generators will yield a considerable gain in performance. These results also demonstrated that that social interaction during hypothesis generation degrades performance; a better course would be to use a synthetic pooling of hypotheses such as that done in the group study (8) and the aiding study (4). Depending upon the importance of the problem, synthetic groups of varying sizes can be used, and the pooled hypothesis sets of large groups result in a dramatic improvement in performance (8).

Debiasing to encourage alternate schemata. If the hypothesis generator is encouraged to try to think of another schema which might explain the data, the hypothesis sets are less biased by pre-existing hypotheses (6). This procedure should be routinely employed as it cost almost nothing to use.

Debiasing plausibility estimates. Steps which can be taken to reduce the bias in plausibility estimates are to help the hypothesis generator populate the set of unspecified hypotheses (2). Not only does this reduce the bias in these estimates, but it might be expected to encourage the hypotheses generator to continue to search memory beyond the

point were such searches normally stop.

Selecting good hypothesis generators. Finally, it seems possible to select good hypothesis generators by means of tests which measure divergent thinking, and our study on this topic (5) suggests that such paper-and-pencil tests are effective predictors of hypothesis generation performance.

Improving problem analysis and definition in act generation.

One of the characteristics of ill-defined problems (Taylor, 1974) is that the decision maker often has to analyze the problem before starting work on it. We believe that one of the first steps that a decision maker takes in analyzing a problem is to define it. Problem definition involves the identification of goals, problem constraints, and "operators" or "control variables" which can transform the present state into the goal state (Newell & Simon, 1972). The impoverished act generation performance of the typical subject may be due to incomplete problem analysis and definition. For example, subjects in the Parking problem almost always thought of building more parking spaces, but only 37% thought of a way to use the existing parking space more effectively, and only 23% thought of carpooling. Data such as these suggests that the typical subject defines the problem too narrowly, thus limiting the variety of actions that can be generated. Why don't subjects think of the obvious?

Our first study in this project (14) examined why subjects sometimes did not generate what seemed to us to be obvious solutions to the problem. We examined two possible explanations. The first explanation is that our subjects were simply ignorant of parking solutions; what were obvious solutions to us did not exist in their memories in any form.

The second explanation was that the information necessary to generate these obvious actions is available in memory, but inaccessible to the subjects because of the way they defined the problem. Perhaps the subjects are incapable of generalizing from the problem that they are working on to other similar problems because they did not recognize the similarity of the Parking problem to these other common problems. For example, one very effective action that only 10% of the subjects generated is to paint the lines of the parking spaces closer together to exploit the change to smaller cars in the University community. Subjects probably did not have this solution in memory directly as a solution to the Parking problem. However, they all must use other versions of this strategy on a frequent basis to make room for a book on a crowded bookshelf, or to fit another passenger in a car. Why were they unable to make this inductive leap?

In the first study in this series we investigated the effect of supplying the subjects with either generic strategies for solving the problem, or specific instances of these generic strategies. We were interested to see if subjects possess the information to implement the generic strategies if these strategies are suggested to them. We were also interested to see if subjects could generalize the specific instances of the generic strategies to create other, related solutions which are based on the same generic idea.

The subjects generated all the actions that they could think of in the Parking problem. Immediately after they said that they could think of

nothing else, they were given either generic or specific instances of the same generic cues. Table 6.1 shows the results.

TABLE 6.1
ACT GENERATION PERFORMANCE IN THE CUEING EXPERIMENT

	BEFORE CUEING	GAIN AFTER CUEING
LIMBS GENERATED (6 POSSIBLE)		
GENERIC CUES	4.0	1.7
SPECIFIC CUES	3.1 NS	0.5 P<.01
BRANCHES GENERATED (40 POSSIBLE)		
GENERIC CUES	8.2	3.4
SPECIFIC CUES	7.2 NS	1.3 P<.01
PERFORMANCE SCORE		
GENERIC CUES	439.7	172.4
SPECIFIC CUES	397.0 NS	65.2 P<.001

As can be seen in table 6.1, the two groups were approximately equal in performance prior to cueing. The new solutions generated after cueing are shown in the right-hand column. The generic cue group generated significantly more limbs, branches, and scored higher in terms of the performance score discussed in Chapter 5. The generic subjects clearly are capable of implementing a generic strategy if it is suggested to them, so the information necessary to generate more solutions is available in some form if they are given the kernel of the idea. When we said, in effect, "Can you think of a way to get more parking without building more parking spaces?", the subjects could think of a solution such as painting the lines in a parking lot closer together.

Notice the striking inability of the subjects given specific instances of these same generic cues to generalize these cues and discover the kernel of the cue and exploit it. When we said, in effect, "Can you think of a similar solution to painting the lines closer together in the parking lots?", most subjects could not think of redesigning the lot to increase the number of cars that can be parked, or segregating cars in the lots according to size. Most of the actions generated by the specific cued subjects involved minor embellishments of the specific cues, which did not increase their scores.

These results were obtained just a few minutes after the subjects claimed that they could think of nothing else. They demonstrate that at least some of the earlier failures to generate the "obvious" are due to the subject's inability to access information when it is stored as solutions to other problems. If we "hit them over the head" with a generic cue, they were usually able to generate an instance of it, but they usually were unable to extract the kernel of a specific cue and use it to generate other related ideas.

The second study in this series explored the "incubation effect" (Gick

& Holyoak, 1979). Subjects were asked to come back a week later for "another experiment" when they had completed the Parking problem. When they returned, they resumed work on the Parking problem. A weeks "incubation" did improve performance, but most of the gain was due to elaboration of ideas generated in the first session. Most of the actions generated were new branches on the limbs discovered in the first session; the average subject generated 0.9 new limbs in the second session. Apparently the passage of a week does not cause a noticeable redefinition of the problem.

The pattern of results obtained is consistent with the notion that subjects do not perform an exhaustive analysis of the problem leading to a complete problem definition. To the extent that the problem is incompletely defined, subjects may lack the "generic" cues that will aid them in generating a wide variety of actions.

The third study in this series explored the effects of training in problem analysis and definition on act generation performance. Subjects in an organizational training group were given brief training in strategies to analyze shortage problems (the Parking problem is a shortage problem), and exercised these strategies by working on a shortage problem involving starvation in India. Then they were invited to use these same strategies to analyze the actions that they had previously generated in a previous session. Two other groups were used. One group spent a comparable amount of time memorizing actions generated in a previous session, and the third control group was simply asked to return for another session, and on their arrival immediately started working on the Parking problem.

Both the organizational training group and the "memorization" group performed significantly better than the control group, as shown in table 6.2. The training group generated approximately twice as many new limbs as did the control, but the difference between the training group and the memorization group was not statistically reliable.

TABLE 6.2
MAJOR CATEGORIES GENERATED IN EXPERIMENT 2
(SIX CATEGORIES ARE POSSIBLE)

	DAY 1	DAY 2 GAIN	
ORGANIZATIONAL TRAINING	4.0	1.5 ---|	
		|-NS--|	
MEMORIZATION CONTROL	4.0	1.0----| |	P<.05
		|	
CONTROL	3.7	0.7----------|	

However, when we examined the frequencies of generating various categories of actions, we found that the organization group was significantly better than the memorization group as shown in table 6.3. As can be seen from inspecting this table, the control subjects spent most of their time "dreaming up" more variations of ideas they had generated in session 1, concentrating on alternate transportation, and diffe ent places to plant parking lots. However, subjects trained in problem analysis and organization tended to generate more actions to use the present parking space more effectively, reducing the number of people who want to park, and

various indirect strategies for solving the problem. Therefore, our training was effective in getting the subjects to analyze the problem more completely.

TABLE 6.3
SESSION 2 FREQUENCIES OF GENERATING ACTIONS IN SELECTED CATEGORIES

MAJOR CATEGORY	ORGANIZATION	MEMORIZATION	CONTROL
ALTERNATE FORMS OF TRANSPORTATION	37	29	72
BUILD MORE PARKING	34	52	58
USE EXISTING PARKING MORE EFFECTIVELY	15	16	10
REDUCE NUMBER OF PEOPLE WHO PARK	43	26	20
INDIRECT STRATEGIES	17	10	3

(ORGANIZATIONAL GROUP SIG. DIFF. FROM MEMORIZATION GROUP P<.001)

When considering the three studies in this project as a whole, we get the impression that our subjects have the ingredients to bake a cake, but no recipe. Training in problem analysis and organization improves the accessibility of possible solutions to the problem. Subjects do possess the information to create instances of generic strategies, but apparently do not discover all of the generic strategies in the absence of training. Their ability to exploit past experience seems to be limited by their difficulty in extracting the generic kernels from other, related ideas.

Other ways of increasing act generation performance.

Synthetic groups. As previously discussed in Chapter 6, pooling the responses of several act generators is a very effective way of getting a more complete set of possible actions for decision making. It is of the utmost importance, however, that these groups not be allowed to interact socially during the actual act generation, as socially interacting groups are little or no better than a single individual. Information exchange such as we used, or that used in the Delphi procedure, may be of benefit, but should not be expected to to produce enormous gains in performance.

Selecting act generators on the basis of divergent thinking ability. Also, as discussed in Chapter 6, selecting act generators on the basis of divergent thinking ability is very effective. If individuals can be found who excel at this ability, their performance will be several times better than the typical, unselected individual.

Overall recommendation.

Our research on act generation suggests that the typical act generator is impoverished, generating only a small subset of the acts worth considering. If the importance of the problem warrants it, we suggest simultaneously using all three techniques that we have discovered lead to better act generation. First, we recommend that individuals be selected for good divergent thinking ability, the higher the better. Second, these individuals should be trained in problem analysis and definition. Third,

they should work in small, non-interacting groups, and their ideas pooled. Although we have not studied these three techniques in use simultaneously, we see no reason why they should not be effective in concert. The net result, we believe, will be marked improvement in performance.

CHAPTER 8. SUMMARY OF WHAT HAS BEEN LEARNED

Presented below is a summary of major accomplishments and conclusions developed in our hypothesis generation contract and our act and outcome generation contract as an aid to those who wish to get an overview of our conclusions. As with any compact summary, these conclusions are not completely qualified, and the reader is referred to the earlier parts of this monograph, or even better, to the original technical reports for more complete discussion and qualification. Pertinent technical reports are indicated by a number, for example, (18). Following the summary conclusions, a section is presented which discusses the major conclusion from both projects.

Hypothesis generation:

1. One goal of hypothesis generation is to provide a pool of hypotheses that are potential explanations for a set of data. A model for retrieving hypotheses from memory was constructed that fits the data well. It appears that a hypothesis need not be associated in memory with all data for retrieval to take place, nor are hypotheses that are only associated with a single datum typically retrieved. It was estimated that hypothesis are retrieved from memory if tagged by two or three data (1).

2. If hypotheses are retrieved from memory using part of the data, a "consistency check" is performed where the newly retrieved hypothesis is checked for consistency with any data not used in its retrieval (3).

3. Consistency checking appears to be a rapid, logical checking process involving high-speed semantic verification and terminates if a hypothesis is found to be logically inconsistent with the data (3).

4. Consistency checking results in a number of hypotheses (approximately two, in the task used) being rejected as logically inconsistent before the first consistent hypothesis is found and the hypothesis generator is aware of some rejected hypotheses (3).

5. Hypothesis retrieval from memory seems to be an activation process where many hypotheses are activated for further consideration and evaluation. Consistency checking is the first part of the evaluation process which narrows the set of hypotheses the hypothesis generator considers (3).

6. Hypotheses that are retrieved and checked for consistency form a pool of hypotheses that will be processed further for plausibility if the task warrants it. **This pool is quite incomplete, and only contains the correct hypothesis about half the time** (1, 2, 3, 4, 5, 7, 9).

7. The process of plausibility assessment is the process where the decision maker decides if the hypothesis in question is sufficiently plausible to be considered as a candidate explanation for the available data. It involves a judgment of the relative likelihood of the hypothesis in question with respect to other hypotheses that have been generated, and also with respect to hypotheses that have not been generated (1, 2).

8. Subjects have only a very rough idea of which hypotheses in their hypothesis sets are the most plausible candidates. Their orderings the hypotheses in terms of likelihood, or estimates of the likelihood, are only weakly related to the veridical orderings or values (1).

9. Subject's plausibility estimates are excessively certain, sometimes by as much as a factor of three. This excessive certainty is probably due to the unavailability of hypotheses that were not generated. This is true for experimenter-supplied hypotheses, or subject-generated hypotheses (1, 7).

10. The hypothesis generator typically generated supplemental hypotheses if new data arrives that makes the existing hypothesis set less plausible. An increasingly strict, sliding criterion for admission of new hypotheses to the set seems to be used. New hypotheses are admitted mainly if they are "leading contenders", that is, if they are close competitors with the best hypotheses that are already in the set (2, 7).

11. Group hypothesis generation is markedly superior to that of individuals only if the subjects are in a nominal or synthetic group. Socially interacting subjects are somewhat better than individuals, but quite inferior to nominal groups. These conclusions are based on a model that partitions group performance into social and informational components (8).

12. Interpretations are made of the data used in hypothesis generation, and these schematic interpretations are sometimes used as retrieval cues rather than the data themselves (6).

13. An artificial memory can be developed to aid the hypothesis generation process. The aid investigated used a artificial memory based on our model of hypothesis retrieval process. While aids based on other retrieval schemes might perform better, the aid investigated, if used as the sole source of hypotheses, was as good as the aided subject (4).

14. There were a number of replications of the finding that both expert and non-expert hypothesis generator seem to suffer from the same cognitive deficiencies. That is, both groups exhibited impoverished hypothesis sets, and believed that these sets were much more complete than they actually were (See 9 for a summary.).

15. It is possible to predict hypotnesis generation ability to some extent. Hypothesis generation ability seems to be related to divergent thinking ability, and does not appear to bear much relation to inductive reasoning, achievement, general mental ability, or episodic memory. There seems to be two additive components involved in divergent thinking: the analysis of the problem into its implicit dimensions, and the retrieval from memory using these implicit dimensions. Subjects who score above the median on both of these two components do 67% better than subjects who score below the median (5).

Act and outcome generation:

16. Act generation is the process of creating possible actions that may solve a decision problem. Outcome generation is the process of specifying possible outcomes of actions. Act generation of our typical subjects can also be described as impoverished; subjects typically generate two or three ideas worth implementing in situations where there were typically 30 to 40 possible ideas that could be considered (10).

17. A technique for calculating a performance score was developed that combines both quality and quantity of generated actions. This score is a lower-bound estimate of optimal performance, and is useful for measuring both the breadth and depth of a subject's performance (10).

18. The generality of conclusions reached using the performance score was examined by using either experimenter-generated hierarchical representations of the decision problem, or representations based on cluster analysis. In addition, the source of the utility estimates used in the score was examined. The conclusions were similar, irrespective of how the score was calculated (10).

19. It appears that the limited act generation performance of subjects is not due to lack of motivation. Substantial incentives for good performance in terms of quality, or quantity did not result in appreciable gains in performance (12).

20. Subject's post-experimental estimates of the number of good actions that existed that they could not think of are about 2.5 to 5.0. This result, replicated in several studies, suggests that they believe that they have thought of nearly everything worth considering, when this is far from the case (12).

21. The subjective representation of the problem space in act generation problems was studied using multidimensional scaling and cluster analysis. Subjects apparently did not see the problem studied as a member of a generic class of shortage problems, but rather in fairly concrete terms, being concerned with quite specific strategies for solving these problems (11).

22. Interacting small groups have the possibility of exchanging information in a synergistic fashion and thereby improving their performance. The additive model developed previously for small group research in hypothesis generation was extended to allow estimates of the size of the information interchange component. Information interchange was found to result in a 6.9% improvement in performance. Most of the "synergism" that occurred was minor variations of the other person's ideas; the average utility of acts was not increased by information exchange. Synthetic groups were again found to be quite superior to interacting groups (13).

23. The decision maker's "frame" or "perspective" was not found to have much of an effect in laboratory simulations. This negative result should be interpreted with caution, since the simulation may have lacked conditions necessary for these effects to occur (14).

24. The act of explanation, or creating a causal scenario, causes

substantial and persistent changes in the way that the decision maker views a decision problem. The number of success and failure outcomes generated in an outcome generation task do not depend on the nature of the initial explanation made, as reported previously. However, the causal factors used in scenarios, the importance weightings of causal factors, and subject's predictions about how these factors would turn out in the future all varied with the type of initial scenario constructed (15).

25. Apparently, the changes in the way a decision maker views a problem after constructing an initial scenario is due at least in part to initial inferences that the subjects make from the case history. These changes do not seem to depend on selective encoding of the case history. The inferences which were used to support the initial explanation apparently persist and continue to be used in constructing alternate causal scenarios leading to different outcomes (15).

26. Divergent thinking ability apparently plays an important role in act generation as well as in hypothesis generation. In a study of expertise in act generation, it was found that graduate student "experts" displayed excellent act generation performance, quite unlike that of an unselected subject. A second experiment was performed where the expertise of the graduate students should have been irrelevant, and the graduate students continued to show the same fine performance. As the graduate students scored about twice as high as undergraduates on measurements of divergent thinking, these results demonstrate the importance of divergent thinking ability in act generation, and suggest that the result in the initial study was probably due superior divergent thinking ability. Clearly, the earlier conclusion regarding impoverished act generation ability should be qualified with regard to individuals who are excellent divergent thinkers (17).

27. Subjects can usually implement a generic strategy if they are cued with the kernel idea. However, if they are cued with a specific instance of that generic strategy, they are only rarely able to discover the generic strategy and generate other, related acts based on that strategy (16).

28. The "incubation effect" is observed if a problem is set aside for a period of time, and then work is continued. Such a passage of time apparently does not cause subjects to reanalyze the problem again. Most acts generated after a one-week rest were minor variations of the same generic ideas. Subjects rarely rethink the problem when resuming work on it (16).

29. Training in problem analysis and definition helps subjects generate a wider variety of actions, particularly those involving indirect strategies for solving the problem. This effect apparently occurs because they use a broader and more general problem definition (16).

30. A number of suggestions were made for improving predecision generation performance of hypotheses, acts and outcomes. In our opinion, the most effective way of improving performance in this area would be to do all of the following **simultaneously**: 1) Select generators who have excellent divergent thinking ability, 2) Train them in problem analysis and definition, and 3) have them generate acts or hypotheses in small, non-interacting groups (18).

The "Fat and Happy" Hypothesis and Act Generator.

If we single out the most important conclusion of our research, the one with the broadest implications to decision theory, we arrive at the following conclusion:

Hypothesis generation. One major conclusion supported by this research is that sets of hypotheses generated by our subjects were impoverished, but subjects estimated that these sets were more complete than they actually were. Similar results have been obtained using a wide variety of tasks, several experimental strategies, and several response modes. Although some variables do effect estimates of the extent of hypothesis generation deficiencies, we have found no exceptions to the general conclusions that subjects generate impoverished hypothesis sets and overestimate their completeness.

During this project we have employed a variety of hypothesis generation tasks, partially to determine if our results were task-specific. We employed tasks where subjects generated hypotheses about the majors of undergraduates, occupations of skilled workmen, and identities of States of the Union (1, 2, 4, 9). Other tasks involved generating the identity of animals (3), and defects in an automobile (7). Two experiments used problems where the object was to generate hypotheses about an unknown geographical area (5, 6). In all of those experiments where a measure of hypothesis generation performance was obtained, subjects generated impoverished hypothesis sets. In all of those experiments where plausibility estimates were obtained, subjects were excessive in their assessments of the completeness of their hypothesis sets.

The same general conclusions that were reached using college students seem to be justified for expert subjects (4, 7). Although this variable was investigated in only two studies, the results suggest that experts and non-experts have similar difficulties.

In one study, it was shown that plausibility estimates were excessive irrespective of whether the subjects were judging hypothesis sets that they had generated or hypothesis sets supplied by the experimenter. In this same study, it was shown that the plausibility estimation measurement technique used in many of these studies produced much the same results as probability estimation.

Act generation performance. Similar conclusions were reached for act generation process. Although only two tasks were used because of the extensive effort necessary to develop performance measures, a replicable pattern of results was found in 5 studies (10, 12, 13, 16, 17). Subjects usually could generate two or three actions which were good candidates for possible adoption in situations where there were twenty or thirty actions that could be considered. This effect does not appear to be due to lack of incentive, an emphasis on quality or quantity, the source of the hierarchical structure, or the utility estimates used in the analysis.

Also, as in hypothesis generation, subjects believed that their set of actions were much more complete that it actually was. They believed that 2.5 to 5 good ideas still remained to be generated, when in fact there were about 20 to 30 actions that could have been considered.

The only exception we've encountered in regard to this general picture is the performance of the exceptionally good divergent thinker. These individuals are rarely found, being in the upper few percent of the general population in regard to this ability, but their performance approaches optimal performance (17).

These results, taken as a whole, present a rather unflattering picture of the hypothesis or act generator. Hypothesis or act generators may feel "fat and happy" about the completeness of their hypothesis or act sets, when the available data about their performance suggests that they should feel "thin and worried." Generated hypothesis sets lack important hypotheses and generated act sets lack important actions, yet when these sets are evaluated, the hypothesis or act generator feels that they are more complete that they really are.

Our data suggests that the explanation for the "fat and happy" syndrome lies in deficiencies in the memory search process. The subjects' inability to access all plausible hypotheses or all effective actions available in memory seems to be the underlying cause of both poor generation and the feeling that these sets are almost complete. The paradox is that these results suggest that hypothesis and act generators may be unaware of their deficiencies because the difficulty in retrieving hypotheses and acts from memory also affects the evaluative process where they assess the completeness of their performance.

CHAPTER 9. RECOMMENDATIONS FOR FURTHER RESEARCH IN PREDECISION

It probably would be counter-productive to make an encyclopedic list of recommendations for research for projects as large as those described in this monograph. Each study we performed raised more questions than it answered. Some of these questions are discussed in the individual technical reports. Only the most general recommendations are presented in this chapter.

The desirability of more research in predecision processes.

Given the advanced state of the art in decision theory, and these preliminary results in hypothesis, act, and outcome generation, we believe that the investigation of predecision processes should receive a priority equal to, or perhaps greater, than that of traditional decision theory. We believe this because it appears that more improvement in decision making can result from an improved understanding of predecision processes than from an equal expenditure on further refinement of the optimization techniques of traditional decision theory. Decision theory as a topic has received hundreds, or perhaps thousands of experimenter years of investigation. We have spent only six, yet our results suggest that the problem structuring of decision makers is so incomplete that it seems rather pointless to spend further effort developing optimization techniques for what may well be incomplete models until predecision problem structuring is better understood.

Based on our research findings, we have three primary recommendations, each of which is discussed below.

First, we suggest that other experimenters, preferably those who are skeptical about our general conclusions and findings, be invited to confirm or disconfirm our major results and conclusions with other tasks and in other contexts. While we have had the advantage of examining the raw data in many similar studies, we will be the first to acknowledge that it is possible that we may possess some of the biases that we accuse our subjects of having. Results obtained in a single laboratory are unlikely to make a major impact on the decision theory field until they are independently confirmed. While this process has started, as witnessed, for example, by the work of Pitz, et al. (1980) and Thompson (1983), it should be accelerated.

Second, our efforts have only scratched the surface of a topic which we believe will grow to be as large as the decision theory topic. Most of our research has been devoted to hypothesis and act generation, with relatively little attention being paid to outcome generation. Although we are quite proud of what we have accomplished in this area, it is at best a beginning. Much of our effort was devoted to developing new experimental paradigms and measurement techniques, and hopefully research that builds on our work will be more straightforward, and involve fewer false starts. We do not believe that our results are definitive on any of the topics that we investigated, and feel that this area of research is wide open and waiting for conquest.

Third, there are important predecision topics that have received relatively little or no attention. For example, our tentative theory of

problem detection was included in this monograph with some fear and trepidation because we have done no empirical research on this topic, and are unavare of any other research done in this area from a decision-making perspective. We included this theory for the sake of completeness as it clearly is one of the most important of the predecision topics (cf. Corbin, 1980). We could have included a chapter on problem analysis and definition, another unexplored area. However, our thinking on this topic is little advanced from that described in our single study on this topic (16). Research on both problem detection and on problem analysis and definition will be our next several topics for research.

10. REFERENCES

Corbin, R. Decisions that may not get made. In Wallsten, T. (Ed.) **Cognitive Processes in choice and decision behavior.** Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1980.

Dawes, R. & Corrigan, B. Linear models in decision making. **Psychological Bulletin**, 1974, **81**, 95-106.

de Groot, A. **Thought and Choice in Chess.** The Hague: Mouton, 1965.

Einhorn, H. J., & Hogarth, R. M. Uncertainty and causality in practical inference. **Center for Decision Research**, Graduate School of Business, University of Chicago, April, 1981.

Fischoff, B. Hindsight ≠ foresight: The effects of outcome knowledge on judgment under uncertainty. **Journal of Experimental Psychology: Human Perception and Performance**, 1975, **1**, 288-299.

Gentner, D. & Stevens, A. **Mental Models.** Hillsdale, N.J.: Lawrence Erlbaum Associates, 1983.

Gick, M. E. & Holyoak, K. Problem solving and creativity. In A. L. Glass, K. J. Holyoak, & J. L. Santa (Eds.) **Cognition.** Reading, MA: Addison-Wesley Publishing Co., 1979.

Kahneman, D., & Tversky, A. The simulation heuristic. **Stanford University, Department of Psychology**, (Tech. Report # 5), 1981.

Lindstone, H. & Turoff, M. (Eds.). **The Delphi method: techniques and applications.** Reading, Mass.: Addison-Wesley, 1975.

Newell, A. & Simon, H. **Human Problem Solving.** Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Osborn, A. F. **Applied imagination.** New York: Scribner's, 1957.

Pitz, G., Sachs, N., & Heerboth, J. Procedures for eliciting choices in the analysis of individual decisions. **Organizational Behavior and Human Performance**, 1980, **26**, 396-408.

Raiffa, H. **Decision Analysis.** Reading, Mass. Addison Wesley, 1968.

Roberts, F. **Measurement Theory.** Reading Mass.: Addison-Wesley, 1979.

Rohlf, F., Kishpaugh, J. & Kirk, D. Numerical Taxonomy System of Multivariate Statistical Programs, October, 1979.

Schank, R., & Ableson, R. **Scripts, plans, goals, and understanding.** Hillsdale, N.J.: Erlbaum, 1977.

Shepard, R. Representation of structure in similarity data: problems and prospects. **Psychometrica**, 1974, **39**, 373-421.

Slovic, P. & Fischoff, B. On the psychology of experimental surprises. **Journal of Experimental Psychology: Human Perception and Performance**, 1977, 35, 544-551.

Stein, M. I. **Stimulating creativity**, New York: Academic Press, 1975.

Taylor, R. Nature of problem ill-structuredness: Implications for problem formulation and solution. **Decision Sciences**, 1974, 5, 632-643.

Thompson, P. Some models of hypothesis generation. Unpublished Phd. Dissertation, University of North Carolina, Chapel Hill, N.C., 1983.

Tversky, A., & Kahneman, D. Causal schemata in judgments under uncertainty. In M. Fishbein (Ed.), **Progress in Social Psychology**, Hillsdale, N.J.: Erlbaum, 1980.

von Neuman, J. & Morgenstern, O. **Theory of games and economic behavior** (2nd ed.). Princeton, N.J.: Princeton University Press, 1947.

11. TECHNICAL REPORTS WITH ABSTRACTS

The hypothesis generation contract (N00014-77-C-0615):

1. Gettys, C., Fisher, S., and Mehle, T. Hypothesis generation and Plausibility assessment (Tech. Rep. TR 15-10-78). Norman, Ok.: University of Oklahoma, **Decision Processes Laboratory**, July 1979.

A hypothesis generation model is described which consists of two subprocesses. Hypotheses are retrieved from memory using several data as retrieval cues in the hypothesis retrieval sub-process. These hypotheses are then evaluated by a plausibility assessment sub-process. Two experiments are described. A memory retrieval experiment examined hypothesis retrieval from memory using multiple data. A memory-tagging model is described which predicts the probability of multi-data hypothesis retrieval. Performance in this task was poor; subjects rarely generated an adequate hypothesis set. A second plausibility assessment experiment was performed where subjects estimated the plausibility of specified hypotheses using varying amounts of data. Plausibility assessments for specified hypotheses were usually extreme in comparison to the posterior odds calculated by Bayes' theorem. This result was also attributed to deficiencies in hypothesis retrieval from memory.

2. Mehle, T., Gettys, C., Manning, C., Baca, S., and Fisher, S. The availability explanation of excessive plausibility assessments (Tech. Rep. TR 30-7-79). Norman, Ok.: University of Oklahoma, **Decision Processes Laboratory**, July 1979.

The assessment of hypotheses in hypothesis generation involves a comparison between those hypotheses that have been generated (specified) and those that are not generated (unspecified). This study investigated the "availability explanation" (Tversky and Kahneman, 1973) for subjects' overconfidence in estimating the probability of specified hypotheses. The conjecture is that subjects have difficulty retrieving unspecified hypotheses; a complete set of candidate unspecified hypotheses is unavailable during assessment. Therefore, the underpopulated set of unspecified hypotheses is regarded as less probable and the specified set is regarded as more probable. A control group in this study replicated previous findings of overconfidence for specified hypotheses. Two manipulations to increase the availability of unspecified hypotheses were investigated. One manipulation involved explicitly requesting subjects to populate the unspecified set. The other manipulation consisted of computer presentation of candidate unspecified hypotheses. Although in a normative sense, neither manipulation should have affected judgments, results indicated that assessment overconfidence for both experimental groups was reduced. These results support our conjecture that the availability heuristic is at least partially responsible for subjects' excessive behavior in evaluating specified hypotheses.

3. Fisher, S., Gettys, C., Manning, C., Mehle, T., and Baca, S. Consistency checking in hypothesis generation (Tech. Rep. 29-7-79). Norman, Ok.: University of Oklahoma, **Decision Processes Laboratory**, July 1979.

Three experiments were performed to provide evidence that the generation of hypotheses in response to multiple data may involve two different cognitive processes. First, a candidate hypothesis may be retrieved or activated in memory in response to only part of the available data. This candidate hypothesis may then be checked for consistency against the remaining data. This latter process is called "consistency checking." Experiment 1 was performed to provide evidence that consistency checking occurs during hypothesis generation. Subjects were able to recognize hypotheses which were retrieved during a hypothesis generation problem but not emitted as hypothesis responses, suggesting that consistency checking was responsible for the rejected hypotheses. Experiment 2 indicated that the amount of time needed to process an additional datum in a consistency checking task was less than an estimate of the time needed to process an additional datum in hypothesis retrieval. The results suggest that consistency checking is a high-speed verification process rather than a slower search process. Experiment 3 was performed to provide evidence that consistency checking is a self-terminating process. Subjects' latencies depended upon the position of a disconfirming datum within a data set, supporting this conjecture. The results generally confirmed the existence of a high-speed verification process in hypothesis generation and also suggest that the generation of hypotheses in response to multiple data occurs as a result of dual processes.

4. Gettys, C., Mehle, T., Baca, S., Fisher, S., and Manning, C. A memory retrieval aid for hypothesis generation (Tech. Rep. TR 27-7-79). Norman, Ok.:University of Oklahoma, **Decision Processes Laboratory**, July 1979.

Hypothesis generation consists of retrieving explanations for data from memory, and assessing these explanations for plausibility. Previous research has established that human hypothesis generation performance is deficient in both hypothesis retrieval and assessment. This study investigates an aid for the hypothesis retrieval process which is based on a model for hypothesis retrieval developed by Gettys, Fisher, and Mehle (1978). A computer simulates the human hypothesis retrieval process by searching an enriched associative memory which contains the associations of a number of individuals in the form of lists of hypotheses for each datum. When the data of a decision problem become known, the appropriate lists are searched by the computer. Hypotheses that are common to most or all of the lists are suggested to the user, who assesses them for plausibility. An experiment was performed to determine the utility of the aid for both expert and non-expert users. The aid produced a substantial gain in performance for both groups of users, suggesting that further development of the aid would be worthwhile in decision situations which are repeated often enough to warrant the creation of an enhanced artificial memory. Also discussed are several techniques for implementing the aid, and determining the maximum gain in performance that the aid can produce.

5. Manning, C., Gettys, C., Nicewander, A., Fisher, S., and Mehle, T. Predicting individual differences in hypothesis generation (Tech. Rep. TR 28-7-79). Norman, Ok.: University of Oklahoma, **Decision Processes Laboratory**, July 1979.

Two experiments were performed to determine the extent to which individual differences in hypothesis generation could be predicted. In the first experiment, several published tests of creativity were used as predictors of hypothesis generation ability. The Alternate Uses test was the best predictor of hypothesis generation performance. In a second experiment, measures of achievement, general mental ability, and information were included with Alternate Uses as predictors of performance. Again Alternate Uses was the best predictor of performance. Several variants of the Alternate Uses test were also employed to isolate the components of hypothesis generation. It was found that two components were involved: retrieval of implicit dimensions of the objects and retrieval of uses when the dimensions are explicitly provided. The latter component was found to be by far the most important. It was concluded that good hypothesis generators have skills that enable them to effectively retrieve information stored in memory.

6. Manning, C., and Gettys, C. The effect of a previously-generated hypothesis on hypothesis generation performance (Tech. Rep. TR 8-5-80). Norman, Ok.: University of Oklahoma, **Decision Processes Laboratory**, August 1980.

An experiment was performed to determine what effects exposure to a previously generated hypothesis would have on subsequent hypothesis generation. The results showed that hypothesis generation performance is relatively unchanged if the previously-generated hypothesis is consistent with a salient interpretation of the data. However, if the previously-generated hypothesis is consistent with a relatively unusual interpretation of the data, then subjects use both the interpretation that is consistent with the hypothesis and the more commonly used interpretation as cues to retrieve hypotheses. In this case, resulting hypothesis sets included more varied types of hypotheses. Instructions to consider other interpretations of the data also resulted in subjects' generating richer hypothesis sets.

7. Mehle, T. Hypothesis generation in an automobile malfunction inference task (Tech. Rep. TR 25-2-80). Norman, Ok.: University of Oklahoma, **Decision Processes Laboratory**, February 1980.

Expert and novice subjects generated hypotheses in an automobile troubleshooting inference task. Data collected included subjects' verbal protocols during the inference tasks and subjects' estimates of the probabilities of their generated sets of hypotheses. Analyses indicated that both expert and novice subjects had difficulty generating complete sets of hypotheses and were overconfident in their subjective estimates of the probabilities of generated hypotheses.

8. Casey, J., Mehle, T., and Gettys, C. A partition of group performance into informational and social components in a hypothesis generation task (Tech. Rep. TR 3-3-80) Norman, Ok.: University of Oklahoma, **Decision Processes Laboratory**, August 1980.

A technique is presented for partitioning group performance into two components: a component due to the increased information possessed by the group and a component representing the change in performance due to social interaction. The hypothesis-generation performance of individuals working alone was compared to the performance of interacting groups of four. The particular task employed permitted calculations of the veridical probabilities of generated sets of hypotheses. Analyses of results were based on a new method, obtained by pooling hypothesis sets from individual subjects to obtain "synthetic" groups. This method permits direct comparisons of interacting and synthetic groups' hypothesis-generation performance. Using this method, we found that groups of four subjects were equivalent to synthetic groups of 1.8 subjects.

9. Gettys, C., Manning, C., Mehle, T., and Fisher, S. Hypothesis generation: A final report of three years of research (TR 15-10-80). Norman, Ok.: University of Oklahoma, **Decision Processes Laboratory**, October 1980.

This final report summarizes 14 experiments conducted over a three-year period. First discussed is a hypothesis generation model and research which addresses the model. Several major findings were obtained: 1) Hypothesis retrieval from memory is impoverished. Hypothesis generators are not able to retrieve all relevant hypotheses from memory that should be considered in a decision problem. 2) Hypotheses that are retrieved from memory are first checked for logical consistency with the data. Those hypotheses that are logically consistent may be assessed further for plausibility. 3) Hypothesis generators think that collections of hypotheses which they generated are much more complete than they actually are.

The next section discusses research on hypothesis generation performance. Topics include protocol analysis, group hypothesis generation, the biasing effects of schemata, individual differences in hypothesis generation, and generalizing to expert populations.

A third section is devoted to a survey of research relevant to aiding the hypothesis generation process. An artificial aid for retrieving hypotheses from memory is discussed. Also discussed are other ways of improving hypothesis generation performance.

The general conclusion of this project is that both the failure to retrieve enough hypotheses from memory and the subjects' belief that these collections of hypotheses are more complete than they actually are can be traced to deficiencies in the memory retrieval process.

The act and outcome generation contract (N00014-80-C-0639):

10. Gettys, C., Manning, C. & Casey, J. An evaluation of human act generation performance (TR 15-8-81). Norman, Ok.: University of Oklahoma, **Decision Processes Laboratory**, August 1981.

A series of experiments addressed the adequacy of act generation performance, an important precursor to problem structuring. Each of two decision problems was studied by a series of three experiments. In the first experiment, subjects were given a realistic decision problem and were asked to respond with any act occurring to them. In the second experiment, the acts suggested were evaluated by different subjects for feasibility. In a third experiment, additional subjects estimated the utility of the acts judged feasible. The act generation performance of subjects was evaluated using two techniques. First, a decision tree was generated by the experimenters by combining the acts suggested by all subjects. The decision tree generated by each subject was compared with the experimenter-generated tree. It was found that subjects failed to generate important limbs and branches of the group decision tree. Second, the quality of the trees generated by individual subjects was evaluated by an opportunity loss calculation. This calculation provided an estimate of the potential cost of failing to generate limbs and branches of the decision tree. The opportunity loss analysis suggested that the failure to generate a complete tree could be costly.

11. Manning, C. Describing the representation of decision problems: An application of multidimensional scaling and cluster analysis (TR 15-12-81). Norman, Ok.: University of Oklahoma, **Decision Processes Laboratory**, December, 1981.

The purpose of this study was to describe the important representations for an example of a common class of decision problems, facing a shortage of a commodity. Describing potential problem representations is important because decision problems are typically ill-structured (Taylor, 1974), and a decision maker's representation of a problem is not obvious to the experimenter. Describing the dimensions along which a group of subjects judged the similarity of potential solutions to a problem should give insight into various ways in which the problem may be represented. This will provide a basis for additional research on the processes involved in the generation of act solutions and their associated outcomes.

Multidimensional scaling and cluster analysis were used to analyze the similarity of 43 acts suggested to solve the parking problem at Oklahoma University. In Experiment 1, sixty subjects rated the similarity of a set of randomly chosen act pairs. The similarity judgments were averaged across subjects and submitted to the ALSCAL procedure of SAS. A three dimensional solution was identified as most appropriate. In Experiment 2, fifty subjects rated randomly chosen subsets of the same acts on twelve bipolar scales which represented potential ways of representing a problem. Three scales suggested generic strategies for solving the problem. Four scales suggested problem-solving strategies specific to the parking problem. One scale suggested a personal goal which might be fulfilled by employing an

action. Four scales were potential measures of the acts' utility. The scale ratings obtained in Experiment 2 were averaged across subjects, then regressed on the three dimensional solution derived from multidimensional scaling to objectively describe the dimensions. The three dimensions were found to most closely resemble specific strategies for solving the parking problem. Dimension 1 was identified as "involves alternate forms of transportation". Dimension 2 was identified as "involves rescheduling activities" and "changes current priorities". Dimension 3 was identified as "requires building new facilities".

Hierarchical cluster analysis was used to analyze the similarity judgments to examine neighborhoods of acts in the three dimensional space to determine whether an alternative interpretation of the relationships between acts might be obtained. Seven clusters were identified. Four clusters were specific instances of a more general category "increase the amount of space available". Another cluster was the category "involves alternate forms of transportation". Two other clusters involved rescheduling activities and enforcing current parking regulations more strictly.

The three dimensions derived from multidimensional scaling and the set of clusters obtained from cluster analysis seem to describe alternative strategies for solving the parking problem from which individual decision makers might sample when representing the problem. Although in real-world decision problems, the problem space is unstructured, these results suggest that a limited number of constructs may sufficiently describe the important problem representations decision makers employ to interpret a problem.

12. Pliske, R., Gettys, F., Manning, C. & Casey, J. Act generation performance: the effects of incentive (TR 15-8-82). Norman, Ok.: University of Oklahoma, **Decision Processes Laboratory**, August 1982.

Two experiments explored the generalizability of earlier research which indicated that human act generation performance was impoverished. Subjects were given a realistic decision problem and were asked to generate actions which could be taken to solve the problem. Subjects in two incentive conditions were offered monetary rewards for generating additional actions. Subjects in one condition were rewarded for the sheer quantity of actions produced and subjects in the other condition were rewarded for the quality of the actions produced. In a second experiment, both expert and naive subjects judged the quality of the actions produced by subjects in the first experiment. The results replicate earlier research in that most subjects generated relatively few actions and they also failed to generate important actions as rated by both expert and naive judges. There were no significant differences between the performance of subjects in the incentive conditions and subjects in the control condition. Thus, even when subjects are given substantial monetary incentives to generate additional actions, their act generation performance is impoverished. Differences in the act generation performance of the "quantity" and "quality" incentive conditions are discussed.

13. Casey, J., Gettys, C., Pliske, R. & Mehle, T. A partition of small group performance into informational and social components (TR 30-8-82). Norman, Ok.: University of Oklahoma, **Decision Processes Laboratory**, August 1982.

New theoretical and methodological techniques for partitioning and identifying the sources of performance differences between groups and individuals in hypothesis and act generation tasks are presented in two experiments. Experiment 1 presents a two-component model which separates group performance into informational and social components. The model proposes that the pooling of information in an interacting group (the information component) is mediated by the social factors (e.g., level of arousal, cohesiveness, etc.) which are present in a given situation (the social component). Interacting groups were found to be inferior to nominal groups in an hypothesis generation task. Thus, in Experiment 1, the social component was found to have a negative effect on the information component. Experiment 2 further partitions the social component into a social information component which accounts for the additional information which becomes available as a result of group interaction and a social, non-informational component which consists of purely social factors. The social information component estimates the synergistic effect of group interaction on information retrieval. The social informational component was estimated by including a group of subjects who exchanged ideas (information) via computers but had no social interaction. The "information exchange" group was found to be somewhat superior to a nominal group in an act generation task, and both of these groups were superior to an interacting group. Experiment 2 illustrates that even when the social, non-informational component has a negative effect on the informational component, the social information component may have a positive effect.

14. Manning, C. The Role of a Decision Maker's Perspective in the Generation and Assessment of Actions in a Conflict Situation (TR 15-9-82). Norman, Ok.: University of Oklahoma, **Decision Processes Laboratory**, September 1982.

Two experiments were performed to assess the influence of perspective and information on the generation of actions an opponent might take to resolve a conflict. Both experiments employed a problem in which guerrilla forces captured the French embassy in a hypothetical South American country and took the personnel hostage. In the first experiment, subjects were assigned the perspective of a guerrilla, a hostage, or an advisor to the President of France. Subjects generated five actions the French government was most likely to take to resolve the conflict, ranked the actions, then provided likelihood estimates and estimates of the French government's preferences for a specified set of actions. No large differences in performance resulted from manipulating perspective. However, some subtle differences were observed. Hostage subjects generated acts more likely to benefit both the guerrillas and the French than subjects in other conditions. All Guerrilla subjects generated at least one military action, while some subjects in the other perspective conditions failed to generate any.

Experiment 2 was performed to assess the effect of providing both a perspective and information about an opponent's objectives on the generation of actions the opponent might take to resolve a conflict.

Subjects in one Guerrilla condition read irrelevant information about the geography of France, subjects in another Guerrilla condition were asked to imagine the French government's objectives, and subjects in a third Guerrilla condition were provided with an explicit description of the French government's objectives. Another set of subjects assigned the French perspective was used as a control condition. Again, no major differences were found in act generation, but some subtle differences were observed. The Guerrilla subjects who read explicit information about the French government's objectives generated acts that were more beneficial to the French than subjects in the other Guerrilla conditions. Guerrilla subjects reading irrelevant information about France generated acts that tended to benefit both parties more than the acts generated by French subjects. In neither experiment did subjects differ in their estimates of the likelihood with which the French government might take a specified set of actions or in their estimates of the French government's preferences for a specified set of actions.

These results may suggest that perspective has only a limited influence on the generation and assessment of actions an opponent might take to resolve a conflict. Without further research, it is difficult to determine whether perspective impairs a decision maker's performance in a conflict situation or whether its influence is only salient in hindsight.

15. Pliske, R. & Gettys, C. The role of causal explanation in outcome generation (TR 8-2-83). Norman, Ok.: University of Oklahoma, **Decision Processes Laboratory**, August 1983.

It is assumed that Decision makers generate possible outcomes for action by creating a mental model, ie. a causal schemata which represent the decision maker's model of the way the world works. Some causative factors are seen as relevant, and others are seen as irrelevant. Those relevant causal factors that are included in the mental model form a casual field, and the causal field determines to a large extent the outcomes that are generated. Therefore, when the decision maker first attempts to generate outcomes for an act, a causal field is created, and this causal field may persist throughout the outcome generation task. The persistence of the causal field in the decision maker's thinking may make it difficult to create other, alternate mental models which might enable the decision maker to anticipate other outcomes for that act.

The present investigation examines the persistence of initial causal fields, and the cognitive mechanisms that may be responsible for this persistence. In the first study of this series, subjects were asked to explain one of several outcomes selected by the experimenter thus defining a causal field. Then they made predictions about the future outcome of the decision problem, identified factors in the causal field, generated alternate outcomes and estimated their likelihood, and made judgments about what factors would be important in determining the future. Subjects tended to focus on the same factors that were present in their initial explanation when generating additional outcomes, and their predictions about future events were biased by their initial explanation. However, they tended to generate the same numbers of success and failure outcomes, and their estimates of the likelihoods of these outcomes was also uninfluenced by the initial explanation they made. These results suggest the importance of the initial causal field has in outcome generation. A second study explored why

the causal field persists. The persistence is not due to selective encoding of the task information, but rather seems to be due to persistence of inferences that the subjects made from the task information when making their initial explanation.

16. Gettys, C.,Kelley, M., Pliske, R. & Beckstead, J. Problem analysis and definition in act generation (TR 8-8-83). Norman, Ok.: University of Oklahoma, **Decision Processes Laboratory**, August 1983.

Three experiments are reported which provide converging evidence suggesting that problem analysis and definition is an important component in generating actions that might solve a problem. Subjects in the first experiment were given two types of cues to help them create solutions to a typical shortage problem. In one condition these cues were generic strategies for solving the problem, whereas in the other condition, specific implementations of these generic strategies were used as cues. Subjects were able to translate the generic cues into specific implementations as expected, but were relatively unsuccessful at extracting the generic "kernels" from cues that were in the form of specific implementations and exploiting variations of these ideas. The second experiment explored the "incubation" phenomena by having subjects resume generating possible solutions to a problem one week after their initial attempt. It was found that problem reorganization rarely occurred between the first and second sessions, and that most of the ideas generated in the second session were elaborations or variations of first-session ideas. The third experiment examined the effects of explicit training in problem analysis and definition. Subjects who received this training showed an improved ability to generate examples of most of the generic solutions to the problem, and tended to generate more indirect solutions to the problem.

17. Engelmann, P. & Gettys, C. Ability and expertise in act generation (TR30-9-83). Norman, Ok.: University of Oklahoma, **Decision Processes Laboratory**, November 1983.

Act generation is a process used by decision makers to create a set of possible actions that might solve a problem. Since previous research had shown college students to generate incomplete sets of possible actions in act generation, the sets of actions generated by experts were examined in the first of two experiments to see if they were more complete. In the first of the two experiments, graduate psychology students were given an act generation task on a subject at which they were expert. Verbal behavior was recorded to aid in the description of expert performance.

In the second experiment the same graduate psychology students were given a task at which their expertise should be of little or no value and were compared to a group of undergraduates. Measures of act generation performance in both experiments included measures of quantity and quality of actions generated.

Graduate psychology students serving as experts in the first experiment excelled in terms of the quality and the quantity of the generated actions. Their performance was markedly superior to the performance found of non-experts in previous experiments on act generation.

In the second experiment, where expertise was not an issue, graduate psychology students again excelled as compared to the undergraduates. One clue that may account for the large performance differences observed between the two groups in the second experiment is divergent thinking ability. This ability, as measured by Guilford's "Alternate Uses" test, was approximately twice as high for the graduate student subjects as compared to the undergraduates.

Since excellent act generation performance of graduate psychology students was found in tasks at which they were either expert or non-expert, divergent intellectual ability was implicated as the source their of excellence. In conclusion, while high intellectual ability was shown to be valuable in generating a nearly exhaustive set of actions, the issue of the effect of expertise on act generation performance remains unsettled.

18. Gettys, C. Research and theory on predecision processes (TR 11-30-83). Norman, Ok.: University of Oklahoma, Decision Processes Laboratory, November 1983.

This monograph discusses six years of research and theory building at the Decision Processes Laboratory concerned with predecision processes, the cognitive processes that occur prior to making the actual decision. These processes include problem detection, the process by which the decision maker decides that a problem exists; act generation, the process of creating candidate acts that might solve the problem; hypothesis generation, where various states of the world are identified that might affect the outcomes of various actions; and outcome generation, a process where the possible results or outcomes of actions are generated.

There are nine substantive chapters in the monograph. The first five chapters are concerned with modeling the various predecision processes and describe the empirical research that addresses these models. Chapter 6 is devoted to research on various topics such as schemata, causal explanation, small group research, individual differences, and expertise in various predecision processes. Chapter 7 discusses recommendations for improving predecision performance, including specific attempts to aid the decision maker, and chapter 8 presents, in summary form, the major conclusions of this program of research. In a chapter 9, general suggestions are made for further research in the area. Also included are titles and abstracts for all technical reports produced in both contracts.

November 1983

OFFICE OF NAVAL RESEARCH

Engineering Psychology Group

TECHNICAL REPORTS DISTRIBUTION LIST

OSD

CAPT Paul R. Chatelier
Office of the Deputy Under Secretary
of Defense
OUSDRE (E&LS)
Pentagon, Room 3D129
Washington, D. C. 20301

Dr. Dennis Leedom
Office of the Deputy Under Secretary
of Defense (C^3I)
Pentagon
Washington, D. C. 20301

Department of the Navy

Engineering Psychology Group
Office of Naval Research
Code 442 EP
Arlington, VA 22217 (2 cys.)

Aviation & Aerospace Technology
Programs
Code 210
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

CDR. Paul Girard
Code 250
Office of Naval Research
800 North Quincy Street
Arlington, VA 22211

Physiology & Neuro-Biology Program
Office of Naval Research
Code 441NB
800 North Quincy Street
Arlington, VA 22217

Deparment of the Navy

Dr. Andrew Rechnitzer
Office of the Chief of Naval
Operations, OP952F
Naval Oceanography Division
Washington, D.C. 20350

Manpower, Personnel & Training
Programs
Code 270
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Mathematics Group
Code 411-MA
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Statistics and Probability Group
Code 411-S&P
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Information Sciences Division
Code 433
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

CDR Kent S. Hull
Helicopter/VTOL Human Factors Office
NASA-Ames Research Center MS 239-21
Moffett Field, CA 94035

Dr. Carl E. Englund
Naval Health Research Center
Environmental Physiology
P.O. Box 85122
San Diego, CA 92138

Department of the Navy

Special Assistant for Marine Corps Matters
Code 100M
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Mr. R. Lawson
ONR Detachment
1030 East Green Street
Pasadena, CA 91106

CDR James Offutt, Officer-in-Charge
ONR Detachment
1030 East Green Street
Pasadena, CA 91106

Director
Naval Research Laboratory
Technical Information Division
Code 2627
Washington, D.C. 20375

Dr. Michael Melich
Communications Sciences Division
Code 7500
Naval Research Laboratory
Washington, D. C. 20375

Dr. J. S. Lawson
Naval Electronic Systems Command
NELEX-06T
Washington, D. C. 20360

Dr. Neil McAlister
Office of Chief of Naval Operations
Command and Control
OP-094H
Washington, D. C. 20350

Dr. Mark L. Montroll
DTNSRDC - Carderock
Code 1965
Bethesda, MD 20084

Department of the Navy

Dr. Robert G. Smith
Office of the Chief of Naval Operations, OP987H
Personnel Logistics Plans
Washington, D. C. 20350

Combat Control Systems Department
Code 35
Naval Underwater Systems Center
Newport, RI 02840

Human Factors Department
Code N-71
Naval Training Equipment Center
Orlando, FL 32813

Dr. Alfred F. Smode
Training Analysis and Evaluation Group
Naval Training & Equipment Center
Orlando, FL 32813

CDR Norman E. Lane
Code N-7A
Naval Training Equipment Center
Orlando, FL 32813

Dr. Gary Poock
Operations Research Department
Naval Postgraduate School
Monterey, CA 93940

Dean of Research Administration
Naval Postgraduate School
Monterey, CA 93940

Mr. H. Talkington
Ocean Engineering Department
Naval Ocean Systems Center
San Diego, CA 92152

Human Factors Engineering
Code 8231
Naval Ocean Systems Center
San Diego, CA 92152

Department of the Navy

Mr. Paul Heckman
Naval Ocean Systems Center
San Diego, CA 92152

Dr. Ross Pepper
Naval Ocean Systems Center
Hawaii Laboratory
P. O. Box 997
Kailua, HI 96734

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps
Code RD-1
Washington, D. C. 20380

Dr. L. Chmura
Naval Research Laboratory
Code 7592
Computer Sciences & Systems
Washington, D. C. 20375

Office of the Chief of Naval
Operations (OP-115)
Washington, D.C. 20350

Professor Douglas E. Hunter
Defense Intelligence College
Washington, D.C. 20374

CDR C. Hutchins
Code 55
Naval Postgraduate School
Monterey, CA 93940

Human Factors Technology Administrator
Office of Naval Technology
Code MAT 0722
800 N. Quincy Street
Arlington, VA 22217

Commander
Naval Air Systems Command
Human Factors Programs
NAVAIR 334A
Washington, D. C. 20361

Department of the Navy

Commander
Naval Air Systems Command
Crew Station Design
NAVAIR 5313
Washington, D. C. 20361

Mr. Philip Andrews
Naval Sea Systems Command
NAVSEA 61R2
Washington, D. C. 20362

Commander
Naval Electronics Systems Command
Human Factors Engineering Branch
Code 81323
Washington, D. C. 20360

Larry Olmstead
Naval Surface Weapons Center
NSWC/DL
Code N-32
Dahlgren, VA 22448

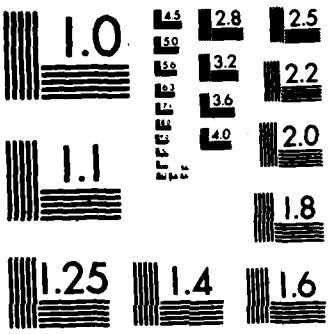
Mr. Milon Essoglou
Naval Facilities Engineering Command
R&D Plans and Programs
Code 03[illegible]
Hoffman Building II
Alexandria, VA 22332

CAPT Robert Biersner
Naval Medical R&D Command
Code 44
Naval Medical Center
Bethesda, MD 20014

Dr. Arthur Bachrach
Behavioral Sciences Department
Naval Medical Research Institute
Bethesda, MD 20014

Dr. George Moeller
Human Factors Engineering Branch
Submarine Medical Research Lab
Naval Submarine Base
Groton, CT 06340

AD-A137 962 RESEARCH AND THEORY ON PREDECISION PROCESSES(U) 2/2
OKLAHOMA UNIV NORMAN DECISION PROCESSES LAB C F GETTYS
30 NOV 83 TR-11-30-83 N00014-80-C-0639
UNCLASSIFIED F/G 5/10 NL



Department of the Navy

Head
Aerospace Psychology Department
Code L5
Naval Aerospace Medical Research Lab
Pensacola, FL 32508

Commanding Officer
Naval Health Research Center
San Diego, CA 92152

Dr. Jerry Tobias
Auditory Research Branch
Submarine Medical Research Lab
Naval Submarine Base
Groton, CT 06340

Navy Personnel Research and
Development Center
Planning & Appraisal Division
San Diego, CA 92152

Dr. Robert Blanchard
Navy Personnel Research and
Development Center
Command and Support Systems
San Diego, CA 92152

CDR J. Funaro
Human Factors Engineering Division
Naval Air Development Center
Warminster, PA 18974

Mr. Stephen Merriman
Human Factors Engineering Division
Naval Air Development Center
Warminster, PA 18974

Mr. Jeffrey Grossman
Human Factors Branch
Code 3152
Naval Weapons Center
China Lake, CA 93555

Human Factors Engineering Branch
Code 4023
Pacific Missile Test Center
Point Mugu, CA 93042

Department of the Navy

Dean of the Academic Departments
U. S. Naval Academy
Annapolis, MD 21402

Dr. W. Moroney
Human Factors Section
Systems Engineering Test
Directorate
U. S. Naval Air Test Center
Patuxent River, MD 20670

Human Factor Engineering Branch
Naval Ship Research and Development
Center, Annapolis Division
Annapolis, MD 21402

Dr. Harry Crisp
Code N 51
Combat Systems Department
Naval Surface Weapons Center
Dahlgren, VA 22448

Mr. John Quirk
Naval Coastal Systems Laboratory
Code 712
Panama City, FL 32401

Department of the Army

Dr. Edgar M. Johnson
Technical Director
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Technical Director
U.S. Army Human Engineering Labs
Aberdeen Proving Ground, MD 21005

Director, Organizations and
Systems Research Laboratory
U. S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

November 1983

Department of the Army

Mr. J. Barber
HQS, Department of the Army
DAPE-MBR
Washington, D. C. 20310

Department of the Air Force

U. S. Air Force Office of Scientific Research
Life Sciences Directorate, NL
Bolling Air Force Base
Washington, D. C. 20332

AFHRL/LRS TDC
Attn: Susan Ewing
Wright-Patterson AFB, OH 45433

Chief, Systems Engineering Branch
Human Engineering Division
USAF AMRL/HES
Wright-Patterson AFB, OH 45433

Dr. Earl Alluisi
Chief Scientist
AFHRL/CCN
Brooks Air Force Base, TX 78235

Foreign Addressees

Dr. Daniel Kahneman
University of British Columbia
Department of Psychology
Vancouver, BC V6T 1W5
Canada

Dr. Kenneth Gardner
Applied Psychology Unit
Admiralty Marine Technology Establishment
Teddington, Middlesex TW11 OLN
England

Foreign Addressees

Director, Human Factors Wing
Defence & Civil Institute of Environmental Medicine
P. O. Box 2000
Downsview, Ontario M3M 3B9
Canada

Dr. A. D. Baddeley
Director, Applied Psychology Unit
Medical Research Council
15 Chaucer Road
Cambridge, CB2 2EF England

Other Government Agencies

Defense Technical Information Center
Cameron Station, Bldg. 5
Alexandria, VA 22314 (12 copies)

Dr. Clinton Kelly
Defense Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, VA 22209

Dr. M. Montemerlo
Human Factors & Simulation Technology, RTE-6
NASA HQS
Washington, D.C. 20546

Other Organizations

Ms. Denise Benel
Essex Corporation
333 N. Fairfax Street
Alexandria, VA 22314

Dr. Andrew P. Sage
School of Engineering and Applied Science
University of Virginia
Charlottesville, VA 22901

Other Organizations

Dr. Robert R. Mackie
Human Factors Research Division
Canyon Research Group
5775 Dawson Avenue
Goleta, CA 93017

Dr. Amos Tversky
Department of Psychology
Stanford University
Stanford, CA 94305

Dr. H. McI. Parsons
Essex Corporation
333 N. Fairfax
Alexandria, VA 22314

Dr. Jesse Orlansky
Institute for Defense Analyses
1801 N. Beauregard Street
Alexandria, VA 22311

Dr. J. O. Chinnis, Jr.
Decision Science Consortium, Inc.
7700 Leesburg Pike
Suite 421
Falls Church, VA 22043

Dr. T. B. Sheridan
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139

Dr. Paul E. Lehner
PAR Technology Corp.
P.O. Box 2005
Reston, VA 22090

Dr. Paul Slovic
Decision Research
1201 Oak Street
Eugene, OR 97401

Dr. Harry Snyder
Department of Industrial Engineering
Virginia Polytechnic Institute and
State University
Blacksburg, VA 24061

Other Organizations

Dr. Ralph Dusek
Administrative Officer
Scientific Affairs Office
American Psychological Association
1200 17th Street, N. W.
Washington, D. C. 20036

Dr. Robert T. Hennessy
NAS - National Research Council (COHF)
2101 Constitution Avenue, N. W.
Washington, D. C. 20418

Dr. Amos Freedy
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91364

Dr. Robert C. Williges
Department of Industrial Engineering
and OR
Virginia Polytechnic Institute and
State University
130 Whittemore Hall
Blacksburg, VA 24061

Dr. Meredith P. Crawford
American Psychological Association
Office of Educational Affairs
1200 17th Street, N. W.
Washington, D. C. 20036

Dr. Deborah Boehm-Davis
General Electric Company
Information & Data Systems
1755 Jefferson Davis Highway
Arlington, VA 22202

Dr. Howard E. Clark
NAS-NRC
Commission on Engrg. & Tech. Systems
2101 Constitution Ave., N.W.
Washington, D.C. 20418

Dr. Robert Fox
Department of Psychology
Vanderbilt University
Nashville, TN 37240

Other Organizations

Dr. Charles Gettys
Department of Psychology
University of Oklahoma
455 West Lindsey
Norman, OK 73069

Dr. Kenneth Hammond
Institute of Behavioral Science
University of Colorado
Boulder, CO 80309

Dr. James H. Howard, Jr.
Department of Psychology
Catholic University
Washington, D. C. 20064

Dr. William Howell
Department of Psychology
Rice University
Houston, TX 77001

Dr. Christopher Wickens
Department of Psychology
University of Illinois
Urbana, IL 61801

Mr. Edward M. Connelly
Performance Measurement
Associates, Inc.
410 Pine Street, S. E.
Suite 300
Vienna, VA 22180

Professor Michael Athans
Room 35-406
Massachusetts Institute of
Technology
Cambridge, MA 02139

Dr. Edward R. Jones
Chief, Human Factors Engineering
McDonnell-Douglas Astronautics Co.
St. Louis Division
Box 516
St. Louis, MO 63166

Other Organizations

Dr. Babur M. Pulat
Department of Industrial Engineering
North Carolina A&T State University
Greensboro, NC 27411

Dr. Lola Lopes
Information Sciences Division
Department of Psychology
University of Wisconsin
Madison, WI 53706

National Security Agency
ATTN: N-32, Marie Goldberg
9800 Savage Road
Ft. Meade, MD 20722

Dr. Stanley M. Roscoe
New Mexico State University
Box 5095
Las Cruces, NM 88003

Mr. Joseph G. Wohl
Alphatech, Inc.
3 New England Executive Park
Burlington, MA 01803

Dr. Marvin Cohen
Decision Science Consortium, Inc.
Suite 721
7700 Leesburg Pike
Falls Church, VA 22043

Dr. Wayne Zachary
Analytics, Inc.
2500 Maryland Road
Willow Grove, PA 19090

Dr. William R. Uttal
Institute for Social Research
University of Michigan
Ann Arbor, MI 48109

Dr. William B. Rouse
School of Industrial and Systems
Engineering
Georgia Institute of Technology
Atlanta, GA 30332

Other Organizations

Dr. Richard Pew
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02238

Dr. Hillel Einhorn
Graduate School of Business
University of Chicago
1101 E. 58th Street
Chicago, IL 60637

Dr. Douglas Towne
University of Southern California
Behavioral Technology Lab
3716 S. Hope Street
Los Angeles, CA 90007

Dr. David J. Getty
Bolt Beranek & Newman, Inc.
50 Moulton street
Cambridge, MA 02238

Dr. John Payne
Graduate School of Business
Administration
Duke University
Durham, NC 27706

Dr. Baruch Fischhoff
Decision Research
1201 Oak Street
Eugene, OR 97401

Dr. Alan Morse
Intelligent Software Systems Inc.
Amherst Fields Research Park
529 Belchertown Rd.
Amherst, MA 01002

Dr. J. Miller
Florida Institute of Oceanography
University of South Florida
St. Petersburg, FL 33701

END

FILMED

3-84

DTIC