

General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.
- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.
- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.
- This document is paginated as submitted by the original source.
- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.



PB94-123205

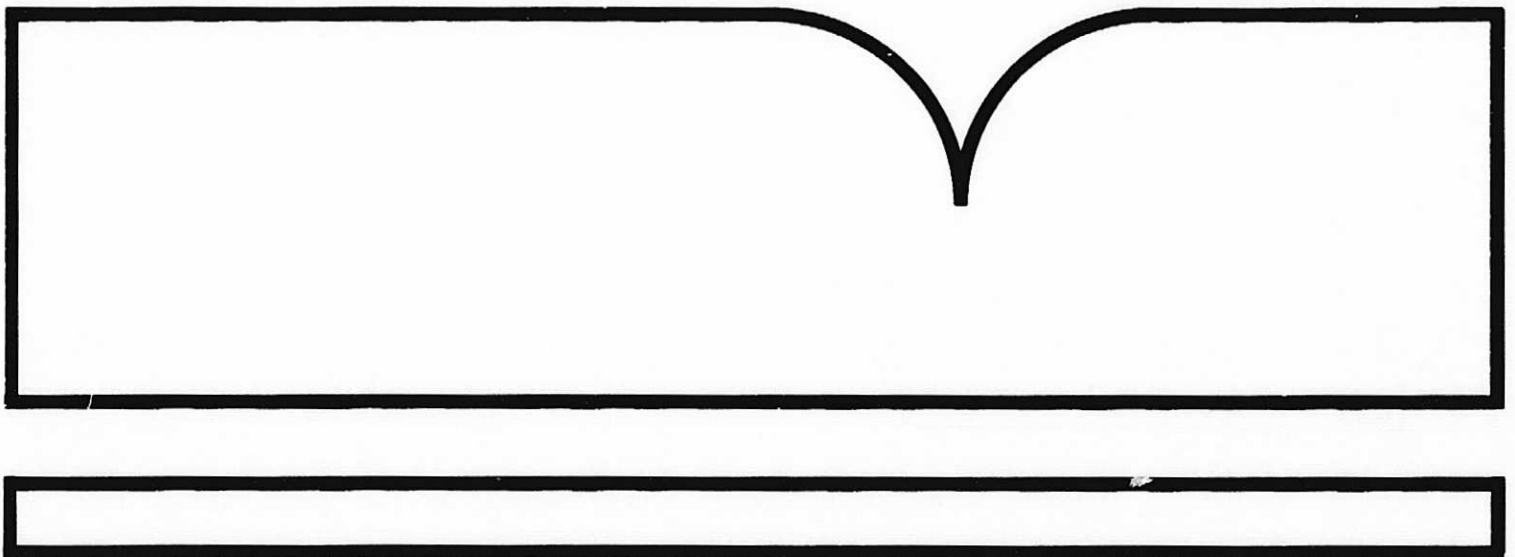
Human Performance Models for Computer-Aided Engineering

National Research Council, Washington, DC

Prepared for:


National Aeronautics and Space Administration, Washington, DC

1989



U.S. DEPARTMENT OF COMMERCE
National Technical Information Service

NTIS

 PB94-123205	2. Report Date: 1989	3. Report Type And Dates Covered:	
4. Title And Subtitle: Human performance models for computer-aided engineering		5. Funding Numbers: Grant no. NAG-2-407	
6. Author(s): Editors: Jerome I. Elkind, Stuart K. Card; Julian Hochberg; Beverly Messick Huey			
7. Performing Organization Names And Addresses: National Research Council Commission on Behavioral and Social Sciences and Education Committee on Human Factors Panel on Pilot Performance Models in a Computer-Aided Design Facility		8. Performing Organization Report Number:	
9. Sponsoring/Monitoring Agency Name(s) And Address(es): National Aeronautics and Space Administration		10. Sponsoring/Monitoring Agency Report Number:	
11. Supplementary Notes:			
12a. Distribution/Availability Statement: unlimited		12b. Distribution Code:	
13. Abstract: This report discusses a topic important to the field of computational human factors: models of human performance and their use in computer-based engineering facilities for the design of complex systems. It focuses on a particular human factors design problem -- the design of cockpit systems for advanced helicopters -- and on a particular aspect of human performance -- vision and related cognitive functions. By focusing in this way, the authors were able to address the selected topics in some depth and develop findings and recommendations that they believe have application to many other aspects of human performance and to other design domains.			
14. Subject Terms: human engineering; data processing; computer-aided engineering;		15. Number Of Pages: 324	
		16. Price Code:	
17. Report Security Classification:	18. Page Security Classification:	19. Abstract Security Classification:	20. Media:

NAS
NAE
IOM



Human Performance Models for Computer-Aided Engineering

1

Human Performance Models for
Computer-Aided Engineering
by [illegible]
[illegible]

Human Performance Models for Computer-Aided Engineering

Jerome I. Elkind, Stuart K. Card,
Julian Hochberg, and Beverly Messick Huey, *editors*

Panel on Pilot Performance Models in a Computer-Aided Design Facility
Committee on Human Factors
Commission on Behavioral and Social Sciences and Education
National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C. 1989

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard to appropriate balance.

This report has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The National Academy of Sciences is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Frank Press is president of the National Academy of Sciences.

The National Academy of Engineering was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Robert M. White is president of the National Academy of Engineering.

The Institute of Medicine was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Samuel O. Thier is president of the Institute of Medicine.

The National Research Council was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Frank Press and Dr. Robert White are chairman and vice chairman, respectively, of the National Research Council.

This work relates to the U.S. Army, National Aeronautics and Space Administration Grant No. NAG-2-407. However, the content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

The United States Government has at least a royalty-free, nonexclusive and irrevocable license throughout the world for government purposes to publish, translate, reproduce, deliver, perform, dispose of, and to authorize others so as to do, all or any portion of this work.

Available from:
Committee on Human Factors
National Research Council
2101 Constitution Avenue N.W.
Washington, D.C. 20418

Printed in the United States of America

PANEL ON PILOT PERFORMANCE MODELS

JEROME I. ELKIND (Chair), Xerox Corporation, Sunnyvale,
California

DEBORAH BOEHM-DAVIS, Department of Psychology, George
Mason University

STUART K. CARD, Xerox Corporation, Palo Alto, California

RENWICK E. CURRY, Tycho Systems, Inc., Palo Alto, California

DOUGLAS HARRIS, Anacapa Sciences, Inc., Santa Clara, Utah

JULIAN HOCHBERG, Department of Psychology, Columbia
University

RICHARD PEW, Experimental Psychology Department, Bolt,
Beranek and Newman Laboratories, Cambridge, Massachusetts

ANDREW B. WATSON, NASA Ames Research Center, Moffett
Field, California

CHRISTOPHER WICKENS, Aviation Research Laboratory,
University of Illinois, Savoy, Illinois

HAROLD P. VAN COTT, Study Director

STANLEY DEUTSCH, Study Director (1985-1987)

BEVERLY M. HUEY, Staff Officer

PILOT PERFORMANCE MODELING RESOURCE GROUP

IRVING BIEDERMAN, Department of Psychology, University of
Minnesota

MYRON BRAUNSTEIN, Department of Social Sciences,
University of California, Irvine

LYNN COOPER, Department of Psychology, University of Arizona

BARUCH FISCHHOFF, Carnegie-Mellon University

WALTER SCHNEIDER, University of Pittsburgh

JAMES TODD, Department of Psychology, Brandeis University

DAVID WOODS, Department of Industrial and Systems
Engineering, Ohio State University

GREG ZACHARIAS, Charles River Analytics, Inc., Cambridge,
Massachusetts

COMMITTEE ON HUMAN FACTORS

- DOUGLAS H. HARRIS (Chair), Anacapa Sciences, Santa Clara,
Utah
- MOHAMED M. AYOUB, Institute of Biotechnology, Texas Tech
University
- JEROME I. ELKIND, Information Systems, Xerox Corporation,
Sunnyvale, California
- MIRIAN M. GRADDICK, Human Resources Strategy and Systems
Department, AT&T, Basking Ridge, New Jersey
- OSCAR GRUSKY, Department of Sociology, University of
California, Los Angeles
- JULIAN HOCHBERG, Department of Psychology, Columbia
University
- THOMAS K. LANDAUER, Information Sciences Division, Bell
Communication Research, Morristown, New Jersey
- NEVILLE MORAY, Department of Mechanical and Industrial
Engineering, University of Illinois
- RAYMOND NICKERSON, Bolt, Beranek and Newman
Laboratories, Cambridge, Massachusetts
- JUDITH REITMAN OLSON, Graduate School of Business
Administration, University of Michigan
- THOMAS B. SHERIDAN, Department of Mechanical Engineering,
Massachusetts Institute of Technology
- CHRISTOPHER I. WICKENS, Aviation Research Laboratory,
University of Illinois, Savoy
- ROBERT C. WILLIGES, Department of Industrial Engineering
and Operations Research, Virginia Polytechnic Institute and
State University
- HAROLD P. VAN COTT, Study Director
- BEVERLY M. HUEY, Staff Officer
- ELIZABETH F. NEILSEN, Research/Staff Assistant
- AUDREY E. HINSMAN, Administrative Secretary

Contents

FOREWORD	xiii
----------	------

PREFACE	xv
---------	----

PART I

1	INTRODUCTION	3
	Helicopter Flight Problems and Applications of Human Performance Models, 9	
	Detectability and Visibility, 9	
	Surface and Motion Estimation, 10	
	Object Recognition, 11	
	Hetero-Ocular Vision, 12	
	Workload and Pilot Performance, 13	
	Decision Theory, 14	
	Memory Overload, 14	
	Skill Acquisition, 15	
	Human Error, 15	
	References, 16	
2	PREVIEW OF MODELS	17
	Framework, 17	
	Assessment of Models, 19	
3	USE AND INTEGRATION OF MODELS	23
	Design Process, 24	

8	MOTION-BASED STATE ESTIMATION AND SHAPE MODELING	106
	GREG ZACHARIAS	
	Introduction and Summary, 106	
	Framework for Motion-Based State Estimation and Shape Modeling, 108	
	Review of Research in Motion-Based State Estimation and Shape Modeling, 112	
	Model Applications and Limitations, 119	
	Future Research, 121	
	References, 123	
9	REAL-TIME HUMAN IMAGE UNDERSTANDING IN PILOT PERFORMANCE MODELS	126
	IRVING BIEDERMAN	
	Theories of Object Recognition, 127	
	Model-Based Matching: Lowe's SCERPO and Ullman's Alignment Models, 131	
	Perception of Multiobject Displays, 138	
	References, 142	
10	MANIPULATION OF VISUAL INFORMATION	144
	LYNN A. COOPER	
	Summary, 144	
	Introduction, 145	
	Transformations on Information Presented in a Static Visual Display, 146	
	Memory for Positions in a Sequence of Static Displays, 150	
	Extrapolation of Perceptually Driven Spatial Transformations, 151	
	Judgments of Object Structure From Partial Views, 153	
	Future Research, 154	
	References, 155	
11	COMBINING VIEWS	159
	JULIAN HOCHBERG	
	Integration of Successive Views, 159	
	Binocular Combination, 162	
	References, 163	
12	AFTERWORD	166

PART III

13	INTRODUCTION TO COGNITION MODELS	171
14	COGNITIVE ARCHITECTURES	173
	STUART K. CARD AND ALLEN NEWELL	
	Symbolist Architectures, 174	
	Connectionist Models, 176	
	References, 178	
15	RESOURCE MANAGEMENT AND TIME-SHARING	180
	CHRISTOPHER D. WICKENS	
	Overview, 180	
	Serial Allocation, 182	
	Parallel Allocation, 187	
	Serial Competition, 188	
	Parallel Competition, 189	
	Synthesis of the Optimal Model, 193	
	Conclusion, 197	
	References, 198	
16	MODELS OF WORKING MEMORY	203
	STUART K. CARD	
	Phenomena of Working Memory, 204	
	Models of Working Memory, 208	
	References, 212	
17	TRAINING MODELS TO ESTIMATE TRAINING COSTS FOR NEW SYSTEMS	215
	WALTER SCHNEIDER	
	Overview, 215	
	Skill Development, 216	
	Models for Predicting Human Performance, 220	
	Engineering Guidance Without an All-Inclusive Model, 226	
	Use of Rapid Prototyping and Quick Empirical Evaluations, 227	
	Needed Research, 228	
	References, 229	
18	MODELING SCENARIOS FOR ACTION	223
	STUART K. CARD	
	Fixed Scenarios, 233	

	Scenarios With Simple Contingencies, 240	
	Modeling More Complex Scenarios, 242	
	References, 245	
19	MODELING AND PREDICTING HUMAN ERROR	248
	DAVID D. WOODS	
	Introduction, 248	
	Error Modeling, 253	
	References, 269	
20	MODELING DECISION MAKING FOR SYSTEM DESIGN	276
	BARUCH FISCHHOFF	
	Why Decision Making Seems Easy to Model—	
	Sometimes, 277	
	Implication for Modeling Operator Performance, 279	
	Modeling Without Optimality, 282	
	Making Behavior More Model-like, 285	
	Testing the Limits of Decision Making, 287	
	References, 288	
21	KNOWLEDGE ELICITATION AND REPRESENTATION	291
	DEBORAH A. BOEHM-DAVIS	
	Knowledge Elicitation, 291	
	Knowledge Representation, 294	
	Mental Models and Design Decisions, 297	
	References, 298	
22	AFTERWORD	299

PART IV

23	FINDINGS AND RECOMMENDATIONS	303
	Desirable Attributes and Types of Models, 304	
	Adequacy of Models for the A ³ I Design Facility, 305	
	Validation, 306	
	Need for Access to Human Factors Data Base, 306	
	Broader Context of Computational Human Factors, 307	
	Importance of the System Design Context for Research on Models, 307	
	Focusing the A ³ I Program, 308	
	Providing a Framework and a Box of Tools, 309	

Foreword

The Committee on Human Factors was established in October 1980 by the Commission on Behavioral and Social Sciences and Education of the National Research Council. The committee is sponsored by the Army Research Institute for the Behavioral and Social Sciences, the Office of Naval Research, the Air Force Office of Scientific Research, the National Aeronautics and Space Administration, the National Science Foundation, and the Army Advanced Systems Research Office. The principal objectives of the committee are to provide new perspectives on theoretical and methodological issues, to identify basic research needed to expand and strengthen the scientific basis of human factors, and to attract scientists both within and outside the field for interactive communication and performance of the necessary research. The goal of the committee is to provide a solid foundation of research as a base on which effective human factors practices can build.

Human factors issues arise in every domain in which humans interact with the products of a technological society. To perform its role effectively, the committee draws on experts from a range of scientific and engineering disciplines. Members of the committee include specialists in such fields as psychology, engineering, biomechanics, physiology, medicine, cognitive sciences, machine intelligence, computer sciences, sociology, education, and human factors engineering. Other disciplines are represented in the working groups, workshops, and symposia. Each of these contributes to the basic data, theory, and methods required to improve the scientific basis of human factors.

Preface

The Panel on Pilot Performance Models for Computer-Aided Engineering was formed by the National Research Council (NRC) in response to a request from the Army Advanced Systems Research Office. The National Aeronautics and Space Administration (NASA) Ames Research Center asked the NRC to conduct a study that would provide advice and guidance in a number of areas important for the Army-NASA Aircrew/Aircraft Integration (A³I) program which is developing a prototype of a human factors computer-aided engineering (CAE) facility for the design of helicopter cockpits. This study was conducted under the auspices of the Committee on Human Factors within the National Research Council's Commission on Behavioral and Social Sciences and Education.

The objectives of the study were to review current models of human performance; to identify those that would be most useful for the CAE facility; to identify limitations of the models; to provide guidance for the use of these models in the CAE facility; and to recommend research on models and modeling that might overcome existing limitations. The panel focused its attention on the visual and associated cognitive functions required of pilots in the operation of advanced helicopters, which often fly under low-visibility and low-altitude conditions. By limiting the scope of the study in this way, the panel was able to address an important domain of human performance models (vision and associated cognition) in some depth and to gain an understanding of the prospects and problems of using such models in a CAE facility for helicopter design. In addition, the

The panel appreciates the cooperation, support and advice it received from many individuals at the NASA Ames Research Center, including Loren Hayworth, Jerry Murray, James Hartzell, James Larimer, and David Nagel. It is also grateful to Jim Vorhees, Rik Warren, and William Rouse for their presentations to the panel. It is especially appreciative of the contributions of the NRC staff who assisted in beginning this study, nurtured it throughout its execution, and assisted greatly in bringing it to completion. The study began under Stanley Deutsch, who was the second study director of the Committee on Human Factors, and was completed under the auspices of Harold Van Cott, the current study director. Both helped organize the study and participated in the work of the panel. Beverly Huey, Committee on Human Factors staff officer, not only organized and participated in the meetings of the panel, but willingly and effectively took on the task of assembling this report and contributed to its editing. Audrey Hinsman and Carole Foote provided secretarial assistance in preparing this document for review.

Finally, I want to personally acknowledge and thank the panel members for their extensive contributions, their many thoughtful position papers, and their gracious collaboration throughout this study. Special thanks to Julian Hochberg and Stuart Card who undertook the major tasks of organizing and editing the collection of reviews of vision and cognition models and who were enormously influential in determining the course of this study.

Jerome I. Elkind, Chair
Panel on Pilot Performance
Models in CAD/CAE Facilities

Part I

1

Introduction

This report discusses a topic important to the field of computational human factors: models of human performance and their use in computer-based engineering facilities for the design of complex systems. It focuses on a particular human factors design problem—the design of cockpit systems for advanced helicopters—and on a particular aspect of human performance—vision and related cognitive functions. By focusing in this way, the authors were able to address the selected topics in some depth and develop findings and recommendations that they believe have application to many other aspects of human performance and to other design domains.

The report is addressed to human factors professionals and others interested in human performance models, human factors design methodology, and design tools. It describes some of the key vision-related problems of helicopter flight and cockpit design as a way of introducing the reader to the design domain on which the report is focused. It discusses issues in the integration of models into a computer-based human factors design facility and the use of such a facility in the design process, and it reviews existing models of vision and cognition with special attention to their use in a computer-based design facility. It concludes with a set of findings about the adequacy of existing models for a computational human factors facility and a related set of recommendations for research that is needed to provide a stronger foundation of models upon which to base such a facility.

A model is a representation or description of all or part of an object or process. There are many different types of models and they are developed for a variety of reasons. In a design context, models can be considered to be a “thing” of which we ask questions about some aspect of a design. Models of human performance have long been used in the human factors design of complex systems to answer questions

Analytic models represent human performance mathematically, typically in terms of algebraic or differential equations. Both the form of the equations and their parameters are of interest to the psychologist and the designer. Analytic models often provide concise descriptions and even "laws" governing human behavior that are of enormous value in the design process.

Some models attempt to represent specific human processes, usually by simulation, and as a result are known as process models. Others attempt to predict only human output without claiming to be good representations of the human processes involved, and are known as performance models. Models of the processes used by the human to accomplish the task under study are more powerful than those that just describe the observed external behavior (outputs) because they are more likely to be applicable to a wider range of tasks and conditions.

Most models in the literature are descriptive in the sense that they were developed to describe observed human behavior, performance, or processes. A few, however, are prescriptive in the sense that they prescribe how the human should perform if he were to behave in a rational way that takes into account the information available, the constraints that exist, the risks, rewards, and objectives. Some rational models are based on strong theories of optimality, such as those that have been developed in the fields of control, decisions, and signal detection, and are known as ideal observer or ideal operator models. We will often refer to prescriptive models as rational action or normative models.

Until fairly recently, most human performance models were numerical or quantitative and lent themselves to classical, numerically-based computation. As a result of progress in artificial intelligence and cognitive science, a substantial body of non-numerical, qualitative, but calculable models, has been developed. These models are necessary for representing cognitive behavior. Although they are qualitative, they are computational and, as such, are amenable for inclusion in a computer-based engineering facility. Many of the reviews of vision and cognition in this report address qualitative models.

Models can represent behavior at different levels and with different amounts of detail. There are mission-level models that attempt to encompass the whole mission or major mission segments by representing human behavior at a high level of abstraction. Such models are concerned, typically, with issues such as the workload on the human operator. Models can address entire human subsystems, such

as vision or motor control, or be focused on a part of a complex task or of a human subsystem. There is the goal of building models that tie together detailed models of several human subsystems to obtain a "complete" representation of human behavior in a complex system. However, most comprehensive models contain little detail about specific aspects of human performance, reflecting the harsh reality of the trade-off of breadth against depth.

Most existing models of human performance were developed with a simple task in mind, but there have been numerous efforts to build more comprehensive models that attempt to represent more complex behavior, often by assembling and integrating simple task models within a uniform framework. As a result of decades of research, a large collection of models now exists for many aspects of human perceptual, motor, cognitive, and biomechanical performance. The extent to which these simple task models can be usefully integrated to represent more comprehensive behavior depends upon the nature of the gaps in the coverage of the models and on the completeness of the linkages among them. Both of these problems are addressed in the reviews of models in Parts II and III of this report.

Much of the progress in modeling that has occurred in recent years has been due to the remarkable increase that has occurred in the power of mainframe and desk top computer systems and in the ability to network large numbers of computers together. This increase in computational power has made more comprehensive and complete models, as well as large scale simulation models and models of cognitive processes, practical. This has made it easier to apply models of all types to the problems of system analysis and design, and has fostered advances in software technology, most notably in the areas of human interface design and in the construction of very large modular software systems that are critical to dealing with complex models.

The advances in computer technology have also made possible the development of very important computer-aided engineering (CAE) tools for a number of different disciplines such as mechanical, VLSI (very large scale integrated), electronic, architectural, and aircraft design. These tools have greatly increased the efficiency of the design process and the quality of the resulting designs, largely by enabling the designer to work rapidly, construct a model of the system being designed, and carry out computations on that model to predict and analyze its performance under a wide range of conditions.

It is not surprising, given all these developments, that growing interest has emerged in applying computational modeling and engineering techniques to the human factors design of complex systems. Underlying this interest is the belief that from the collection of existing computational models of human performance, a sufficiently comprehensive set could be assembled in a CAE facility to make feasible a computer-based human factors design methodology for complex human-machine systems. Such systems could be used to formulate and evaluate alternatives for allocating functions to human operators, for the design of human-machine interfaces, and for the design of machine characteristics.

In this technological context a joint program was initiated by the Army and the NASA Ames Research Center in 1985 (Corker, Davis, Papazian, and Pew, 1986) with the objective of developing a computer-based methodology and a set of tools focused on the design of advanced helicopter cockpit systems, a challenging example of human factors design of particular interest to these organizations. This program, called A³I (Army-NASA Aircrew/Aircraft Integration), is developing a prototype human factors computer-aided engineering (HF/CAE) facility to investigate problems of computational design methodology and to demonstrate the utility of the methodology and of the facility itself. The HF/CAE facility will incorporate models of human performance together with other data and tools useful for human factors design and will make them accessible to trained design practitioners for use in actual design problems. The project hopes to demonstrate that it is possible for designers to explore many more design alternatives than they can now and to make better evaluations of these designs before they are committed to the costly and time-consuming construction of prototype hardware and software. Although the A³I CAE system is directed toward the design of advanced helicopter cockpit systems, the system itself and the concepts and technology upon which it is based have broad application to the development of computational human factors design methodology for complex human-machine systems.

In 1985 NASA requested the Committee on Human Factors of the National Research Council to conduct a study to provide advice and guidance for the development of the human factors aspects of the HF/CAE facility. The purpose of the study was to review current models of human performance, identify those that would be most useful for the purposes of the CAE facility, identify limitations of these models, provide guidance for the use of these models in the

CAE facility, and recommend research on models and modeling that might overcome these limitations. The focus of this study was to be the perceptual and control tasks required of a single pilot in advanced helicopter operation in low-altitude (i.e., nap of the earth) and low-visibility (e.g., nighttime) missions, which are very demanding flight conditions.

As the panel began its work and acquired a better understanding of helicopter piloting and cockpit design problems, it became apparent that the overwhelmingly dominant problems, in terms of human factors, under the low-altitude, low-visibility flight conditions have to do with human vision, particularly the interpretation of visual information, and the use of visual aids and displays designed to assist the pilot in obtaining information necessary for the successful completion of a mission. The panel also concluded that much was known about the state of manual control models, especially since the Committee on Human Factors had earlier initiated another study of human performance models of complex dynamic systems (Baron and Kruser, in press). For these reasons, the panel decided to focus its attention on models of vision and on those aspects of cognition that interact with the human visual system in the helicopter flight task. It undertook to review the state of models in these areas, to recommend how they might be used in design and integrated into the CAE facility, to propose how they could be integrated into such a facility, and to suggest research that might make models more useful in the future for CAE-based design by eliminating the gaps and limitations of currently available models. Although its study focused on vision, the panel believed that an in-depth study of this area not only would provide useful guidance about vision to the A³I project, but would provide broader insights into the potential problems of attempting to incorporate models from the psychological and human factors literature into a computer-based design tool such as the HF/CAE facility.

To conduct this study, the panel assembled a number of experts from the fields of vision, cognition, perception, performance modeling, aviation psychology, decision theory, system design, manual control, and related fields. The results of its work are reported here.

HELICOPTER FLIGHT PROBLEMS AND APPLICATIONS OF HUMAN PERFORMANCE MODELS

Helicopter operation is difficult, and performing low-altitude, low-visibility missions with a single-person crew places very severe demands on the pilot. Analysis and design of the helicopter cockpit system and the missions it is to perform must be thorough to ensure that the missions are indeed possible and that the cockpit system, especially visual aids and displays, facilitates successful accomplishment of required flight tasks. If the A³I project and others based on similar concepts are successful, this analysis and design will be accomplished by using CAE facilities and design methodologies based on the use of human performance models of the type discussed in this report.

This chapter attempts to give the reader a concrete, intuitive feel for the application of human performance models to the design of advanced helicopters and other highly automated vehicles. A sequence of vignettes is presented, each of which is a brief episode illustrating an important practical problem that can arise from a limitation in the perceptual and cognitive capabilities of the pilot, which might be solved through design based on human performance models. The kinds of models that might be used to characterize pilot capabilities are described, along with the way in which they might be used for design. Reference is made to chapters of this report in which these models and their application are discussed more fully.

DETECTABILITY AND VISIBILITY (CHAPTER 5)

Parched by summer drought, the pine, oak, and eucalyptus of northern California's Santa Cruz Mountains erupt in flames, and many remote mountain households are threatened. By afternoon, access roads to some homes have been cut off by encroaching flames, and rescue helicopters are summoned. As dusk approaches, the pilot's ability to navigate and identify his destination deteriorates. Soon the forest below dissolves into a sea of gray. What is most worrisome is that the pilot can no longer scan the landscape for suspended wires, the cause of a disproportionate number of rotorcraft accidents. The pilot pulls down a visor on which is mounted a sophisticated night vision system, and at once the earth below appears alive with light. The crisp detail of the imagery enables the pilot to avoid an oncoming power line and to detect the white clapboard corner of the threatened home.

The design of the night vision system was aided by computer models of the pilot's visual system. Engineers could predict, in advance of construction, whether the pilot equipped with a candidate system would have adequate resolution, contrast, and temporal dynamics to detect the targets critical to mission success. In particular, the visibility of threatening wires in both central and peripheral view could be calculated accurately. Equipped with this information, design decisions can be made to optimize the performance of the viewer. Without adequate computer models, designers would be forced into a repeated cycle of design, prototype fabrication, and field test—each step costly and time-consuming. Computer models also provided a further insight: even with an optimal viewer design, not all threatening wires can be detected. In certain cases, the wire may simply not be visible enough to the human visual system. This suggested the need for a vision aid that could either enhance wirelike features of the visual image or automatically detect their presence and notify the pilot.

SURFACE AND MOTION ESTIMATION (CHAPTER 8)

Attempting to evade enemy radar, a helicopter pilot approaches a target at high speed and low altitude. Hugging the rolling desert terrain contours at this speed, the pilot must react instantly to changes in the terrain, which is especially difficult because it is dusk and shading cues are absent. Suddenly, sagebrush that previously dotted the landscape is no longer there, and the terrain below becomes a featureless, untextured sheet. The pilot immediately engages a ground contour synthesizer (GCS) and instead of shapeless terrain, the full depth of the undulating desert floor is revealed.

This illusion is made possible by the helmet-mounted display (HMD) of a computer-generated image (CGI), texturing superimposed on the view of the terrain below. The GCS design draws heavily on human models of self-motion and object shape perception, describing pilot performance in dynamic visual environments. Early in the design state, mission planners and human factors engineers used these models to identify mission segments and environmental conditions that could pose significant problems to the pilot flying the baseline vehicle configuration. A variety of augmentation schemes were then proposed and evaluated, again by using pilot models to rank the expected performance improvements. The GCS scheme was selected for further evaluation, and the design engineers outlined a

basic architecture consisting of a laser range finder driving a CGI on a head-tracked HMD. Using some of the available perceptual models, the display and controls group then evaluated a wide range of design factors, such as field-of-view, texture density, ranging accuracy, and update rates, and narrowed their full-scale simulation evaluations to a small number of promising designs. This allowed them to focus on system "tuning" well before committing to prototype hardware.

OBJECT RECOGNITION (CHAPTER 9)

To locate a missing vehicle, a pilot is flying a rescue reconnaissance mission under threat of hostile ground fire. In a standard defensive precaution, the pilot must "pop up" briefly from behind each protective hill, survey the scene from that short vantage, and immediately drop down again behind the hill. In that momentary survey, he must determine whether the missing vehicle is present, what potentially hostile objects are present, and the position and orientation of each relative to the terrain and objects.

The cockpit designer has considerable control over the ease with which a pilot can perform this type of perceptual task. The shape of the windscreen and the distribution of occluding structural components limit the size of the continuous field of view. Also critical are the visual parameters of artificial displays such as night vision or other video and computer-generated imaging devices. These parameters include the spatial resolution, contrast, and gray scale or number of colors used to depict objects and features; the field of view encompassed by the display; the display refresh rate; and the rate at which a viewing camera or sensor's direction can be changed. If image enhancement algorithms are used, they may hinder or facilitate rapid object recognition by interacting with stimulus features of the patterns being reproduced. Although not enough is known to develop computational models that will take parameters such as these into account in modeling object recognition, enough is now known to help the designer make quantitative assessments. Research is underway to expand this knowledge.

Even for direct vision unobstructed by parts of the aircraft, the rapid survey task may be a formidable one that requires assistance. Target objects and their context are likely to be viewed from an unforeseen direction and with unanticipated partial occlusions from other objects and features of the terrain. This is especially likely when the pilot's course has been tortuous in order to take advantage

of whatever cover the terrain provides. Synthesized schematic previews of the vista, using navigational data and terrain data bases, which show anticipated target objects viewed from several positions and reduced to their essential features, may improve task performance. The design of such displays can call upon the growing knowledge about the component processes involved in the rapid recognition of scenes and objects, and about mental rotation.

HETERO-OCULAR VISION (CHAPTER 11)

As illustrated in previous vignettes, conditions of haze or darkness that would otherwise make low-altitude helicopter flight impossible can be at least partly overcome by vision-augmenting devices. While this makes flying possible when it would otherwise be impossible, it may impose a new set of demands on the pilot. One system in active use, the pilot night vision system (PNVS), is a helmet-mounted monacle that presents the right eye with both a video picture of the environment as scanned by a forward looking infrared (FLIR) sensor and an array of symbols that reflect the state of the vehicle. The left eye is free to view the world directly. This system is usable, making nap of the earth (NOE) flights possible under conditions of very low visibility, but has severe drawbacks in its present form. It is difficult to learn, demanding and fatiguing to use, and interferes with normal involvement of the two eyes. Some of these problems are structural, such as the fact that the FLIR sensor (which moves in response to the pilot's head movements) is substantially offset in viewpoint, but another set of problems arises because the pilot must attend to the disparate information received by the two eyes.

In general, when the two eyes receive different views that cannot combine to form a single scene, binocular rivalry results: at each small region in the combined field of view, the control of one eye or the other is visible, but not both. One eye will occasionally and for a short time dominate to the exclusion of the other: thus, a small dot in one eye's view will be visible almost continually if it falls against a blank field in the corresponding part of the other eye's view. A piecemeal alternation between fragments of the two views is, however, the more general occurrence. Which view prevails in any region depends on the stimulus conditions (contrast, sharpness of contour, etc.) in each of the two corresponding hetero-ocular regions and on the gaze directions and states of the two eyes (e.g., their adaptation and accommodation). Although the pilot must try to

attend to the eye that offers the information needed at any moment, this can apparently be done only by closing or otherwise diminishing the effect of the other eye, by changing the relative gaze directions of the two eyes so as to bring different regions into correspondence, or by physically changing the monocular video display in some way.

Control over the physical characteristics (luminance, contrast, temporal and spatial discontinuities, etc.) can thus give the pilot more control over the rate and bias of the rivalry. With the growing knowledge in the field of binocular rivalry, it seems reasonable to aim for models that will allow the rate of rivalry, and its effects on information under various designs and viewing conditions, to be evaluated.

Even without rivalry, there are problems associated with the use of two eyes as separate channels of information. For example, the pilot is denied the binocular information about depth that is normally so important for judging near distances. Yet rivalry seems to be the most troublesome aspect of the hetero-ocular procedure, and one that should prove relatively easy to ameliorate by proper design. When that is done, the optimal hetero-ocular method can then be compared to alternative ways of presenting the various channels of information.

WORKLOAD AND PILOT PERFORMANCE (CHAPTER 15)

A helicopter is flying "nap of the earth" below treetop level in the dim illumination of twilight. The pilot listens to the copilot call out landmarks that must be located and aimed at all the while judging altitude, adjusting speed, and assuring clearance from ground obstacles. While mentally computing the distance from a rendezvous point, the pilot receives a radio communication describing the relative locations of other aircraft in the area. An alarm sounds, indicating a potential fault in the tail rotor engine.

How well will the pilot be able to integrate and time-share these various activities? Will the auditory alert be noticed while navigational instructions are being encoded? How will the difficulty of resolving landmarks in the twilight degrade the pilot's ability to visualize the spatial layout of helicopters in the area or comprehend verbal communications? How will all of these cognitive activities degrade the ability to fly? To attempt to answer these questions, the designer will need workload models that will predict the interference between these activities as a function of their similarity to each other

and their difficulty. With this information, the designer can make intelligent decisions regarding crew complement, information displays, and decision-making aids.

DECISION THEORY (CHAPTER 20)

Returning to home base and low on fuel, the pilot receives a transmission requesting assistance elsewhere. Is there enough fuel? In the past, this could be determined only by difficult mental calculations involving fuel, airspeed, and wind velocity. The pilot consults a new display that shows an ellipse superimposed on a map of the local area. The ellipse encloses the points that can be reached given the current fuel, airspeed, and wind velocity. On the basis of this simple, accessible display, the pilot makes a rapid decision.

Design of the display was assisted by models of human decision making. These models suggest that reducing uncertainty leads to better performance; thus, this display should improve performance in estimating whether the pilot can reach a given destination. However, because the display is concrete and precise, pilots may attribute excessive accuracy to the readings, and thus unduly reduce their margin of safety, thereby actually increasing the chance that they will run out of fuel. Use of models of decision making, along with simulations of hypothetical missions, could assist in answering this type of question.

MEMORY OVERLOAD (CHAPTER 16)

As the pilot pursues his rescue mission through fire and smoke, radio communications must be maintained with other helicopters, air traffic control, and teams from the fire and police departments. In previous cockpits, setting many precise communication frequencies had been a difficult manual and mental task. The pilot is fortunate in having a new interface that places the memory burden on the computer rather than the human.

The design of the new communications interface was guided by models of human memory integrated into a system for simulation of the rotorcraft mission. The engineers sought to understand whether the previous system imposed too much mental workload on the pilot and how much improvement would arise from several new (and more expensive) proposed designs. Extensive simulations of previous designs showed that frequent confusion of radio frequencies occurred,

and that memory overload degraded performance in other tasks, such as looking for threatening wires. By contrast, simulations of the new design demonstrated that it effectively eliminated confusion and reduced the overall memory workload.

SKILL ACQUISITION (CHAPTER 17)

The pilot enters way point information into the navigational computer using the navigation keyboard. Then the flight computer is used to enter fuel consumption information. Keyboards for the two computers are laid out somewhat differently. Despite repeated use, the pilot cannot get the "feel" for entering the data and must look to the cockpit much longer than safety would allow.

Having multiple keyboards creates both cognitive and motor difficulties. The skills to be acquired in using one keyboard interfere with the skills to be acquired in using the other. The pilot's use of navigation and flight computers must be so highly practiced that data entry tasks can be performed "almost without thinking." This level of skilled performance is known as the achievement of automaticity. The presence of multiple keyboards, however, prevents the pilot from being able to acquire automaticity in performing either task and increases the probability that an error will be made in entering data. Models of skill acquisition cannot yet provide direct specifications for keyboards and displays that would promote automaticity. However, the simplest models in the form of guidelines suggest that minimizing the number of alternative methods allowed for data input will promote automaticity.

HUMAN ERROR (CHAPTER 19)

Guiding a highly automated commercial aircraft toward the airport, the pilot moves the controls so as to produce an appropriate descent toward the runway. Unknown to the pilot, the aircraft is in an automatic control mode, and the pilot's action has no effect. Because the pilot's intended path and that executed by the automatic systems are very similar, there are few indications that anything is wrong. Only when an unusual condition occurs, for example, a request from air traffic control that the pilot use an alternate runway, do problems begin. The aircraft does not respond to control actions, and it takes some time for the pilot to realize the source of the problem.

This is an example of a mode error. When systems have multiple modes, people may confuse which mode the system/interface is in

and take actions for an inappropriate mode. The control, navigation, communication, and weapon systems aboard modern aircraft have numerous system modes; thus, the potential exists for many mode errors.

Mode error correction strategies fall into two categories: one is to design the systems with a minimal number of modes (Norman, 1983); the other is to provide a perceptual cue of the current mode of the system. Labels are generally not sufficient; a salient background field may be more effective (Monk, 1986). Models of human information processing, and specific models of mode error, may identify interfaces that are prone to mode error and assist in the design of new systems that are immune to this potentially disastrous flaw.

REFERENCES

- Baron, S., and Kruser, D., eds.
in press *Human Performance Modeling*. Committee on Human Factors, National Research Council. Washington, D.C.: National Academy Press.
- Corker, K., Davis, L., Papazian, B., and Pew, R.
1986 *Development of an Advanced Task Analysis Methodology and Demonstration for Army-NASA Aircrew/Aircraft Integration*. Report No. 6124. Cambridge, MA: Bolt Beranek and Newman, Inc.
- Monk, A.
1986 Mode errors: A user-centered analysis and some preventive measures using keying-contingent sound. *International Journal of Man-Machine Studies* 24:313-327.
- Norman, D.A.
1983 Design rules based on analyses of human error. *Communications of the ACM* 26:254-258.

2

Preview of Models

Parts II and III of this report contain chapters that review models of vision and cognition which are important for the analysis and simulation of pilot performance of the visual tasks encountered in low-altitude, low-visibility helicopter operations. The models discussed in these chapters are important candidates for inclusion in a human factors computer-aided engineering (HF/CAE) facility.

FRAMEWORK

In selecting and organizing the models reviewed, the authors had in mind the general framework and functional decomposition shown in Figure 2-1. Even though the chapters in Parts II and III and the individual models discussed in these do not follow this framework rigidly, it has been useful for organizing the discussion of this complex field.

The framework of Figure 2-1 is aimed toward a full simulation model of the visual system. This system is modeled as a serial set of processes starting with early vision. Eye fixation, although shown in the figure, is only treated statistically, and the details of the eye movement process are covered superficially. The inputs to the early vision models are direct physical measures of the visual scene, and thus these models and those that build from them are image driven. The framework assumes that the outputs of models at one stage provide the inputs needed by those at the next stage in progression from early vision to form perception, three-dimensional structure through motion, state-variable estimation, object recognition, mental manipulation of information and finally to combination of views. The later stages of vision are recognized as being cognitive, and are shown as being within the envelope of the cognitive system. Later visual

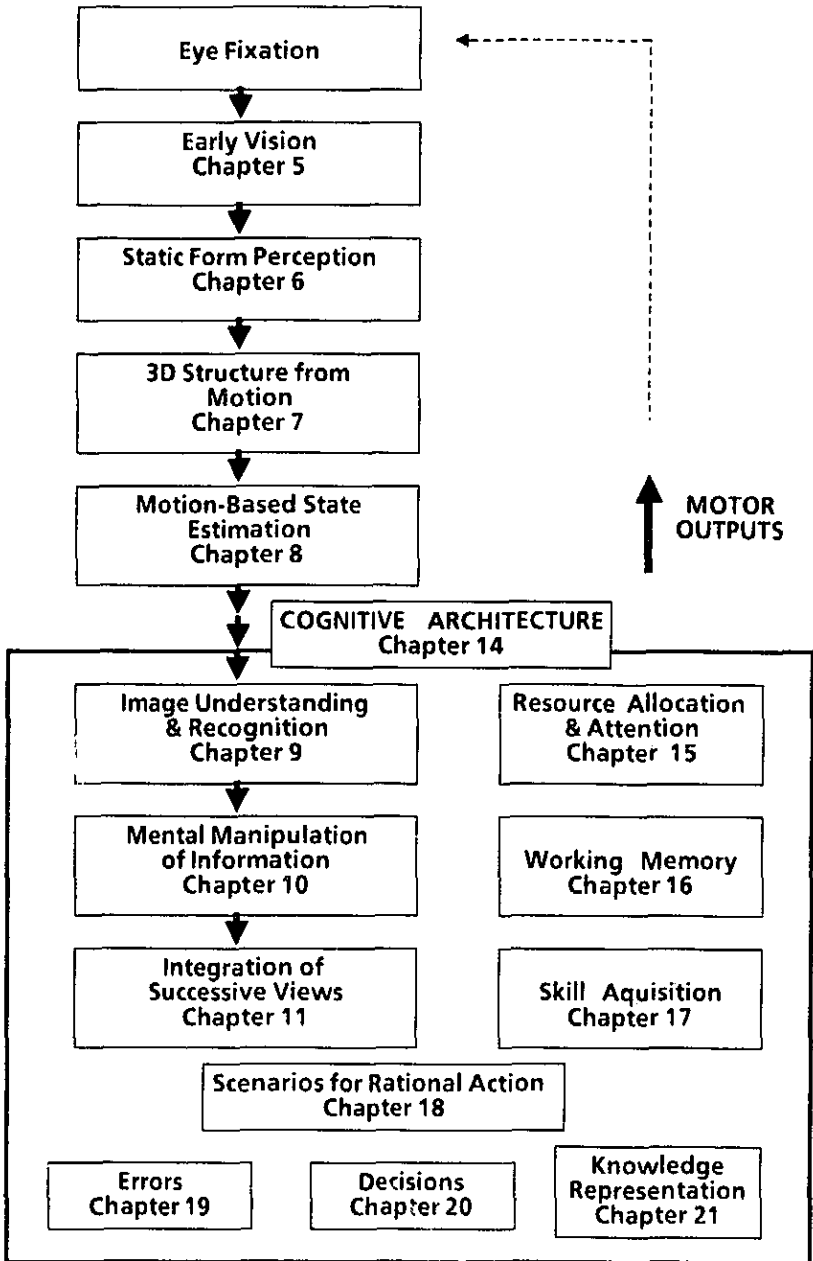


FIGURE 2-1 Framework for models of vision and cognition.

and many cognitive processes, especially those that determine what will be attended to, also influence earlier visual processes, although these effects are not shown in the framework. This linear framework has proven to be a useful way of organizing the discussion of vision even though it is clearly an oversimplification.

For cognition we lack a well-developed architecture to structure simply the flow of information and interaction among the functional components of cognitive processing. Rather, we have found it useful to differentiate between models of mechanisms of the human cognitive architecture and models of rational action. The section on cognitive models begins with a review of models for the architecture of human information processing. We then discuss several component mechanisms of the cognitive architecture, namely resource allocation and attention, working memory, and learning. The rest of the section focuses on models of rational action, first addressing models that are based on scenarios consisting of the actions the pilot is required to perform to execute a specified mission. Three other types of rational action models are treated in this section: errors, decisions, and representation of knowledge. The later stages of vision that are included within cognition belong mostly within the rational action grouping. This collection of topics does not provide complete coverage of all the cognitive functions involved in helicopter flight or even of those just dealing with the visual tasks of flight, but it is a large and important subset of those functions.

ASSESSMENT OF MODELS

The reviews of Parts II and III cover a large domain and a great number of models. A rough estimate of the number is provided by the bibliographic citations in the review papers, of which there are about 600, equally divided between vision and cognition reviews. While there is often considerable overlap among models in a functional area, it is quite clear that the designers of a human factors design facility must comprehend a very large collection of models if they are to provide complete coverage of just these two aspects of human performance. Moreover, the users of the design facility must have a significant level of understanding of the models in order to recognize their limitations and to interpret correctly the results from applying them in design. Coping with this complexity and numerology will be a challenge.

There is a strong bias in the vision reviews toward simulation models that can in principle be connected together to simulate human vision interacting with the physical environments encountered in helicopter flight. Most of these are descriptive modes, but a few are normative, such as those dealing with motion-based state estimation (Chapter 8). A considerable number of models, especially those of higher level visual processes and cognition, are taken from the artificial intelligence (AI) literature and were not developed as models of either human process or performance. Rather, they are machine (computer) implementations of functions required for constructing complete machine vision systems. They have been included in this collection of models because they represent the only currently available computational implementations of certain functions. Although there is considerable controversy on this point, one can argue that it is better, perhaps necessary, for a complete simulation of human vision to have some representation of these functions in the HF/CAE facility rather than to leave them completely unaccounted for. The psychology and AI communities have developed increased interest in investigating how well these machine implementations represent human behavior and how they and the concepts incorporated in them can be adapted to model human behavior. Examples of machine implementation models can be found in Chapters 7 (structure from motion) and 9 (real-time human image understanding).

The models in the cognitive section are more disjoint, and there is no attempt to provide a complete cognitive simulation that could interact with the physical environment. Doing so is well beyond the state of the art for most of the vision tasks confronting pilots. However, in many areas of cognition the models make close, although separate, contact with the physical or operational environment of flight. They do this at several fairly well-defined levels of abstraction or aggregation. For example, the models for scenario-based actions provide a basis for addressing problems of mission planning and feasibility by focusing on the workload that the mission imposes on the pilot. These models make very crude approximations about the human performance of individual actions, but provide very useful tools for answering high-level questions about mission alternatives and crew task assignments. At a different level of abstraction, models of resource allocation and attention use parameters of the physical environment to predict how visual fixations are distributed among instruments on a panel. Many of the models within the cognitive realm are predicated on the notion of rational action and thus are

prescriptive to some extent. Machine implementations have had a strong influence on these models of cognition, but there has been a considerable effort to fit these ideas into the framework of what is known about human performance.

In reading the following chapters, it becomes clear that a large number of models are relevant and potentially useful to advanced helicopter design. However, the models discussed have many limitations that will affect the ease with which they can be used for computer-based human factors design. The collection of models is fragmentary. Some areas are not covered by existing models, and in many areas the models that do exist have major gaps. The linkages among models are a particular source of concern. In many areas the models for one set of processes do not readily couple to models for other related processes. This makes it difficult to implement a complete simulation of either the visual subsystem or the cognitive functions associated with vision. This problem is exacerbated by the lack of a satisfactory architecture for human information processing that would provide a strong framework for integrating cognitive functions. Finally, many of the individual models and integrated subsets of models discussed have not been well validated against human performance data, and as mentioned earlier, some are not based upon human behavior but are drawn from machine implementations whose authors never aspired to model human behavior. When validation is poor or lacking, the validity of the simulation of which the model is a part and of the analysis performed with the aid of the model is open to question. Although validation is difficult enough for models of single tasks, it is an even more difficult problem in models of composite behavior. Nonetheless, in the absence of validation, doubt is cast on the correctness of analyses and designs based on the use of models.

Thus, one is led to the conclusion that a complete detailed model of human visual performance is not feasible given the current state of models. There are, however, many important questions about vision that can be answered with the aid of existing models if the focus is on simple tasks or on simplified abstractions of more complex tasks. For example, the detectability and legibility of simple targets can be estimated using models of early vision (Chapter 5), and the performance of the pilot in estimating system state variables can be predicted using the models of Chapter 8. Although not discussed in this report, there are also good models of viewability that can be used to evaluate whether or not the pilot can even view displays and

other objects in a proposed design. Some of the models discussed in Chapter 15 (resource allocation and attention) can be used to obtain useful estimates of the attentional demands on pilot vision which, in turn, can be used to answer questions about panel layout. The scenario-based mission analysis methods of Chapter 18, based on models of workload, lend themselves to interactive computer implementations and are a substantial improvement over current static techniques. The models of error, learning, and decisions provide insights about aspects of human performance important for design and, if applied with careful attention to their limitations, could be useful in a design facility.

To summarize, we are far from having a complete set of models for representing human vision and related cognition, but there are a number of important types of questions that current models would help answer. There is a reasonable expectation that integrating these models in a computer design facility could make the existing portfolio of models more accessible to designers than they are today, enable their wider use in design, and lead to improvements in the design process and the resulting designs. It would also provide the base from which more capable and complete design facilities could evolve. It could provide a driving force for extending models in directions that would make them even more useful for design. It would almost certainly raise a number of interesting theoretical questions about models, modeling, and their application. We discuss these issues more fully in the next chapter which is about integration and use.

3

Use and Integration of Models

The purpose of the human factors computer-aided engineering (HF/CAE) facility is to improve the process by which complex piloted aircraft systems are designed and, thereby, improve the quality of the designs. The basic premise underlying the HF/CAE program is that better system designs will result from enabling designers to explore more design alternatives and to evaluate these designs before constructing costly and time-consuming prototype hardware. Models such as those discussed in Part II are central to improving the quality of the evaluation process. By making models and other information and facilities more accessible to designers, the HF/CAE facility can increase the range of variables and the number of alternatives that can be explored.

When implemented, the HF/CAE facility will be a tool—one hopes a key tool—in the design process. In aircraft cockpit design, well-established and complex design processes exist into which this facility or tool must fit. These processes have evolved over many years and are unlikely to change rapidly. As a result, the HF/CAE facility must work well with the existing design processes; otherwise, it will not be accepted. In time, as it proves successful, it will lead to changes in the design processes in which it has been embedded, but these changes will come primarily from successful application of the facility and the improved designs that result from its use.

A detailed discussion of cockpit design methodology is beyond the scope of this report. To understand some of the key issues involved in the use and integration of the models discussed here, it is helpful to summarize the nature of the processes currently being used for aircraft cockpit design and give some examples of the analyses that are performed in the course of design.

DESIGN PROCESS

The design of helicopter cockpit systems is a complex process involving a large number of people representing many different disciplines and constituencies. They work in a design space that is large but has many constraints, and attempt to satisfy a large set of interacting and often contradictory requirements. Satisfying these requirements is almost never easy and is, in some cases, impossible. The amount of analysis, simulation, information, and data that must be considered in the design is great. A large number of potential designs must be explored before a final configuration is developed and adopted.

Cockpit design begins as a top-down process with an analysis of system and mission requirements and the development and analysis of mission scenarios. This leads to the identification of functions that must be performed by the system and to the successive decomposition of these functions into the procedures and then into the individual tasks that must be performed to accomplish the required mission scenarios.

In a complex system with a complicated set of requirements, the task of the designer focuses first on developing a thorough understanding of the problem that the system is supposed to solve, on the requirements themselves, and on their implications for system design. The structure of the requirements must be understood, and in particular, those requirements that critically drive the design and critically interact with other requirements must be identified. For such design problems, the goal is usually to satisfy a set of requirements, not to optimize performance because optimization is too difficult. In fact, people do not apprehend the amount of effort that designers expend to avoid and eliminate sources of catastrophe. Thus, the goal of design is often to make the system adequate without blunders.

Although, in principle, design starts out at the "top" with the analysis of requirements and missions, it does not unfold as a purely top-down process. Detailed design or analysis of lower level functions, tasks, and proposed solutions is required to determine whether higher level functional decompositions or procedural definitions are acceptable. Alternative approaches must be conceived, trial solutions developed, detailed analyses completed, and results communicated to other members of the design team so that the impact on other functions can be understood. Real design requires both top-down decomposition and bottom-up synthesis and analysis.

Most system designs are not "clean sheet" but rather start with certain critical components or elements of the design prespecified and constrained. Design features that come later must conform with the decisions already made if unacceptable penalties of cost and delay are to be avoided. In helicopter design, human performance is both critical and constrained, and it makes sense to take account of human constraints early in the design, as a reference to which other decisions must conform. This suggests a design process that is user centered, in which the support of user roles in the system is a major driver of the design. However, current design practice usually relegates definition and support of human roles to later stages of design. One of the reasons for this lack of an early focus on user roles is the lack of methods for considering user roles early in the design process (Rouse and Cody, in press). An HF/CAE facility should help remedy this.

User-centered design moves from the system and mission requirements to a characterization of the role of the crew in terms of the general tasks that it will perform. It proceeds by assessing the demands that these tasks impose upon the crew in terms of critical performance requirements and workload. Information requirements and control actions required for these tasks must be determined, and techniques for providing this information and eliciting the appropriate responses must be developed. Obstacles to the satisfactory performance of the tasks by the crew must then be identified and appropriate revisions made to the configuration. To complete a design in this manner, the designer clearly requires a strong support system.

There have been many studies of the design process and of what designers of aircraft systems actually do. Rouse and Cody (in press) found that designers spend most of their time consulting with other individuals working on their project and doing individual problem solving, analysis, and synthesis. Much of this time is spent studying and interpreting system requirements. Little time is spent consulting formal printed materials. Most information is obtained from informal contacts with people close at hand. The circumscribed nature of personal technical interactions is a well-known phenomenon (Allen, 1977). Clearly, the primary support system for current design practices is the designer's colleague group. The HF/CAE facility must be designed so that it augments, not replaces, this group in addition to providing technical tools for analysis and design.

TOOLBOX FRAMEWORK

It should be apparent from the preceding discussion that the HF/CAE facility will be used to examine pilot performance in a variety of situations and from a variety of viewpoints. The facility must accommodate many different kinds of human performance models and data. It should be constructed so as to enable a skilled designer to move flexibly through the design process, asking and answering specific questions as they arise, and to iterate previous design decisions as new information, constraints, or interactions among elements of the design become prominent.

This type of use suggests that the facility should be a framework for integrating an evolving collection of tools and data that is placed at the disposal of the designer and that, ultimately, embodies the design itself. This collection will include tools for doing simulation at several levels; for static analysis; for accessing data bases of guidelines, case studies, and behavioral data; and for conducting rapid experimentation (another application of simulation). There should also be tools for adding new models and data to the facility, both as part of the design process and as part of the process of maintaining and enhancing the facility. It is up to the designer to make good use of this collection of tools and to determine when to employ particular tools, how to use the results, and how to proceed through the design process.

Many of the tools in the collection will be devoted to understanding the mission and its operational requirements and the crew's role in meeting these requirements. These mission analysis tools should allow the designer to design prototype mission scenarios, to perform task analyses, and to determine workload as well as regions of overload or interference between modalities and tasks. Other tools in the collection will be aimed at detailed design in which the ability to construct and evaluate prototypes or simulations of prototype devices, displays, layouts, etc., is important. These detailed design tasks would benefit from access to a rich collection of simulation and analytic models of various types and to human factors data bases, all of which would help the designer assess the impact of proposed designs on user performance, user loading, and overall mission success.

The principal types of human performance evaluation tools that should be included in the HF/CAE toolbox are listed below:

- Complete Pilot Simulation Models
- Mission Level Simulation Models

Partial Simulation Models
Static Analytic Models
Guidelines, Data, Case Histories
Rapid Experimentation Facilities

Three levels of simulation models are identified in this list ranging from a complete pilot model ("megasimulation") to simulation models of different aspects of human performance. The static (non-simulation) analytic models presumably cover a wide range of human performance. Data bases consisting of guidelines, case histories, and human performance data, as well as facilities for rapid experimentation by human operators are also included.

Building one megamodel that ties together models of all relevant aspects of human performance and aspires to be a complete simulation of pilot behavior is theoretically possible. Such a model would be able to answer all human performance questions. However, it is clearly impractical and unrealistic to build such a model today. As is apparent from the discussion in Parts II and III, current models are not complete enough to support this approach. The validity of such a simulation would be limited by the weakest element in any of its components. Even if a megasimulation model could be developed, it is not clear that a design system should be based entirely upon such a model. Most people who have experience with systems that have taken this approach find them cumbersome and awkward to use. Among other things, this results from having to specify a large amount of information to use the model for even the most trivial of questions. Thus, a complete simulation model does not appear to be a practical basis for the HF/CAE facility now or in the near future.

The other types of models listed above are practicable today even though megasimulation is not. Mission level simulation models attempt to encompass the entire mission (or major segments of a mission) by using models of human and system performance at a high level of abstraction. Individual human functions are approximated either statistically or deterministically as discrete decisions or actions; and cognitive, perceptual, or motor processes are not represented. Mission level simulation models are useful for showing how a mission will unfold and for estimating its probability of success. They are also useful for performing task analysis to determine the critical parts of a mission, where the demands on the pilot are high and the sensitivity of pilot load to mission parameters is great. Mission level analysis is useful as a starting point for the analysis of pilot

performance and for identifying where more detailed analysis should be directed.

Partial simulation models have less breadth than mission level models but more depth in specific aspects of human performance. A partial simulation model attempts to represent a single aspect of human performance with enough detail so as to be useful for answering specific performance questions and carrying out detailed analysis. Because such models are simulations of human performance, they can be coupled to the external environment and run in a closed-loop mode that reveals the dynamics of interaction between a pilot and the environment, for example, vehicle flight path. As discussed in Part II, some lower levels of vision can be represented by simulation models of this type that will be used to answer questions about fixation, detection, and recognition performance. Biomechanical models, although not discussed in this report, are partial simulation models that can be used to determine the viewability and accessibility of displays and controls in a cockpit design. Similarly, control theoretical models can be used to simulate pilot control performance in a variety of situations.

For many aspects of human performance, there are no models that are complete enough or of the correct form for simulating pilot performance; however, models do exist which provide static analytic descriptions of specific types of performance. These models are useful for carrying out static analyses of specific aspects of pilot performance and for estimating parameters of that performance. Classical examples of such models are Fitts law for predicting the time to point to a target as a function of distance and size, signal detection models, and models for predicting instrument scanning patterns. Many of the models discussed in Parts II and III are of this type, and a large collection of such models reported in the literature is potentially useful to the designer.

It is also important that the computer-aided design/computer-aided engineering (CAD/CAE) cockpit design facility be the repository for description of the resulting design decisions as they are being made. This description should incorporate graphical layout and detailed design decisions expressed in graphic form, as well as a narrative rationale that permits an audit trail of the state of the design at each stage. By embodying this description in the same facility as the design tools, those models or analyses that require access to extended aspects of cockpit design will have the data available

electronically and will not have to reenter specific parameters of the design.

Often, experimental results and design principles have not been reduced to analytic form. The human factors and psychology literature contains a wealth of information in the form of data, guidelines, and case histories that is of great potential value to the designer and would be even more useful if readily accessible. These data, guidelines, and histories form an important knowledge base that can be used for making design decisions that are in accord with established practice or previous experiences and for evaluating a design to determine its consistency with guidelines and principles. The HF/CAE facility can serve an important function in providing access to such information and facilitating its use in the design and evaluation process.

Finally, a facility that contains a rich collection of simulation tools for representing the vehicle under design and simulation models of various aspects of human performance is a powerful tool for conducting rapid experiments to answer specific questions within the context of the missions for which the system is being designed. For example, pilots can be asked to fly parts of a mission and the acceptability of their performance can be measured, visual scenes can be constructed, and the detectability of a specific object can be determined.

SELECTING TOOLS AND MODELS

The HF/CAE facility will be an evolving set of tools based on a growing body of models. It is important that the initial set of tools and models be chosen well because they will have a large influence on the success of any effort to develop a design facility. The goal should be to choose an initial set of tools that will make the design process better in some important way. It is probably a good strategy to focus on improving the design process rather than on improving designs, since it is easier to see how tools change the process than it is to see how they change the designs.

In selecting the tools it is useful to think in terms of the kinds of engineering analyses that are required for a design, the questions that need to be answered in the course of these analyses, and the models that might help answer these questions. Analyses should be chosen that are important to the design, required by the design process, and difficult or time-consuming to do. Questions should be chosen that

are not easily or well answered by current design methods and for which models exist that can provide insights important to answering these questions. It is not necessary to do a perfect job with these analyses or questions, it is only necessary to do significantly better than is currently possible.

Although we are far from having a complete simulation model of human vision, the reviews in Parts II and III indicate that we have many models that do in fact provide useful insights relevant to answering a number of important design questions. Performance that depends primarily on aspects of early vision and estimation of aircraft state from two-dimensional optical flow information, and some related response in certain restricted cases, seem tractable with current models. For time sharing and workload, practical models are available. Current scenario techniques can be extended and applied to good advantage. There are approaches that could be taken to predict errors that have some limited usefulness. The models of decision making behavior have potential near-term application to system design.

This limited portfolio of models can support analyses in a number of areas. Much can be done with instrument panel layout, viewability of displays, and their visibility and legibility, and with target detection. The state estimation models, in conjunction with models of human control performance, can be used in a variety of analyses of vehicle flight control performance. The scenario techniques and workload models support a variety of analyses of mission feasibility, task analyses of crew workload and allocation of functions among the crew. Some limited error prediction analysis can be done. Finally, analysis of pilot decision performance is feasible, but care is needed to take account of the special characteristics of human decision behavior. Selection from among the set of supportable analyses should be done only with good knowledge of the practices followed by experienced cockpit design teams. They are the customers for the design facility and must be willing and interested in using it for their work.

Once the analyses that are to be supported have been chosen, it is then possible to think about the way in which a tool should be designed to support and enhance each different type of analysis. Each tool should integrate the most appropriate methods and models available for answering the questions central to this analysis. If experience from other disciplines is a guide, the initial set of tools will be rather crude and limited in the breadth and depth of analyses that they address, but over time they will improve provided they are

actually put to use. The feedback from real application is compelling and drive the evolution of the systems like this, if the initial version of the system turns out to be interesting enough to attract and reward early users.

The following section illustrates how the HF/CAE facility might be used to carry out some of the engineering analyses required to design a helicopter cockpit. In the course of so doing it gives some examples of how the tools and the types of models incorporated into the facility might be used.

ENGINEERING ANALYSES

The state of human performance models relevant to the design of a CAE workstation for helicopter cockpit design is reviewed in Parts II and III of this report. The useful incorporation of analytic models into design and engineering methodology is itself a step requiring substantial effort and insight. It is beyond the scope of this report to address design methodologies for the computer-aided engineering of cockpits; however, it is useful to sketch briefly a few possible applications of human performance models in design. Such applications give a flavor of the enterprise and emphasize the point that the selection of models may depend deeply on which factors matter greatly in design (and, therefore, must not be compromised) and which matter very little (and can, therefore, be largely approximated.)

One way to envision such engineering use is to consider the design outputs of other engineering models. During design, engineering models are often employed to perform analyses, some more or less standard, others unique to a particular question. On a CAE workstation, these analyses may be reflected in cathode-ray tube (CRT) displays as designers explore variants and "what if" questions. Eventually, the most important paths of the analyses would be included as pages in engineering design documents. Human engineering models might also be expected to lead to analyses that eventually become pages in engineering manuals and, hence, part of the technical documentation for the device being designed.

Examples of standard engineering analyses, taken from the operator's manual of an AH-64A helicopter (U.S. Army, 1984) are given in Figure 3-1. Figure 3-1(a) shows regions of danger indicated by crosshatched lines surrounding the helicopter, which are the results of various engineering analyses. Although conceptually simple, these analyses establish such factors as clearances required for successful

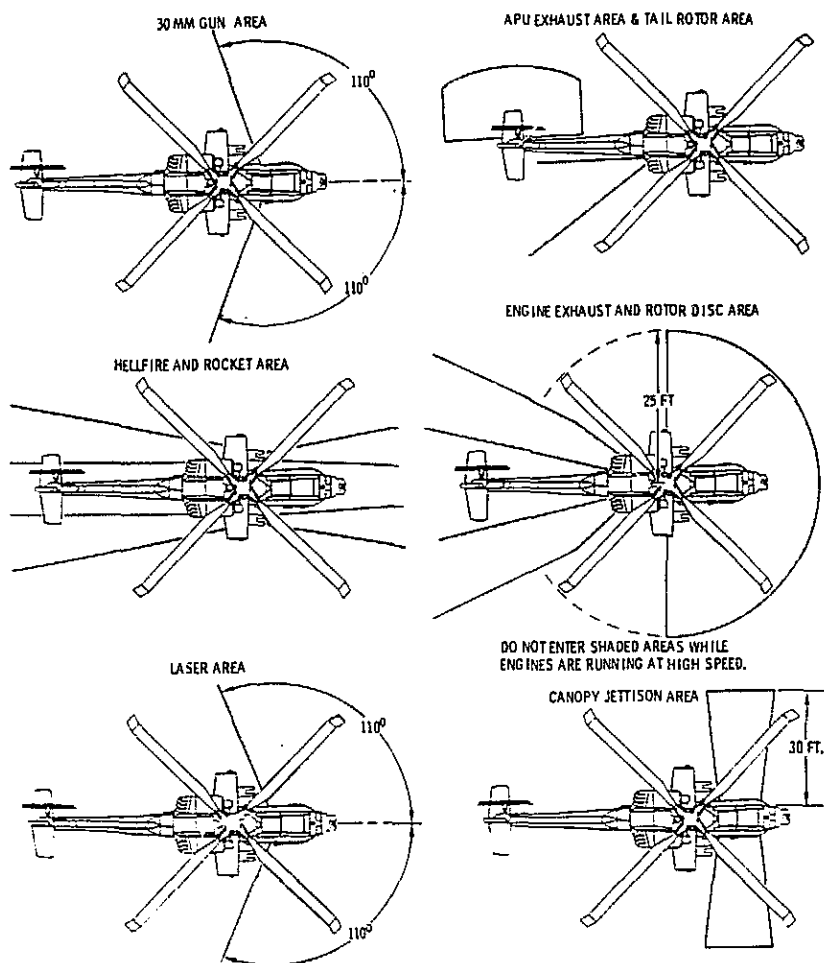


FIGURE 3-1(a) Danger areas. SOURCE: U.S. Army (1984).

canopy jettison and dangerous areas for service personnel. Figure 3-1(b) summarizes another set of engineering analyses, in this case airspeed operating limits. This diagram enables calculation of maximum airspeed, given pressure and altitude. Figure 3-1(c) is the result of an engineering analysis of the ability to land the helicopter in case of engine failure as a function of flight parameters. Figure 3-1(d) summarizes weight calculations for the helicopter's components.

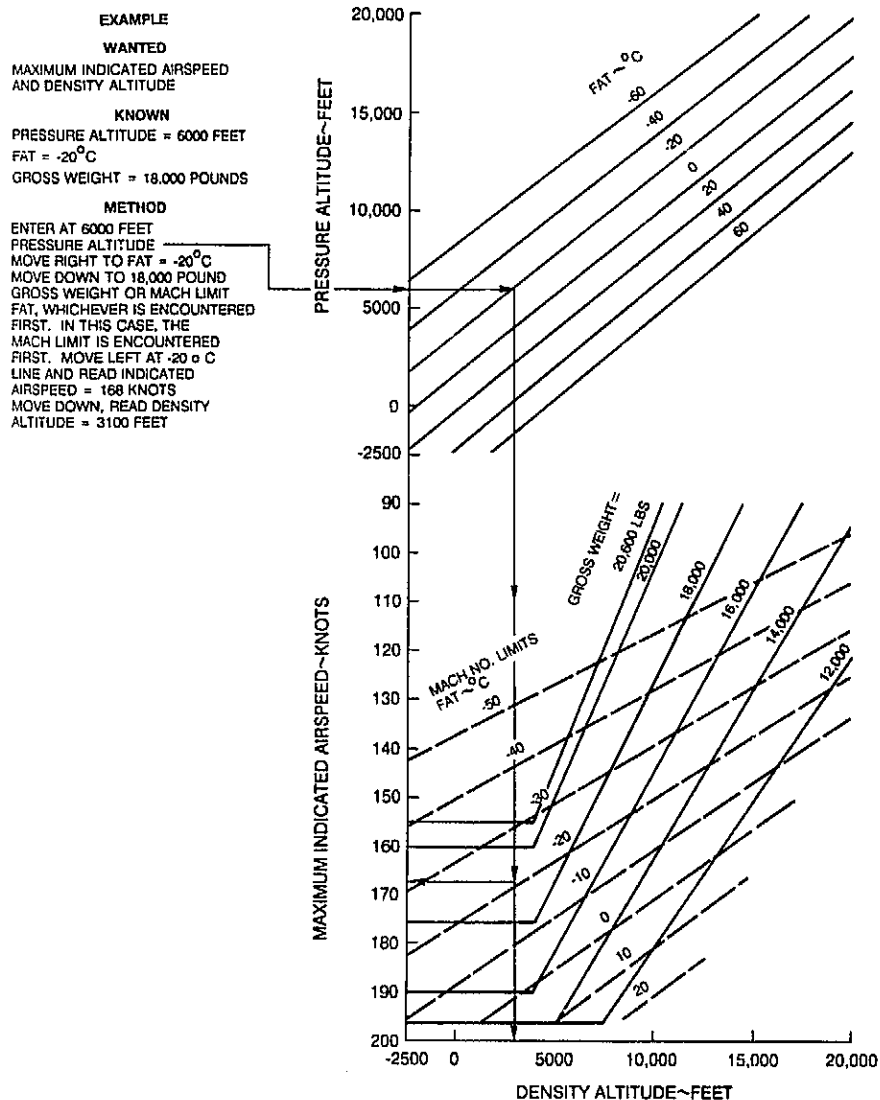


FIGURE 3-1(b) Airspeed operating limits chart (100% rotor RPM). SOURCE: U.S. Army (1984).

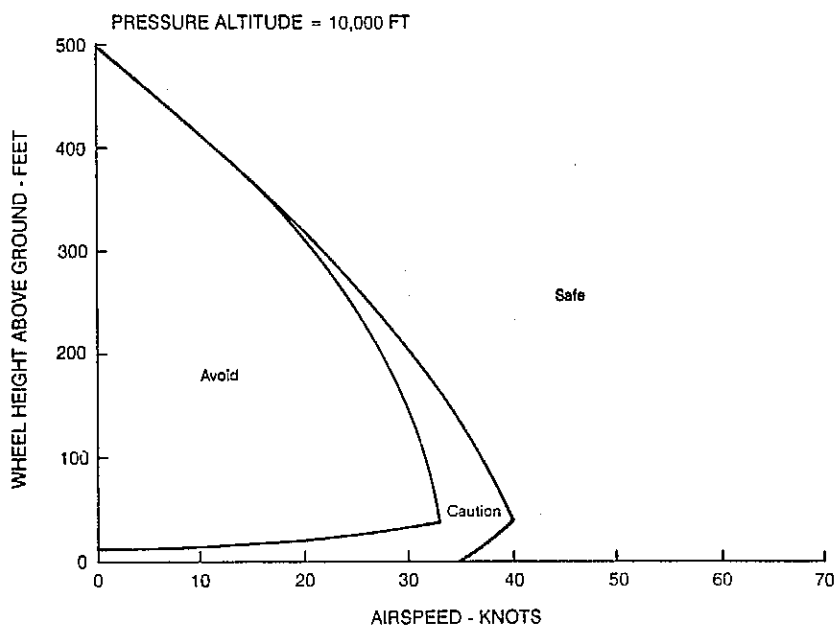
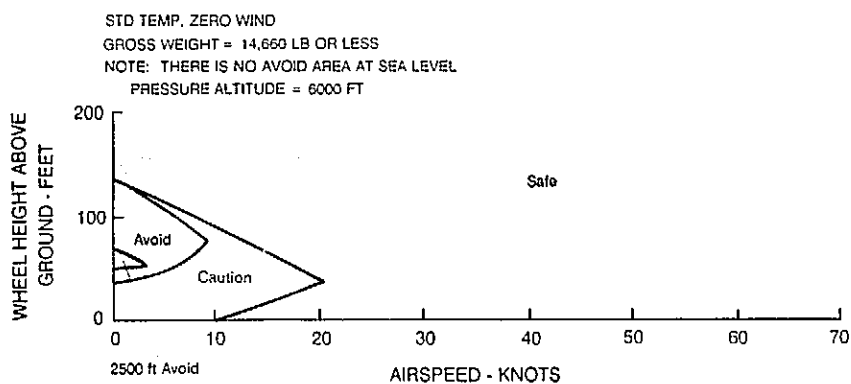


FIGURE 3-1(c) Height velocity plots. SOURCE: U.S. Army (1984).

22 November 1976

Page 1-50

MIL-STD-1374 PART 1		GROUP WEIGHT STATEMENT WEIGHT EMPTY		Page _____
Name _____				Model _____
Date _____				Report _____
1	WING GROUP			22.1.4
2	BASIC STRUCTURE CENTER SECTION			
3	INTERMEDIATE PANEL			
4	OUTER PANEL			
5	GLOVE			
6	SECONDARY STRUCTURE Bmt Wing Fold Weight 2.5 lbs			16.4
7	AIRBORNS Bmt Balance Weight lbs			22.1.4
8	FLAPS - TRAILING EDGE			15.7
9	LEADING EDGE			
10	SLATS			
11	SPOILERS			
12				
13				
14	ROTOR GROUP			22.1.5
15	BLADE ASSEMBLY			
16	HUB & HINGE Bmt Blade Fold Weight 5.0 lbs			22.1.5
17				
18				
19	TAR GROUP			22.1.6
20	BASIC & SECONDARY STRUCT - STABILIZER			
21	FPA Bmt Control			22.1.6
22	VENTRAL			
23	ELEVATOR Bmt Balance Weight lbs			22.1.6
24	RUDERS Bmt Balance Weight lbs			22.1.6
25	TAR ROTOR - BLADES			22.1.6
26	HUB & HINGE			22.1.6
27				
28	BODY GROUP			22.1.7
29	BASIC STRUCTURE FUSELAGE & HULL			
30	BOOMS			22.1.7
31	SECONDARY STRUCTURE FUSELAGE & HULL			22.1.7
32	BOOMS			22.1.7
33	SPIDERS			22.1.7
34	DOORS RAMPS PANELS & MSC			22.1.7
35				
36				
37	ALIGNING GEAR GROUP Type 1			22.1.8
38	LOCATION	Running Gear Arrest Gear	Structure	Controls
39	Wing Gear			
40	Wing Gear			
41	Wing Gear			
42	Wing Gear			
43				
44				
45	ENGINE SECTION & NACELLE GROUP			22.1.9
46	BODY INTERNAL			
47	EXTERNAL			
48	WING INBOARD			
49	OUTBOARD			
50				
51				
52	AIR INDUCTION SYSTEM			
53	DOORS PANELS & MSC			
54				
55				
56	TOTAL STRUCTURE (To Be Strapped Forward)			22.1.9

*Change in Plans & Sketch for Water Type Gear

FIGURE 3-1(d) Group weight statement. SOURCE: U.S. Army (1984).

In the spirit of these analyses, a small sample of analyses is now considered that might be informed by available or reasonably attainable human performance models.

Mission Level Scenario Generation

Many analyses depend on some method of generating pseudo behavior that can serve as a stand-in for what actual behavior would be in an operational environment. Models for scenarios and time lines are discussed in Parts II and III.

The traditional way of generating pseudo behavior is to stipulate a mission, then have the analyst imagine how the actors would behave within the constraints of the scenario and equipment. This process is so labor intensive that it is impractical to repeat for small design variants. Because of the expense of redoing the analysis, time lines are out of date with respect to changes in the system design, rendering them less useful than they otherwise might be (but see Aldrich, Szabo, and Bierbaum, 1988, for examples of where, at substantial expense, time line analyses are used to investigate major system design trade-offs).

The use of a computational environment for such analyses in a CAE workstation eases this constraint. Ideally, such a system would have modules corresponding to

- (1) external environment:
 - terrain,
 - external agents;
- (2) design:
 - methods/doctrine (procedures for accomplishing goals),
 - abstract display and control functionality,
 - crew and automation function assignments,
 - display and control methods (procedures for accomplishing goals using equipment),
 - panel layouts; and
- (3) pilot description:
 - pilot models.

A scenario would be generated automatically from a high-level mission statement (e.g., load 1000 pounds of fire-fighting equipment, take off from forest service camp A, fly through valley B, deliver equipment to fire camp C, return to base). Variants of the scenarios would be generated by making changes to the modules (e.g., changing displays or even details of terrain). A set of 100 basic scenarios,

for example, might be used for the development of a cockpit. Variations in design would be flown against these 100 scenarios, perhaps computed and analyzed overnight. Because the design itself would be altered in a CAE system by modifying its machine representation, this representation might be used, without the labor-intensive operations now generally required, as input to the scenario analysis. In this way the analyses would stay current with the design.

An automatic scenario generation system would require a sophisticated planning model (see Part III). Currently, this is probably feasible only for simplified situations, such as air-to-air combat among two or a few aircraft, which has apparently been done in the AASPEN system. On the other hand, it probably is feasible to improve upon rigid time lines by using a model of the GOMS sort, as described by Corker, Davis, Papazian, and Pew (1986) (see Chapters 15 and 18). Figure 3-2 is a fragment of an analysis page showing methods and doctrine that might be part of a typical mission analysis. This analysis would be used to generate actions for a time line, rather than their being generated directly by hand. The same method might also be applied to linking external mission level tasks to the abstract display and control functionality, as well as to the display and control methods for actually reading displays and manipulating controls.

Time-Line Analyses

Figure 3-3 shows a possible time line generated from the operators in Figure 3-1. For each task the time line specifies a set of four user-defined vectors (eventually these would be stored in a table look up). These vectors are

1. task priority—based on an expected value calculation, by borrowing directly from algorithms included in the PROCURU model;
2. opportunity window—a duration of time within which the task could be rescheduled if required (see below);
3. estimated completion time for discrete tasks; and
4. demand level—a vector quantity for each task that may be borrowed directly (initially) from data obtained by McCracken and Aldrich (1984).

This time line generates useful information that is the input to several other analyses. In a CAE system, the time line itself could be examined by using the class of tools often associated with

```

EVADE-MISSILE-FIRE
. JINK-OR-TURN-TAIL
.. JINK-AND-HIDE
... CHOOSE-JINK-DIRECTION
... JINK
.... WAIT-FOR-JINK
.... FLY-CONTOUR-AND-STOP
... STABILIZE-CRAFT
... HIDE
.... GO-TO-ELEVATION-AND-STOP
.... STABILIZE-CRAFT
.. TURN-TAIL-AND-HIDE
... TURN-TAIL
... STABILIZE-CRAFT
... HIDE
.... GO-TO-ELEVATION-AND-STOP
.... STABILIZE-CRAFT
. LEAVE-EVALUATE-DECISION
.. EXIT-FIRING-POSITION
... SELECT-EXIT-PATH
... NOE-AND-STOP
.. EVALUATE-DAMAGE
. CONTINUE-MISSION-DECISION
.. RETURN-TO-BASE
... CHOOSE-RETURN-PATH
... COORDINATE-AND-RETURN
.... COORDINATE-AND-RETURN-SEQUENCE
..... GIVE-EGRESS-COMMAND
..... SEND-RADIO-MESSAGE
..... PLAN-AND-RETURN
..... AWAIT-MESSAGE
.... NOE-APPROACH
..... NOE-ADJUSTMENT

```

FIGURE 3-2 Operator/suboperator summary of mission level methods.

displays an analysis of the time line as a tree on its side. The height of each box is proportional to the percentage of time used in that operation. To simplify the diagram, only those operations that use 5 percent or more of the time are shown. Users can expand each node on the graph to get a subanalysis. The display can be set to count the number of operation invocations instead of time, workload, or

SEQUENCE	TASK
1	CHOOSE-JINK-DIRECTION
2	WAIT-FOR-JINK
3	FLY-CONTOUR-AND-STOP
4	STABILIZE-CRAFT
5	TURN-TAIL
6	STABILIZE-CRAFT
7	GO-TO-ELEVATION-AND-STOP
8	STABILIZE-CRAFT




FIGURE 3-3 Possible time line generated from operators in Figure 3-1.

memory load or to cumulate the information in various ways. Users are, therefore, able to dynamically explore where system bottlenecks exist and to adjust the design and rerun the analysis to compare the differences.

Workload Analysis

The structure of the workload analysis is based upon combining particular features of the PROCURU, human operator simulator (HOS), Siegal and Wolf, and workload index (WINDEX) models described in Part III. In particular, the analysis makes the distinction between the sequential-scheduling aspects of multiple task environments and the concurrent-parallel aspects of those environments. The principal input is a time line analysis, along with various user-defined parameters described below, whereas its outputs are

- a workload profile over time, which may be used to gauge overall mission difficulty and assess the workload reduction resulting from training or automation; and

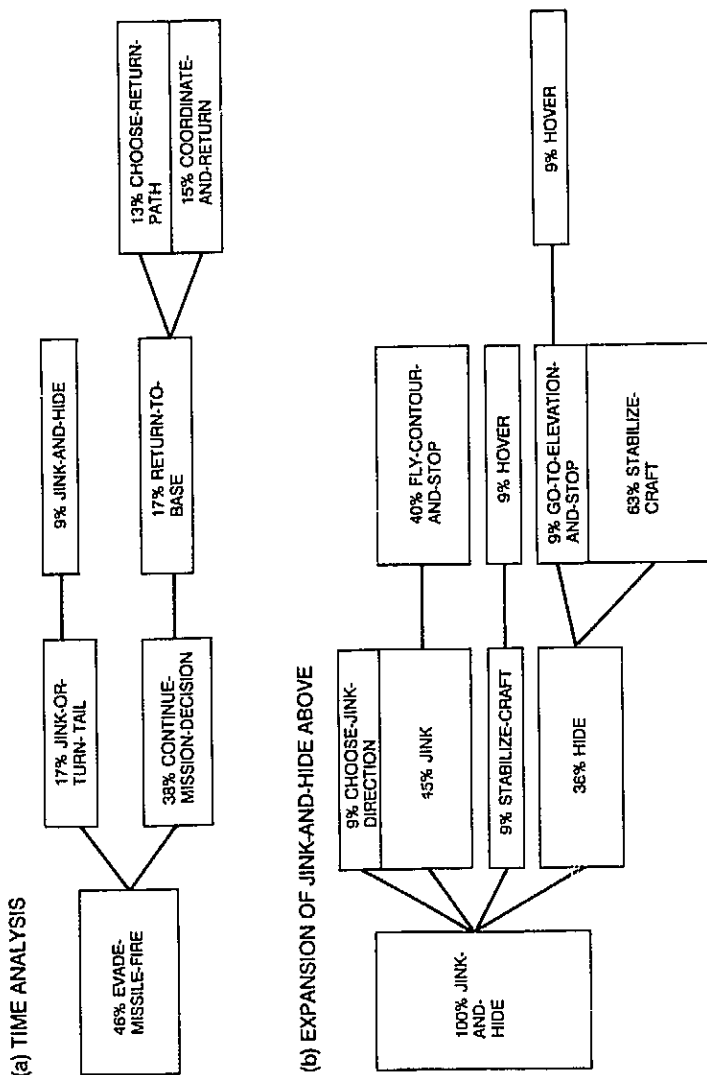


FIGURE 3-4 Analysis of time usage from time line. (a) Part of analysis generated automatically by system. Analyst has used workstation to expand one of the boxes for more detailed analysis in (b).

- specific performance measures on some tasks.

Workload Analysis Model (WLAM) Structure

The structure of the WLAM is shown in Figure 3-5. A task time line, which could be required as input for analysis, was described earlier. The demand vector can be modified in important ways as suggested below. The demand level vector has two important components:

1. the processing resource structures demanded by the task—an entry is required in at least one of the columns of Figure 3-6 (this is a modification of North's 1985 WINDEX model); and
2. the demands for resources within each channel—these demands can change from 0 to maximum (e.g., 10) as a function of the task and the task characteristics. Thus, for example, the demands of helicopter flight control will increase in the visual scene channel from hovering in clear visibility (2-3) to hovering over featureless terrain at twilight (6-8). The demands of continuous manual control will increase with turbulence level.

The demand levels of all tasks to be performed concurrently are input to a WINDEX-based workload computation (see below) whose output is a scalar value of workload (WL) computed at one point in time. This value is compared against a "maximum workload" criterion (WL_m). If $WL < WL_m$, the situation moves to the next time point and WL is recomputed. If $WL > WL_m$ (workload is excessive), then rescheduling is carried out. This logic simulates an operator's strategy of task shedding when demands become excessive. All tasks are checked according to their priority levels, and those of lowest priority are abandoned and placed in a task queue.

Tasks in the queue are then joined by new tasks on the time line, and these must compete with each other for reentry to the workload matrix. The highest priority task in the queue will enter the matrix if

- it has higher priority than tasks already in the queue, or
- the workload computed with its inclusion does not exceed WL_m . Discrete tasks leave the queue after their completion.

Task priorities may be governed by (1) user-defined baseline values (e.g., stability control has a higher priority than communications) and (2) time passage (e.g., a postponed task may gain priority with the passage of time). This gain can be modeled by a function that

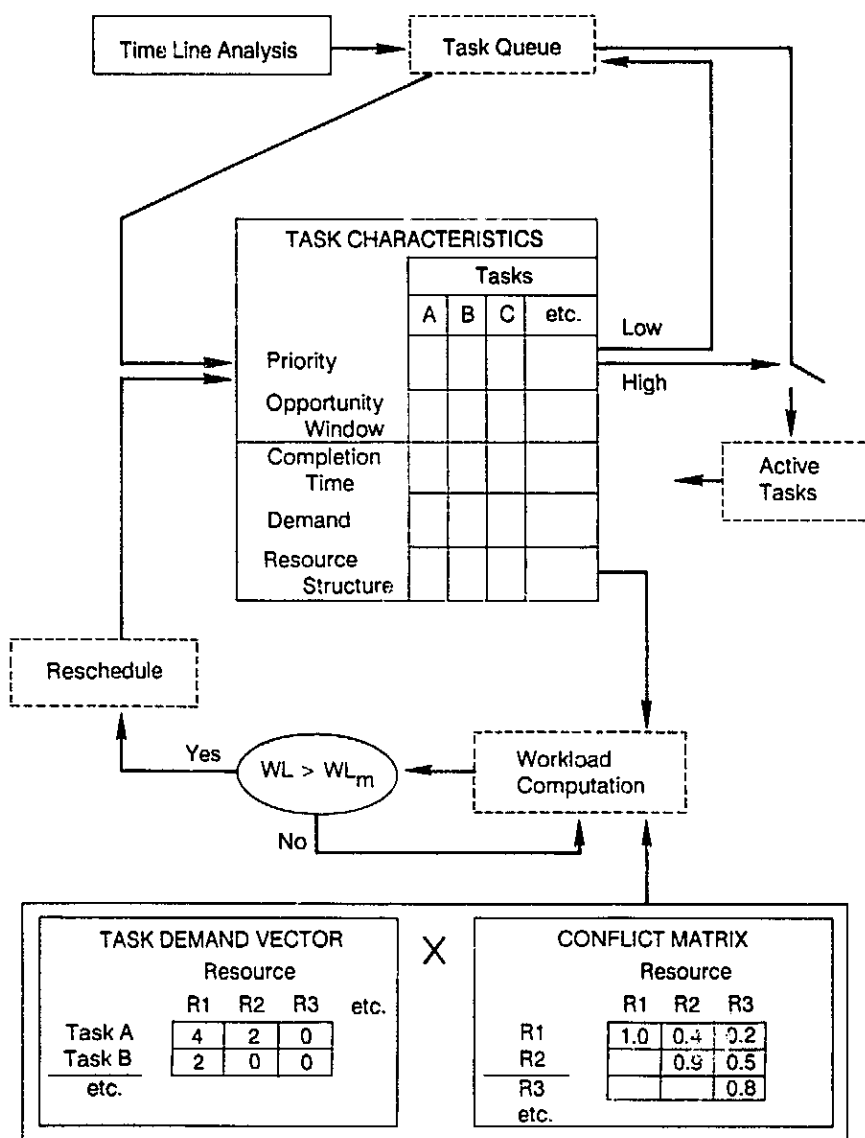


FIGURE 3-5 Workload analysis logic.

	CHANNEL								ABBREVIATIONS		
	Visual		Auditory		Cognitive		Response		W	World	Visual
	I		V		S		K		I	Instruments	Auditory
	V		Vr		Vr		St		V	Voice	Cognitive
	V		Vr		Vr		St		K	Keyboard	Response
	V		Vr		Vr		St		St	Stick	Response
	V		Vr		Vr		St		V	Voice	Response
	V		Vr		Vr		St		V	Voice	Response
Task A											
Task B											
Task C											
Task D											
Task E											

FIGURE 3-6 Workload channels by task matrix.

increases linearly from the baseline value to a maximum value until the opportunity window is passed, then it is reset to zero.

The passage of time may also influence the demand level for certain tasks. For example, responding to a request for data entry will increase in demand as the time passes because of the increased working memory load (or decreased reliability) of the material over time.

Workload Computation

Any task can be identified by the set of demand values in Figure 3-6. The interference of this task performed concurrently with a second one is calculated by summing the demand values within each column and multiplying (or adding) each sum to which both tasks contribute to a resource conflict value. Examples of values, shown in Figure 3-7, range from 0 to 1 (1 to 10 if addition is used) and are based upon assumptions from multiple-resource theory. For example, it heavily penalizes two tasks that may compete for common processing stages (manual data entry while controlling), codes (voice control while rehearsing communications information), or display modalities (requirement to target search while map reading). This computation is carried out across all nonzero cells of the 8 by 8 matrix, and workload is set as the sum across these cells. The aggregate conflict value may be thought of as a penalty that is subtracted, in a manner inversely proportional to the priority value, from the performance of each task in a pair.

Continuous Tasks

It is clear that stability/flight path control will be a continuous entry in the task matrix (except as replaced by autopilot). The modeler should probably also be aware that planning is a continuous task as well as one that is modulated over time according to the depth of planning.

Outputs

In addition to the workload analysis, it is possible to make more specific predictions of task performance. These include measures of task delay for discrete tasks and are equal to the service time plus time spent in the queue. Service time itself may be modified by workload calculations. It may be lengthened in inverse proportion to

CHANNEL									
Visual			Auditory		Cognitive			Response	
W	I	V	S	Vr	K	St	V		
0.7	1.0	0.4	0.6	0.3	0.3	0.3	0.1	World	
	0.9	0.6	0.6	0.5	0.5	0.3	0.2	Instruments	
		1.0	0.5	0.9	0.4	0.2	0.9	Voice	
			0.8	0.8	0.6	0.6	0.3	Spatial	
				0.9	0.3	0.3	0.8	Verbal	
					0.9	0.8	0.7	Keyboard	
						0.7	0.4	Stick	
							1.0	Voice	
								Response	

FIGURE 3-7 Resource conflict matrix. The precise values require experimental validation.

the resources allocated. The output may also include degradations in the quality of performance (e.g., loss in flight control resolution or reduction of expected accuracy level of discrete task). Each task in competition for resources required for other concurrently performed tasks will be penalized proportionally to (1) the amount of resource competition and (2) its priority value relative to the competition. Thus, two time-shared tasks of equal priority will suffer equally if they suffer at all.

For visual tasks, the percentage of resources allocated to visual channels may be a fundamental parameter passed to the visual performance models. Techniques from optimal control models can be used to derive a signal-to-noise ratio for resolution of these visual inputs as a function of the resources allocated.

Model Simplification

The workload analysis model may be simplified for exercise in any of a number of directions. First, it may be made into an open-loop model by breaking the feedback loop after the workload computation in Figure 3-5. Hence, no scheduling or prioritizing logic would be employed other than that which is inherently built into the fixed time line provided as input. Second, assumptions regarding changing priorities or demand levels with the passage of time can be abandoned. Third, assumptions regarding the resource competition between concurrent tasks can be simplified along any number of lines, as suggested by the discussions in Part III. In particular, the number of resources or channels assumed to modulate task competition can be reduced from the eight shown in Figures 3-6 and 3-7 to one. In this case there is no conflict matrix, and demand values can simply be added across tasks. An example of a two-level vector might be one that assigns task resources to one of two categories: perceptual-cognitive or response.

At this point it appears that the model is relatively modular, so simplifications of one sort do not distort the operation of other components of the model.

Model Exercise

Rather than fully describing the exercise for each of the two problems—communications and pop-up—a brief description is presented of some of the implications of the model for the two problems performed concurrently. First, during preparation for unmasking, planning activity would be particularly heavy, thereby imposing conflict with concurrent tasks to the extent that the latter are perceptual-cognitive (e.g., a penalty would be applied to understanding communications). Additional high penalties would be imposed on continuous manual control if a stable hover was required in turbulent conditions with small margin for deviation (e.g., among the trees). This demand would penalize heavily any tasks requiring keyboard data entry. After unmasking, heavy resources are demanded by the task of visual scanning, which interfere extensively with the visual aspects of flight stabilization (maintaining position and altitude over ground). These perceptual-cognitive demands will not however, greatly disrupt response tasks such as voice output (e.g., reporting targets) or keyboard data entry. Disruption of the keyboard task should be reduced further if voice, rather than keyboard, is used for data entry. If a secondary perceptual or cognitive task is imposed at this time (e.g., determining fuel status or dealing with an instrument advisory), performance of this task would be postponed until the workload of one or both of the higher-priority tasks of flight path control and target identification were reduced below criterion level. Perceptual resource demands of target acquisition would be dictated by measures of scene complexity and target-background similarity (e.g., feature overlap). These measures should be provided by parameters passed from the visual models. Cognitive resource demands of this task would be governed by measures of target identity and location uncertainty, as well as by the number of relevant targets to be located. Quantitative demands of communications would be linearly related to message length and working memory load.

The quantitative value of the fraction of resources allocated to flight path control and to target detection would be passed back to the visual models.

Display Layout Analysis

The pilot's usual strategy for scanning an instrument display is driven directly by his information needs. Important information

channels or those delivering high bandwidth information will be fixated frequently, whereas rapid transmissions may be observed between pairs of information channels that are associated with highly cross-correlated information (e.g., rate of climb and altitude). Studies by Fitts and his colleagues (Fitts, Jones, and Milton, 1950) and by Senders (1964, 1983) have confirmed these assertions. Furthermore, human engineering applications of these conclusions by McRuer, Jex, Clement, and Graham (1967) have demonstrated that display layouts which are guided by analysis of fixation frequency and transition probability can result in improved pilot-vehicle performance. Quite simply, information sources that are fixated frequently should be located near the center or top of the display. Those between which transitions occur frequently should be located in close proximity to each other. The concerns for close spatial proximity are guided not so much by the time required for visual scanning as by cognitive organizational factors related to confusion of display elements and target search. Hence, the design guidelines are equally applicable to the design of heads-up displays (HUDs) and helmet-mounted displays in which actual eye movement is less of a concern. However, it should be noted that peripheral motion and guidance information may not suffer from the constraints of visual scanning.

Besides fixation frequency and transition probability, six additional constraints must be considered in the analysis of display layout, particularly for helicopter design.

(1) The requirement to scan instruments is unlikely to replace outside-the-cockpit viewing. Hence, primary concern must focus on the view outside as the primary flight instrument.

(2) Clustering of instruments in terms of system organization facilitates interpretability. Thus, displays pertaining to the same physical system, or the same functional system, should be displayed contiguously (Goodstein, 1981). Although only two dimensions of physical space are available to define contiguity, these may be augmented by the use of color codes that define physical or functional similarity.

(3) Optimal scanning patterns may differ between normal system functioning and system abnormality. During the former, operators will sample from one each of a cluster of correlated instruments, because sampling from other members of the same cluster provides redundant information. During failure, however, operators will be more likely to sample sequentially within a cluster (Moray, 1986).

(4) Display organization and clustering should also be guided by stimulus-response compatibility. Hence, displays that provide information relevant to left-handed controls should be positioned to the left of those providing information relevant to right-handed controls (Hartzell, Dunbar, Beveridge, and Cortilla, 1983).

(5) Some success at reducing the number of separate displays to be scanned can be accomplished through object integration in which two or more dimensions of quantitative or categorical information are represented as dimensions of a single object (Barnett and Wickens, 1988).

(6) When display space is at a premium, computer-callable displays—although sometimes necessary—should be incorporated with considerable caution (Moray, 1981). Replacement of valuable physical real estate by logical circuitry to make displays callable on command will add the perceptual-motor (or speech) demands necessary to call up the particular displays and will increase potential memory and cognitive loads associated with knowing where one is in a menu structure.

The steps necessary to accomplish this analysis are outlined in Figure 3-8 and proceed as follows. From the time line analysis, an information analysis is produced that provides a second-by-second profile of the information necessary to perform the tasks. The N channels along which such information may be displayed can be placed in three N by N matrices that represent three different dimensions of what is called "task proximity," shown on the right side of Figure 3-8:

(1) Correlational proximity is based on the product moment correlations between state values sampled within a four-second window.

(2) Functional proximity is based on the model user's decision of the extent to which two indicators must be integrated/compared in performing a task (Boles and Wickens, 1987).

(3) Physical proximity is based on the similarity between the physical sources of the two indicators of each pair of displayed sources. Thus two indicators of rotor functioning are more similar than one of rotor functioning and one of navigational functioning.

In addition to these three task proximity matrices, each channel is associated with a value representing

- (4) Frequency of use,
- (5) Proximity to the windscreen, and

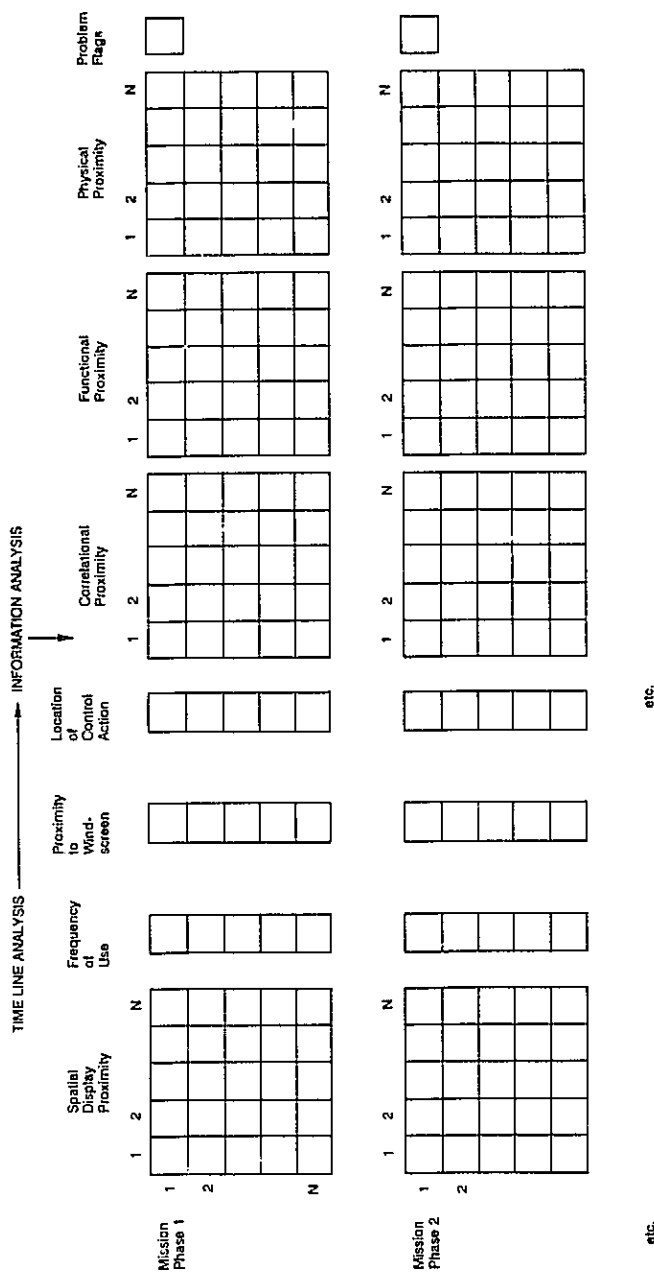


FIGURE 3-8 Cluster and frequency display and control analysis.

(6) Associated relevant control actions.

The final matrix is defined by (7) the spatial proximity between pairs of displays in a particular configuration. Display proximity may be measured in centimeters. Alternatively, it should probably be measured by the number of intervening displays between relevant pairs (hence, adjacent displays would be assigned a value of zero). Two displays configured as dimensions of a single object would also have a display proximity of zero.

These data will not, by themselves, dictate an organizational format that minimizes the distance (correlational, functional, physical, and responsive) between all related pairs of instruments, although in theory they could be made to do so. However, criterion values of distance along any combination of the three distance metrics (functional, physical, and correlational) can be set, and a cockpit configuration that is generated by designer's intuition can be checked against these criteria to establish if the physical distance between any particular pair violates the maximum distance criterion. (A pair of instruments might be said to have this violation if they have a task distance less than X , but are located with a physical distance separation greater than Y). Similar criteria can be set for violations of stimulus-response compatibility or excessive distance of high information displays from the outside-the-cockpit view. These criteria may be weighted by the frequency of information use. This scheme will allow the designer to alter the design in response to severe violations of proximity and allow reconfiguration through rapid prototyping.

Naturally, a tool of this sort is only as effective as the information analysis that provides input to it and the talent of the designer or expert who codes correlational, physical, and functional proximity. Analysis output then would consist of listing all pairs of displays that violate user-defined proximity criteria and all single displays that violate (1) stimulus-response compatibility and (2) frequency criteria for distance from the visual window.

DISCUSSION

Different demands are placed on a model of human vision or cognition when it is used for engineering analysis than when it is used by a scientist to fully characterize a visual or cognitive process. To characterize a process or mechanism, a model should be "deep" but usually need not be wide, because scientific models are typically

concerned with small, subtle effects. They are intended to explicate a mechanism that can be identified by an empirical signature (observable phenomena), even if the empirical signature is small. It is occasionally even useful to fit model parameters backward from the data the model is to explain. By contrast, for engineering use, models that are broadly applicable and robust are required. They must generally be applied from an analysis of the situation they are to model, with no fitting of parameters (zero-parameter predictions). As Woods and Roth (1986, p. 29) in their study of models for nuclear power plant operators say,

In part, the integration of heterogeneous concepts to model a complex domain represents a heuristic to deal with a tradeoff between the formal, applicable, and scope dimensions of models. In general for the behavioral sciences, the more formal a model, the narrower the coverage of and applicability to real world tasks.

Approximate models that trade precision for broad applicability are often appropriate here, but it should be noted that the validity of a model is logically prior to its ease of application. Easily applicable, but wrong, models are still wrong and may be worse than no model at all.

AFTERWORD

The suggestion has been made that the HF/CAE facility be considered as a framework for a set of tools to design helicopter and, presumably, other aircraft cockpits. In developing this framework it is important to provide for evolutionary growth of the facility and to foster the acceptance of tools and models from many sources. There is the opportunity to use the HF/CAE facility as a vehicle for stimulating the development, evaluation, and refinement of tools and models for design by a large community of students, researchers, and practitioners in the human factors and aircraft design field. To do so, the facility must be "open" in that its interfaces and programming conventions should be available to groups outside of NASA. If this is done, there is a good chance to build a large community of contributors that collectively will help develop the facility into a widely useful tool.

Attention is focused in this report on models for use in answering questions about human performance relevant to helicopter design. In building the facility, the focus can be on design models or design questions. Although this report is clearly directed toward models, it is

probably more important for the design of the facility to concentrate on developing tools that help answer design questions. These tools will call on models that have been discussed, but it is the answers that are needed. The models are a means to this end. The questions will lead to the selection and prioritization of the models that should be incorporated into the facility.

Finally, it is well known that designers have been reluctant to use human performance models, possibly because these models are unfamiliar to them. Familiarity now depends heavily upon personal or at least colleague group knowledge. To introduce a new collection of tools like those to be incorporated in the HF/CAE facility into a design community requires that careful attention be paid to methods for promoting acceptance by that community in a reasonable time.

REFERENCES

- Aldrich, T.B., Szabo, S.M., and Bierbaum, C.R.
1988 The development and application of models to predict operator workload during system design. In G. MacMillan, ed., *Human Performance Models*. NATO AGARD Symposium, Orlando, FL.
- Allen, T.T.
1977 *Managing the Flow of Technology: Technology Transfer and the Dissemination of Technological Information within the R&D Organization*. Cambridge, MA: The MIT Press.
- Barnett, B.J., and Wickens, C.
1988 Display proximity in multicue information integration: The benefit of boxes. *Human Factors* 30:15-24.
- Boles, D., and Wickens, C.
1987 Display formatting in information integration and nonintegration tasks. *Human Factors* 29:395-406.
- Corker, K., Davis, L., Papazian, B., and Pew, R.
1986 *Development of an Advanced Task Analysis Methodology and Demonstration for Army-NASA Aircrew/Aircraft Integration* Report No. 6124. Cambridge, MA: Bolt Beranek and Newman, Inc.
- Fitts, P.M., Jones, R.E., and Milton, J.L.
1950 Eye movements of aircraft pilots during instrument landing approaches. *Aeronautical Engineering Review* 9:1-5.
- Goodstein, L.P.
1981 Summary of workshop discussion. In J. Rasmussen and W.B. Rouse, eds., *Human Detection and Diagnosis of System Failures*. New York: Plenum.
- Hartzell, E.J., Dunbar, S., Beveridge, R., and Cortilla, R.
1983 Helicopter pilot response latency as a function of the spatial arrangement of instruments and controls. *Proceedings of the 18th Annual Conference on Manual Control*. Dayton, OH: Wright-Patterson Air Force Base.

- McCracken, J., and Aldrich, T.B.
1984 *Analysis of Selected LHX Mission Functions*. Technical Note ASI 479-024-84(B). CA: Anacapa Sciences, Inc.
- McRuer, D.T., Jex, H.R., Clement, W.F., and Graham, D.
1967 *Development of a Systems Analysis Theory of Manual Control Displays*. TR-163-1. Systems Technology, Inc.
- Moray, N.
1981 Feedback and the control of skilled behavior. In D.H. Holding, ed., *Human Skills*. London: John Wiley and Sons, Ltd.
1986 Monitoring behavior and supervisory control. In K.R. Boff, L. Kaufman, and J.P. Thomas, eds., *Handbook of Perception and Human Performance*. New York: John Wiley and Sons.
- North, R.
1985 WINDEX: A workload index for interactive crew station evaluation. *Proceedings NAECON*. New York: IEEE.
- Rouse, W.B., and Cody, W.
in press Designer's criteria for choosing human performance models. In C.R. McMillan, ed., *Applications of Human Performance to Systems Design*. New York: Plenum.
- Senders, J.W.
1964 The human operator as a monitor and controller of multi-degree of freedom systems. *IEEE Transactions on Human Factors in Electronics* HFE-5:2-5.
1983 *Visual Scanning Processes*. Netherlands: University of Tilburg Press.
- U.S. Army
1984 *Operators Manual for Army AH-64A Helicopter*. Technical Manual TM 55-1520-238-10. Headquarters, Department of the Army.
- Woods, D., and Roth, E.
1986 *Models of Cognitive Behavior in Nuclear Power Plant Personnel*. NUREG/CR-4532. Washington, D.C.: U.S. Nuclear Regulatory Commission.

Part II

4

Introduction to Vision Models

At least four requirements must be met in order to develop a computer-based model of the sensory performance of a pilot. First, it must be possible to make quantitative or definite predictions of sensory responses to measured input characteristics for appropriate situations. This is a challenging task and one which daunts even the most intrepid computer scientist who aspires to devising a machine vision system equivalent to human perceptual performance.

Second, it must be possible to provide the computer with direct physical measures of sensory input as opposed to encodings that can only be made by a human observer. Third, it must be possible to provide the components of the predictive system in compatible computer algorithms. Fourth, the model should, in principle, be image driven, which means that it should be able to respond to the physical characteristics of the visual displays that confront the hypothetical pilot. This requirement engages some of the most difficult problems faced by computer scientists in their attempts to devise machine vision systems that achieve results equivalent to human perceptual performance. The question of empirical or principled psychophysical prediction of human perceptual behavior is, therefore, only one of the two major problems, actively investigated but by no means completely solved, on which the efficacy of pilot modeling must depend.

A great deal of quantitative and qualitative psychophysical information about human vision now exists that should be accessible to designers. However, this information does not yet meet the four requirements for a computer-based model of sensory performance. Thus, one cannot do justice to such information in this report, noting only where appropriate that it exists (Boff and Lincoln, 1986). In the attempt to evaluate whether computerized pilot performance

modeling is presently feasible, the chapters in Part II focus on what seem to be critical aspects of visual input and perception.

The chapters in Part II move in a rough progression: from early vision; through form perception, three-dimensional structure through motion, flight estimation and object perception from flow patterns, object recognition, and mental manipulation of objects; to a final combination of views. These chapters offer a substantial, but incomplete, sample of the visual tasks for which human performance models are needed that can be implemented now or in the near future. These models should be able to predict the information a pilot retrieves from the visual environment and hand these predictions on for further processing by decision and performance models.

Models are identified and referenced in all the chapters, but only Watson and Zacharias (see Chapters 5 and 8) seem comfortable with the proposition that existing models now or soon can be used to simulate pilot visual performance in the domains with which their chapters are concerned.

The models on early vision in Chapter 5 (as opposed to handbook wisdom, no matter how quantitative) are in better shape than the models in any of the other chapters. The kinds of questions addressed are indicated by the chapters' subheadings. Although models that might be image driven fare better in early vision than in the more perceptual areas, much remains to be done. Thus, although some models described therein could be used in simulation in which the attempts at performance modeling ask special questions about target visibility or the legibility of specific signals or symbols, in most cases these models would have to be queried not by the cognitive group's questions but through intervening higher perceptual questions. Also, important gaps exist. For example, there is as yet no explicit bridge between models of two-dimensional velocity detection and the extraction of three-dimensional structure from such two-dimensional velocity fields, a central issue in pilot performance. Again, although attention and expectation are admittedly important in at least such tasks as wire detection, these factors have not yet been embodied in attempts to apply early vision models to specific problems.

Todd and Braunstein (Chapters 6 and 7) are pessimistic about the use of any currently implementable models, calling for new models and for systematic gathering of the data on which to base them. Todd, discussing shading cues of form, argues for the construction of expert systems on vision that would guide human factors work now and provide a more solid base for future models. Hochberg

(see Chapter 11) agrees with this suggestion. Braunstein, reviewing attempts to recover three-dimensional structure from moving two-dimensional displays, holds that no current models will work and a way is needed to make present knowledge, as well as new research, more accessible to the designer and to human factors practitioners.

On the other hand, Zacharias (see Chapter 8), dealing with the extraction of flight state estimation from the same two-dimensional displays and using essentially the same set of models, believes that those models should be employed and presents a schematic system for the simulation effort. The disagreement is primarily one of emphasis, rather than of analysis or even evaluation. Zacharias, like the other authors, repeatedly notes the incompleteness of the models, the serious constraints on the conditions in which they can be applied, and most important, the paucity of data in which predictions from the models are validated against human performance. (Indeed, few such validations are mentioned.) However, he argues that the nap of the earth (NOE) situation provides stimulus arrays to which some models may well be applicable (see Model Applications and Limitations section in Chapter 8, page 119) within limits that will impose cautions on their use.

In Chapter 9, Biederman presents a compelling account of object perception—recognition-by-components—outlining material that should be of great importance to the human factors of NOE missions, but a great deal of experimental and theoretical work remains to be done. There are no currently implementable models that recognize objects as well as humans do or by similar processes. However, this has recently become an area of intense activity, and implementations by Biederman and his colleagues represent promising developments. One difference between many of the machine vision models and human performance is that humans are very much affected by orientation in the aircraft but only moderately affected by rotation in depth. Processing time and effort are needed for recognizing inverted objects, so that mental manipulation of visual information becomes important to object recognition and to navigation. That is the topic of Chapter 10 by Cooper. Although some models are mentioned there, they are poor candidates for human performance model development at this time.

Chapter 11, by Hochberg, outlines two sets of problems that arise, particularly with regard to the artificial displays currently used in NOE flying: (1) viewers must combine successive partial views of scenes, presented piecemeal in displays, into coherent schematic

representations of scenes, events, or objects; and (2) the views of two eyes when disparate, may combine in stereoscopic combination, in alternating dominance, or in piecemeal rivalry. Although no models suitable for simulating pilot performance currently exist for either problem area, it does not seem impossible to achieve models for limited aspects of these problems (i.e., those aspects that belong most properly in early vision).

5

Models in Early Vision

ANDREW B. WATSON

OVERVIEW

Early vision refers to those stages of vision that involve the capture, preprocessing, and coding of visual information, but do not involve interpretation or other cognitive processing of visual information. A number of models of parts of early vision are reviewed here: temporal dynamics, spatial processing, and motion processing.

For present purposes, a model is defined as a simulation of some physical system, typically as a set of mathematical expressions or computer programs, that produces explicit predictions. For purposes of comparison, models may be rated according to breadth, depth, and accuracy, as well as whether they predict competence or performance. A competence model describes how a task is done, whereas a performance model describes how well the task is accomplished, without necessarily indicating how it is done. Models may also be distinguished by their degree of validation, their implementation, the nature of their inputs and outputs, their domain of operation, restrictions on their operation, and their applications.

In the domain of spatial vision, a number of models have been implemented which are reasonably broad and accurate, but shallow. Most are concerned only with detection and discrimination, and then only of rather specific simple targets. However, they provide the basis for a fairly general and competent model of visibility, that is, of what can and cannot be seen. The generality of these models could be increased by a more thorough treatment of masking. At the level of coding or representation, there are a number of interesting and plausible approaches, but little of this work has been validated by experiment.

In the temporal domain, current models generally predict the visibility of temporal fluctuations in luminance or contrast. The models are highly developed, accurate, and relatively well validated, but shallow and narrow in domain. Integration with spatial models, motion models, and models of light adaptation would considerably extend their domain and utility.

As in other domains that have been considered, models that predict the visibility of moving signals are well developed and do not pose serious implementation problems. Models of higher-level estimation, including several that estimate local velocity at several scales, have been implemented but are more speculative. Nonetheless, they may be of considerable value. They already incorporate the visibility aspect, as well as many known properties of human motion sensing. They are thus more than simple models of competence, although less than complete models of performance.

The perceptual process can be partitioned into three segments: filtering, coding, and interpretation. Filtering determines what information is captured and what is lost, either within the total system or within a particular stream or channel. Coding describes how specific visual mechanisms represent particular components of visual information. Interpretation describes how the coded information—perhaps from numerous sources, including memory—is used to determine the state of objects in the visible world.

The models reviewed in this chapter deal largely with the filtering stage, slightly with the coding stage, and hardly at all with interpretation. The models that inspire the most confidence are clearly those at the earliest stages. Indeed, there is no obstacle to the creation of a fairly comprehensive model of visibility that would incorporate spatial, temporal, and motion sensitivities, as well as the effects of mean luminance and location in the visual field. Work on the later stages is vigorous, but there are currently no convincing models of coding and interpretation.

INTRODUCTION

A sad consequence of the expanding knowledge of human vision is the increasing compartmentalization of vision science. Early vision has come to refer to those stages that involve the capture, preprocessing, and perhaps the coding of visual information, but do not include interpretation or other cognitive processes. Fortunately, the precise border between early and late vision is not of great consequence.

This chapter considers a number of models of parts of early vision, spatial processing, temporal sensitivity, and motion processing. It begins with a discussion of models *per se*: what they are, how they may be integrated, and how they are related to simulation.

WHAT IS A MODEL?

The word "model" has many definitions, and several are to be found even within the pages of this report. For the purposes of this chapter, a model is defined as a simulation of some physical system. Although this simulation might take mechanical or electronic form, it is typically a set of mathematical expressions or computer programs. A defining characteristic, however, is that it produces explicit outcomes. This, therefore, excludes qualitative, intuitive, or purely conceptual descriptions of a process.

Beyond this, models can be distinguished along many dimensions (Watson, 1987c). How large a piece of reality do they encompass in breadth (one receptor versus the complete set of receptors) and in depth (ranging from the optics of the eye to behavioral performance) and with what accuracy do they mimic that reality? How explicit is the model? Are models of both competence and performance of interest? Vision science has numerous modest, ad hoc models of small components of performance, mostly in a form less explicit than computer code. Because these model fragments are not likely to be useful for simulating interesting segments of reality, this chapter is confined to a few explicit models of sizable parts of the system.

MODEL ATTRIBUTES

To get a better grasp of the capacities and limitations of existing vision models it is useful to determine the following attributes for each model:

- **Validation:** Are there data demonstrating agreement between the model and human performance, or the superiority of the model over other models?
- **Implementation:** Does the model exist in the form of computer programs? If not, could the programs be developed easily?
- **Input:** Distinction is made between models that are *image driven* and those that are *parameter driven*. In the former, the input is an image or sequence of digital images derived, for example, from a camera and digitizer, and the output is some prediction of human

performance relative to those images. In the latter, the output may be the same but the input is some set of parameters, for example, the coordinates of points or the amplitude and phase of a sinusoidal grating. While the latter sort of model may be useful, it begs the question of how the visual system derives the parameters from the image data. This relates to the issue of generality of input. An image input is quite general, whereas a model accepting an amplitude and phase as an input has no natural way of treating a natural image. For each model then, one may ask: Is it image driven or parameter driven? What are the parameters? As noted, a model typically simulates some piece of the chain from sensation to action. As one moves further along this chain, it becomes less and less likely that the input can be drawn directly from the physical environment (i.e., be image driven). This means that these later models depend critically on the assumptions regarding their input, which is some internal state not known to exist.

- **Outputs:** Outputs can be represented in terms either of human performance in a well-specified task or of some observer knowledge of the observed world. A disadvantage of the latter is that it requires an additional step to actually predict performance, whereas a disadvantage of the former is that it cannot predict any task other than the one for which it was designed.

- **Restrictions:** What, if any, are the restrictions on the application of the model, beyond those implicit in the characterization of the inputs and outputs?

- **Applications:** How might the model be used to simulate pilot performance? Although this general question is considered in more detail in the section on integration, models can play a role as a component in some larger integrated simulation of the pilot or as a discrete simulation of some isolated fragment of performance.

- **Domain:** It is useful to categorize the various models of low-level vision in terms of the primary input variables with which they are concerned: temporal, spatial, and motion. For each domain, this chapter presents a sequence of models, usually proceeding upward in terms of the complexity of dimensionality of the inputs and outputs.

SPATIAL VISION

A spatial model can be defined as that which proceeds from an input defined primarily in spatial terms (e.g., as a static luminance image), to some human performance relative to that input or to

some estimates of the spatial configuration of surfaces, textures, or objects. Such models may be useful for predicting the visibility of information and for describing human representation of visual spatial information. Models of the earliest stages of spatial vision are concerned primarily with the detection of luminance contrast and the discrimination of simple spatial imagery.

Wilson and colleagues have developed a series of models for detection and discrimination of spatial patterns (Wilson and Bergen, 1979; Wilson and Gelb, 1984). The essence of these models is a set of sensors with specific spatial and temporal weighting functions (receptive fields), and a specific nonlinear output function for each receptive field. There are a number of different sizes of receptive fields, all of which grow with increasing distance from the fovea. Inputs are usually one-dimensional continuous luminance patterns (e.g., vertical lines or gratings), and output is a small set of numbers (between approximately 4 and 150, depending on the model) that are the sensor responses. Rules are given for converting these numbers into performance on various tasks, such as detection and discrimination of various patterns. The models predict¹ a wide variety of data, such as contrast sensitivity, effects of frequency adaptation, and frequency discrimination. There has been little independent validation of the models. Shortcomings of these models include the following: (1) they operate only on one-dimensional images, (2) there is no general scheme for predicting performance from sensor responses, and (3) more complex tasks would require a more complete specification of the sensor set, because the small number of sensors defined clearly does not capture all the visual information (Nielsen and Wandell, 1986).

Burbeck and Kelly (1980) predict thresholds for sinusoids in space and time by means of a filter that is characterized in both spatial and temporal dimensions. No mechanism is provided for extending the predictions to arbitrary targets. Because this is a "single-channel" model, it cannot predict second-order effectors due to multiple channels (Watson, 1982).

Carlson and Cohen (1980) have a model designed primarily to predict the visibility of artifacts (such as blur and aliasing) in television displays. The input is a one-dimensional image decomposed into several bands of spatial frequency which are then perturbed by noise,

¹ Here and elsewhere in the text, the phrase "the model predicts" is used to mean "the model generates a prediction, which may or may not be correct."

squared, and integrated. Later modifications introduce change in spatial scale with eccentricity (Carlson and Klopfenstein, 1985). The model has been applied with some success to predict one-dimensional hyperacuity thresholds (thresholds for visual elements smaller than the dimensions of a human cone).

Klein and Levi (1985) have created a model to interpret hyperacuity thresholds. It accepts one-dimensional spatial waveforms and generates a space-frequency diagram or viewprint which, with further processing, predicts certain detection and discrimination thresholds. Like Wilson's models, it incorporates multiple sizes of receptive fields and a nonlinear contrast-response function. A distinctive feature of the model is that some phase information is discarded by combining the response energies of odd and even receptive fields. Shortcomings of the model are (1) it operates only in one dimension, and (2) there is no general scheme for predicting performance from sensor responses.

Watt and Morgan (1985) have developed a model that transforms a one-dimensional spatial waveform into an ordered list of "primitives," such as regions of signed response and inactivity. These primitives have been related in somewhat indirect ways to human performance in detection and discrimination of contrast and blur. There are difficulties in extending the model to two dimensions (Watt, 1987). It seems doubtful that the primitives suggested provide a complete description of the visible image information. Programs exist for this model.

Geisler and Davila (1985) have developed a model of detection and discrimination based on an ideal observer and the known properties of the visual optics and receptors. It can predict detection and discrimination of arbitrary foveal two-dimensional spatial patterns. With additional assumptions, it can predict color discriminations. Its predictions often agree in form with human performance but typically differ in absolute sensitivity by about 1.5 log units. Programs for this model are available (Geisler, 1987). Because the model deals only with losses of information at the very earliest stages of vision prior to the electrical response of the receptors, it cannot predict phenomena due to losses of information later in the system. It is nevertheless a powerful and general first approximation to a description of human visual sensitivity.

Watson and colleagues (Ahumada and Watson, 1985; Nielsen, Watson, and Ahumada, 1985; Watson, 1983), have a model which accepts arbitrary two-dimensional spatial images and transforms them to an internal feature vector of many thousands of elements, which

are then acted on by an uncertain ideal observer. Each element of the feature vector is the response of a sensor with a Gabor-shaped receptive field. The model predicts contrast detection thresholds and discrimination thresholds for arbitrary two-dimensional images placed anywhere in the visual field. It has a modest amount of validation and has been implemented on various machines. A simplified version has been used as the basis of a scheme for image coding (Watson 1987a,b). Shortcomings are that the model (1) functions only at threshold; (2) is large and cumbersome; and (3) like the model of Geisler and Davilla (1985), fails to predict "higher-level" discrimination (Nielsen et al., 1985).

Whereas the preceding models are concerned primarily with detection and discrimination, a number of models have been advanced that purport to describe the coding or representational properties of early spatial vision. Examples are the zero-crossing representation proposed by Marr and Hildreth (1980), the MIRAGE model of Watt and Morgan (1985) mentioned above, the CORTEX transform of Watson (1987a,b), and the boundary contour and feature contour systems described by Grossberg (1987). All of these schemes are somewhat speculative at this point, and none has compelling evidence in its favor. Nevertheless, if they were to be made more explicit and linked more closely to performance, and perhaps to physiology, a clearer picture in this area may emerge. One general difficulty is that the visual system is not a serial sequence of processing modules but rather several parallel streams. Since each stream may require a different model, it is essential to know which performance is due to which stream.

There is also a large body of work in the computer vision literature on "early vision," that deals with feature detection and representational schemes, but little of this work relates directly to a model of human performance. The work is nevertheless a useful source of ideas concerning the computational and functional aspects of early vision.

Prospects

At the level of early vision, models are expected to provide predictions of simple detection and discrimination, and perhaps some indication of how the spatial image is represented internally or, put another way, how primitive image properties are estimated. Most of the models considered in this review are concerned simply with

detection and discrimination, and only with rather specific targets. However, they provide the basis for a fairly general and competent model of *visibility*, that is, of what can and cannot be seen. At the level of representation, little work has been closely tied to human performance, but this is an active area of research that will also benefit from synergy with research on artificial vision.

TEMPORAL SENSITIVITY

The temporal nature of a stimulus has important effects on visibility and discriminability, and models in this area attempt to account for these effects. Input is typically a continuous time waveform specifying the brightness or contrast of an image, and output is the detection threshold for that waveform. Internally, most models have the form of a linear filter or filters, whose parameters may depend in nonlinear ways on the adapting luminance, followed by some point nonlinearity, further integration, and a threshold (De Lange, 1952; Fourtes and Hodgkin, 1964; Kelly, 1961; Rashbass, 1970; Sperling and Sondhi, 1968). A review of this early work is provided by Watson (1986).

Kelly (1971a,b) has introduced refinements that allow the spatial configuration of the target to control the amount of low-frequency attenuation. Roufs (1972) has developed a quite complete analytic formulation, whose parameters are estimated from extensive data on thresholds for pulses and sinusoids at various adapting luminances. Watson (1979, 1986; Miller, 1984) has emphasized the role of probability summation over time and has attempted to test his model against a wide range of aperiodic waveforms.

The visibility of temporal signals is not separable from their spatial configuration, so that purely temporal models are of limited practical use. Several efforts have been made to combine both spatial and temporal models of visibility. As noted above, some of the models have parameters that are controlled by spatial configuration. Burbeck and Kelly (1980) have derived a spatial temporal filter that is reported to account for thresholds for spatial temporal sinusoids. Watson, Ahumada, and Farrell (1986) have shown how a very simple first-order model of spatial temporal visibility can be derived by assuming approximate separability.

Most of the simple temporal models are available in explicit form, usually as mathematical expressions but occasionally as computer programs (Watson, 1986). They are "image driven" in the sense

that their input is luminance or contrast over time. The author is aware of no fully implemented, image-driven computer programs that include both spatial dimensions, the time dimension, and the effects of adapting luminance. However, the principles for constructing such a program are clear.

Prospects

Models of temporal sensitivity are highly developed, accurate, and relatively well validated, but narrow in domain. Integration with spatial models and with models of light adaptation would extend their usefulness considerably.

MOTION PROCESSING

A model of early motion sensing is defined as one that proceeds from the visual input to some human performance or to estimates of the three-dimensional motion parameters (and confidence measures) of objects in the recent visual field, and of the self relative to those objects. No existing models satisfy this definition completely, but many address aspects of it. In particular, there are models of motion detection, of one-dimensional direction estimation, and of two-dimensional velocity estimation.

At the earliest level, models of motion detection exist (Burbeck and Kelly, 1980; Watson, 1986). These allow one to compute the probability of detection of spatial temporal perturbations in luminance (see preceding section). As such, they are not specifically "motion" models, but, nonetheless, serve to predict the visibility of moving images. They do not estimate motion parameters. Each suffers from various restrictions, and neither has been fully implemented; but each could be expanded, generalized, and implemented without extraordinary effort. Both have a substantial amount of empirical validation.

At the next level are models of one-dimensional direction estimation (i.e., discriminating one of two possible directions of a moving pattern) (Adelson and Bergen, 1985; Van Santen and Sperling, 1984, 1985; Watson and Ahumada, 1983). Typically these are models of a single sensor, which accept an input with one spatial and one temporal dimension. The spatial temporal receptive field of the sensor is arranged so as to respond to only one-dimensional direction of motion. Thus a pair of units, tuned for opposite directions, predicts

the apparent one-dimensional direction of the image at one location. A distinctive feature of these models is that each sensor is tuned for a band of spatial frequency, so that the motion sensing is carried out at several scales in parallel. The models have some validation and either have been implemented or could be with moderate effort.

Next are models that estimate two-dimensional velocity fields over a space-time image (i.e., discriminating the two-dimensional direction of a moving pattern from any possible direction) (Heeger, 1987; Watson and Ahumada, 1985). Input is a sequence of discrete two-dimensional images, output is several velocity flow-field sequences, one at each of several spatial scales. Each vector within a flow-field is an estimate of the two-dimensional velocity of image components at a particular resolution and location. Both models operate by first applying a set of local sensors tuned for different directions and then resolving the set of responses into a single estimate of local velocity. The models are based on several well-validated aspects of human visual function, but neither has much validation. The models have been implemented.

Another model in this general class is that of Marr and Ullman (1981), which computes approximate direction at the locations of edges. The basic algorithm is based on a comparison of spatial and temporal gradients (Fennema and Thompson, 1979), which does not fare well at motion discontinuities or textures (Kearney, Thompson, and Boley, 1987) and does not agree with the spatial frequency tuning of human perception. It has been implemented by Batalia and Ullman (1979), but there is little published validation.

Beyond this point, models do not usually begin at the image level, but rather at the level of defined points or contours. For example, there are the various algorithms that derive, from a set of corresponding two-dimensional projected points in several frames, the three-dimensional structure and motion of the objects on which the points lie (see Chapters 7 and 8). These models come largely from the machine vision literature and are often concerned only tangentially with human performance.

Prospects

As in the other domains, models that predict the visibility of moving signals are quite well developed and do not pose serious implementation problems. Models of higher-level estimation are much more speculative. Nonetheless, they may be of considerable value.

They already incorporate the visibility aspect and were designed to simulate at least some of the evident properties of human vision. They are thus more than simple models of competence, although less than complete models of performance. Future models may benefit from rapid advances in the knowledge of the physiology of motion pathways (Emerson, Citron, Vaughn, and Klein, 1987; Movshon, Adelson, Gizzi, and Newsome, 1986). Finally, there are efforts underway to link these low-level models to higher-level estimates of three-dimensional object motion (Zacharias, Caglayan, and Sinacori, 1985).

SUMMARY

It is useful to partition early vision into three processes: filtering, coding, and interpretation. Filtering determines what information is captured and what is lost, either within the total system or within a particular stream or channel. Examples are the spatial filter expressed in the contrast sensitivity function, the temporal filter expressed in the temporal contrast sensitivity function, and the spectral luminosity function that describes how well each wavelength contributes to luminance. Coding describes how specific visual mechanisms represent particular components of visual information. For example, a motion sensor may represent the velocity at a particular location. Interpretation describes how the coded information—perhaps from many sources including memory—is used to deduce the state of objects in the visible world.

The models reviewed here deal primarily with the filtering stage, rather than with coding or interpretation. The models that inspire the most confidence are those of the earliest stages. Indeed, no obstacle exists to the creation of a fairly comprehensive model of visibility incorporating spatial, temporal, and motion sensitivities, and the effects of mean luminance and location in the visual field.

REFERENCES

- Adelson, E., and Bergen, J.
1985 Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A* 2:284-299.
- Ahumada, Jr., A., and Watson, A.
1985 Equivalent noise model for contrast detection and discrimination. *Journal of the Optical Society of America A* 2:1133-1139.

- Batalia, J., and Ullman, S.
1979 Motion detection and analysis. Pp. 69-75 in L.S. Bauman, ed., *Image Understanding*. Arlington, Virginia: Science Applications.
- Burbeck, C., and Kelly, D.
1980 Spatiotemporal characteristics of visual mechanisms: Excitatory-inhibitory model. *Journal of the Optical Society of America A* 70:1121-1126.
- Carlson, C., and Cohen, R.
1980 A simple psychophysical model for predicting the visibility of displayed information. *Proceedings of the Society for Information Display* 21:229-245.
- Carlson, C., and Klopfenstein, R.
1985 Spatial-frequency model for hyperacuity. *Journal of the Optical Society of America A* 2:1747-1751.
- De Lange, H.
1952 Experiments on flicker and some calculations on an electrical analogue of the foveal systems. *Physica* 18:935-950.
- Emerson, R., Citron, M., Vaughn, W., and Klein, S.
1987 Nonlinear directionally selective subunits in complex cells of cat striate cortex. *Journal of Neurophysiology* 58:33-65.
- Fennema, C.L., and Thompson, W.B.
1979 Velocity determination in scenes containing several moving objects. *Computer Graphics and Image Processing* 9:301-315.
- Fourtes, M., and Hodgkin, A.
1964 Changes in the time scale and sensitivity in the ommatidia of limulus. *Journal of Physiology* 172:239-263.
- Geisler, W.
1987 Software package SDE. *Vision Research* 27:1.
- Geisler, W., and Davila, K.
1985 Ideal discriminators in spatial vision: Two-point stimuli. *Journal of the Optical Society of America A* 2:1483-1497.
- Grossberg, S.
1987 Cortical dynamics of three-dimensional form, color, and brightness perception: I. Monocular theory. *Perception and Psychophysics* 41:87-116.
- Heeger, D.J.
1987 Model for the Extraction of Image Flow. *Journal of the Optical Society of America A* 4:1455-1471.
- Kearney, J.K., Thompson, W.B., and Boley, D.L.
1987 Optical flow estimation: An error analysis of gradient-based methods with local optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-9:229-244.
- Kelly, D.
1961 Visual responses to time-dependent stimuli. II. Single-channel model of the photopic visual system. *Journal of the Optical Society of America A* 51:747-754.
1971a Theory of flicker and transient responses. I. Uniform fields. *Journal of the Optical Society of America A* 61:537-546.
1971b Theory of flicker and transient responses. II. Counterphase gratings. *Journal of the Optical Society of America A* 61:632-640.

- Klein, S., and Levi, D.
1985 Hyperacuity thresholds of 1 second: Theoretical predictions and empirical validation. *Journal of the Optical Society of America* 2:1170-1190.
- Marr, D., and Hildreth, E.
1980 Theory of edge detection. *Proceedings of the Royal Society of London B*:207:187-217.
- Marr, D., and Ullman, S.
1981 Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society of London B*:211:151-180.
- Miller, L.
1984 Predictions of Observers' Contrast Sensitivity Responses to Various Types of Temporal Waveforms. M.S. Thesis, San Jose State University.
- Movshon, J., Adelson, E., Gizzi, M., and Newsome, W.
1986 The analysis of moving visual patterns. *Experimental Brain Research Supplement* 11:117-152.
- Nielsen, K., and Wandell, B.
1986 Analysis of the linear component of spatial models. *Investigative Ophthalmology and Visual Science* 27:226.
- Nielsen, K., Watson, A., and Ahumada, Jr., A.
1985 Application of a computable model of human spatial vision to phase discrimination. *Journal of the Optical Society of America A* 2:1600-1606.
- Rashbass, C.
1970 The visibility of transient changes of luminance. *Journal of Physiology* 210:165-186.
- Roufs, J.
1972 Dynamic properties of vision. II. Theoretical relationships between flicker and flash thresholds. *Vision Research* 12:279-292.
- Sperling, G., and Sondhi, M.
1968 Model for visual luminance discrimination and flicker detection. *Journal of the Optical Society of America A* 58:1133-1145.
- Van Santen, J., and Sperling, G.
1984 A temporal covariance model of human motion perception. *Journal of the Optical Society of America A* 1:451-473.
1985 Elaborated Reichardt detectors. *Journal of the Optical Society of America A* 2:300-321.
- Watson, A.
1979 Probability summation over time. *Vision Research* 19:515-522.
1982 Summation of grating patches indicates many types of detector at one retinal location. *Vision Research* 22:17-25
1983 Detection and recognition of simple spatial forms. In O. Braddick and A. Sleight, eds., *Physical and Biological Processing of Images*. Berlin: Springer-Verlag.
1986 Temporal sensitivity. In K. Boff, L. Kaufman, and J. Thomas, eds., *Handbook of Perception and Human Performance*. New York: Wiley.
1987a Efficiency of an image code based on human vision. *Journal of the Optical Society of America A* 4:2401-2417.

- 1987b The cortex transform: Rapid computation of simulated neural images. *Computer Vision, Graphics, and Image Processing* 39:311-327.
- 1987c The role of vision models in human factors. *Night Vision*. Committee on Vision, National Research Council. Washington, D.C.: National Academy Press.
- Watson, A., and Ahumada, Jr., A.
- 1983 A look at motion in the frequency domain. In J. Tsotsos, ed., *Motion: Perception and Representation*. New York: Association for Computing Machinery.
- 1985 Model of human visual-motion sensing. *Journal of the Optical Society of America A* 2:322-342.
- Watson, A., Ahumada, Jr., A., and Farrell, J.
- 1986 Window of visibility: Psychophysical theory of fidelity in time-sampled visual motion displays. *Journal of the Optical Society of America A* 3:300-307.
- Watt, R.
- 1987 Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus. *Journal of the Optical Society of America A* 4:2006-2021.
- Watt, R., and Morgan, M.
- 1985 A theory of the primitive spatial code in human vision. *Vision Research* 25:1661-1674.
- Wilson, H., and Bergen, J.
- 1979 A four mechanisms model for threshold spatial vision. *Vision Research* 19:19-33.
- Wilson, H., and Gelb, D.
- 1984 Modified line-element theory for spatial frequency and width discrimination. *Journal of the Optical Society of America A* 1:124-131.
- Zacharias, G., Caglayan, A., and Sinacori, J.
- 1985 Visual cuing model for terrain-following applications. *Journal of Guidance* 8:201-207.

6

Models of Static Form Perception

JAMES T. TODD

The purpose of this chapter is to review current theoretical analyses of how human observers (e.g., pilots) are able to determine the three-dimensional structures of objects and surfaces in the surrounding environment from statically presented patterns of light intensity. The process of image formation is considered first, that is to say, the way in which the reflection of light by surfaces in the environment produces a structured pattern of stimulation at the point of observation. Existing methods for analyzing different aspects of optical structure are then examined, with careful attention paid to the assumptions about image formation that must be satisfied for these analyses to perform as advertised. The discussion is organized according to the complexity of optical structure being analyzed. First, perceptions of shape and surface quality from smooth variations in image shading are discussed. Next, the detection of abrupt discontinuities in shading and the manner in which they must be organized and categorized are considered. Finally, current models are reviewed for the analysis of surface shape from different types of image discontinuities including texture, reflectance contours, occlusion contours, and the edges of plane-faced polyhedra.

IMAGE GENERATION

The amount of light that reflects from a surface in any given direction depends on a variety of physical variables including the orientation, roughness, and chemical composition of the surface, as well as the positions and spectral compositions of the sources of illumination. Most recent analyses of image formation (Blinn, 1977; Blinn and Newell, 1976; Cook and Torrance, 1981; Kay and Greenberg, 1979; Phong, 1975; Whitted, 1980) model the reflection of light

as a linear combination of two separate components, referred to as diffuse and specular reflection. The diffuse component of reflection refers to light that is scattered equally in all directions. It originates from multiple surface reflections on a rough surface or from internal scattering when the incident light is able to penetrate beneath the surface. The intensity of diffusely reflected light at any given point of observation depends on the surface albedo and the angle of illumination. However, because the reflected light is scattered in all directions, its intensity is independent of viewing position. The specular component represents the highlights produced by the mirrorlike properties of shiny surfaces in which reflected light is concentrated in a particular direction. The image intensity of a specular surface will vary with viewing position and can be modeled for a variety of surface materials by using the Beckman distribution function (Beckman and Spizzichino, 1963).

Analysis of image shading becomes considerably more complex when the environment is cluttered with many different objects because the amount of light reflected from one surface can be influenced dramatically by the presence of another. One such effect is the appearance of cast shadows which occur when light rays headed toward a visible surface are occluded by an opaque object. A related phenomenon occurs when transparent surfaces are observed. For example, consider the pattern of image shading produced by a pool of clear water. Some light rays are reflected from the surface of the pool. Others are transmitted through the water and reflected from the bottom. Both sets of reflected rays eventually combine to determine the pattern of image intensity at a point of observation. Another way in which patterns of shading can be affected by surface interactions is through the process of indirect illumination. Whenever a surface is illuminated, some of the incident energy is reflected in many directions and can illuminate other objects in exactly the same way as direct illumination from a luminous body such as the sun.

IMAGE ANALYSIS

In a model of image formation, the structure of the environment is given and the resulting pattern of light intensity (i.e., image) at a point of observation must be computed. In a model of visual perception, however, the problem is reversed: that is, a pattern of light

intensity is given and the structure of the environment that produced it must be determined. Because mapping from environmental structure to images is "many-to-one," inverse mapping cannot be computed uniquely without applying additional constraints to the solution.

Shape from Shading

The lowest possible level of image structure that could contain information about an object's three-dimensional form is the quantity of reflected light from each local region, commonly referred to as image shading. One possible model for determining an object's shape from shading has been developed by Horn and his coworkers (Horn, 1975, 1977, 1981; Ikeuchi and Horn, 1981). The goal of this analysis is to determine the local orientation of a visible surface region from the intensity of its reflected light. To constrain the solution the model assumes that (1) the direction of illumination is known, (2) the spectral composition of the light source is known, (3) the surface has a homogeneous reflectance, (4) its albedo is known, (5) there are no specular highlights, (6) shadows are cast on the surface, (7) there is no indirect illumination, (8) there is no transparency, and (9) the surface is smooth. Horn has shown that whenever these assumptions are satisfied, the local surface orientation can be computed by solving a set of differential equations. The problem, of course, in applying this model in an uncontrolled natural environment is that the required assumptions are seldom, if ever, satisfied.

Pentland (1982, 1984b) and, more recently, Lee and Rosenfeld (1983, 1985) have developed an alternative approach to Horn's that does not require prior knowledge of the direction of illumination or the surface albedo. To compute the direction of illumination, these models assume that all possible surface orientations occur with equal frequency throughout the observed scene. The models also differ from Horn's in the manner in which local surface orientation is computed. Horn's model uses the known albedo and illumination to establish a mapping between individual intensity values and their corresponding surface orientations. The analyses of Pentland and of Lee and Rosenfeld, in contrast, use the gradient of intensity to compute surface orientation: the magnitude of the gradient is used to estimate surface slant, and the direction of the gradient is used to

estimate surface tilt. This assumes, however, that the observed surface region is locally spherical (i.e., that the magnitude of curvature is equal in all directions).

Pentland (1984a, 1986) has also extended this analysis to eliminate the assumption of surface smoothness. If the roughness of the surface is governed by a spatially isotropic fractal function, which seems to be the case for many naturally occurring surfaces, then the average orientation of the surface can be determined from the statistical variations in image intensity. The tilt of the surface is specified by the direction in which there is the highest frequency of variation in intensity, and the slant is estimated by the magnitude of this frequency relative to the average value within a more globally defined region.

Although there are several important differences among these models for computing shape from shading, a few critical assumptions are shared by all, namely, that the observed surface has a homogeneous reflectance with homogeneous illumination and that there are no transparencies or specular highlights. It is important to keep in mind that these assumptions are frequently violated under natural viewing conditions, and there is some psychophysical evidence to suggest that such violations may have little or no effect on the perception of shape from shading by actual human observers (Beck, Prazdny, and Ivry, 1984; Gilchrist, 1979; Hagen, 1976; Metelli, 1974; Todd and Mingolla, 1983). In one recent experiment, for example, Mingolla and Todd (1986) obtained observers' local orientation judgments for simulated ellipsoid surfaces with differing reflectance functions. On the basis of existing theory it would be reasonable to predict that the addition of specular highlights in a display should increase the error in observers' judgments. That was not the case, however. Indeed, there was a small but statistically significant improvement in performance as the proportional contribution of the specular component to image intensity increased.

Surface Quality from Shading

Another important issue that must be considered in the analysis of image shading concerns the determination of surface quality, including such distinctions as matte versus shiny, rough versus smooth, opaque versus transparent, and light versus dark. Human observers are in fact quite good at identifying different surface materials under a broad range of conditions. Most of the existing work in this

area has focused primarily on the perception of surface reflectance (e.g., albedo and color) under conditions of varying illumination (i.e., the phenomenon of lightness and color constancy). There are many different models of this phenomenon (Brill, 1979; Land, 1986; Land and McCann, 1971; Wandell and Maloney, 1984; Weinberg, 1976), all of which compare the spectral characteristics in any given unit area with other unit areas throughout the entire scene. The models differ in terms of the extent to which they can tolerate variations in illumination, and whether or not they require prior knowledge about the reflectances of certain objects to provide a reference for subsequent calculations.

A fundamental assumption in all of these models of lightness and color constancy is that the observed surfaces in a scene are completely opaque. This does not seem to be the case, however, for actual human observers. There is considerable psychophysical evidence that observers can perceive the transparency of a surface under appropriate experimental conditions. Models of this phenomenon, proposed by Metelli (1974) and by Beck et al. (1984), can successfully predict the perceived transparency of a surface from the lightness values in several neighboring regions subject to certain configurational constraints.

Some research has also been reported on the perception of surface roughness. Pentland (1984a, 1986) has argued that the fractal dimension of any given image region provides potential information about the roughness of the depicted surface in that region. This hypothesis seems to be supported by observers' judgments of surface roughness: if the fractal dimension of an image is 2 (i.e., the same as its topological dimension), the surface is perceived as smooth. As the fractal dimension is made larger than 2, the apparent roughness of the surface increases.

Edge Detection

Many of the existing techniques for computing the structure of the environment from visual images do not work directly from image intensities but are designed instead to interpret image contours or edges, which are defined by abrupt changes in intensity. The detection of image contours has thus become an active area of research in both human and machine vision.

The most common method of contour detection is to convolve the image with an appropriate set of locally applied "edge operators," and much evidence suggests that a similar strategy has evolved in

biological systems. A variety of different local operators have been proposed as possible mechanisms for the process of edge detection (Canny, 1986; Grimson and Hildreth, 1985; Haralick, 1984; Marr and Hildreth, 1980; Torre and Poggio, 1986). Most of these operators are generically quite similar. The image is first smoothed at a particular scale by convolving it with a regularizing function (e.g., a Gaussian or a Gabor); then a differentiation operator (e.g., the Laplacian) is applied to detect rapid changes in intensity.

A fundamental problem with this general method of edge detection is that the response of each local operator combines the effects of many different image properties such as edge position, orientation, and contrast, which must ultimately be disentangled. It is also not clear how a population of local operators could produce the global patterns of organization so characteristic of human perception. Some researchers have attempted to address these issues by using parallel distributed networks of neural elements (Walters, 1986a,b; Zucker, 1986). The most well-developed model of this genre has been proposed in a series of recent articles by Grossberg and Mingolla (1985a,b; 1987). Their model uses an ingenious combination of competitive and cooperative interactions to sharpen and organize the outputs of local operators, and has been employed to simulate a surprisingly broad range of psychophysical phenomena in the areas of pattern and form perception.

Another important issue that has received only limited attention in the analysis of image contours is the problem of contour classification. Image contours can arise from a variety of physical phenomena including changes in surface reflectance, changes in illumination (e.g., shadows), specular highlights, abrupt discontinuities in surface geometry (e.g., the edges of polyhedra), and the occlusion of one part of a surface by another. Existing techniques for determining the three-dimensional form of a surface from patterns of image contours inevitably assume that the process of contour classification has already been performed.

Reflectance Contours

One possible source of information about the three-dimensional form of a visible surface is provided by the overall pattern of reflection contours. Suppose, for example, that the reflectance contours on a surface form small bounded regions called texture elements (e.g., the spots on a leopard). The projected sizes and shapes of these texture

elements would vary systematically with the surface geometry. In particular, the projected size of each element would decrease with the square of its distance from the point of observation, and the projected shape of each element would be compressed (i.e., foreshortened) by increasing surface orientation relative to the direction of gaze. Thus, any systematic variation over space in the projected sizes and shapes of bounded texture elements provides potential information about the geometry of an observed surface.

The first computational analyses of this type of texture pattern were developed by Gibson and his associates in the 1950s (Gibson, 1950; Purdy, 1958). These analyses assume that the observed surface is planar and that its distribution of texture elements is stochastically regular (i.e., that the texture elements within equal areas of a surface have comparable distributions of size, shape, and density). Whenever these assumptions are satisfied, the gradients of size, shape, or density in the optical projection of the texture pattern can be used to determine the orientation of the surface in three-dimensional space.

More recent analyses have attempted to analyze the three dimensional structures of curved surfaces from patterns of optical texture. One approach adopted by Witkin (1981) assumes that the texture elements are approximately circular and viewed from a sufficiently long viewing distance to approximate a parallel projection. Surface orientation in that case can be computed from the foreshortening of each element in the visual image. Another approach adopted by Stevens (1981a) assumes that each element is approximately circular, of known size, and viewed under strong polar projection. Under these conditions, the depth of each texture element can be determined by the length of its optical projection.

There is little evidence to suggest that any of these models have much in common with the processes of human perception. For example, in a recent series of experiments, Todd and Akerstrom (1987) asked observers to estimate the perceived eccentricity of simulated ellipsoid surfaces that were depicted by using various types of texture patterns. The results demonstrated that the perception of a curved surface can be achieved under a variety of theoretically anomalous conditions including both parallel and polar projections, as well as displays in which all of the projected texture elements have constant length, constant foreshortening, or constant area. Todd and Akerstrom proposed that changes in the perceived depth of a surface are determined by smooth variations over space in the widths of its projected texture elements. A specific implementation of this analysis

was proposed, based on the neural network model of Grossberg and Mingolla (1987), which provides a close fit to the psychophysical data over a wide range of experimental conditions.

A second method of deriving the three-dimensional form of a surface from its projected pattern of reflectance contours focuses on the nature of contour intersections rather than the regions that are bounded by contours (i.e., the texture elements). One such model, recently proposed by Stevens (1981b, 1983), assumes that the contours on a surface are restricted to lines of maximum and minimum curvature. Based on this assumption, it is possible to determine the local Gaussian curvature of a surface in the neighborhood of a contour intersection. If one of the contours projects to a straight line, the surface is parabolic (i.e., a plane or a cylinder). If the contours project to curves of the same sign, the surface is elliptic (i.e., locally concave or convex); and if the contours project to curves of opposite sign, the surface is hyperbolic (i.e., saddle shaped). There is some psychophysical evidence that human observers may employ such a strategy when presented with simple patterns of two intersecting curves (Ivry and Cohen, 1987). However, the utility of the analysis for a model of pilot performance seems dubious at best, because the reflectance contours encountered in an unconstrained natural environment are seldom restricted to lines of principal curvature.

Occlusion Contours

Another possible source of information about the three-dimensional form of a visible surface comes from patterns of occlusion contours, which arise in images when one part of an object is partially hidden behind another. In a recent mathematical analysis, Koenderink and van Doorn (1976, 1982) have shown that the curvature of an occlusion contour with respect to its attached surface region provides potential information about the Gaussian curvature of the surface in that region. If the occlusion contour is convex, then the corresponding surface region to which it projects must be elliptic. On the other hand, if the occlusion contour is concave, the corresponding surface region must be hyperbolic. This analysis assumes that the observed surface is smooth (i.e., the edges of polyhedra require a different type of analysis, considered below) and that the region of a surface to which the contour is attached is clearly specified. There is some psychophysical evidence to suggest that in the absence of other information, an occlusion contour is perceptually attached to

the visible surface region directly below it. Under these conditions an inversion of the image produces a corresponding change in the perceived curvature of the depicted surface.

Identifiable Features

Some analyses of visual information are designed to operate only after more primitive analyses have identified the optical projections of specific features of environmental structure. For example, Sedgwick (1973, 1983) has shown that when an object is in contact with an unbounded, planar ground surface, its height above the ground is visually specified by a relationship between its projected size and its position relative to the horizon. To apply this analysis, however, it is necessary to distinguish among the optic elements that correspond to the object, the ground, and the sky. This is not a trivial requirement. Although human observers apparently have little difficulty identifying bounded regions within a cone of visual solid angles, there are at present no adequate theories of how this is accomplished.

Other analyses in the field of artificial intelligence are also designed to operate on identifiable features. Indeed, many of the scene-analysis programs used in computer vision research receive coded representations of line drawings as inputs rather than real visual images. This is typically justified by assuming that some earlier process has identified the lines and junctions on a visual projection surface that correspond to the edges and vertices of opaque, plane-faced polyhedra in three-dimensional space. One famous program written by Guzman (1968) classifies line junctions on a visual projection surface into a relatively small number of categories. The result of this classification is generally ambiguous, because each type of line junction can have many possible three-dimensional interpretations. However, because there are severe topological constraints on the way in which the line junctions can be connected to one another, the number of possible interpretations for the entire configuration is reduced dramatically. Subsequent research has demonstrated that any remaining ambiguities can often be eliminated by taking into account additional constraints on the relative orientations of lines on the projection surface (Macworth, 1977) or the projected boundaries of cast shadows (Waltz, 1975).

POTENTIAL APPLICATIONS

The models of static form perception described in this chapter have a variety of potential applications for facilitating pilot performance. This is particularly true for helicopter flight maneuvers in which the aircraft remains near ground level except for short periods of time when it must pop up for a brief glimpse of the surrounding terrain before returning to the relative safety of a more concealed position. During this brief period of unmasking, the pilot must obtain vital information about the visual scene, including the presence of potentially hostile targets and the structure of the surrounding landscape.

For example, one important way in which theories of static form perception could help improve pilot performance in this context is in the design of computer-generated visual displays. This application is especially relevant for the next generation of potentially windowless helicopters in which all information for piloting the aircraft will be provided by cockpit instrumentation. It is of obvious importance in this type of environment that the appropriate information be presented with the greatest possible perceptual salience. This can only be achieved, however, with a thorough understanding of the mechanisms of human perception.

Another important application for theories of static form perception is the design of automatic aids. Because the perceptual capabilities of human observers are far superior to any machine vision system developed to date, the existing technology can probably be improved significantly by copying some of the proven methods of analysis that nature has developed through the process of evolution.

The primary stumbling block for achieving these potential applications is that existing models have been designed with little regard to the properties of human vision and therefore, have only minimal value in predicting pilot performance or in optimizing the perceptual salience of cockpit instrumentation. Moreover, as documented in the present summary of these models, they are typically derived from highly restrictive assumptions that would seldom be satisfied in the natural imagery encountered by real pilots. This lack of generality is a serious shortcoming that limits the utility of existing models for the design of automatic aids (e.g., a machine vision system for target recognition).

These limitations are sufficiently severe that no adequate models of static form perception are expected to appear in the foreseeable future. The most feasible strategy for developing a useful model in

the near term is likely to involve some sort of expert system. This could best be facilitated by a substantial increase in psychophysical research on the visual perception of three-dimensional form. The benefits of this research will be twofold. Its most immediate benefit would be to provide useful human factors guidelines for optimizing the perceptual salience of cockpit displays depicting surfaces in three-dimensional space. It would also give a longer term benefit by providing a more solid empirical foundation for the development of future computational models.

REFERENCES

- Beck, J., Prazdny, K., and Ivry, R.
1984 The perception of transparency with achromatic colors. *Perception and Psychophysics* 35:407-422.
- Beckman, P., and Spizzichino, A.
1963 *The Scattering of Electromagnetic Waves from Rough Surfaces*. New York: MacMillan.
- Blinn, J.F.
1977 Models of light reflection for computer synthesized pictures. *Computer Graphics* 11:192-198.
- Blinn, J.F., and Newell, M.E.
1976 Texture and reflection in computer generated images. *Communications of the ACM* 19:542-547.
- Brill, M.H.
1979 Further features of the illuminant-invariant assessment of chromatic relations. *Journal of Theoretical Biology* 78:305-308.
- Canny, J.
1986 A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8:679-699.
- Cook, R.L., and Torrance, K.E.
1981 A reflectance model for computer graphics. *Computer Graphics* 15:307-316.
- Gibson, J.J.
1950 *The Perception of the Visual World*. Boston: Houghton Mifflin.
- Gilchrist, A.L.
1979 The perception of surface blacks and whites. *Scientific American* 240:112-124.
- Grimson, S.E.L., and Hildreth, E.C.
1985 Comments on "Digital step edges from zero crossings of second directional derivatives." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 7:121-127.
- Grossberg, S., and Mingolla, E.
1985a Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review* 92:173-211.
1985b Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perception and Psychophysics* 38:141-171.

- 1987 Neural dynamics of surface perception: Boundary-webs, illuminants and shape from shading. *Communications on Vision, Graphics and Image Processing* 37:116-165.
- Guzman, A.
1968 Computer Recognition of Three-dimensional Objects in a Visual Scene. Ph.D. Thesis, MAC-TR-59, Project MAC. Massachusetts Institute of Technology, Cambridge, MA.
- Hagen, M.
1976 The development of sensitivity to cast and attached shadows in pictures as information for the direction of the source of illumination. *Perception and Psychophysics* 20:25-28.
- Haralick, R.M.
1984 Digital step edges from zero crossing of second directional derivatives. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6:58-68.
- Horn, B.K.P.
1975 Obtaining shape from shading information. In P.H. Winston, ed., *The Psychology of Computer Vision*. New York: McGraw-Hill.
1977 Understanding image intensities. *Artificial Intelligence* 8:201-231.
1981 Hill-shading and the reflectance map. *Proceedings of the IEEE* 19:14-47.
- Ikeuchi, K., and Horn, B.K.P.
1981 Numerical shape from shading and occluding contours. *Artificial Intelligence* 17:141-184.
- Ivry, R., and Cohen, A.
1987 The perception of doubly-curved surfaces from intersecting contours. *Perception and Psychophysics* 41(4):293-302.
- Kay, D.S., and Greenberg, D.
1979 Transparency for computer synthesized images. *Computer Graphics* 13:158-165.
- Koenderink, J.J., and van Doorn, A.J.
1976 The singularities of the visual mapping. *Biological Cybernetics* 24:51-59.
1982 The shape of smooth objects and the way contours end. *Perception* 11:129-137.
- Land, E.
1986 Recent advances in retinex theory. *Vision Research* 26:7-22.
- Land, E.H., and McCann, J.J.
1971 Lightness and retinex theory. *Journal of the Optical Society of America* 61:1-11.
- Lee, C.H., and Rosenfeld, A.
1983 Albedo estimation for scene segmentation. *Pattern Recognition Letter* 1:155-160.
1985 Improved methods of estimating shape from shading using the light source coordinate system. *Artificial Intelligence* 26:125-143.
- Macworth, A.K.
1977 How to see a simple world: An exegesis of some computer programs for scene analysis. In E. Elcock and D. Michie, eds., *Machine Intelligence*. Chichester: Ellis Horwood.

- Marr, D., and Hildreth, E.
1980 Theory of edge detection. *Proceedings of the Royal Society of London B* 207:187-217.
- Metelli, F.
1974 The perception of transparency. *Scientific American* 230:90-98.
- Mingolla, E., and Todd, J.T.
1986 Perception of solid shape from shading. *Biological Cybernetics* 53:137-151.
- Pentland, A.P.
1982 Finding the direction of illumination. *Journal of the Optical Society of America* 72:448-455.
1984a Fractal-based description of natural scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6:661-674.
1984b Local shape analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6:170-187.
1986 Shading into texture. *Artificial Intelligence* 29:147-170.
- Phong, B.T.
1975 Illumination for computer-generated pictures. *Communications of the ACM* 18:311-317.
- Purdy, W.C.
1958 The hypothesis of psychophysical correspondence in space perception. *Dissertation Abstracts International* 19:1454-1455. (University Microfilms No. 58-5594).
- Sedgwick, H.A.
1973 The Visible Horizon. Ph.D. Thesis. Cornell University, Ithaca, NY.
1983 Environment-centered representation of spatial layout: Available visual information from texture and perspective. Pp. 425-458 in J. Beck, B. Hope, and A. Rosenfeld, eds., *Human Machine Vision*. New York: Academic Press.
- Stevens, K.A.
1981a The information content of texture gradients. *Biological Cybernetics* 42:95-105.
1981b The visual interpretation of surface contours. *Artificial Intelligence* 17:47-73.
1983 The line of curvature constraint and the interpretation of 3-D shape from parallel surface contours. *Eighth Annual International Joint Conference on Artificial Intelligence* 1057-1061.
- Todd, J.T., and Akerstrom, R.
1987 Perception of three-dimensional form from patterns of optical texture. *Journal of Experimental Psychology: Human Perception and Performance* 13(2):242-255.
- Todd, J.T., and Mingolla, E.
1983 The perception of surface curvature and direction of illumination from patterns of shading. *Journal of Experimental Psychology: Human Perception and Performance* 9:583-595.
- Torre, V., and Poggio, T.A.
1986 On edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8:47-163.

Walters, D.K.W.

- 1986a A computer vision model based on psychophysical experiments. Pp. 88-120 in H.C. Nusbaum, ed., *Pattern Recognition by Humans and Machines*. New York: Academic Press.

1986b Selection and use of image features for segmentation of boundary images. *Proceedings of the IEEE CVPR*.

1987 Generating semantic descriptions from drawings of scenes with windows. In P. Winston, ed., *The Psychology of Computer Vision*. New York: McGraw-Hill.

Walters, D.K.W. and Maloney, L.T.

- 1986 Computational methods for color identification. *Journal of the Optical Society of America* 1:1231A.

Werner, J.S.

- 1970 The geometry of colors. *General Relativity and Gravitation* 7:135-169.

Whitted, T.

- 1980 An improved illumination model for shaded display. *Communications of the ACM* 23:343-349.

Witkin, A.P.

- 1981 Recovering surface shape and orientation from texture. *Artificial Intelligence* 17:17-45.

Zucker, S.

- 1986 Early process for orientation selection and grouping. Pp. 170-200 in A.P. Pentland, ed., *From Pixels to Predicates*. Norwood, NJ: Ablex.

7

Structure From Motion

MYRON L. BRAUNSTEIN

OVERVIEW

The changing images resulting from relative motion between an observer and surfaces in the environment constitute an important source of information about the shapes, orientations, and relative distances of these surfaces. The perception of three-dimensional structure from motion is relevant to virtually all tasks involving vision of the environment outside the cockpit, whether direct or provided optically or electronically. For some problems such as wire detection, minimum visibility considerations are primary and structure from motion is likely to play a minor role at best. For other tasks, structure from motion is likely to be a major source of much of the needed information, especially if one considers the interaction of motion with other sources of depth information, such as occluding contours, texture, shading, and binocular disparity. Structure from motion is likely to be important in visual navigation, the perception of the distal scene, and the identification of objects and landmarks. In unmasking, target detection, and masking maneuvers, structure from motion considerations would be important in determining what is perceived in the visual scene and how quickly it is perceived.

Many empirical results in the literature on visual perception that would be useful in making predictions about the role of structure from motion in the perception of a visual scene under various conditions. However, most of these findings are not captured by models in a way that would allow specific outputs to be predicted from specific inputs. A number of models describe the information potentially available about three-dimensional structure from motion. Some describe the minimum numbers of points and views required to recover structure under various constraints; and others describe components of optic

flow that are informative about structure and motion. As currently formulated, however, these models are not directly applicable to the specific flight tasks mentioned above. There appear to be two reasons for this. The first is that the existing models deal with situations that do not approach the complexity found in most real-world visual scenes. Second, for reasons detailed in this report, the validity of general computational models of structure from motion has not been determined for human observers. If a very simple scene was involved, such as three small lights on the ground being observed during a fixed-axis rotation of the helicopter, on the basis of structure-from-motion proofs, one might predict that a human observer could judge the relative distances between the points. Even with a model that states that three distinct views are necessary for that output to result from that input, no quantitative predictions of accuracy could be made because the concept of "distinct views" is not clearly defined and the relationship between the output of the model—recovering the three-dimensional structure—and the output required of the pilot would have to be determined.

Overall, it must be concluded that a great deal of additional research on complex perceptual processes, such as structure from motion, is needed before model-based input-output relationships can be determined. Additional models need to be developed that are based on a systematic consideration of human psychophysical data and provide specific quantitative predictions of human performance. Models must be developed that integrate different perceptual modules and relate multiple stages in perceptual processes, where interdependence among these stages is likely. Psychophysical research should be extended to consider more complex surfaces, and additional methods must be developed to provide quantitative measures of the recovery of three-dimensional structure from motion. In laboratory settings, display technology is required which will allow researchers to study the effects of the subtle variations in velocity and other display parameters that affect perception in direct vision. On the positive side, there has been major progress in both model building and psychophysical research in this area over the last few years. As understanding of the complexities involved in the perception of real three-dimensional scenes continues to increase, it should be possible to move in the direction of useful models that are validated against human behavior.

INTRODUCTION

The psychophysical data from studies of the recovery of structure from motion, in broad terms, indicate that human observers can recover the three-dimensional shapes of environmental objects and the relative distances of surfaces on the basis of very little stimulus information—brief exposures and small numbers of visible points. There are often ambiguities involving depth reversal, at least in laboratory situations devised to study structure from motion in isolation; and in some cases there are predictable errors in judging relative depth. The small amount of information required to recover three-dimensional structure and the achievement of solutions that contain relative depth ambiguities (reversals of the sign of depth) are also found in various models of the recovery of structure from motion.

These broad relationships between the mathematical models that have been proposed and the psychophysical data are encouraging. At present, however, no general theories of the recovery of structure from motion have been sufficiently tested against behavioral measures to allow for any degree of confidence that they represent human performance. There are models developed in specific experimental contexts that have been tested rigorously, and more general models that have proved compatible with existing data on human performance. Most general models of the recovery of structure from motion, however, although they provide rigorous descriptions of what a visual system can theoretically recover from dynamic images, have not been tested adequately against human performance.

Some of the reasons for this lack of rigorously tested models are historical and are being overcome at the present time. Early laboratory research on structure from motion had its origins in a phenomenological tradition (Metzger, 1934; Wallach and O'Connell, 1953), and only recently have models been used to make quantitative predictions of laboratory results in this area of research (Braunstein, 1972; Todd, 1982). On the other hand, there has been enormous progress in the development of computational models of vision in a little over 10 years. Testing these models with human subjects would appear to be an important direction for psychophysical research, which should result in the availability of at least some experimentally validated models of the recovery of structure from motion by human observers. Attempts to validate computational models, however, have been slowed by several serious difficulties.

First, virtually all general computational models specify competence rather than predict performance. In this context, competence refers to the knowledge a subject should be able to acquire about the three-dimensional environment from two-dimensional images, whereas performance refers to the behavior of a subject in specific observable tasks. (See Ullman, 1986, for additional discussion of competence versus performance in structure-from-motion models.) Competence, however, cannot be studied directly in the laboratory. The psychophysicist can only measure performance. For example, Ullman (1979) has shown that it is theoretically possible to recover the three-dimensional positions of four noncoplanar points from three distinct views (orthographic projections). To determine whether a human observer has this competence, however, some task must be given the observer on which performance can be measured. Linking the predicted competence to a measure of performance is not straightforward, and various tasks are likely to result in varying decrements in performance relative to the predicted competence, or even in enhancements of performance over the expected competence (Braunstein, Hoffman, Shapiro, Andersen, and Bennett, 1987).

A second difficulty arises because some models that are applicable only when certain conditions are met do not incorporate these prior conditions. Structure-from-motion models typically require that the correspondence problem has been solved. This is the problem of matching the points in successive two-dimensional projections of an object moving relative to the eye, so that the points in one view are correctly paired with the points in another view. A correct pairing means that the matched points are both projections of the same point in the three-dimensional scene. The assumption that the correspondence problem has been solved is a very reasonable one, because correspondence is not usually a problem in human vision. However, when one attempts an experimental test of a structure-from-motion theory, conditions must be used in which false correspondences are avoided. This requirement severely restricts the conditions under which the theory can be tested and, in most cases, eliminates any possibility of a general test. Consider a structure-from-motion theorem in which the only constraint is rigidity. A general test would require that any degree of rotation be allowed between the successive views. To meet the correspondence restriction, either the degree of rotation between the views must be severely limited or severe restrictions must be placed on the location of the points and the axis of rotation. These limitations may serve

as additional constraints that can be used by the human observer to recover the three-dimensional structure.

The third difficulty may be the most serious because there is no way, in principle, to overcome it. If a behavioral task is to be used to determine whether three-dimensional structure has been recovered from two-dimensional images, the information for performing this task must be present in those images. This means that it is not possible, in principle, to determine whether the task has been performed by recovering the three-dimensional structure or by using the information in the two-dimensional images in some other way that did not require recovering the three-dimensional structure. Often one must rely on pragmatic arguments (that any direct two-dimensional processing of a particular stimulus would be too difficult) or phenomenology (that the subjects reported three-dimensional perceptions). These arguments can be made most convincingly when there is an inverse relationship between the detectability of the two-dimensional information and the recovery of three-dimensional structure. This seems to be the case for motion parallax (Braunstein and Tittle, in press) where judgments of relative distance become more accurate as the difference in the projected velocities of the nearest and farthest texture elements decreases. This is, however, an unusual case. It is far more common for the information that leads to more accurate judgments about three-dimensional structure to be positively correlated with differences in the two-dimensional images (Braunstein et al., 1987).

A fourth difficulty in psychophysical testing of mathematical models of the recovery of structure from motion is that the most precise psychophysical methods available may be inappropriate to some of the important questions. Psychophysical procedures can be classified into two types. The most familiar measures discriminative abilities of the observer, such as the ability to detect minimal differences in illumination. To obtain optimum performance, highly trained subjects are generally used and these subjects are usually given feedback. The use of feedback implies that there is a correct answer. This type of procedure is informative about what a human observer can do and usually provides very precise, quantitative data.

This is not the only possible question, however. It is often important to address questions about how something appears to an observer, how it is categorized, or what perception normally occurs. This is the question of what an observer does do, rather than can do. Often there is no objective basis for specifying a correct answer, and

even if there were, the use of feedback would be contrary to the purpose of the study. The methods used in this type of research involve categorization and judgments of similarity, rather than minimal discriminations. Research in color vision has traditionally used these methods. It is important to note that the distinction between discrimination and categorization methods does not imply a distinction between automatic and cognitive processing. Consider the example presented earlier in which subjects were asked to judge which texture elements are nearer and which are more distant in a motion parallax display. A discrimination paradigm could be used in which subjects are expected to give the correct answer and are given feedback, or a categorization paradigm could be used in which the emphasis is on categorizing the appearance of the stimuli, with no indication that there is a correct answer and no use of feedback. In the latter paradigm, as noted earlier, subjects are more accurate when the velocity differences are small (within the range that has been studied). When the differences are large, they are noticed as two-dimensional velocity differences. Indeed, observers knowledgeable about motion parallax report surprise that the faster moving elements (in two-dimensional displays) sometimes look further away than the slower moving elements in such displays. It is likely, for these cases, that a discrimination task with feedback would result in judgments in accordance with the proximal velocities, even though the subjective experience might not be in accordance with these velocities. Such data would be likely to indicate that subjects made more accurate discriminations as the velocity difference increased, which is the opposite of the results obtained when subjects are asked to categorize the appearance of the stimuli. A discrimination task using feedback may thus provide misleading information about how a subject would respond to motion parallax information in a real-world situation.

The result of these historical trends and inherent difficulties is the existence of a body of rigorous mathematical models of the recovery of structure from motion, as well as a body of experimental literature on human performance, but very little evidence concerning the applicability of the mathematical models to human performance.

Although progress has been slow in developing psychophysical tests for mathematical models of the recovery of structure from motion, important progress has been made in one area—the testing of constraints underlying mathematical models. Two constraints that

are central to most models are rigidity and correspondence. The former constraint has been studied by a number of investigators (Braunstein and Andersen, 1984, 1986; Schwartz and Sperling, 1983; Todd, 1982, 1984). The latter has been thoroughly investigated by Todd (1985). A third constraint that is central to almost all optical flow models—smoothness of the velocity field—has been studied recently by Andersen (1988). In all three cases the constraints have been found to be of less general applicability than most current models indicate. Perception is often not in accord with predictions based on rigidity, and three-dimensional structure may be recovered as easily in nonrigid as in rigid configurations. Structure can be recovered in the presence of severe violations of correspondence and of smoothness of the velocity field. These findings suggest that human perception is more flexible than current models indicate. Progress in relaxing assumptions such as rigidity is being made in current computational research (especially Koenderink and van Doorn, 1986, summarized below).

MODELS

There are a large number of analyses of the information available in dynamic two-dimensional projections for the recovery of various aspects of the three-dimensional environment. The following discussion includes examples of such analyses representing different types of approaches, for which attempts have been made to relate the analyses to human vision. Other analyses, developed primarily in an artificial intelligence (AI) context, are not covered, although they have sometimes included discussions of possible relationships to human vision.

There are basically two types of analyses: discrete points and views analyses, and optical flow analyses. The discrete points and views analyses consider minimum numbers of texture elements or feature points and minimum numbers of views or frames required to recover depth information under varying environmental constraints. Most, but not all, of these analyses have employed the mathematics of orthographic projection, employing projective properties that are not dependent on variations in viewing distance. Optical flow analyses, on the other hand, use the instantaneous projected velocity field or acceleration field as the basis for recovering information about depth relationships. Almost all of these analyses employ the geometry of polar perspective and depend on the effects of variations in

viewing distance. One exception is Hoffman's (1982) analysis, which recovers local surface orientation from orthographic projections by using velocity and acceleration fields.

The best known of the discrete points and views analyses is Ullman's (1979) proof for three orthographic views of four noncoplanar points. The proof proceeds essentially as follows. An assumption is made that the points are rigidly connected, that is the three-dimensional interpoint distance between each pair of points is constant across views. This assumption is expressed in a set of simultaneous equations. If the rigidity assumption is true, there will be exactly two solutions (reflections about the line of sight). Otherwise, there will be no solutions. In other proofs the number of points and views required has been reduced by introducing additional constraints such as planarity (Hoffman and Flinchbaugh, 1982), fixed axis of rotation (Hoffman and Bennett, 1986), and constant angular velocity (Hoffman and Bennett, 1985).

As indicated earlier, psychophysical data suggest that models based on strict rigidity may not be general enough to account for the recovery of structure from motion by human observers. Ullman (1984) has proposed a model that seeks to overcome some of the objections to a strict rigidity-based analysis. This incremental rigidity scheme maintains an internal model of the structure of a moving object that consists of the estimated three-dimensional coordinates of points on the object. The model is continually updated as new positions of image features are considered. Initially, the object is assumed to be flat, if no other cues to three-dimensional structure are present. Otherwise, its initial structure may be determined by other cues available, from stereopsis, shading, texture, or perspective. As each new view of the moving object appears, the algorithm computes a new set of three-dimensional coordinates for points on the object that maximizes the rigidity in the transformation from the current model to the new positions. This is achieved by minimizing the change in the three-dimensional distances between points in the model. Thus, the algorithm interprets the changing two-dimensional image as the projection of a moving three-dimensional object that changes as little as possible from one moment to the next. Through a process of repeatedly considering new views of objects in motion and updating the current model of their structure, the algorithm builds and maintains a three-dimensional model of the objects. If objects deform over time, the three-dimensional model computed by the algorithm also changes over time (Hildreth and Koch, 1987). Although

the incremental rigidity scheme may seem plausible as a process that could handle situations in which a precise rigidity solution cannot be computed due to visual noise or deformations in the object over time, the validity of this scheme as a model of human behavior has not yet been demonstrated. Research in progress by Hildreth and her colleagues attempts to test incremental rigidity against psychophysical data.

The optical flow approach has taken a number of forms. Some analyses concentrate on a particular aspect of the flow field that seems especially important for biological vision. Lee (1980), for example, emphasizes the ratio of the projected velocity of a texture element to its projected radial distance from the point of fixation. This ratio provides relative distance information. Lee points out that absolute distance information can be recovered if this relative information is scaled according to some measure of the observer (such as eye height) or of the observer's motion (such as the observer's velocity when the motion in the optical array is self-generated). This seems to fit with the concept that relationships between the environment and parts of the helicopter are used in distance and speed judgments (Murray and Hayworth, personal communication). This chapter is concerned primarily with the use of optical flow to recover the structure of the three-dimensional environment. The use of optical flow to estimate parameters of observer motion is discussed in Chapter 8.

Analyses of the optical flow field often divide the flow field into components. This division into components is generally in accordance with established geometric concepts and not based initially on perceptual considerations. However, a number of papers suggest possible relationships between the geometric components and the use of optical flow in perception. Optical flow may be divided into divergence (div), curl, and deformation (def) components, where div describes expansion and contraction in the image plane, curl describes rotation in the image plane, and def describes shearing motion (expansion in one dimension with an area-preserving contraction in the orthogonal dimension) in the image plane (see Koenderink, 1986, for a review). These two-dimensional components can be related to four components of three-dimensional motion—translational and rotational components along the line of sight, as well as translational and rotational components perpendicular to the line of sight. The relationship between the two-dimensional and three-dimensional components has been discussed by Koenderink and van Doorn (1986).

The same four categories of three-dimensional motion have been used to classify psychophysical research (e.g., Braunstein, 1978).

Longuet-Higgins and Prazdny (1980) present two methods for recovering the gradient of a surface and the motion of the eye relative to a surface from optical flow. The first method uses the velocity field generated by an observer moving relative to a stationary scene. It requires that there be at least two points with the same visual direction at different distances, which by definition means that all the points cannot be part of the same flow field. This requirement is a very reasonable one for the helicopter environment, although it has received very little attention in the laboratory. (See Andersen and Braunstein, 1985, and Andersen, 1988, for laboratory studies of this type of stimulus.) A second analysis is presented for cases in which this requirement is not met or in which there are a number of objects in rigid motion. In this analysis, a separate computation is required for each rigid object, and these computations require access to both the first and the second derivatives of points in the motion field. Longuet-Higgins and Prazdny discuss the possibility that the human visual system possesses channels for the analysis of four flow-field derivatives: dilation (divergence or *div*), two components of shear (deformation or *def*), and vorticity (*curl*). Although, as they note, some evidence for dilation channels exists (Regan and Beverley, 1978), there is as yet no evidence for direct sensitivity to deformation or *curl*.

Prazdny's (1983) analysis of optical flows as a source of information about the three-dimensional environment is based on the following assumptions: (1) the availability of velocity vectors at a set of retinal loci, (2) motion relative to a rigid environment, and (3) metric information about the positions of these loci relative to a two-dimensional reference frame. The retinal velocity of a visible point is resolved into three components. Two components are due to rotation of the object relative to the observer, and one is due to translation. One of the rotational components is due to rotation parallel to the image plane; the other, to rotation perpendicular to the image plane. Prazdny shows that the instantaneous projected velocity field contains information about the relative depths of two retinal points which are projections of points that are rigidly connected in three-dimensional (points on the same rigid object or stationary points relative to a moving observer). This extends earlier work by Gibson, Olum, and Rosenblatt (1955), Lee (1980), and Clocksin (1980) from pure translation to curvilinear motion. Local surface orientation is

also computed at a given point. Prazdny notes that the quality of relative depth and surface orientation decays with distance. Finally, noting studies which indicate that perceptual continuity may be given precedence over the information in optical flow, Prazdny (1983, p. 257) remarks that "the theoretical existence of information in itself does not guarantee that it will be used."

Koenderink and van Doorn (1986) have presented an analysis in which depth and shape are obtained from optical flow without an assumption of global rigidity. Instead, bending deformations are allowed to occur at dihedral edges between triangular facets on a polyhedral surface. A solution is obtained from two views by using only the def component of the image. This solution has a fourfold ambiguity. It is reduced to twofold ambiguity by the use of curl, with the remaining ambiguity one of relief—whether a dihedral edge is concave or convex. The relief ambiguity can be overcome by repeating the analysis from a different vantage point. This method recovers shape from two views of seven points, with no four points rigidly connected. There is no evidence about the applicability of this analysis to human vision, but Koenderink and van Doorn report a demonstration that is suggestive of a relationship. The type of fourfold ambiguity that occurs in the analysis, prior to the use of curl, appeared to occur during observation of simulated bending polyhedrons, suggesting that the human observer may use the analysis based on def but does not use curl to reduce the ambiguity.

There are a large number of optical flow analyses by other investigators, mostly emphasizing machine vision but often alluding to possible relationships with human vision. Especially notable is a series of reports by Waxman and his collaborators, on optical flow alone and on optical flow combined with stereopsis (for example, Waxman, 1984; Waxman and Duncan, 1985; Waxman and Ullman, 1983).

CONCLUSION

In conclusion, general mathematical analyses are available for determining the minimum amount of visual information required to recover three-dimensional structure from two-dimensional images, for an observer moving relative to a rigid environment, for an environment with multiple rigid objects, and for an environment in which bending deformations are present. Incorporating these methods into a model of pilot performance would provide an indication

of what an ideal observer, under specific assumptions, might be able to accomplish, but this would not necessarily match the capabilities of a human observer. The human observer might not do as well, or might do better by combining sources of information and by using environmental constraints that have not been incorporated into a particular theoretical analysis.

What is known at present is that optical flow provides information for the relative distances of surfaces from the observer and may provide absolute distance information. Human observers probably do not use all of the available information but, in some situations, appear to be able to use amounts of information close to the theoretical minima. The type of questions one would like to be able to answer is: Given stimuli that are above detection thresholds for luminance contrast and motion, what surfaces will be detected, and how quickly and accurately will they be detected? At the present time there is insufficient knowledge about how well existing models match human performance for specific sources of information, and how different sources of information are integrated to make general predictions from models. The development of models that are more directly testable against human performance, and of more precise behavioral measures to study judgments for which the most precise psychophysical methods available are unsuitable, should result in progress at least in the identification of part-task models. An important intermediate step, which will be of value in developing models which are applicable to human behavior and is important in its own right, is the organization of the vast body of empirical data that have accumulated in the study of structure from motion and related perceptual issues, especially over the last 10 years. The answers to many important design questions are likely to be found in these data if they become accessible to designers.

RESEARCH NEEDS: STRUCTURE FROM MOTION

The discussion of structure from motion models in Chapter 2 indicated that the applicability of existing models to pilot performance is limited by two factors. First, the models are theoretical accounts that specify the information about three-dimensional structure that a vision system might recover from two-dimensional images. The models have not been successfully validated against human behavior. Second, most of the models are concerned with displays that are too simple to be of interest in developing models of pilot performance. A

number of steps could be taken to facilitate the development of valid and useful models of human perception of structure from motion:

(1) Models should be developed on the basis of a systematic consideration of human psychophysical data. This is the established approach in many other areas of investigation, but existing structure-from-motion models are based primarily on mathematical analyses of what information a vision system might recover under varying constraints. These models are often not directly testable against human performance (as noted in Chapter 2) and, when tested, prove to be inadequate models of human vision. Psychophysical research must be brought in at the model development stage. This requires collaborative efforts by researchers specializing in the development of theoretical models and those specializing in human psychophysics, and the training of researchers who combine these specialties. Some of this is happening now, but not nearly enough of these combined efforts are occurring to provide the valid models of human perception of structure from motion that are needed.

(2) Models have to be developed, or current models extended, to include predictions about human performance on behavioral tasks. This need is closely related to the first one. Because many current models have been developed in an AI context, they often specify competence rather than predict performance. Rather than specifying what an observer should know about the three-dimensional environment, a testable model should specify what judgments an observer is able to make and, even better, the accuracy with which the observer can make these judgments. This requires further theoretical development, again involving a combination of expertise in computational theory and psychophysical research to assure a match between the types of behavior predicted by the models and the types of responses that can be elicited from human subjects.

(3) Models must be developed that combine different perceptual "modules." Likely, more interaction exists among perceptual processes than can be found in most current theoretical accounts. There is a paucity of data and of general models of human perception that combine structure from motion with other types of depth information, such as stereopsis, occluding contours, shading, and texture. It is unlikely that performance can be predicted in complex scenes until these interactions are understood.

There is also a need for an improved understanding of how different types of information about three-dimensional structure are combined. Some sources of depth information, such as orthographic

projections of rotations about axes other than the line of sight, are informative about depth relationships within objects, providing three-dimensional shape in object-relative coordinates. Other types of information, such as polar projections of translations in depth, are informative about the locations of objects in the three-dimensional environment relative to the observer. The manner in which information about three-dimensional shape is combined with information about relative distance to form a unified perception of the three-dimensional environment is an issue requiring further investigation.

(4) Models must be developed which integrate stages in the recovery of structure from motion that are now treated separately. Although it is appropriate in a theoretical analysis to make assumptions about the results of earlier stages, such as an assumption that the correspondence problem has been solved, a testable model should take into account the restrictions on the stimulus domain implied by these assumptions.

(5) Research on structure from motion should be extended to more complex surfaces than the spheres and cylinders typically studied. Work along these lines is just beginning to appear (Andersen, 1988; Landy, Sperling, Doshier, and Perkins, 1987).

(6) Additional psychophysical methods are needed to study the recovery of three-dimensional structure from motion and from other sources of information combined with motion. For some important research issues, such as the interpolation of perceived surface structure between visible features on complex surfaces, methods would be useful that provide some of the advantages of feedback without having a predetermined correct response. Interactive graphics methods offer some excellent possibilities. The subject can be required to adjust a display until it meets a specific criterion (e.g., apparent smoothness of a surface). Although the experimenter would specify the criterion, the subject would decide when the criterion has been met. Interactive graphics provide responses that can be measured precisely and include a form of reinforcement (the display appearing correct to the subject) without the use of an externally determined correct response.

(7) Related to the need for the development of psychophysical methods that take advantage of such technologies as interactive graphics is a need for more extensive use of high-resolution displays in research on human motion perception. Attempts at precise testing of models may be misleading if displays provide only gross approximations of the visual stimulus. The issue here is not the usual "fidelity"

issue of how much information should be displayed, but the issue of how accurately the displayed information must be represented.

REFERENCES

- Andersen, G.J.
 1988 The detection of structure and depth from optical flow: Violations of the smooth velocity constraint. *Investigative Ophthalmology and Visual Science* 29:(3, Supplement).
- Andersen, G.J., and Braunstein, M.L.
 1985 Induced self-motion in central vision. *Journal of Experimental Psychology: Human Perception and Performance* 11:122-132.
- Braunstein, M.L.
 1972 Perception of rotation in depth: A process process model. *Psychological Review* 79:510-524.
 1978 Perception of motion. In E. Carterette and M. Friedman, eds., *Handbook of Perception Vol. VIII*. New York: Academic Press.
- Braunstein, M.L., and Andersen, G.J.
 1984 A counterexample to the rigidity assumption in the perception of structure from motion. *Perception* 13:213-217.
 1986 Testing the rigidity assumption: A reply to Ullman. *Perception* 15:641-644.
- Braunstein, M.L., Hoffman, D.D., Shapiro, L.R., Andersen, G.J., and Bennett, B.M.
 1987 Minimum points and views for the recovery of three-dimensional structure. *Journal of Experimental Psychology: Human Perception and Performance* 13:335-343.
- Braunstein, M.L., and Tittle, J.S.
 in press The observer-relative velocity field as the basis for effective motion parallax. *Journal of Experimental Psychology: Human Perception and Performance*.
- Cocksins, W.F.
 1980 Perception of surface slant and edge labels from optical flow: A computational approach. *Perception* 9:253-269.
- Gibson, J.J., Olum, P., and Rosenblatt, F.
 1955 Parallax and perspective during aircraft landings. *American Journal of Psychology* 68:372-385.
- Hildreth, E.C., and Koch, C.
 1987 The analysis of visual motion: From computational theory to neuronal mechanisms. *Annual Review of Neuroscience* 10:477-533.
- Hoffman, D.
 1982 Inferring local surface orientation from motion fields. *Journal of the Optical Society of America* 72:888-892.
- Hoffman, D., and Bennett, B.
 1985 Inferring the relative 3-D positions of two moving points. *Journal of the Optical Society of America* 75:350-533.
 1986 The computation of structure from fixed axis motion: Rigid structures. *Biological Cybernetics* 54:1-13.
- Hoffman, D., and Flinchbaugh, B.
 1982 The interpretation of biological motion. *Biological Cybernetics* 42:197-204.

- Koenderink, J.J.
1986 Optic flow. *Vision Research* 26:161-180.
- Koenderink, J.J., and van Doorn, A.J.
1986 Depth and shape from differential perspective in the presence of bending deformations. *Journal of the Optical Society of America* 3:242-249.
- Landy, M.S., Sperling, G., Doshier, B.A., and Perkins, M.E.
1987 Structure from what kinds of motion? *Investigative Ophthalmology and Visual Science* 28:(3, Supplement) 233.
- Lee, D.N.
1980 The optic flow field: The foundation of vision. *Philosophical Transactions of the Royal Society of London* B-290:169-179.
- Longuet-Higgins, H.C., and Prazdny, K.
1980 The interpretation of a moving retinal image. *Proceedings of the Royal Society of London* B-208:385-397.
- Metzger, W.
1934 Tiefenerscheinungen in optischen Bewegungsfeldern. *Psychologische Forschung* 20:195-260.
- Prazdny, K.
1983 On the information in optic flows. *Computer Vision, Graphics, and Image Processing* 22:239-259.
- Regan, D., and Beverly, K.I.
1978 Looming detectors in the human visual pathway. *Vision Research* 18:415-421.
- Schwartz, B.J., and Sperling, G.
1983 Nonrigid 3-D percepts from 2-D representations of rigid objects. *Investigative Ophthalmology and Visual Science* 24:(3, Supplement) 239.
- Todd, J.
1982 Visual information about rigid and nonrigid motion: A geometric analysis. *Journal of Experimental Psychology: Human Perception and Performance* 8:238-252.
1984 The perception of three-dimensional structure from rigid and non-rigid motion. *Perception and Psychophysics* 36:97-103.
1985 The perception of structure from motion: Is projective correspondence of moving elements a necessary condition? *Journal of Experimental Psychology: Human Perception and Performance* 11:689-710.
- Ullman, S.
1979 *The Interpretation of Visual Motion*. Cambridge, MA: MIT Press.
1984 Maximizing rigidity: The incremental recovery of 3-D structure from rigid and nonrigid motion. *Perception* 13:255-274.
1986 Competence, performance, and the rigidity assumption. *Perception* 15:644-646.
- Wallach, H., and O'Connell, D.N.
1953 The kinetic depth effect. *Journal of Experimental Psychology* 45:360-368.
- Waxman, A.M.
1984 *An Image Flow Paradigm*. Technical Report No. 45. University of Maryland, Center for Automation Research.

Waxman, A.M., and Duncan, J.H.

1985 *Binocular Image Flows: Steps Toward Stereo-Motion Fusion*. Technical Report No. 119. University of Maryland, Center for Automation Research.

Waxman, A.M., and Ullman, S.

1983 *Surface Structure and 3-D Motion from Image Flow: A Kinematic Analysis*. Technical Report No. 24. University of Maryland, Center for Automation Research.

Motion-Based State Estimation and Shape Modeling

GREG ZACHARIAS

INTRODUCTION AND SUMMARY

A number of candidate models and algorithms for motion-based state estimation and shape modeling are reviewed in this chapter, along with the problem of "state and structure" through motion. To provide a framework for discussion, an overall end-to-end process-oriented structure for modeling the generation of state and shape estimates from dynamic visual images is described. Three major processing functions are identified: (1) flow-field estimation, which generates vector flow-field estimates on the basis of the temporal dynamics and spatial characteristics of the image time history; (2) state-time estimation, which generates estimates of observer rotational and translational egomotion states, and an estimate of the instantaneous field of "impact times" defining a scaled three-dimensional depth map of the imaged scene; and (3) object shape modeling, which accounts for the depth map via appropriate selection, parameterization, and localization of object models. Some of the attributes of this processing structure are discussed in terms of its compartmentalization, its ability to help in the identification of information flow and information reliability, and its potential for coupling with other process-oriented models of human perception and performance.

Within this framework, a number of candidate models and algorithms are reviewed. In the area of flow-field estimation, models can be categorized as feature-, gradient-, or frequency-based. A number of attractive models are found in the latter two categories: some because of their potential for simulating errors in human flow perception, and others because of their natural linkage to human visual frequency selectivity. In the area of state-time estimation, both

quasi-static and dynamic algorithms are reviewed. Models capable of extracting the egomotion information needed to subserve visually guided locomotion are identified, along with dynamic estimation approaches which could account for the human's expectations of state evolution in active control situations. Initial validation studies have begun the task of matching model predictions of perceptual egomotion errors with those seen empirically. In the area of object shape modeling only a few models are identified because the focus is on approaches to "assembling" three-dimensional objects from basic observer-centered depth information. Additional work in this area can be found in Chapters 6 and 7.

A number of areas exist in which some of these models could be applied to understanding and aiding the helicopter nap of the earth (NOE) mission, specifically in visually guided flight control. Potential areas include

- prediction of flight path control precision and speed-height trade-offs under different workload levels;
- identification of concurrent visual environments (texture, shape, occlusion boundaries, etc.) and maneuver envelopes (position, attitude, and their rates) that are likely to cause disorientation or illusion;
- evaluation of display aids to augment "weak" outside-the-window cues, under adverse visual conditions (e.g., fog, smoke) or under conditions of high workload that demand visual attention sharing; and
- development and evaluation of novel dynamic pictorial displays to provide integrated situational information in a natural visual format.

A range of other application areas can be considered special cases of these, as discussed below.

Three basic areas that require additional research can be identified. First, current algorithms must be enhanced to provide sufficient generality to deal with complex visual scenes and with the noisy and less-than-ideal visual environments that might characterize an NOE mission. Second, current models require more validation against past and current psychophysical data, with particular attention paid to the current research focus on "active psychophysics." Finally, there is a need to begin integrating these perceptual models with existing control-decision models, to begin to address the essential closed-loop nature of visually guided flight. Coupling with control models can

subserve flight path performance assessment, whereas coupling with decision-theoretic models can support predictions of the pilot's situational awareness; obviously, other model couplings could subserve the analysis and prediction of other visually driven mission tasks in like fashion.

The following section contains an overall integrative structure for comparing the variety of motion-based vision algorithms and models considered here. Then individual reviews of models and algorithms are provided, followed by model applications to the helicopter NOE mission problem. Finally, future areas of research are identified.

FRAMEWORK FOR MOTION-BASED STATE ESTIMATION AND SHAPE MODELING

Description of Framework

Figure 8-1 illustrates, in block diagram form, an overall end-to-end process-oriented structure for motion-based state estimation and shape modeling. Although fairly simple, the structure attempts to identify the information flow (via the lines) and the processing functions (via the blocks) presumed present in human visual processing of dynamic imagery.

The processing begins with the generation of an image sequence: a simple monocular two-dimensional gray level function $I(x,y,t)$, defined over the imaging surface (x,y) , and varying with time t . This intensity function can be considered essentially continuous in space and time, spatially sampled (pixellated), or temporally sampled (frame by frame), or both. For discussion purposes and later computational reasons, both spatial and temporal discreteness are assumed, which means that a discrete sequence of pixellated image frames is available for processing. It is a little difficult to justify the temporal sampling, but easier for spatial sampling, given the existing retinal photoreceptor array (e.g., see Williams and Collier, 1983; Yellot, 1983). Note also that no consideration is given to the potential contribution of color in this processing description.

This image sequence is then processed by a flow-field computation block to generate a corresponding two-dimensional "flow field," a vector field which specifies the instantaneous angular rate of the line of sight (LOS) of each imaged point in the field-of-view (FOV). The computed field is temporally sampled with a new two-dimensional flow-field frame computed at every image frame time. The computed

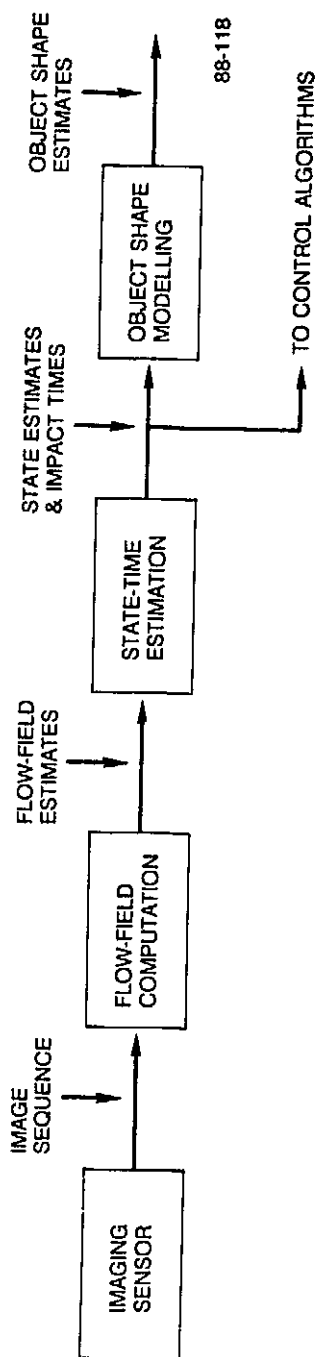


FIGURE 8-1 Framework for motion-based state estimation and shape modeling.

field is also presumed to be spatially sampled because of image pixelation, so that the flow is computed only at each pixel. In effect, the computed field provides a temporally and spatially sampled version of the continuous flow-field associated with observer-scene relative motion. The flow field can be specified in the image sensor plane as a set of two-dimensional in-plane vectors (the conventional approach) or as a set of three-dimensional angular LOS rate vectors, defined in an arbitrary observer-referenced coordinate frame. This latter specification allows for a definition of "optic flow" that is independent of imaging plane orientation and FOV, in the Gibson tradition.

The resulting flow-field is then processed by a state-time estimation block to generate, at each frame time, estimates of the fundamental observer states that can be inferred from the input image sequence: instantaneous aim point, angular velocity, and a two-dimensional vector field of "impact times" defining the directions and transit times to imaged points in the FOV. Because the number of input flow-field vectors is likely to be large (roughly equaling the number of image pixels) in comparison with the number of unknown observer states (two in heading and three in angular velocity), the translation-rotation state estimation problem is overdetermined. A least-squares estimation approach can provide a simple means for dealing with this situation, while simultaneously minimizing the effects of flow-field estimation error propagation. The resulting estimates of aim point and angular velocity can then be used to compute the impact time vector field, which defines an observer-centered, spatially sampled, speed-scaled replica of the imaged scene. In effect, the impact time field provides a scaled three-dimensional "depth map" of the imaged scene.

The resulting impact time vector set can then be processed by an object shape modeling block to select a "best-fit" object model and generate corresponding parameter estimates for the selected model. One way of accomplishing this is via another least-squares estimation process, in which a fixed-form object model is adjusted, via its intrinsic parametric specifiers and via extrinsic scaling, rotation, and translation, to obtain a best fit to the estimated impact time vector set over the full FOV of the imaged scene. This type of processing has the potential for significant data compression, yielding a small set of object parameters from a large number of impact time vectors. Subsequent iteration over an internal dictionary of generic object shapes could then provide optimized shape modeling over the range of dictionary objects known to the observer.

Some Comments on the Processing Framework

Some comments on the general structure of this overall process are in order. First, note that an explicit structure has been proposed, consisting of three separate processing blocks, for the extraction of motion-based state estimates and object shape estimates. In particular, a flow-field computation block has been identified, on the presumption that such processing is separate from other information-processing requirements it may subserve such as state estimation. Whether this reflects reality, of course, is unclear at this time. For example, the flow-field estimates could be processed directly to obtain estimates of local surface curvature, via the method proposed by Koenderink and van Doorn (1986), to directly subserve the object shape modeling function, while bypassing the intermediate step of impact time estimation. The explicit separation of processing functions proposed, however, allows for such a "processing shortcut," while still providing an end-to-end integrative framework via the information flow links between blocks.

Second, it is appropriate to note that these links specify the information base needed to determine both competence and performance in a model, a requirement identified by Watson (Chapter 5). In this context, competence is determined by an explicit specification of the input and output variables of each block. For example, for the state-time estimation block, an input set of the flow-field vectors and an output set of the aim point, angular velocity, and impact time vectors are specified. Performance is determined by an explicit specification of error propagation in these variables. Thus, for the same block, the way in which errors in the input flow-field vectors lead to errors in the output state-time vectors is specified. This can be done via brute force Monte Carlo simulation techniques or more elegant (but limited) covariance propagation techniques. The main point is, however, that by specifying both competence and performance, error propagation can be modeled from end to end, and "high-level" human output performance statistics can be generated (e.g., false alarm and missed detection statistics for target discrimination) on the basis of "low-level" front-end sensory-perceptual characteristics (e.g., simple foreground-background relative motion detection performance). Some of these issues are discussed further by Braunstein (Chapter 7).

Third, note that to model such "higher-level" performance, it is necessary to add one or more blocks to model the generation of external measurable control or response activity, driven by internal

estimates of image flow, observer state, or object shape. Thus, for example, to model the pilot's detection of a simple ridge line, one could append a detection block to process the object model block output and choose an appropriate set of utility functions to weight false alarm and missed detection costs. Alternatively, to model the pilot's visually guided terrain-following flight control performance, one could append a flight control block to process the state estimator output and generate appropriate pilot control actions to drive the vehicle flight control system (FCS). Full loop closure would be provided here via a vehicle dynamics model driving a scene generation model, which would then feed back to the pilot's "imaging sensor" shown at the left of Figure 8-1. Clearly, quite complicated (and less verifiable) models can be built up in this fashion to begin attacking some of the performance questions of interest in this report. However, to be able to build and verify such models requires a basic separation of processing functions (into blocks) and an explicit specification of information flow (between blocks).

REVIEW OF RESEARCH IN MOTION-BASED STATE ESTIMATION AND SHAPE MODELING

Some of the more recent work conducted in motion-based state estimation and shape modeling is reviewed briefly here. According to the framework introduced earlier and illustrated in Figure 8-1, this section is organized into three broad areas: (1) flow-field computation algorithms and techniques, (2) state and impact time estimation, and (3) flow-based object shape modeling.

Flow-Field Computation

First, studies concerned with the definition of the optic flow-field and the development of algorithms for estimating it are reviewed briefly. Models can be categorized as feature-based, gradient-based, or frequency-based. Because feature-based algorithms provide flow estimates at only a small number of points in the FOV and also suffer from a frame-to-frame correspondence problem, only the latter two model categories will be considered here. An additional discussion of feature-based approaches can be found in the Chapter 7.

Prazdny (1983) reviews earlier work in the perceptual psychology community relative to the basic information "contained" in the optic flow-field. This reference discusses how six degree-of-freedom (DOF)

motion of the observer with respect to the observed object gives rise to the optic flow seen in a specified two-dimensional imaging plane. The discussion focuses on the "forward transformation" from motion state to observed flow, and provides only a qualitative discussion of the "inverse transformation" from observed flow to estimated motion state. No explicit algorithms for flow computation or state estimation are presented.

Rieger's (1983) discussion is similar. The equations for flow due to six DOF motion are given, but they are specialized to a particular axis system. The information "contained in" these equations is discussed (i.e., flow-field measurements), but no algorithms are presented for flow computation or state estimation.

Horn and Schunck (1981) and Schunck (1983) concentrate on the computation of the flow-field itself from input image sequences that change with time, due to the imager's (or object's) motion. The basic algorithm is presented in Horn and Schunck (1981). In brief, the algorithm first computes, at each pixel, the temporal and spatial gradients of the image intensity function. These are then combined in a flow constraint equation. A flow field solution is then found which, in a least-squares sense, best satisfies that equation, while at the same time maintaining a reasonable "smoothness" of the flow over the imager field-of-view. The technique requires no knowledge of the structure of the visual world, has none of the critical reliance on image "features" often found in other flow algorithms, and is particularly well suited to situations in which the flow-field evolves with time. It does have its share of problems, however, dealing with nonorthographic projections and scenes in which occluding surfaces are present.

Further work on the algorithm is presented in Schunck (1983), which identifies the basic problems inherent in processing image sequences having object occlusion boundaries. Attempts at occlusion edge detection and regional smoothing are presented in this reference, but the results are generally unsatisfactory in terms of improving the algorithm's ability to work with occlusion-induced flow shear.

A very different approach to flow-field estimation is presented by Watson and Ahumada (1983, 1985), motivated by their considerations of human motion perception and its dependence on spatial frequency content. By working in the three-dimensional spatial temporal frequency domain (defined by the moving image's two spatial frequency axes and one temporal frequency axis) and introducing localized direction-sensitive Gabor filter "sensors," they construct a

spatially distributed estimator of the flow-field, tuned to a particular spatial frequency in the visual bandwidth of interest. Summing outputs across a set of these tuned field estimators, they then obtain a full bandwidth vector field defining the flow. Preliminary simulations of the algorithm demonstrate the potential for modeling human psychophysical performance in discrimination, perception of pattern coherence, and the like, but further validation is needed to match or explain the additional psychophysical data available.

A modified version of this approach is described and applied by Heeger (1987) to a variety of synthetic and natural textured image sequences. For cases involving simple image translation, it is shown how the image signal-to-noise (S/N) ratio drives the flow-field estimation errors. However, for realistic images, Heeger (1987) notes that the primary source of error comes from the fact that "the model assumes image translation, ignoring motion [occlusion] boundaries, accelerations, deformations (rotation, divergence, shear), and motion transparency." Unfortunately, these same limitations apply to many of the other reviewed algorithms and models. Heeger closes the paper with a preliminary comparison of model simulation results with human psychophysical data on the coherence of sine-grating patterns.

Jain (1984) presents an algorithm for identifying the relative motion parameters of independently moving objects in an imager's field-of-view. The algorithm is restricted to translational motion and requires knowledge of the imager motion parameters. However, it may prove to be of utility in "segmenting" more complex scenes for modeling human performance in complex visual environments.

Kahn (1985) and Mutch and Thompson (1985) present algorithms for detecting occlusion edges in dynamic scenes, algorithms which may be directly applicable to enhanced flow-field computation models. Kahn (1985) proposes an algorithm to estimate edge direction and speed, based on spatial temporal variations in intensity recorded over a triangular pixel triplet. The approach is restricted to constant velocity straight edges, and no study of image noise susceptibility is presented. Mutch and Thompson (1985) propose a token (feature) matching approach to detecting and characterizing occlusion edges: any lost (created) tokens imply membership on an occluded (occluding) surface; the boundary between surfaces must then be the occlusion edge. An example is given and several limitations of the approach are discussed.

Adiv (1985) takes a more general approach to flow-field computation in assuming that the flow-field arises because of observer motion through an environment of several objects that may be in relative motion to one another. The computation algorithm first computes the flow-field in "segments" where, within each segment, the flow computation is assumed to arise from the motion of a single planar surface. The algorithm then groups segments under the assumption that the observed flow is due to the motion of a single (larger and connected) moving surface. The results show successful discrimination of occluding objects and their relative motion, in the face of sparse and noisy sequential image inputs. The results promise a significant improvement in flow-field computation capabilities, for more complex scenes.

Additional work in scene segmentation is provided by Murray and Buxton (1987). A global optimization criterion is used to decide optimum segmentation for a given (assumed) number of objects. Convergence is slow, however, and not well predicted. More work on this approach appears to be needed.

Terzopoulos (1986) describes the application of general multi-grid relaxation methods to a number of image-processing problems, one of which is flow-field computation. He demonstrates that dynamic superpixelation (going from a coarse to a fine pixel grid) can be used to significantly improve the convergence characteristics of the basic Horn and Schunck (1981) algorithm. Results indicate that an order-of-magnitude reduction in computation time can be expected, to obtain the same level of flow-field estimation accuracy.

Additional work attempting to improve fundamental gradient-based methods, such as that of Horn and Schunck (1981), is presented by Kearney, Thompson, and Boley (1987) and by Nagel and Enkelmann (1986). Kearney et al. (1987) use perturbation methods to specify error propagation characteristics due to errors in the underlying gradient estimates and note that the dominant error source is due to occlusion-induced flow-field discontinuities, which violate underlying continuity assumptions. Nagel and Enkelmann (1986) introduce the notion of an "oriented smoothness constraint" to help handle such problems, and with some additional "heuristic modifications," they demonstrate reasonable flow-field estimation performance in a case involving foreground-background relative motion.

State and Impact Time Estimation

Studies concerned with the estimation of dynamic state information, based on flow-field input measurements, are now reviewed briefly. Zacharias (1982) derives basic equations defining the flow-field for general six degrees of freedom (DOF) observer motion and arbitrary imager geometry. The constraint equations relating flow to state are given and then transformed to yield equivalent constraint equations in terms of the potentially inferable states: heading, angular velocity, and impact time. A discussion follows concerning minimal flow measurement counts and observer geometries which ensure "solvability." However, an algorithm that solves for or estimates observer states, given the flow-field measurements, is not provided.

Ullman (1983) reviews earlier discussions regarding solvability of the flow-field equations. The discussion focuses on translational motion and considers the implications of orthographic versus perspective projections. The discussion is qualitative, however, and no algorithms are given.

Rieger and Lawton (1983) present an algorithm, based on earlier work by Lawton (1982), for determining imager heading from the optic flow pattern, for arbitrary six DOF imager motion. The basic approach relies on the fact that along an occlusion boundary, any discrete changes in the flow-field are attributable solely to the translational contribution to flow, not the rotational. This allows for a separation of the two contributions and subsequent stepwise solution of each. Results are presented which demonstrate how the algorithm makes use of the flow discontinuity and successfully estimates heading with sparse pixellation and noisy images.

Bruss and Horn (1983) formally define the flow-field based estimation problem for general six DOF motion and a planar imager geometry. Using a least-squares criterion, they derive a set of nonlinear constraint equations that must be satisfied by the estimated states. These are solved for simple three DOF translation and for simple three DOF rotation, but no results are presented for the general six DOF case. The lack of a consistent vector-matrix notation leads to significant difficulties in following the derivation and interpreting the results. Only analytic results are given, many in the form of equations specifying necessary conditions; no simulations of estimator performance are presented.

Broida and Chellappa (1986) consider dynamic estimation of the states of a rigid two-dimensional body undergoing planar four

DOF translation and rotation. The measurements are the in-screen locations of a small set of object feature points, in this particular case two points. An extended Kalman filter (Gelb, 1974) is used to estimate the object's translational and rotational states. Monte Carlo simulations are used to generate estimation error statistics and demonstrate the algorithm's basic ability to dynamically infer four DOF states from two feature points. The overall approach appears to have significant potential for incorporating a knowledge of known observer dynamics and provides for filtering, over time, of the generated state estimates. However, the approach does require significant updating if it is to be extended to estimating six DOF motion parameters with a measurement set several orders of magnitude larger than the small feature set used in the study.

Merhav and Bresler (1986a,b) also describe a dynamic filtering approach to the state estimation problem. They apply Kalman filtering to the flow-field estimation problem as well, but only along a simple "raster line" in the field of view. Considerable potential appears to exist for improved low-noise estimation over that seen in conventional quasi-static modeling efforts.

Mitiche (1986) presents the most recent rederivation of the flow constraint equations. Vector matrix notation is not used, so the essential structure of the constraint is not apparent. The assertion is made that only four feature points are required to estimate a six DOF state, which is false as demonstrated by Zacharias (1982). An algorithm is presented for solving for the state, but the author notes that the "solution," obtained via numerical search, is highly dependent on the initial guess. No consideration is given to the basic problem of estimating state with a highly redundant set of noisy flow-field measurements.

Zacharias, Caglayan, and Sinacori (1983a,b) derive a flow-field based state estimator for general six DOF motion and arbitrary imager motion. The estimator minimizes a quadratic cost function based on the flow constraint equation residuals and, by using redundant and noisy flow-field measurements, estimates observer heading, angular velocity, and impact time to all observed points in the imager FOV. Monte Carlo simulations are conducted to generate estimation error statistics to demonstrate performance sensitivity as a function of pixel count and flow-field noise level. The model is used to simulate a simple human aim point estimation task, and model estimation accuracy is shown to provide a reasonable match to that obtained experimentally.

Waxman and Sinha (1986) introduce the concept of "dynamic stereo" as an algorithm for extracting passive ranging information in complex dynamic scenes. When an observer is moving relative to a group of objects, each of which is moving relative to the other, impact times are no longer simple functions of relative range and observer speed. Waxman and Sinha show how two imagers, moving relative to each other by a known amount, can be used to generate a "difference flow" that, in turn, can be used to extract the desired scene depth information. Their emphasis is on computer vision for autonomous vehicle navigation, but their results may be applicable to understanding human head movement strategies when confronted with complex dynamic scenes.

Flow-Based Object Shape Modeling

Some studies that focus on the estimation of observed object shape, based on flow-field input measurements, can now be summarized.

Clocksion (1980) describes how the flow-field depends on viewed object shape. The discussion is primarily qualitative but points out how (1) discontinuities in the flow-field indicate the presence of occluding surfaces; (2) discontinuities in the flow-field gradient* indicate the presence of concave or convex "cusps" on the surface; and (3) values taken on by the flow-field Laplacian reflect the orientation of the viewed object surface normal. The discussion centers on the "forward transformation" from object properties to flow-field characteristics; no algorithms are presented for the "inverse transformation" from observed flow to inferred object shape.

Hoffman (1980) presents a similar discussion of how observer state and object shape act in concert to determine not only the observed flow-field but also the first spatial derivative of this flow-field. Constraint equations relating observer state to flow and flow rate are derived for the special case of an orthographic projection geometry. An argument is given for the "solvability" of these equations, but no solutions or algorithms for solutions are given. Also unaddressed is the problem of reliably computing the spatial derivative of a noisy flow-field.

*That is, the spatial rate of change of the flow-field, as a function of the line-of-sight in the observer FOV.

Horn (1984) discusses the extended Gaussian image (EGI) and its applicability to the representation of object surface shape. The EGI for an N -faced polyhedron consists of the complete set of N scaled face normals, where the scaling reflects face area. The EGI, in theory, provides all the information needed to reconstruct the object itself, including orientation. Extension to a continuous, smoothly curved object is relatively straightforward. The potential for deriving the EGI from the impact time vector set, generated by the present state-time estimator, recommends its further study and evaluation for three-dimensional object modeling.

Bolle and Cooper (1986) formulate an object modeling and location algorithm that processes range measurements to generate object surface patch primitives (planar, cylindrical, and spherical patches). These, in turn, are "assembled" to form a more complex object, to support three-dimensional object range estimation. The problem formulation and estimation algorithm is presented, and two simulations are given. The method appears to have considerable promise for object shape modeling, but additional evaluation with more realistic ranging data is called for.

MODEL APPLICATIONS AND LIMITATIONS

A number of areas can be identified in which some of the above models could be applied to understanding and aiding the helicopter NOE mission. There are also a number of limitations in the use of these models, before reasonable confidence can be placed in their predictive abilities. Some of these areas are described briefly in the following paragraphs.

The area in which the greatest contribution could be made appears to be in visually guided flight control. The NOE mission provides few classic "geometric" cues to orientation or location, such as occur in a turn to final or an approach to landing. Rather, the cues are likely to be dominated by unstructured texture (e.g., treetop leaves) and by dynamic rather than static attributes, so that flow-field cues can dominate static shading or textural gradient cues. A motion-based model of state and shape estimation would thus appear to be particularly appropriate here.

Earlier an overall model structure was outlined which, given a dynamic textured input image sequence, can generate several of the required translational and rotational state estimates needed for

flight path control: heading, angular rate, and orientation with respect to the terrain-treetop surface (Murray and Hayworth, personal communication). The key, of course, is selecting or developing the component submodels needed to flesh out the overall structure and ensuring that they not only generate the appropriate informational variables (satisfying the earlier competence requirement) but also model human pilot capabilities and limitations (satisfying the earlier performance requirement).

In the area of flow-field estimation, a number of submodels could be considered. The frequency-domain approaches of Watson and Ahumada (1983, 1985) and Heeger (1987) are attractive because of their linkage to frequency-selective processing by the nervous system. The gradient-based approaches of Horn and Schunck (1981) and others, although more computer vision oriented, are also attractive because of the ease with which they can be used to simulate flow generation and error propagation. Both general approaches, however, require further development to deal with occlusion boundaries and rotational motion; considerably more validation against psychophysical data is also necessary.

In the area of state-time estimation, a number of submodels could be considered. The quasi-static estimator of Zacharias et al. (1983a,b) generates the required heading, angular velocity, and impact time estimates needed for the NOE task, and initial validation studies have begun to match model predictions with earlier psychophysical data (e.g., Warren, 1976). More recent work by Broida and Chellappa (1986) and Merhav and Bresler (1986a,b) demonstrates how modern dynamic estimation theory (in the form of Kalman filtering) can be brought to bear on the problem of generating dynamic state estimates, where the observer accelerates and constantly changes the visual flow. This potential for dynamic filtering suggests a linkage with dynamic models of the human pilot, which noted below.

Finally, in the area of object shape modeling, only a few candidates have been identified for consideration. For the NOE mission, sophisticated algorithms are not required because the flight control task needs only rough estimates of upcoming terrain shape to provide the preview information necessary for short-term flight path planning and anticipation of upcoming maneuvers. Thus, it may suffice to build on the work cited earlier by Horn (1984), using extended Gaussian image (EGI) models, or the work by Bolle and Cooper

(1986), who construct objects from surface patch primitives. Naturally, the appropriateness of any of these candidates ultimately rests on how well they match measured human perceptual performance in the NOE environment. Additional candidate models for object shape estimation can be found in Chapters 6 and 7 for static and dynamic situations respectively.

The following are four potential areas for model applications, within the context of visually guided flight control:

- prediction of flight path control precision and speed-height trade-offs under different workload;
- identification of concurrent visual environments (texture, shape, occlusion boundaries, etc.) and maneuver envelopes (position attitude, and their rates) that are likely to cause disorientation or illusion;
- evaluation of display aids to augment "weak" outside-the-window cues, under adverse visual conditions (e.g., fog, smoke) or under conditions of high workload that demand visual attention-sharing; and
- development and evaluation of novel dynamic pictorial displays to provide integrated situational information in a natural visual format.

A range of other applications can be considered special cases of the above. For example, the unmask-mask maneuver sequence performed for target acquisition can be analyzed in a model context with regard to maneuver precision (item 1), cue insufficiency (item 2), or utility of display aids (item 3). Likewise, evaluation of navigation performance could consider the likelihood of correct identification of a topographic way point (item 2), or evaluation of dynamic visual warning displays might be considered in the context of new pictorial formats (item 4). It seems likely that other such specialized flight tasks could be categorized as a special case of one of the four general application areas identified above.

FUTURE RESEARCH

Three basic areas requiring additional research can be identified: (1) enhancement of current algorithms for added robustness, (2) validation of model performance predictions against psychophysical data, and (3) integration of perceptual models with control-decision models.

Current algorithms still do not demonstrate sufficient generality to deal with complex scenes, nor do they demonstrate adequate robustness in less-than-ideal visual environments. Except under restricted viewing conditions, performance lags behind that of human observers. Thus, algorithm improvements are called for. In flow-field computation, work needs to be done in dealing with field discontinuities due to occlusion boundaries, as well as with "distortions" that arise from complex motion patterns. In state-time estimation, dynamic filtering approaches should be explored to see how prior expectations of egomotion can improve performance reliability. Finally, in object shape modeling, effort should be placed on integrating other static and dynamic sources of object edge-surface information. In the meantime, judgment must be exercised in applying any model in situations exceeding its original development assumptions.

An inadequate amount of model validation has been conducted, probably because of the difficulty of "model-tuned" experiments that generate the needed metrics for model versus data comparisons. However, a growing empirical data base is being generated in parallel with the modeling effort, and advantage should be taken of it. For example, the earlier work by Lee (1976) in perception of impact time could be compared with the accuracy of time-depth estimates generated by a number of the models. More recent work by Owen, Warren, and their colleagues (Owen and Warren, 1982; Owen, Warren, Jensen, Mangold, and Hettinger, 1981; Warren, 1976), evaluating the effect of flow-field attributes on the perception of egomotion, could serve as the direct basis for validating flow-field based state estimation models. One such example using the data generated by Warren (1976) is given in Zacharias et al. (1983a,b); others are clearly called for. Finally, it should be noted that there are ongoing empirical efforts (at NASA Ames, U.S. Air Force Army Aeromedical Laboratory, etc.) in both the passive and the active psychophysics of flow-field induced egomotion. Clearly, the modeling community should be taking advantage of this growing base of empirical data.

Finally, there is a need to begin integrating these perceptual models with existing control-decision models, to begin to address the essential closed-loop nature of visually guided flight. The discussion here has focused on end-to-end open-loop processing, but one should begin to consider loop closures generated by active control of the visual environment by the pilot. One approach studied (Brun and Zacharias, 1986; Zacharias, 1985) involves coupling a flow field

based perceptual model with a modern control model of the pilot-vehicle system (Kleinman, Baron, and Levison, 1971), to yield a closed-loop model that supports predictions of visually guided flight control performance. The potential exists for similar couplings to decision-theoretic models, to support predictions of, for example, detection probabilities of a topographic way point under conditions of limited visibility. Obviously, other model couplings could subserve the analysis of other visually driven flight tasks in like fashion.

REFERENCES

- Adiv, G.
1985 Determining three dimensional motion and structure from optical flow generated by several moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-7:384-401.
- Bolle, R.M., and Cooper, D.B.
1986 On optimally combining pieces of information, with application to estimating 3-D complex-object position and range data. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8:619-638.
- Broida, T.J., and Chellappa, R.
1986 Estimation of object motion parameters from noisy images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8:90-99.
- Brun, H.M., and Zacharias, G.L.
1986 Model-based methodology for terrain-following display design. Technical Report No. R8603. MA: Charles River Analytics Inc.
- Bruss, A.R., and Horn, B.K.P.
1983 Passive navigation. *Computer Vision, Graphics, and Image Processing* 21:3-20.
- Clocksinn, W.F.
1980 Perception of surface slant and edge labels from optical flow: A computational approach. *Perception* 9:253-269.
- Gelb, A.
1974 *Applied Optimal Estimation*. MA: MIT Press.
- Heeger, D.J.
1987 Model for the extraction of image flow. *Journal of the Optical Society of America A* 4:1455-1471.
- Hoffman, D.D.
1980 Inferring shape from motion fields. (AI Memo 592). MIT, Cambridge, MA: Artificial Intelligence Laboratory.
- Horn, B.K.P.
1984 Extended gaussian images. *Artificial Intelligence* 72:1671-1686.
- Horn, B.K.P., and Schunck, B.G.
1981 Determining optical flow. *Artificial Intelligence* 17:185-203.
- Jain, R.C.
1984 Segmentation of frame sequences obtained by a moving observer. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-6:624-629.

- Kahn, P.
1985 Local determination of a moving contrast edge. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-7:402-409.
- Kearney, J.K., Thompson, W.B., and Boley, D.L.
1987 Optical flow estimation: An error analysis of gradient-based methods with local optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-9:229-244.
- Kleinman, D.L., Baron, S., and Levison, W.H.
1971 A control theoretic approach to manned-vehicle systems analysis. *IEEE Transactions on Auto Control* AC-16.
- Koenderink, J.J., and van Doorn, A.J.
1986 Depth and shape from differential perspective in the presence of bending deformations. *Journal of the Optical Society of America* 3:242-249.
- Lawton, D.T.
1982 Motion analysis via local translational processing. *IEEE Conference Proceedings* 70. NY, NY: IEEE.
- Lee, D.N.
1976 A theory of visual control of braking based on information about time to collision. *Perception* 5:437-459.
- Merhav, S.J., and Bresler, Y.
1986a,b On-line vehicle motion estimation from visual terrain information, Parts I & II. *IEEE Transactions on Aerospace and Electronic Systems* AES-22:583-603.
- Mitiche, A.
1986 On kineopsis and computation of structure and motion. *IEEE Transactions on Pattern Analysis of Machine Intelligence* PAMI-8:109-112.
- Murray, D.W., and Buxton, B.F.
1987 Scene segmentation from visual motion using global optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-9:220-228.
- Mutch, K.M., and Thompson, W.B.
1985 Analysis of accretion and deletion at boundaries in dynamic scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-7:133-138.
- Nagel, H., and Enkelmann, W.
1986 An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8:565-593.
- Owen, D.H., and Warren, R.
1982 Perceptually relevant metrics for the margin of aviation safety: A consideration of global optical flow and texture variables. *Proceedings of the Conference on Vision as a Factor in Military Aircraft Mishaps*. San Antonio, TX: USAF School of Aerospace Medicine.
- Owen, D.H., Warren, R., Jensen, R.S., Mangold, S.J., and Hettinger, L.L.
1981 Optical information for detecting loss in one's own forward speed. *Acta Psychologica* 48:203-213.
- Prazdny, K.
1983 On the information in optical flow. *Computer Vision, Graphics, and Processing* 22:239-259.

- Rieger, J.H.
1983 Information in optical flows induced by curved paths of observation. *Journal of the Optical Society of America A* 73:339-344.
- Rieger, J.H., and Lawton, D.T.
1983 Determining the instantaneous axis of translation from optic flow generated by arbitrary sensor motion. *Association of Computing Machinery* 33-41.
- Schunck, B.G.
1983 Motion Segmentation and Estimation. Sc.D. Thesis. Dept. of Electrical Engineering and Computer Science. Massachusetts Institute of Technology, Cambridge, MA.
- Terzopoulos, D.
1986 Image analysis using multigrid relaxation methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8:129-139.
- Ullman, S.
1983 Computational studies in the interpretation of structure and motion: Summary and extension. (AI Memo 706). Cambridge, MA: Artificial Intelligence Laboratory.
- Warren, R.
1976 The perception of egomotion. *Journal of Experimental Psychology: Human Perception and Performance* 2(3):448-456.
- Watson, A.B., and Ahumada, A.J.
1983 A look at motion in the frequency domain. NASA Technical Memorandum No. 84352. CA: National Aeronautical and Space Administration.
- 1985 Model of human visual-motion sensing. *Journal of the Optical Society of America A* 2:322-342.
- Waxman, A.M., and Sinha, S.S.
1986 Dynamic stereo: Passive ranging to moving objects from relative image flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8:406-412.
- Williams, D.R., and Collier, R.
1983 Consequences of spatial sampling by a human photoreceptor mosaic. *Science* 221:385-387.
- Yellott, J.I., Jr.
1983 Spectral consequences of photoreceptor sampling in the rhesus retina. *Science* 221:382-385.
- Zacharias, G.L.
1982 *Flow-Field Cueing Conditions for Inferring Observer Self-Motion*. Report No. 5118. Cambridge, MA: Bolt Beranek and Newman, Inc.
- 1985 Modelling the pilot's use of flight simulator visual cues in a terrain-following task. Technical Report No. R8505. MA: Charles River Analytics Inc.
- Zacharias, G.L., Caglayan, A.K., and Sinacori, J.B.
1983a A model for visual flow-field cueing and self-motion estimation. *Proceedings of the 1983 American Control Conference*. San Francisco, CA.
- 1983b A visual cueing model for terrain-following applications. *Proceedings of the AIAA Flight Technologies Simulation Conference*. Niagara Falls, NY.

Real-Time Human Image Understanding in Pilot Performance Models*

IRVING BIEDERMAN

The need to identify objects provides a major justification for the presence of a human in the cockpit on most aircraft missions. Object recognition is required for the identification of potential targets and the determination of features for navigation. Both laboratory research and commercial film editing practice have established that from a 100 millisecond (msec) exposure of an object or scene, humans can accurately interpret images of objects and scenes never previously experienced. This capacity for real-time identification of objects or scenes is readily evidenced for line drawings, suggesting that much of human recognition is based on shape. Consequently, most accounts of human object recognition have concentrated on how the edges extracted from an image of an object or scene can activate—in real time—an appropriate representation of that object in memory.

Any theory of human object recognition must account for the phenomena that the speed and accuracy of performance often decline when the image is degraded, lacking parts, only moderately occluded, viewed from a novel orientation in depth, or presented as a simple line drawing. Indeed, a major value of placing a human visual system in the cockpit is this remarkable robustness of visual recognition over an extraordinary range of conditions of image perturbation and degradation.

*The research was supported by the Air Force Office of Scientific Research Grants 86-0106 and 88-0231. Correspondence about this chapter should be addressed to Irving Biederman, Department of Psychology, University of Minnesota, Elliott Hall, 75 E. River Road, Minneapolis, MN 55455.

This chapter first presents an overview of recent theoretical work on object recognition and a summary of some of the major empirical findings. Special problems related to the perception of multiobject and scene displays are then discussed. Throughout the chapter, significant gaps in our knowledge are also indicated.

THEORIES OF OBJECT RECOGNITION

The theoretical work reviewed in this section is confined to efforts that display a capability for handling the phenomena of human vision described previously (e.g., robustness for images that might be rotated in depth or degraded). The first model reviewed, and the one discussed most extensively, is the author's recognition by components (RBC) (Biederman, 1987a,b; 1988) because it is the most developed effort addressed to real-time human object recognition. Models developed as machine vision efforts but inspired by characteristics of human recognition (Brooks, 1981; Huttenlocher and Ullman, 1987; Lowe, 1987; Pentland, 1986) are described in a subsequent section and contrasted with RBC. Also considered are the formal characterizations of images based on topological properties (Koenderink, 1987; Pong, Shapiro, and Haralick, 1985), although these efforts have not been developed into recognition models.

Recognition by Viewpoint Invariant Components (RBC)

Decomposition of an Image into Geons

Recognition by components is directed primarily toward offering an account of how humans can rapidly and accurately classify images of objects at a basic level. "Basic level" is the most general level of a class that specifies shape information. The words for this level, such as "giraffe" or "telephone," typically specify almost as much shape information as a subordinate term, such as "articulated giraffe" or "desk phone," respectively. Basic level terms appear earlier in a child's vocabulary and are used far more frequently to refer to a class than either subordinate or superordinate level terms. Superordinate terms, such as "mammals," "instruments," or "modes of transportation" do not specify shape information. When instances of a basic level such as "penguin" or "ostrich" for the class "BIRD" do not share a common shape description with the basic level, they are handled, not as subordinate, but as their own basic level class.

RBC assumes that complex visual entities are decomposed into simple components, typically at regions of matched concavities. Such concavities are almost always produced when parts are arbitrarily joined (Hoffman and Richards, 1985). The resultant components activate the closest fitting member of a particular set of 24 convex or singly concave edge-based volumetric primitives that can be modeled as a family of generalized cones called "geons" (for geometric icons), such as bricks, cylinders, cones, and wedges.

Viewpoint-Invariant and Categorical Origins of Geons

The image properties from which geons are activated are viewpoint invariant (VIP), or nonaccidental, (Lowe, 1984) and highly resistant to degradation. Viewpoint-invariant properties (VIPs) include such characteristics as whether an edge is curved or straight, the type of vertex (fork, arrow, L, or tangent Y) at the termination of edges, and whether pairs of edges are parallel or symmetrical. For example, a cylinder differs from a wedge in that the former has a curved cross section, parallel sides along its axis, and tangent Y vertices that are absent in the wedge. By deriving the geons from simple contrasts in VIPs (such as whether the cross section is straight or curved and sides parallel or not), the geons themselves become invariant under changes in viewpoint and visual noise, and allow objects so represented to possess the same invariance. Geon determination requires only categorical classification of edge characteristics, such as whether the edge is straight or curved, rather than precise metric specification, such as degree of curvature or length. Metric judgments cannot be made with sufficient speed or accuracy by humans to be the controlling processes for real-time human object recognition.

Relation to Brooks's (1981) ACRONYM

Perhaps the closest model to RBC is Brooks' (1981) ACRONYM. Like RBC, ACRONYM posits a generalized cylinder characterization of the parts of objects. Unlike RBC, the critical visual information for ACRONYM is the ellipses and ribbons that characterize the cross section and sides of generalized cylinders. These differ metrically so recognition for ACRONYM depends critically on accurate assessment of such quantities. RBC emphasizes nonaccidental qualitative contrasts of edges and classification of vertices. The origin of the different approaches taken by ACRONYM and RBC may lie in the

former's attempt to classify aerial images, where vertices may not be available, into subordinate types such as different models of airplanes, where the determination of metric variation may, in fact, be required because such classes have similar geon descriptions. In contrast, RBC, seeks to model recognition at the entry (or basic) level only. Subordinate level classifications often depend on metric variations.

Relations Among Geons and Models

Simultaneously with the activation of geons, the relations among joined pairs of geons are also detected. The actual composition of these relations is still under development, but RBC assumes that the relations are also viewpoint invariant and categorical, such as "top of" and "center connected". The same subset of geons represent different objects if they are in different relations to each other. Geons thus play a role highly analogous to the role played by phonemes in speech perception. A description of the input consisting of geons plus relations is termed an object model, and it is assumed to activate a similar type of description in memory. For example, one kind of lamp can be described as a cylinder "centered under the larger end" of a cone. Activation is graded in that the activation of a representation will be slower (and of lower maximum value) when an image description differs in geons or relations from the model stored in memory.

Connectionist Implementations of RBC

Distributed Implementation

A six-layer fully distributed connectionist implementation of RBC is currently being developed (Hummel, Biederman, Gerhardstein, and Hilton, 1988). The model takes as input the end points of edges in the central 4 degrees of the visual field. At the lowest level are units that can detect the orientation and termination of image edges at three spatial scales. A hexagonal array of these units feeds into a single unit at the second hidden level. This layer is trained, through back propagation, to develop distributed representations of the local viewpoint-invariant characterizations (namely, vertices) of the image edges. Layer three codes parallelism; the fourth level codes independent distributed representations of geons, geon orientation;

and geon aspect ratio. The fifth level provides a translationally invariant representation of the geons (so that the same representation might be activated regardless of where the image is located on the retina) and their relations. The sixth layer represents objects.

Local Cascade Implementation of RBC

A local cascade simulation of RBC (Biederman, 1987a, 1988) may actually allow closer evaluation of human factors variables in the pilot performance modeling effort. This model assumes that the time course of object recognition is a cascade of three stages: (1) an initial image feature activation layer; (2) an intermediate geon-determination stage (corresponding to layers 2-4 in the distributed model described in the previous section), in which image features activate nodes corresponding to individual geons; and (3) a final stage in which the nodes represent objects. The image of an object is represented by an image-feature vector which specifies the values for the vertices, edges, and geon relations of the object. The model posits that the earlier geon node of image features is transmitted to the activation of nodes representing objects. A given geon node may transmit activation to all object nodes that contain it and inhibition to nodes for objects in which it is not present. An object node will have excitatory connections from those geon nodes that are compatible with it and inhibitory connections from those nodes that are inconsistent with it. The representation of both geons and objects is local in that a single node is presumed to represent a given geon or object. (Although not yet included, relations among geons will be represented by a set of input nodes directly connected to the object layer.) This model is similar to the McClelland and Rumelhart (1981) model of word perception.

Factors reducing image quality, such as contour interruption from small particles (produced, for example, if the object is behind light foliage), low pass filtering, lowered contrast, or small size, are assumed to affect the activation of the image feature nodes that would affect the activation of the geon nodes. In this case, there should be slow growth in the activation of the geon nodes. However, once all the geon nodes are activated, there should be fast and maximum activation of the object node. Factors affecting the similarity of the image to a representation of the object in memory, such as whether an object is missing parts, is occluded by a large surface, is viewed at an unusual angle, is rotated in the plane, or is an unusual exemplar,

are assumed to affect a later stage where activation from the geon (and relations) layer activates the nodes at the object layer. Under these conditions there should be rapid geon activation for the geons that are in the image but slow activation of the object node. If geons are missing in the image, then there will be less activation relayed to the object layer from the geon layer. In the present context where missions will often be performed under conditions of reduced visibility, much of the nonoptimum perceptual performance will be a consequence of diminished quality at the first (feature) stage, rendering it difficult to determine the geons. However, cases in which objects are occluded by surfaces so that no contours of a part are present in the image, for example, will reduce activation at the object layer. Given the availability of image enhancement and restoration by machine, much of the reason for having a pilot in the cockpit is his capability for second-stage processing.

The model, although somewhat elaborate to present in a condensed verbal form, provides a general basis for combining factors that affect image quality with factors that affect the similarity of the image description to the description of objects in memory. Although convenient for summarizing the effects of variables, the local character of the model renders it less realistic as a detailed characterization of human image understanding compared to the distributed model.

Principles of Geon Recovery

It has been estimated that much of basic level recognition can be handled with a vocabulary of not more than 10^6 object models. Are 24 geons sufficient for modeling this many objects? With 24 geons and four classes of viewpoint invariant relations (giving 108 possible combinations of relations), 1.4 billion 3-geon objects models can be generated. A derivation from this analysis is that 3 geons should suffice for the rapid entry level classification of almost any object.

The theory thus suggests a principle of geon recovery: if an arrangement of two or three geons can be recovered from the image, objects can be recognized quickly even when they are occluded, rotated in depth, novel, extensively degraded, or lacking customary detail, color, and texture.

MODEL-BASED MATCHING: LOWE'S SCERPO AND ULLMAN'S ALIGNMENT MODELS

RBC is a one-way, bottom-up model, proceeding from image to activation of the representation of the object. Edge extraction is assumed to be accomplished by a module that can proceed independently of the later stages, except for likely effects of the viewpoint invariant property of smooth curvature.

Does object recognition always proceed as a largely one-way street? Probably not. When edge extraction is difficult, top-down effects are likely to be revealed. Such effects could be of two types: (1) a general source from the viewpoint-invariant properties of cotermination, parallelism, and symmetry or from the geons themselves, and (2) from object models. The latter route is termed model-based matching. Two detailed proposals for such matching have been advanced recently by Lowe (1987) and by Huttenlocher and Ullman (1987). Both the Lowe model and the Huttenlocher and Ullman model differ from RBC in their allowance for transformations, such as rotation, that place the image in spatial correspondence to the model. RBC dispenses with the requirement for such transformations by positing viewpoint-invariant primitives (the geons) and appeals to such transformations only when the initial activation is unsuccessful.

Lowe's SCERPO

A major difficulty for any implementation of a model of recognition is the large number of possible object models that must be evaluated. Lowe's (1987) SCERPO model offers the possibility of constrained search in reducing the computational load posed by large numbers of models.

Lowe's SCERPO model is primarily directed toward the determination of the orientation and location of objects, even when they are partially occluded by other objects, under conditions in which exact object models are available. The model detects edges by finding sharp changes in image intensity values as reflected in the zero crossings of a $\nabla^2 G$ convolution across a number of scales. The edges are then grouped according to viewpoint-invariant properties of collinearity, parallelism, and cotermination. A central assumption in this effort is the viewpoint consistency constraint: "The locations of all projected (object) features in an image must be consistent with projection from a single viewpoint" (Lowe, 1987, p. 57). From the initial detection

of nonaccidental properties of edges. SCERPO proposes a tentative match to an object at a particular orientation (via the viewpoint consistency constraint) and uses predictions from that object to test for additional object features. These matches provide segments not detected initially by the zero crossings and discard edges that were initially detected but are not part of the object model, such as those produced by glare. Matching proceeds in this iterative fashion. A few of these image features are then tentatively matched against a component of the object model in which the orientation of the object that would maximize the fit of those image features is determined.

SCERPO and the alignment model described in the next section may provide a plausible scheme for characterizing human performance under conditions in which the initial extraction of image edges is uncertain (e.g., conditions of poor visibility), the orientation of an object is unfamiliar, or the object is occluded in an unusual fashion.

Alignment

The Huttenlocher and Ullman (1987) alignment model first reorients all the object models that might be possible matches for the image and tests for the fit of the image against the aligned models in memory. The alignment capitalizes on a recent result: three non-coplanar points are generally sufficient to determine the orientation of any object. In practice the three points are typically viewpoint invariant in that they are selected at a point where there is a cotermi-nation of edges. However, any salient points or even general features such as a "wiggly" region would be sufficient for alignment. Although it appears unlikely that people rotate (align) all possible candidate models in memory prior to matching, the alignment model offers a possible account of those cases in which recognition depends on reorientation of a mental model.

Although both of these models show great promise for machine vision, their applicability for real-time entry level classification remains to be evaluated. Unlike humans, neither the Lowe model nor the Huttenlocher and Ullman model reveals any marked difficulty in handling rotation in the plane relative to rotations in depth. Part of the problem is that relations such as "top of" may be made at a level of description other than that of the coding of the contours themselves. These models also do not readily reveal the similarity among instances revealed in human judgment and discrimination.

For example, the alignment model readily rejects a Saab from a similar looking Volkswagon, although people would have some difficulty making that discrimination. The reason for this is that the model relies on metric differences in curvature and extent—judgments that people perform only with great difficulty. The remarkable ability of humans to classify objects based on similar part structures is not obviously captured by these modeling efforts.

Summary of Distinctions Among the Various Models

The theories considered above can be roughly distinguished in the extent to which they posit decomposition, a limited number of primitives, and spatial transformations:

- **Decomposition into parts:** Brooks, Biederman and Pentland assume that complex images are decomposed into parts (e.g., generalized cylinders). Lowe, as well as Huttenlocher and Ullman, assume matching at the level of individual segments (Lowe) or any salient characterization of the object (Huttenlocher and Ullman).

- **Limited number of primitives:** Biederman assumes a limited number of primitives to characterize the image (or parts). The matching of exact metric variation is assumed by Brooks, Ullman and Huttenlocher, Pentland, and Lowe.

- **Transformations:** Huttenlocher and Ullman, as well as Lowe, assume transformation operations for rotation, size scaling, and deformation. Biederman assumes that depth invariance is provided by viewpoint-invariant properties without rotation. Biederman, Pentland, and Brooks assume different models for significantly differing views of a given object.

These assumptions are clearly not mutually exclusive, and it should generally be possible to construct a more elaborate model by specifying the conditions under which one or the other assumption might be appropriate.

Gaps in Research on Quantitative Modeling of Human Object Recognition

Much remains to be done to achieve a working quantitative model of human image understanding. Two important points are listed here:

- early to intermediate vision: How does one go from presumed early filters (e.g., Gabor detectors) to edge extraction and effects of scale and size?
- relations: How can relations among parts of an object be modeled?

Machine Identification of Targets in Low-Resolution Images

All the human and machine models described above have been applied to images that had sufficient resolution for accurate edge extraction, as noted previously. In many of the operating environments for the pilot performance modeling project, recognition will have to be made under conditions of low visibility, (e.g., darkness or fog) or else from a sensor (e.g., infrared, radar, or a laser range finder) that might have low resolution.

Traditional models of pattern recognition and signal processing attempt to classify an image in terms of any set of image values which can provide a diagnostic set of cues for a particular subset of objects that constitute the relevant domain. No attempt is made with these models to reflect human perceptual performance or intuitions. For example, some investigators have sought to correlate components of the spatial frequency spectra of an image with the output of a sensor. Another attempt correlates a global measure, such as the center of mass of a radar image, with possible object classes. Others capitalize on arbitrary features. For example, if only one object has a hole, then this would be used as a diagnostic cue for that object. None of these efforts have been able to achieve accurate classification when the object was rotated in depth or occluded, or when new instances were added to the set of possibilities.

More relevant are those models that seek to achieve recognition through a classification of topological properties of the images (e.g., Koenderink, 1987; Pong et al., 1985). A smooth surface may be classified as a peak, ridge, saddle, flat, ravine, or pit, for example, by the scheme of Pong et al. (1985). The role of smooth surface characterizations in recognition has not been investigated extensively, but a study by Rock and DaVita (1987) indicated that such characterizations (without a readily available geon model) could not be recognized when viewed from another perspective in depth.

Empirical Studies of Human Image Understanding

An extensive series of experiments on the perception of briefly presented pictures by human observers has provided empirical support for the theory. In these experiments the subject names or verifies briefly presented (100 ms) object pictures. Reaction times and errors are the primary dependent variables. The following are some key results:

- Simple line drawings showing only the edges of the major geons are identified as rapidly as full-color, textured images (Biederman and Ju, 1988). This documents the sufficiency of edge-based descriptions over surface (gray scale) variation in accounting for the initial activation of a representation of an object. In general, humans have difficulty in perceiving three-dimensional structure from smooth gray scale variations (without an attached edge) (Todd and Akerstrom, 1987).

- When only two or three geons of a complex object (such as an airplane or elephant) are visible, recognition can be fast and accurate (although, predictably, not as fast as with the complete image). This supports the derivation of the sufficiency of three geons.

- Complex objects requiring six or more geons to appear complete are not recognized any more slowly than simple objects (such as a flashlight or cup). This is consistent with a model positing parallel activation of the geons in favor of a serial contour tracing process, such as eye movements or the kinds of serial routines posited by Ullman (1984).

- If contour is deleted so that an object's geons cannot be recovered from the image (by deleting cusps for parsing and altering vertices), the object is rendered unrecognizable. If the same or a greater amount of contour is deleted but in such a manner that the geons can be recovered through smooth continuation, objects remain identifiable. This result establishes the necessity of the contours posited by RBC.

- A surprising finding in the previous experiment was the large disruptive effect on error rates and reaction times of interrupting (deleting) contour, such as would occur when an object is viewed behind light foliage, even when the contour could be restored by routines for smooth continuation. This suggests that the routines for contour restoration are not particularly rapid.

- In the studies described in the previous paragraph, the contour that was removed was removed from every geon in the object.

Identification performance is also slowed when objects are missing geons (parts), with the rest of the object intact, which would occur if the object was partially occluded by a solid surface. According to the local connectionist model described previously, the effect of missing or occluded geons is on the matching stage, rather than on the initial determination of the geons.

- From separate studies with familiar objects, it can be concluded that rotation of the object in the plane slows recognition to a much greater extent than rotation in depth (in contrast to most robot vision models). However, it is important to determine if this effect holds for unfamiliar objects. According to RBC, rotation in the plane affects the "top of" relation, but the geon descriptions themselves are largely unaffected by rotation in depth.

Gaps in Empirical and Theoretical Research on Object Recognition

A number of important gaps exist in the research on object recognition.

- **Segmentation:** How is segmentation of an object into its parts achieved? Although part segmentation at regions of matched concavities (cusps) is often subjectively compelling, such that a given edge is grouped with its appropriate geon, what are the algorithms by which this is actually achieved? Although cusps offer a strong basis for segmentation, other factors contribute to segmentation as well. In the absence of a concavity, a variation in a nonaccidental property—the change in parallelism at the junction of the base to the nose cone of a rocket, for example—provides a basis for segmentation. Also, parts tend to be fit to elongated regions that are approximately parallel. Is it even necessary to perform segmentation as a separate step? An alternative account is that the image features in a region activate a geon without an independent segmentation process.

- **Scale:** The human appears to be able to organize the image formation at the appropriate scale, ignoring minor, irrelevant variations in the image. How is this achieved?

- **Edge extraction:** There are acceptable (but not perfect) machine routines for the extraction of edges in an image (e.g., Canny, 1986), but the way in which this is achieved biologically has not been determined. A related problem is how the human manages to distinguish texture and crack edges from boundary edges.

- **Metric variations:** Although there is clear evidence for the rapid use of nonaccidental properties, metric variations also have an

effect. For example, if an object has a cylinder as one of its parts, the cylinder will typically be of some aspect ratio. How does performance degrade with departures in aspect ratio from the original value? What is required is a theory that combines the qualitative nonaccidental contrasts with metric variation. The theories reviewed previously provide some initial progress on this problem. Also, different cortical loci have been implicated for these two classes of visual behavior: the inferior temporal cortex is critical for recognition; the posterior parietal cortex for spatial (metric?) processing (Mishkin and Appenzeller, 1987).

- **Spatial relations:** The edges and vertices that comprise a geon exist in some relation to each other. Similarly, the geons comprising an object are in specified relations to each other, such as "top of" or "side connected." What are these relations, and how are they determined and represented?

- **Degraded images:** Most of the research on object perception has employed displays with clear edges, but people can classify a low pass filtered image. How is this achieved? Is performance predictable from the information available (e.g., blob aspect ratio) from the general model (see below), or is another mode of recognition employed, perhaps topological characteristics? This problem is of particular importance to the pilot performance modeling project.

- **Texture:** Many objects include surface texture in their specification. How is texture to be represented?

PERCEPTION OF MULTIOBJECT DISPLAYS

Objects rarely occur in isolation. In some of the multiobject displays currently envisaged, up to 70 potential targets are displayed in a busy environment. How is object recognition affected by the presence of other visual entities in the display? This problem can be decomposed into several subproblems, as suggested by the following outline.

1. potential uncertainty
 - resolution effects due to retinal eccentricity
 - display load effects independent of eccentricity and camouflage;
2. scene constraints;
3. segmentation effects: camouflage.

Positional Uncertainty

Various sections of the above outline are briefly considered here.

Eccentricity

The problem posed by the presence of multiple objects in a display is that the pilot is uncertain as to which one(s) might be target(s). The most obvious effect is that the target may fall outside of foveal vision as the pilot looks at some other object. Knowing where to look for a target results in dramatically higher detectability than when the target's position is uncertain (Biederman, 1972). The fall off in acuity with increasing eccentricity has been well documented, but surprisingly few studies have measured that effect in the context of viewing of a scene. Biederman, Mezzanotte, Rabinowitz, Francolini, and Plude (1981) showed that there was a rapid decline in target detectability even in the modest region between foveal fixation 0-1° and 6-8° degrees eccentricity. This effect was magnified if the targets were small, camouflaged, or incongruous in the scene. This incongruity effect suggest that humans can rapidly employ scene constraints to bolster their parafoveal performance. This human capacity is likely to be most resistant to automation.

Display Load

With fixed eccentricity, is there an effect of the number of other objects in the display on the detection of a particular object? The search literature (e.g., Treisman and Gelade, 1980) suggests that targets can be detected without any effect of the number of distractors if the target differs from the distractors in a feature not shared by the distractors. Thus there will be no effect on the detection of an X because of the presence of O's in the visual field. Search is then said to be "automatic" (Schneider and Shiffrin, 1977). If the target is defined by a conjunction of independent attributes such as color and shape (e.g., a red X target among green X and red O distractors), then there will be a linear increase in search times as a function of the number of distractors (Treisman and Gelade, 1980) and search is said to be capacity limited or "attentive." The issue for the present case is whether objects generally possess attributes that allow them to be distinguished from other objects, or whether the shape primitives are shared, as suggested by RBC, so that attentive search is required.

Biederman, Blicke, Teitelbaum, Klatsky, and Mezzanote (1988) have demonstrated the latter in that there was a marked linear increase in search reaction time, as well as in errors in the detection of target objects in a 100 msec display of one to six objects arranged at the positions of an imaginary clockface. The large magnitude of these effects suggests a serious limitation on human performance and the critical need for cuing relevant targets and exploiting scene constraints.

This chapter has focused on the processing that occurs during a single visual fixation at a scene (or object). Overall visual performance will consist of a series of saccades as the pilot picks various regions of his visual world (including his displays) to fixate. For the most part these fixations cannot be made at a rate greater than 3 to 4 per second. Whether the pilot has to linger longer than the 250 to 333 msec per fixation will depend on the difficulty of resolving image details and the number of objects in the scene that are not integrated by the scene constraints that are discussed below.

Scene Constraints

When an arrangement of objects does not form a scene, as with the clockface displays in the Biederman et al. (1988) experiment, performance degrades rapidly with increasing display size. At the other extreme are scenes that can be perceived "at a glance," with no obvious increase in recognition latency as a function of the number of entities in the scene. The mystery about such scenes is that the exposure duration required for an accurate, integrated representation of their content is not much longer than that typically required to perceive an individual object. However, the recognition of a visual array as a scene requires not only the identification of the various entities but also a semantic specification of the interactions among the objects and an overall semantic specification of the arrangement (e.g., as a kitchen).

Moreover, the perception of a scene is not, in general, derived from an initial identification of individual objects in that scene. That is, generally we do not first identify a stove, refrigerator, and coffee cup in specified physical relations and then come to a conclusion that we are looking at a kitchen (Biederman, 1988).

Some demonstrations and experiments by Mezzanotte (described in Biederman, 1987b, 1988) suggest a possible basis for understanding rapid scene recognition. Mezzanotte showed that a readily interpretable scene could be constructed from arrangements of single geons that just preserved the overall aspect ratio of the object. In these kinds of scenes, none of the entities, when shown in isolation, could be identified as anything other than a simple volumetric body (e.g., a brick). Most important, Mezzanotte found that such settings were sufficient to cause interference effects on the identification speed of intact objects that were inappropriate to the setting.

The rapid recognition of an arrangement of objects as a scene may be mediated by clusters of the largest geons from a familiar arrangement of interacting objects. For example, a vertical slab appearing behind a large brick is readily interpreted as a desk and chairback. In such cases, the individual geons are insufficient to allow identification of the object. However, just as an arrangement of two or three geons almost always allows identification of an object, an arrangement of two or more geons from different objects may produce a recognizable combination. The cluster acts very much as a large object. If this account is true, fast scene perception should be possible only when such familiar object clusters are present. This account awaits empirical test.

Segmentation

Another effect of the presence of more than one entity in a scene is the possibility that the difficulty of segmenting an object from its background may increase. The potential heterogeneous nature of the source of this difficulty has not been well explored. For example, in some cases the difficulty arises because of reduced differentiation between target and immediately adjacent contour, as when adjacent objects share the same texture. In other cases, the neighborhood is organized in such a way that the target is incorporated into the context, as can be produced with patterns in the Embedded-Figures test.

Gaps in Knowledge of the Perception of Multiobject Displays

Several important gaps exist in our knowledge of multiobject display perception:

- Tests of geon clusters: The geon cluster hypothesis for integrated scene perception requires empirical confirmation.
- Attentional costs: Why is there no evidence for attentional costs in the perception of complex objects compared to simple objects? Are attentional costs balanced by greater information, or does the attention to a region overcome the effect of the number of geons?
- Accessing scene constraints: How can the human's extraordinary knowledge of real-world scenes be represented and accessed efficiently? How might it be exploited to overcome the effect of attentional load?

REFERENCES

- Biederman, I.
 1972 Perceiving real-world scenes. *Science* 177:77-90.
 1987a Matching image edges to object memory. *Proceedings of the First International Conference on Computer Vision* 384-392.
 1987b Recognition-by-components: A theory of human image understanding. *Psychological Review* 94:115-147.
 1988 Aspects and extensions of a theory of human image understanding. In Z. Pylyshyn, ed., *Computational Processes in Human Vision*. New York: Ablex.
- Biederman, I., Blicke, T.W., Teitelbaum, R.C., Klatsky, G.J., and Mezzanotte, R.J.
 1988 Object identification in nonscene displays. *Journal of Experimental Psychology: Human Learning, Memory, and Cognition* 14:456-467.
- Biederman, I., and Ju, G.
 1988 Surface versus edge-based determinants of visual recognition. *Cognitive Psychology* 20:38-64.
- Biederman, I., Mezzanotte, R.J., Rabinowitz, J.C., Francolini, C.M., and Plude, D.
 1981 Detecting the unexpected in photointerpretation. *Human Factors* 23:153-164.
- Brooks, R.A.
 1981 Symbolic reasoning among 3-D models and 2-D images. *Artificial Intelligence* 17:205-244.
- Canny, J.
 1986 A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8(6):679-698.
- Hoffman, D.D., and Richards, W.
 1985 Parts of recognition. *Cognition* 18:65-96.
- Hummel, J.E., Biederman, I., Gerhardstein, P., and Hilton, H.J.
 1988 A connectionist model of Recognition-by-Components. Paper presented at the Neural Architectures for Computer Vision Workshop at the Meetings of the Seventh National Conference on Artificial Intelligence. St. Paul, MN.

- Huttenlocher, D.P., and Ullman, S.
1987 Object recognition using alignment. *Proceedings of the First International Conference on Computer Vision*. London, England: IEEE Computer Society.
- Koenderink, J.J.
1987 An internal representation for solid shape based on the topological properties of the apparent contour. Pp. 257-285 in W. Richards and S. Ullman, eds., *Image Understanding*. Norwood, NJ: Ablex.
- Lowe, D.G.
1984 Perceptual Organization and Visual Recognition. Unpublished doctoral dissertation. Department of Computer Science. Stanford University, Stanford, CA.
1987 The viewpoint consistency constraint. *International Journal of Computer Vision* 1:57-72.
- McClelland, J.L., and Rumelhart, D.E.
1981 An interactive model of context effects in letter perception: An account of basic findings. *Psychological Review* 88:375-407.
- Mishkin, M., and Appenzeller, T.
1987 The anatomy of memory. *Scientific American* 256:80-89.
- Pentland, A.P.
1986 Perceptual organization and the representation of natural form. *Artificial Intelligence* 28:293-331.
- Pong, T.C., Shapiro, L.G., and Haralick, R.M.
1985 Shape estimation from topographic primal sketch. *Pattern Recognition* 18:333-347.
- Rock, I., and DiVita, J.
1987 A case of viewer-centered object perception. *Cognitive Psychology* 19:280-293.
- Schneider, W., and Shiffrin, R.M.
1977 Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review* 84:1-66.
- Todd, J.T., and Akerstrom, R.
1987 Perception of three-dimensional form from patterns of optical texture. *Journal of Experimental Psychology: Human Perception and Performance* 13(2):242-255.
- Treisman, A.M., and Gelade, G.
1980 A feature-integration theory of attention. *Cognitive Psychology* 12:97-136.
- Ullman, S.
1984 Visual routines. *Cognition* 18:97-159.

Manipulation of Visual Information

LYNN A. COOPER

SUMMARY

The ability to transform or manipulate visual information is a perceptual-cognitive skill of central importance in normal perceptual processing and in perceptually driven tasks requiring the use of imagery or the comparison of spatially transformed visual input to a representation of that input in memory. In most cases, image transformation occurs at a level of processing following and relying upon object identification; however, some forms of manipulation of visual information (e.g., integration and transformation of different views of an object) may be involved in the process of identification. For certain pilot performance problems (including those consisting of detection and identification), image manipulation is unlikely to be an important component of operation. For other tasks facing the pilot (including aspects of navigation, localization of a target in a visual array, and comparison of current visual input with previously available views), transformation of visual information may play a central role in performance.

A substantial body of experimental work exists on perceptual and cognitive tasks requiring the transformation of visual information and is briefly reviewed in this report. Research has, for the most part, been directed toward delineating the information-processing consequences of transforming spatial information in terms of time and accuracy constraints on performance. There is considerable evidence concerning the effects of various display and task parameters on the amount of time in which, and the accuracy with which, visual information can be transformed. Furthermore, the process of image transformation can often be shown to conform to highly regular and mathematically straightforward relationships. For cases

in which errors of transformation occur frequently, the magnitude and direction of error often follow a highly predictable pattern. Yet, despite the large body of systematic experimental results, general computational models are still scarce. Those models that have been specified suffer from being stimulus or task specific, and they have generally been based on some single index of task performance.

In sum, although there has been considerable progress in understanding—at a quantitative level—the nature, time course, and limitations on the ability to manipulate spatial information, as well as the various factors that affect different aspects of performance on tasks requiring spatial transformations, as yet no set of large-scale models of image manipulation exists. A further limitation on the applicability of current data and models to pilot performance tasks is the reduced conditions under which data have been obtained in terms of both display richness and concurrent processing demands.

INTRODUCTION

This chapter briefly summarizes relevant empirical research and formal models of performance in laboratory situations that are related to certain pilot performance problems. In particular, the research and models reviewed address perceptual and cognitive capabilities of human observers in transforming or manipulating information presented in the form of a visual display. Most of the research has been directed toward characterizing at a quantitative level the nature and magnitude of errors produced in spatial manipulation tasks. The limits of performance and a delineation of stimulus and task conditions that lead to breakdowns in performance are emphasized. Models of performance and of task conditions are scarce, often taking the form of simple equations to fit observed performance functions. So, although considerable empirical research has been undertaken, there are few computational models available for consideration. In addition, no small leap of faith is required to apply the research reviewed to problems encountered in real-world pilot performance. Experimental work has generally been done in several limited display environments, with little or nothing in the way of additional or competing tasks (except a single judgment about the transformation of a single object) to be performed.

This chapter is divided into four sections. In the first and most substantial section, research on and models for the manipulation of information presented in a static visual display (mental rotation

tasks) are discussed. In the second, relevant work on memory for spatial positions in sequences of static displays is described. In the third, recent research is presented that directly examines the abilities of observers to extrapolate trajectories of projections of objects displayed dynamically. In the fourth section, work on the computation of object structure from partial information about views is considered. The goal is to present a reasonably comprehensive (but not exhaustive) review of relevant literature, highlighting the most significant empirical findings and pointing to models in domains where they have been developed.

TRANSFORMATIONS ON INFORMATION PRESENTED IN A STATIC VISUAL DISPLAY

One of the more robust findings in the literature in cognitive psychology concerns the relationship between performance (measured in time and accuracy) in judging some aspect of a disoriented visual display of an object and the extent of displacement of the object from a canonical or a previously learned position. The amount of time required to determine, for example, whether an object is "standard" or "reflected" in parity increases linearly with the magnitude of the angular difference between the object's displayed orientation and a familiar position. This basic linear relationship between processing time and angular difference holds whether visual stimuli are presented simultaneously (Shepard and Metzler, 1971) or successively (Cooper, 1975)—requiring a comparison of an object with a stored memorial representation, whether the objects transformed are portrayed as two or three-dimensional; whether the rotational transformation itself is in the picture plane or in depth; and, to some extent, regardless of the visual complexity of the objects (Cooper and Podgorny, 1976). Shepard and Cooper (1982) provide a relatively comprehensive, though slightly dated, review of this literature.

This basic finding suggests that the computational cost of mentally transforming a disoriented object can be expressed simply by the linear reaction time function. Although the stimulus parameters discussed above do not, in general, affect the shape of the performance function, they do have discernible effects on both the slope of the function (inferred to measure the rate at which correctional transformations can be carried out) and the intercept (a measure of the time to encode the visual display). Mode of presentation can affect both the slope and the intercept; stimulus complexity and the presence of

landmark features can affect the rate of transformation (Hochberg and Gellman, 1977); and stimulus and transformational dimensionality have questionable effects on both slope and intercept. Estimated rates of mental rotation reported by various investigators for a host of stimulus and presentation conditions range from approximately 60 degrees (for perspective drawings of three-dimensional objects and three-dimensional transformations) to over 500 degrees (for highly practiced subjects transforming simple two-dimensional stimuli) per second.

A theoretical framework that has been proposed to account for these data, which generally takes the linearity of the relation between time and angular displacement as evidence for an internal analog or simulation of the process of physical rotation in the specific sense of passing through intermediate positions in a transformational trajectory that correspond to intermediate stages in the physical rotation of an object, has been demonstrated (Cooper, 1976). The basic finding of Cooper's experiment was that the time to respond to a disoriented object is essentially constant if the object is presented in an expected position, in the sense of being congruent with the currently assumed position of an internal representation of the object that the subject imagined rotating at a particular rate in a particular direction.

Simple linear relations between time for correctional processing and spatial extent have also been reported for transformations other than rotation. Bundesen and Larson (1975), Bundesen, Larson, and Farrell (1981), and Sekular and Nash (1972) have all demonstrated linear relations between the time required to compare two objects of different size and the ratio of the size differences (but see Kubovy and Podgorny, 1981), and combinations of size and rotational transformations contribute additively to comparison times under some circumstances (Bunden et al., 1981). Kosslyn (1973; Kosslyn, Ball and Reiser, 1978) has shown a linear relation between the time required to "mentally scan" from one location to another in an array of objects and the metric distance between the objects in the scan path. Further evidence for the analog nature of translational mental operations is provided by Shulman, Remington, and McLean (1979) in a task requiring the shifting of attention from one location to another.

A host of additional questions that could bear on pilot performance issues can be asked about the nature and time course of correctional mental operations on disoriented or misaligned visual displays. Two that are presently unresolved in the literature but

that have produced some empirical evidence concern (1) whether transformations take time in proportion to proximal or to distal variables and (2) whether transformations of abstract frames of reference can be carried out. With respect to the relative importance of proxima and distal distance, the original mental rotation experiments (Shepard and Cooper, 1982; Shepard and Metzler, 1971) suggest strongly that the relevant distance between two positions over which reaction time increases linearly is the distance between the positions of the two objects in three-dimensional space, rather than the distance between the two objects as projected on the retina (when the two sorts of measured distances are different). Corballis and his associates (Corballis and Roldan, 1975; Corballis, Zbrodoff, and Roldan, 1976) have asked whether mental rotation of a disoriented object occurs to the retinal or the gravitational upright, when the two are different by virtue of head tilt. For visual patterns of familiar objects with an overlearned canonical position in the world, rotation appears to be to gravitational upright, but with unfamiliar complex dot patterns, rotation is carried out to achieve congruence with the retinally defined vertical. Other investigations of the operation of proximally defined versus distally defined distance (in the context of a mental scanning task) indicate that instructions can effectively alter the character of the scan path: when a subject is instructed to imagine scanning between two objects located in three-dimensional space, time increases with distal distance; however, when a subject is instructed to scan from the visual direction of one object to the visual direction of another, time increases linearly with distance in the two-dimensional projection (Pinker, 1980; Pinker and Finke, 1980; Pinker and Kosslyn, 1978).

With respect to the question of whether transformations can be carried out on an abstract frame of reference as opposed to a representation of a particular visual object, experiments by Cooper and Shepard (1973) suggest that such an overall transformation of a coordinate system cannot be done effectively to prepare for the presentation of a disoriented test object. Providing time and the proper information to enable the transformation to be done in advance lowers subsequent reaction time, but the decrease does not change with the magnitude of the angular displacement of the prepared-for position. Subsequent experiments by Jolicoeur (1983) indicate that frames of reference can be transformed in advance when the type of stimulus and type of orientation are known, and the transformation

involves assuming the next in a series of well-defined spatial positions. Note that manipulation of a frame of reference could be an important component of performance in reorienting after "pop-up"; thus, it is important to have a more definitive evaluation of this issue at the basic research level.

In addition to the basic analog model of rotation and related spatial transformations proposed by Shepard, Cooper, and their collaborators, other sorts of models have been offered to account for the data from transformation experiments that assume a discrete representation of a visual object and incremental transformations applied to subparts of the representation (e.g., Anderson, 1978; Just and Carpenter, 1976). The most detailed of these alternative models has been presented by Just and Carpenter (1976) and Carpenter and Just (1978) and is based on an analysis of patterns of eye fixations made during performance of a mental rotation task, similar to that studied by Shepard and Metzler (1971), in which two visual displays differing in orientation are compared with respect to shape. The process model that these investigators propose postulates three successive stages in carrying out transformations on objects presented spatially. In the first "search" stage, sections of the figures that are in potential correspondence are located. In the second "transformation and comparison" stage, segments that are taken to correspond in the two figures are mentally rotated, and a sequence of comparisons is made to determine when the orientations of the segments correspond. The transformations and comparisons are incremental, occurring about every 50 degrees of rotation. In the final "confirmation" stage, a determination is made of whether the other segments of the figure correspond as a result of the transformation. Thus, although this model departs substantially from the analog account, it does fulfill the criterion of an analog process outlined by Cooper (1976) and Cooper and Shepard (1973), but the succession of intermediate positions assumed is by a representation of portions of a visual figure rather than of an integrated representation. More recently, Just and Carpenter (1985) presented a detailed account of performance on a cube comparison task that requires transformations on visual objects. The model is designed to describe differences in performance between individuals of high and low measured spatial aptitude, and it is embodied in a running simulation. The central difference between the two aptitude groups resides in the coordinate system adopted for representing and transforming spatial objects. Note that since this model is designed specifically to account for group differences by

strategy differences, its usefulness in predicting across performance, given a particular stimulus as input, is minimal.

A final example of a model that might be applied to transformations on visual information has recently been proposed by Kosslyn (1987). This qualitative model is a very general account of perceiving and imagining which assumes that different (neural) subsystems encode relations among parts of an object in a categorical fashion (i.e., top-bottom, right-left relations) and in terms of their actual coordinates (metric relations). Presumably, both subsystems are involved in the realignment of disoriented objects, with the categorical relations subsystem enabling comparisons of current relations with stored ones and the coordinate encoding subsystem enabling a precise computation of the position of all parts of an object in space.

MEMORY FOR POSITIONS IN A SEQUENCE OF STATIC DISPLAYS

A second body of empirical work that may be relevant to pilot performance issues addresses accuracy of memory for the last of a series of visual stimuli presented in a sequence of ordered positions with temporal parameters such that the sequence implies directional motion at a particular rate. In this work by Freyd, Finke, and their collaborators (Finke and Shyi, in press; Freyd, 1983, 1987; Freyd and Finke, 1984; Freyd and Johnson, 1987), observers view a sequence of rectangles or dot pattern stimuli discretely presented in successive orientations that specify rotation in the picture plane. Some variable time after offset of a final stimulus in the sequence, a test stimulus is presented, and observers must judge whether or not it is in the same position as the final stimulus in the sequence. The general finding is that errors in memory for the final position are not randomly distributed, but rather have a tendency to occur to test stimuli in positions slightly ahead of the actual position of the final stimulus. This distortion in memory for final position appears to be attributable to the implied motion of the sequence of discrete inducing displays has been called representational momentum. The theoretical framework in which these memory distortions have been cast views representational momentum as very loosely analogous to physical momentum, and there is some evidence for a weak form of this analogy. In particular, the memory distortions increase in proportion to the implied velocity (Finke, Freyd, and Shyi, 1986), and when the applied velocity changes (suggesting decelerating or

accelerating motion), the distortion is related to the final velocity implied by the inducing sequence (Finke et al., 1986).

Freyd and Johnson (1987) have specified quantitative models of the physical process of stopping that predict both the slope of the line relating magnitude of memory distortion to implied velocity and the asymptotic level achieved for different (retention) intervals between the final display in a sequence and the test display. Their preferred physical model, combined with a model that specifies a memory averaging component, does a reasonable job in accounting for data from a series of parametric experiments manipulating inducing interstimulus intervals and retention intervals. Models such as these, based on equations familiar from physics, may be candidates for describing position errors. However, it should be noted that the magnitude of the "error" is small (the largest estimate reported is 2 degrees, and most estimates are well below 1 degree (see Cooper, Gibson, Mowafy, and Tataryn, 1987), and the distortion is revealed by asymmetries in performance functions, rather than by shifts in peaks of responding from the correct position to the distorted position. Furthermore, the estimates of memory shifts are obtained by fitting the data with quadratic equations, which generally do not provide impressive fits.

EXTRAPOLATION OF PERCEPTUALLY DRIVEN SPATIAL TRANSFORMATIONS

A perceptual situation somewhat similar to the memory tasks described above, but which approximates better possible demands on pilot performance, is one in which observers must extrapolate the trajectory of an object shown undergoing a spatial transformation. Cooper (in press) and Cooper et al. (1987) have provided reports of the initial results of such a program of research. In the experimental situation observers view a drawing of a three-dimensional object rotating rigidly; at some randomly determined point in the rotation the object disappears. Some time after the disappearance, the object reappears, and observers must judge whether or not the position of reappearance is at the correct point in the transformational trajectory, if the rotation continued at constant velocity during the blank interval.

The general finding of these experiments is that observers judge as "correct" reappearances, undershoots of the actual position at which the object should reappear. The magnitude of the extrapolation error is approximately 6 degrees of negative displacement from

the correct point of reappearance but increases as the duration of the blank interval increases from 300 to 1200 milliseconds. The extrapolation error is substantial and robust. It is reflected in a true shift in the peak of the response function; that is, the probability of responding "correct" to a displacement of -6 degrees is greater than the probability of responding "correct" to the true position of extrapolated reappearance. The negative shift is obtained both for rotations in the picture plane (in which the projected structure of the object does not change at different reappearance positions) and for rotations in depth (in which the projected structure does not change at different reappearance positions). The extrapolation error does not appear to depend on the amount of immediate exposure to the display, because it occurs in a similar fashion when the blank interval is placed in the first or in the second revolution of the object. Furthermore, over a still limited range of velocities examined, the error does not appear to be influenced substantially by the constant depicted velocity of the rotating object.

These data are not well accounted for by models based on the projected two-dimensional distance between corresponding edges of the object before and after the blank interval; as with the mental rotation work, distally measured distance provides a better account at all values of the blank interval. Furthermore, these data are consistent with those reported by Finke and Shyi (in press), who find that slight undershoots characterize the nature of memory errors when the static, sequential "representational momentum" task is performed with instructions to extrapolate the position of the last display in the sequence in judging the accuracy of the position of the test display. However, Cooper et al. (1987) have reported that performance is extremely poor when subjects are instructed to extrapolate the implied motion of sequences of static displays.

Considerable additional work is needed before the conditions under which extrapolation errors occur and their dependence on stimulus and judgmental factors can be modeled. Other work in which extrapolation of single objects moving at constant velocity has been assessed has generally shown quite accurate performance (e.g., Cooper, 1976; Jagacinski, Johnson and Miller, 1983; Rosenbaum, 1975). However, the experimental situations used in these other investigations differed substantially from those of the Cooper (Cooper, in press; Cooper et al., 1987) experiments.

One general limitation on models proposed to account for data from extrapolation experiments or the "representational momentum"

phenomenon discussed in the previous section concerns the task specificity of such models. That is, most accounts of errors in remembering or extrapolating trajectories of motion make reference to the internalization of some principle or set of principles governing the physical motion of objects in the world. However, which principles of physics are internalized, and how, or under what circumstances, does such internalization occur? Shepard (1984) has offered a general argument and empirical support for the position that the perceptual system (and the cognitive system, in the absence of external perceptual support) has internalized principles of kinematic geometry. Work on the conceptions that naive subjects have concerning the continuing trajectories of moving objects (e.g., Caramazza, McCloskey, and Green, 1981; McCloskey, Washburn, and Felch, 1983) suggests that errors of judgment occur frequently and may be systematic. In the absence of a principled theoretical account of which physical laws are internalized in perceiving, remembering, and reasoning about the motion of objects in space, models of processes like extrapolation and memory for position will necessarily remain specific to the particular display and task features of the experiments in which these processes are assessed.

JUDGMENTS OF OBJECT STRUCTURE FROM PARTIAL VIEWS

One final line of research only marginally related to transformations on visual objects, but potentially relevant to a class of pilot performance problems, concerns the extent to which the structure of visual objects can be determined from partial information. The types of partial information used in these experiments (Cooper, Mowafy, and Stevens, 1986) are those that might be sampled as an observer moves in the environment or views an object in motion, rather than the kind of partial information that occasions low levels of illumination or brief stimulus exposures. Subjects solved problems based on orthographic views of objects and were then asked to make forced choice recognition of isometric views of the objects that would have been formed by the orthographic views shown during problem solving and structurally similar distractor isometrics. Performance on the recognition task was excellent, even though no previous exposure to the isometric views of the objects had occurred. This suggests that the process of reasoning with flat, separated orthographic views

involves the mental construction of a three-dimensional or isometric-like model of the object that is structurally veridical enough to permit discrimination from a similar distractor structure.

Of particular interest for purposes of the present chapter, is that the recognition of correct isometrics was quite accurate, even when those isometrics depicted views of the object that did not correspond to the particular set of orthographics presented during problem solving. That is, subjects could correctly discriminate isometrics that shared only two views in common with the particular orthographics previously displayed at a level almost equal to that obtained when the test isometrics shared all three views in common with the set of orthographics. There is computational cost involved in inferring this "hidden" structure of constructed mental representations of objects: the time required to make the discrimination increased considerably when only two (as opposed to all three) surfaces were shared. In addition, as the number of shared sides fell below two, accuracy also declined until performance was essentially at a chance level when the test isometrics shared no surfaces in common with the isometric that corresponded to the three orthographic projections initially displayed. Results such as these indicate that inferences about the spatial structure of objects not immediately externally available can be made at some level of accuracy. However, both the extent to which underlying mental representations of objects can be characterized as view independent and the nature of constraints on the ability to make these inferences about partially concealed structure remains unclear.

FUTURE RESEARCH

Many of the limitations of existing models and data on manipulations of visual information for application to pilot performance problems have been mentioned in previous sections; these limitations provide some guidelines for future research directions.

First, although considerable experimental data exist that could be useful in partial simulation, static analysis, and rapid experimentation on pilot performance problems, these results have generally been obtained in severely constrained display and task environments. It is commonplace to assert that psychological research should strive to be more "ecologically valid"; in the case of the research reviewed here, even minor modifications of visual displays and performance demands could have substantial consequences for applicability of

data to pilot performance situations. Most of the data that serve as a basis for models of transformations on visual objects in space have considered only single transformations applied to single objects. There are notable exceptions (e.g., the work of Kolers and Perkins, 1969, on transformations on entire lines of text and literature on "cognitive maps," not reviewed in this chapter), but research on transformations on arrays of objects and observers in relation to arrays of objects that might be encountered in a natural scene would seem to be a promising direction. In addition, the transformations studied have generally involved rigid rotation, translation, size scaling, or (rarely) some combination of these simple transformations. The use of multiple and more complex transformations, including nonrigid transformations, would seem important at a theoretical as well as at the practical level of providing more realistic simulations of what a pilot might actually be exposed to. Coupling tasks requiring judgments about transformations on objects or extrapolations of trajectories of motion with additional attention-demanding tasks could also provide information about how concurrent processing demands influence both the time and the accuracy of performance. All of the research directions mentioned constitute fairly natural expansions and extensions of a number of ongoing programs.

In addition to enriching the data base from which models can be developed, a more vigorous modeling effort is required. Furthermore, models should reflect human performance characteristics and provide insight into the nature of internal representations and mechanisms that produce the observed performance. Many current models are qualitative or simple quantitative descriptions of psychophysical functions. Enlarging the scope of these models and providing more comprehensive models of interactions between transformation processes and processes of object identification, for example, is an important goal for future research efforts.

There is a need to provide general and principled theoretical accounts of which kinds of physical processes operating on objects in the world might be internalized by the perceptual system, and of how and why such internalization takes place. Finally, the extent to which accurate anticipation of the transformations of objects might occur in perceptually guided situations, but not in situations requiring reasoning or cognitive activity removed from immediate perceptual input, should be examined both experimentally and theoretically.

REFERENCES

- Anderson, J.R.
 1978 Arguments concerning representations for mental imagery. *Psychological Review* 85:249-277.
- Bundesen, C., and Larsen, A.
 1975 Visual transformation of size. *Journal of Experimental Psychology: Human Perception and Performance* 1:214-220.
- Bundesen, C., Larsen, A., and Farrell, J.
 1981 Visual transformation of size and orientation: Spatial imagery and perceptual processes. In J. Long and A. Baddeley, eds., *Attention and Performance IX*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Caramazza, A., McCloskey, M., and Green, B.
 1981 Naive beliefs in "sophisticated" subjects: Misconceptions about trajectories of objects. *Cognition* 9:117-123.
- Carpenter, P.A., and Just, M.A.
 1978 Eye fixations during mental rotation. In J. Senders, R. Monty, and D. Fisher, eds., *Eye Movements and Psychological Processes II*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Cooper, L.A.
 1975 Mental rotation of random two-dimensional shapes. *Cognitive Psychology* 7:20-43.
 1976 Demonstration of a mental analog of an external rotation. *Perception and Psychophysics* 19:296-302.
 in press Mental models of the structure of visual objects. In B. Shepp and S. Ballesteros, eds., *Object Perception: Structure and Process*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Cooper, L.A., Gibson, B.S., Mowafy, L., and Tataryn, D.J.
 1987 Mental extrapolation of perceptually-driven spatial transformations. Paper presented at the 28th annual meeting of the Psychonomic Society, November 6-9, Seattle, Washington.
- Cooper, L.A., Mowafy, L., and Stevens, J.J.
 1986 Perceiving hidden structure: View-independent recognition of constructed mental representations. Paper presented at the 27th annual meeting of the Psychonomic Society, November 13-16, New Orleans, Louisiana.
- Cooper, L.A., and Podgorny, P.
 1976 Mental transformations and visual comparison processes: Effects of complexity and similarity. *Journal of Experimental Psychology: Human Perception and Performance* 2:503-514.
- Cooper, L.A., and Shepard, R.N.
 1973 Chronometric studies of the rotation of mental images. In W.G. Chase, ed., *Visual Information Processing*. New York: Academic Press.
- Corballis, M.C., and Roldan, C.E.
 1975 Detection of symmetry as a function of angular orientation. *Journal of Experimental Psychology: Human Perception and Performance* 1:221-230.
- Corballis, M.C., Zbrodoff, J., and Roldan, C.E.
 1976 What's up in mental rotation? *Perception and Psychophysics* 19:525-530.

- Finke, R.A., Freyd, J.J., and Shyi, G.C.W.
1986 Implied velocity and acceleration induce transformations of visual memory. *Journal of Experimental Psychology: General* 115:175-188.
- Finke, R.A., and Shyi, G.C.W.
in press Mental extrapolation and representational momentum for complex implied motions. *Journal of Experimental Psychology: Learning, Memory and Cognition* 14:112-120.
- Freyd, J.J.
1983 The mental representation of movement when static stimuli are viewed. *Perception and Psychophysics* 23:575-581.
1987 Dynamic mental representations. *Psychological Review* 94:427-438.
- Freyd, J.J., and Finke, R.A.
1984 Representational momentum. *Journal of Experimental Psychology: Learning, Memory and Cognition* 10:126-132.
- Freyd, J.J., and Johnson, J.Q.
1987 Probing the time course of representational momentum. *Journal of Experimental Psychology: Learning, Memory and Cognition* 13:259-268.
- Hochberg, J., and Gellman, L.
1977 The effect of landmark features on "mental rotation." *Memory & Cognition* 5:23-26.
- Jagacinski, R.J., Johnson, W.W., and Miller, R.A.
1983 Quantifying the cognitive trajectories of extrapolated movements. *Journal of Experimental Psychology: Human Perception and Performance* 9:43-57.
- Jolicoeur, P.
1983 Abstract frames of reference in the identification of disoriented shapes. Research Bulletin, Department of Psychology, University of Saskatchewan, Saskatchewan, Canada.
- Just, M.A., and Carpenter, P.A.
1976 Eye fixations and cognitive processes. *Cognitive Psychology* 8:441-480.
1985 Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review* 92:137-171.
- Kolers, P.A., and Perkins, D.N.
1969 Orientation of letters and their speed of recognition. *Perception and Psychophysics* 5:275-280.
- Kosslyn, S.M.
1973 Scanning visual images: Some structural implications. *Perception and Psychophysics* 14:90-94.
1987 Seeing and imaging in the cerebral hemispheres: A computational approach. *Psychological Review* 94:148-175.
- Kosslyn, S.M., Ball, T.M., and Reiser, B.J.
1978 Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance* 4:47-60.
- Kubovy, M., and Podgorny, P.
1981 Does pattern matching require the normalization of size and orientation? *Perception and Psychophysics* 30:24-28.

- McCloskey, M., Washburn, A., and Felch, L.
1983 Intuitive physics: The straight-down belief and its origin. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 9:636-649.
- Pinker, S.
1980 Mental imagery and the third dimension. *Journal of Experimental Psychology: General* 109:354-371.
- Pinker, S., and Finke, R.A.
1980 Emergent two-dimensional patterns in images rotated in depth. *Journal of Experimental Psychology: Human Perception and Performance* 6:244-264.
- Pinker, S., and Kosslyn, S.M.
1978 The representation and manipulation of three-dimensional space in mental images. *Journal of Mental Imagery* 2:69-84.
- Rosenbaum, D.A.
1975 Perception and extrapolation of velocity and acceleration. *Journal of Experimental Psychology: Human Perception and Performance* 1:395-403.
- Sekular, R., and Nash, D.
1972 Speed of size scaling in human vision. *Psychonomic Science* 27:93-94.
- Shepard, R.N.
1984 Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review* 91:417-447.
- Shepard, R.N., and Cooper, L.A.
1982 *Mental images and their transformations*. Cambridge, Massachusetts: MIT Press (Bradford Books).
- Shepard, R.N., and Metzler, J.
1971 Mental rotation of three-dimensional objects. *Science* 171:701-703.
- Shulman, G.L., Remington, R.W., and Mclean, J.P.
1979 Moving attention through visual space. *Journal of Experimental Psychology: Human Perception and Performance* 5:522-526.

11

Combining Views

JULIAN HOCHBERG

Two quite different kinds of processes for combining views are required in the set of problems associated with nap of the earth (NOE) helicopter design: (1) the integration of successive views into a large or more inclusive perceived layout and (2) the combination of binocular views into a single cyclopean field from which information from the two individual views may or may not be retrievable. Models seem possible in restricted areas of both processes, but none currently meets the criteria listed in the introduction.

INTEGRATION OF SUCCESSIVE VIEWS

The necessity of combining information from successive views pervades virtually all perceptual tasks: in any single glance, the eye provides detailed vision from the larger surrounding periphery. When more information is needed than can be obtained in one glance, the eye moves by ballistic saccades, at rates usually less than 4 per second, bringing to the fovea a preselected part of the field previously seen in peripheral vision. Therefore, information about a single object, layout, or event is usually obtained by means of several glances, each directed at a different place in space.

Ubiquitous though it is, this complex performance concerns the designer and the pilot performance model primarily in three ways:

- In relation to free viewing, the complex movements of the eye, body, and target, compounded by interrupted illumination, may make it difficult or impossible to relate successive visual samples to each other within a coherent directional framework. Some of the potentially offending conditions can be identified and are probably relatively easy to model: for example, the way in which brief flashes of light, or stroboscopic presentations, confuse the registration of

gaze direction (Matin, 1986) and defeat size constancy (Rogowitz, 1984).

Designer remedies may include judicious distribution of landmarks throughout the field of view in question (e.g., the cockpit interior) that are readily distinguished in peripheral vision, can be rapidly identified in search, and form an easily learned spatial map (cf. Finke and Shepard, 1986; Henderson, Pollatsek, and Rayner, 1987; Hochberg and Gellman, 1977; Treisman, 1986). Although there is scattered research literature on many of the components needed to model these processes (see Finke and Shepard, 1986; Humphreys and Quinlan, 1987; Stevens, 1987; Treisman, 1986; Ullman, 1985; Wickens, this volume pp. 191-193 for recent reviews), nothing that approaches an overall model that could be image driven and that would provide quantitative output appears to exist today.

- Given the substantial time required by each glance, the number of saccades, their frequency, and the fixation dwell times that they need must be taken into account in the design wherever visual information is densely packed and widely spread, as in low-redundancy text or alphanumeric arrays that require foveal detail such as instruments or details of the environment. The number, sequence, and distribution of glances executed and the time devoted to each glance are variables that depend complexly on task, stimulus variables, and viewer variables (see reviews by Moray, 1986; Senders, 1983; Wickens, this volume pp. 191-193).

The most active models here are those pursued in attempting to predict dwell time in reading (for reviews, see Carr, 1986; Just and Carpenter, 1980; Rayner, 1978) in a tradition that goes back to Judd and Buswell (1922). These attempts may provide a foundation for, but are not themselves directly applicable to, the informational arrays of cockpit instruments on the helmet display.

- Artificial displays that sample the field of view (or scroll through an alphanumeric array in a saltatory manner), including motion pictures, often do so through markedly discontinuous changes. In movies, these are "cuts", and there is much lore about how to make them comprehensible (Bordwell, 1985; Hochberg, 1986; Monaco, 1977; Reisz and Miller, 1968; Vorkapich, 1972). In computer-generated images (CGI) and in cockpit video, they may reflect low update rates chosen to accommodate limited bandwidth or computer speed (as in simulation, night views, and enhanced terrain displays); or they may occur because of abrupt changes in remote camera direction; and they are often used deliberately to present layouts that

are larger than the display screen (cuts, saltatory scrolls and zooms, etc.) or merely when changing from one array to another.

In normal vision in the world, abrupt view changes result primarily from saccadic eye movements. The problem of combination of views has been approached almost exclusively in terms of using information about those preprogrammed movements to compensate for the image's spatial displacement (see Matin, 1986, for recent review). When the view change is part of the display and not the result of programmed eye movements, some other explanation of our ability to integrate the views is needed.

A first step toward an explanation (but not a model) is Gibson's proposal that the visual overlap between views "specifies" the overall optic array that those views sample (Gibson, 1950, 1979). This does not indicate when and why the process fails, or what kinds of perceptual errors then arise. To do that, a quantitative account is required of the various ways in which such information can be provided and used (e.g., low-level apparent motion, landmarks, swishpan; Hochberg, 1986). No single model has been formulated that will do this. Indeed, several distinct levels are involved in the process: some of the early levels have been modeled, essentially providing for apparent motion (Braddick, 1974; Watson and Ahumada, 1983) between local features that may or may not belong to corresponding objects in the successive views (Braddick, 1974; Hochberg and Brooks, 1974; Kolers and Pomerantz, 1971; Navon, 1976; Or-lansky, 1940). Especially in artificial situations, these can provide for failures and errors of integration and, indeed, may underlie most of the motion picture lore about movie editing and cutting (Brooks, 1984, 1985; Hochberg 1986; Hochberg and Brooks, 1978). It seems likely that the dangers of such errors depend sufficiently on currently measurable stimulus properties, and rest on sufficiently simple and early relationships, that they can be modeled and ameliorated.

However, that cannot tell us what the integrated product of successive glimpses will be. This distinction is analogous to noting that mere spatial knowledge about the location on the page of text at which some set of glances has been directed is not the same as knowing the central idea of that text. Higher processes of object recognition and representation are clearly involved here, and although it seems plausible that models can eventually be developed, as Chapters 9 and 10 show, they do not yet exist.

Moreover, it should be noted that the higher or more complex levels are primary in determining the sequence in normal looking:

where one looks (indeed, whether one looks) is a measure of the course of attention in the execution of an information-gathering task and is more a question of cognition than of vision (see Chapter 15).

In summary, the integration of successive glances cannot be modeled in any overall sense at this time. One can now merely list the classes of error that occur, describe the circumstances in which they are likely to happen, and suggest that models of some of the low-level processes responsible for error may be attainable. Where such errors are likely to be made, it should be noted that film directors have learned that in such cases, enough context—and enough time to assimilate the view change—help to achieve accurate comprehension.

BINOCULAR COMBINATION

Qualitatively speaking, when the two eyes receive disparate views that can be combined into a single layout or scene, "fusion" is said to occur, and the individual views are more or less lost in a single cyclopean percept (although with local disparities that exceed Panum's limit, diplopia can be detected within the otherwise-fused field). When the views cannot be so combined (still speaking loosely), rivalry occurs between them; they alternate—usually locally—in a piecemeal fashion, or more rarely, the view of one eye or the other dominates completely for some usually short time period. Still the most general, but noncomputational, attempt at a model is Sperling's (1970), although sections of that have been pursued in computer science (Marr and Poggio, 1976; Frisby and Mayhew, 1980). Dormant for some decades, attempts to model whether fusion or rivalry will occur are presently being refined from their rather vague starting point. (Blake, in press; Wolf, 1986). They are currently not image driven and have not yet been fully worked out.

When rivalry does occur, which view dominates in any region appears to be determined largely by relatively local measurable stimulus variables (Asher, 1953; Berliner, 1948; Blake, in press; Levelt, 1965), and it should be possible to provide models (perhaps principles, as well as empirical bases) that will account for and predict local dominance. This is not a trivial matter, if pilots continue to be fed different information to each eye and need some auxiliary procedure to bring rivalry under voluntary control. Shifting attention between rivalrous views is a fatiguing task for the pilot, which apparently only gets worse with time and experience (Murray and Hayworth, personal communication).

REFERENCES

- Asher, H.
1953 Suppression theory of binocular vision. *British Journal of Ophthalmology* 37:37-39.
- Berliner, A.
1948 Lectures on visual psychology. Chicago: Professional Press.
- Blake, R.
in press A neural theory of binocular rivalry. *Psychological Review*.
- Bordwell, J.
1985 *Narration in the Fiction Film*. Wisconsin: University of Wisconsin Press.
- Braddick, O.
1974 A short-range process in apparent motion. *Vision Research* 14:519-527.
- Brooks, V.
1984-5 Film, perception and cognitive psychology. *Millenium Film Journal* 14-15:105-126.
- Carr, T.H.
1986 Perceiving visual language. Pp. 20.1-20.92 in K.R. Boff, L. Kaufman, and J.P. Thomas, eds., *Handbook of Perception and Human Performance Volume 2: Cognitive Processes and Performance*. New York: John Wiley and Sons.
- Finke, R.A., and Shepard, R.N.
1986 Visual functions of mental imagery. In K.R. Boff, L. Kaufman, and J.P. Thomas, eds., *Handbook of Perception and Human Performance Volume 2: Cognitive Processes and Performance*. New York: John Wiley and Sons.
- Frisby, J.P., and Mayhew, J.E.W.
1980 Spatial frequency tuned channels: implications for structure and function from psychophysical and computational studies of stereopsis. Pp. 95-113 in N.S. Sutherland and H.C. Longuet-Higgins, eds., *The Psychology of Vision*. London: The Royal Society.
- Gibson, J.J.
1950 *The Perception of the Visual World*. Boston: Houghton-Mifflin.
1979 *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
- Henderson, J.M., Pollatsek, A., and Rayner, K.
1987 Effects of foveal priming and extrafoveal preview on object identification. *Journal of Experimental Psychology: Human Perception and Performance* 13:449-463.
- Hochberg, J.
1986 Representation of motivation and space in video and cinematic displays. Pp. 22.1-22.64 in K.R. Boff, L. Kaufman, and J.P. Thomas, eds., *Handbook of Perception and Human Performance Volume 1: Sensory Processes and Perception*. New York: John Wiley and Sons.
- Hochberg, J., and Brooks, V.
1974 The integration of successive cinematic views of simple scenes. *Bulletin of the Psychonomic Society* 4:263 (Abstract).

- 1978 The perception of motion pictures. In E.C. Carterette and M. Friedman, eds., *Handbook of Perception, Volume 10*. New York: Academic Press.
- Hochberg, J., and Gellman, L.
1977 The effect of landmark features on mental rotation times. *Memory and Cognition* 5:23-26.
- Humphreys, G.W., and Quinlan, P.T.
1987 Normal and pathological processes in visual object constancy. In G.W. Humphreys and M.J. Riddoch, eds., *Visual Object Processing: A Cognitive Neuropsychological Approach*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Judd, C.H., and Buswell, G.T.
1922 Silent reading: A study of the various types. *Supplementary Educational Monographs* 23.
- Just, M.A., and Carpenter, P.A.
1980 A theory of reading: From eye fixations to comprehension. *Psychological Review* 87:329-354.
- Kolers, P., and Pomerantz, J.R.
1971 Figural change in apparent motion. *Journal of Experimental Psychology* 87:99-108.
- Levelt, W.J.M.
1965 On Binocular Rivalry. Ph.D. Dissertation. Soesterberg, The Netherlands: Institute for perception RVO-TNO.
- Marr, D., and Poggio, T.
1976 Cooperative computation of stereo disparity. *Science* 194:283-287.
- Matin, L.
1986 Visual localization and eye movements. Pp. 20.1-20.45 in K.R. Boff, L. Kaufman, and J.P. Thomas, eds., *Handbook of Perception and Human Performance Volume 1: Sensory Processes and Perception*. New York: John Wiley and Sons.
- Monaco, J.
1977 *How to Read a Film*. New York: Oxford University Press.
- Moray, N.
1986 Monitoring behavior and supervisory control. Pp. 40.1-40.51 in K.R. Boff, L. Kaufman, and J.P. Thomas, eds., *Handbook of Perception and Human Performance Volume 2: Cognitive Processes and Performance*. New York: John Wiley and Sons.
- Navon, D.
1976 Irrelevance of figural identity for resolving ambiguities in apparent motion. *Journal of Experimental Psychology: Human Perception and Performance* 2:130-138.
- Orlansky, J.
1940 The effect of similarity and differences in form on apparent visual movement. *Archives of Psychology* 246:85.
- Rayner, K.
1978 Eye movements in reading and information processing. *Psychological Bulletin* 85:616-660.
- Reisz, L., and Miller, G.
1968 *The Technique of Film Editing*. New York: Hastings House.

- Rogowitz, B.E.
1984 The breakdown of size constancy under stroboscopic illumination. In L. Spillman and B.R. Wooten, eds., *Sensory Experience, Adaptation and Perception: Festschrift for Ivo Kohler*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Senders, J.
1983 *Visual Scanning Processes*. Netherlands: University of Tilburg Press.
- Sperling, G.
1970 Binocular vision: A physical and neural theory. *American Journal of Psychology* 83:461-534.
- Stevens, K.A.
1987 Visual object perception from a computational perspective. In G.W. Humphreys and M.J. Riddoch, eds., *Visual Object Processing: A Cognitive Neuropsychological Approach*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Treisman, A.
1986 Properties, parts, and objects. In K.R. Boff, L. Kaufman, and J.P. Thomas, eds., *Handbook of Perception and Human Performance Volume 2: Cognitive Processes and Performance*. New York: John Wiley and Son.
- Ullman, S.
1985 Visual routines. In S. Pinker, ed., *Visual Cognition*. Amsterdam, The Netherlands: Elsevier Science Publishers.
- Vorkapich, S.
1972 A fresh look at the dynamics of film making. *American Cinematographer* 53:182-195.
- Watson, A.B., and Ahumada, A.J.
1983 *A Look at Motion in the Frequency Domain*. NASA Technical Memorandum, TM-84352. National Aeronautical and Space Administration.
- Wolfe, J.M.
1986 Stereopsis and binocular rivalry. *Psychological Review* 93:269-282.

12

Afterword

Do models exist that meet the four requirements listed in the introduction? In Chapter 5, there seem to be some that are in fairly close to usable form and will simulate performance that depends primarily on aspects of early vision. It may also be possible to model aircraft state estimation, and some related responses, from two-dimensional optical flow information in certain restricted cases, as described in Chapter 8, although more validation against human performance is needed. Other processes in early vision, which may affect the integration of successive views and binocular rivalry, as mentioned in Chapter 11, seem amenable to modeling.

A recurrent theme through all the preceding chapters is that of insufficient, totally absent, or failed attempts at validation against human performance. There are two other problems that are similar to each other in their consequences. When "later" or higher-level visual and cognitive functions, which might be thought to rest on the early processes (some psychologists would disagree), are considered in Chapters 6, 7, 9, and 10, the verdict is almost uniformly negative: no usable, valid, image-driven models are within immediate or, in many cases, even fairly close reach. Where the early visual processes are to make their contributions to performance through their effects on higher perceptual processes, the fact that the former can be modeled is somewhat impaired by our present inability to model the latter.

The second problem is this: numerous gaps exist between the functions that can be modeled, so that they do not form a chain or a seamless repertory that can be drawn on automatically in any task. Given the likelihood noted above that many perceptual functions important to pilot performance cannot yet be suitably modeled, it seems clear that a workstation cannot at present be trusted to

perform a full and unsupervised pilot performance model simulation or evaluation of an arbitrary mission in any given cockpit design.

If such limitations on applicability and validity as have been described are accepted however, it would still seem profitable to set up a test system that models what can be modeled. In this way, a more realistic basis will exist for assessing the relative importance of what can and cannot yet be done. This will also enable us to test the nonvisual or cognitive components of any such system.

Part III

Introduction to Cognition Models

At the cognitive level, the architecture that enables us to roughly associate models by stages of processing, as was true for visual models, is no longer available. Furthermore, for the accomplishment of any real task, the functional components of human processing exist in complex interaction with each other, making it difficult to separate out models of the components that are predictive. An added difficulty is that the data structures for models of cognition must often be complex.

Mitigating these difficulties of modeling at the cognitive level are several factors. First, for simple, fast behaviors, say on the order of a second, pieces of the underlying mechanisms of cognitive action show through and can be modeled (although the relationship of these models to interaction in sustained, naturalistic situations may still be problematic). Models of working memory and attention are examples. Second, more complex behavior tends to be in the service of some goal and under constraints in the environment. Detailed studies of the courses of action open in this environment and the "task analysis," "knowledge-oriented," or "rational action" modeling of the environment (and the information possessed about it), together with relatively simple assumptions about the underlying mechanisms of cognitive action, can be used to predict behavior. Decision theory models, problem-solving models, or time line analysis models are examples of this type. The doctrinal and heavily engineered nature of procedures for helicopter flight is an asset here for modeling.

The chapters in Part III move in a progression between these two levels—from models of mechanisms of the human cognitive architecture to models of rational action in a described environment. Thus, some models early in the progression are at the component (or the architectural) level, whereas later models are based almost entirely

Cognitive Architectures

STUART K. CARD AND ALLEN NEWELL

The models in this chapter describe aspects of the human information processing system. Each model attempts to explicate mechanisms of the human processor that give rise to surface behavior. Of course, in reality there is but a single human processor, all of whose mechanisms all fit together. The overall structure of this processor—its architecture—is an object of study in its own right. Recently, the cognitive part of this architecture has been the subject of active study. Most of these proposals are reported in the proceedings of a recent conference on cognitive architectures (CMU, in press), to which the reader is referred. This section does not attempt to review current proposals for cognitive architectures, but rather, gives a brief sketch of the space of alternatives.

Models of architecture are important for several reasons. (These are derived from Newell, Rosenbloom, and Laird, in press): (1) The architecture is the frame in terms of which all processing is done, the locus of structural constraints on cognition; it is a piece of the puzzle in its own right. (2) Gross parameters of the architecture, such as working memory size, can be used to summarize approximately the constraints acting on general cognition. (3) The architecture provides a means of integrating the mechanisms (and reducing their number) identified by other models and of explicating their input, output, and shared resource connections. (4) The architecture is a means of revealing hidden connections and constraints among activities which, on the basis of context and situation, may seem quite distant from each other. (5) The architecture is a means of removing theoretical degrees of freedom from modeling the mechanisms behind specific behaviors; otherwise, the modeling of these is often severely underconstrained.

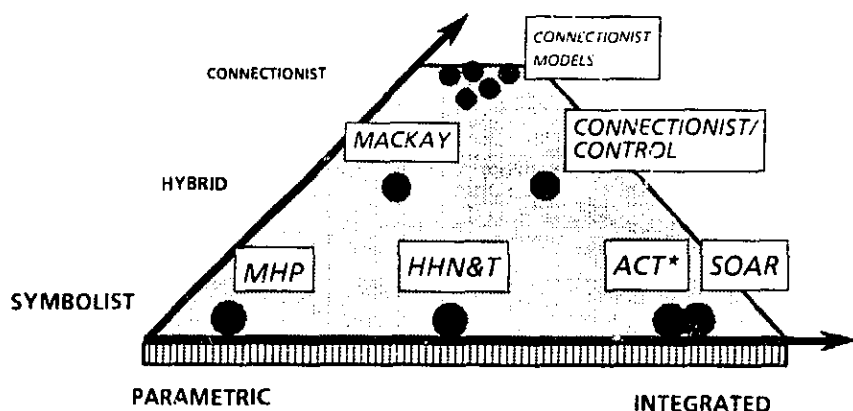


FIGURE 14-1 Schematic diagram of cognitive architectures.

Figure 14-1 is meant to suggest some dimensions of variation in the space of cognitive architectures. To define the space, a few specific architectures are included. The clearest contrast is between those architectures that model human processing in terms of symbolic processing (symbolist architectures) and those that use some sort of subsymbolic processing, represented in graphs with weighted links (connectionist architectures). Intermediate between symbolist and connectionist architectures are proposals that combine some features of both and are, therefore, termed hybrid architectures. The technical development of connectionist architectures is more recent (although they have roots in associationist philosophy of great vintage and the neural models of the 1950s). Because their development is in major flux, they are drawn as a cluster of related models on the horizon and are not differentiated further in Figure 14-1. The symbolist architectures have had more time to reach technical maturity; they are, therefore, drawn in the foreground and further differentiated according to the integration of the architectural mechanisms.

SYMBOLIST ARCHITECTURES

The most integrated of the symbolist architectures are ACT* (Anderson, 1983) and SOAR (Laird, Newell, and Rosenbloom, 1987; Newell, in press). ACT* has three memories: (1) a declarative long-term memory (a semantic net of nodes with weighted links), (2) a procedural long-term memory (condition-action productions),

and (3) a working memory. Elements in both long-term memories have strengths associated with them, and those in declarative long-term memory can have a level of activation associated with them. Working memory is the set of activated elements from declarative long-term memory (including goal structures) plus the set of actions that create new structures in working memory. Activation spreads through declarative memory as a function of element strength. New productions can be created from the effects of previous activity that has made it to declarative long-term memory.

SOAR has two memories, a single long-term memory of productions and a working memory that contains a goal structure, information associated with the goals, preferences about what should be done, perceptual information, and motor commands. Working memory serves as a sort of bus, receiving inputs from sensory elements, exposing these inputs to parallel-acting decoding productions, holding inputs and outputs from cognitive productions, exposing these to parallel-acting encoding productions, and holding outputs for activation of the motor system. All tasks are modeled as searches in some problem space. Productions contribute preferences for the next substeps in this search (choice of goal to work on, choice of state to work on, choice of operator). If there is no clear-cut set of preferences for these choices, the system is at an impasse, leading it to generate a new goal and problem space to solve the impasse itself. This mechanism allows the system to reflect on its own processing, leading it to search both through and among problem spaces. An individual move in a problem space can itself lead to a new problem space to solve the problem of how to make this move. New productions are continuously created that embody the results from successful searches.

At the opposite extreme from integrated symbolist models, such as ACT* and Soar, are models like the model human processor (Card, Moran, and Newell, 1986) that use a few parameters to characterize the architecture instead of detailed interacting mechanisms. The model human processor has four memories (long-term memory, working memory, the visual image store, and the auditory image store) and three processors (cognitive, perceptual, and motor). Each of these is characterized by parameters. For example, the visual image store decays exponentially with a decay constant of 200 milliseconds (msec). Ranges are provided for all the parameters so that upper and lower bounds can be computed to take into account the approximate nature of the analysis and the state of knowledge in the literature.

A set of accompanying laws of behavior (e.g., Fitts's law, Hick's law, Snell's law) augments predictions from first principles.

There also exist symbolist architectures that are intermediate along the parametric-integrated dimension of Figure 14-1. An example is the Holland, Holyoak, Nisbett, and Thagard (1986) theory of induction. General knowledge, in their architecture, is embodied in a set of condition-action rules, which can represent both time-invariant information (dogs are animals) and information about future states of knowledge (if a person annoys a dog, it will growl). These rules form clusters, either explicitly through linking together of the rules or implicitly because the rules share data structures. In particular, rules can be clustered into superordinate categories, which give rise to a default hierarchy of rules (a dog has four legs because a dog is a mammal and a mammal has four legs), together with rules that express exceptions to this default hierarchy. Such sets of rules express the 'mental models' people have about the world. Induction consists of mechanisms for revising the strengths of individual rules and for devising plausible new rules, based on experience. These mechanisms are triggered by failed or successful predictions. Other more or less integrated symbolist architectures exist, largely developed around some particular set of tasks, for example, learning subtraction for VanLehn's (1983) SIERRA architecture or reading for Just and Carpenter's (1987) CAPS architecture.

CONNECTIONIST MODELS

In contrast to the symbolist architectures in which the mind is assumed to be a physical symbol-processing system, connectionist systems are networks of large numbers of interconnected "units." Each unit can have associated with it a certain amount of activation. Connections to other units are given explicit weights (including negative weights). Activation spreads from one unit to another as a function of the weighted links. For example, the function of a typical link might be to multiply the input activation by its weight and then apply a threshold function. A typical unit would sum all of its input activations, then divide this among all its links. The weights on the links are adjustable with experience. Some of the links may represent sensory inputs from the outside world; some may represent output to effectors to the outside world. Units in connectionist models are usually taken to be below the level of a symbol. For example, different

units may represent visual features of a letter such as verticalness or roundedness.

Connectionist models are attractive because they appear to offer the beginnings of a computational architecture that is more neural-like. They seem to show how complex mental operations can be derived from slow, simple mechanisms, and they seem naturally to relate perception to cognition. They have had some initial successes in modeling behavior (see McClelland and Rumelhart, 1986; Rumelhart and McClelland, 1986; Smolensky, 1988). These issues, however, are heavily contested (see Fodor and Pylyshyn, 1988). Connectionist models have been most successful at pattern recognition tasks. It is not known whether they will have adequate computational power to model higher-level cognitive behavior (Smolensky, 1988). At present, the state of connectionist models is changing so rapidly that detailed commentary on particular lines of research would be out of date immediately.

Hybrid architectures, containing both symbolist and connectionist aspects, have also been proposed (e.g., MacKay, 1987; Schneider and Detweiler, 1987; Schneider and Mumme, 1987). For example, in Schneider's connectionist/control architecture, connectionist units model specific sorts of cells in the known neurophysiology. The units are organized into modules, each capable of representing some meaningful piece of information. Modules are connected by bundles of links called message vectors. Information processing occurs by level, with visual feature modules feeding into letter modules, which feed into word modules for example. Each module contains an attenuational control unit and a control box. These allow higher-level modules to control the availability of message vector outputs and also allow a mechanism for sequencing. At the highest level, different modalities are tied together with the output of a context module. Learning and sensory automaticity occur by changing weights between a message vector and an output unit.

The use of one of the aforementioned cognitive architectures for the integration of human performance theory in a computer-aided engineering system could be pursued at the present time as a research project, but not as an engineering component on which other work depends. Furthermore, the project would be feasible for researchers associated with one of the teams now working on cognitive architectures. However, this may change in the next five years, after research on these models has matured. Advances in cognitive architectures can be expected to lead to across-the-board

improvements in the ability to use human performance models for engineering work because they address directly one of the primary difficulties—the complex interactions that occur among interactions for a human engaged in any macrolevel task.

REFERENCES

- Anderson, J.R.
1983 *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.
- Card, S.K., Moran, T.P., and Newell, A.
1986 The model human processor: An engineering model of human performance. In K.R. Boff, L. Kaufman, and J.P. Thomas, eds., *Handbook of Perception and Human Performance*. New York: Wiley.
- Carnegie Mellon University
in press Proceedings of a Conference on Cognitive Architectures. Pittsburgh, PA: CMU.
- Fodor, D., and Pylyshyn, C.W.
1988 Connectionism and cognitive architectures: A critical analysis. *Cognition* 28:3-71.
- Holland, J.H., Holyoak, K.J., Nisbett, R.E., and Thagard, P.R.
1986 *Induction*. Cambridge, MA: MIT Press.
- Just, M.A., and Carpenter, P.A.
1987 *The Psychology of Reading and Language Comprehension*. Boston, MA: Allyn and Bacon.
- Laird, J.E., Newell, A., and Rosenbloom, P.S.
1987 Soar: An architecture for general intelligence. *Artificial Intelligence* 33(1):1-64.
- MacKay, D.G.
1987 *The Organization of Perception and Action: A Theory for Language and Other Cognitive Skills*. New York: Springer-Verlag.
- McClelland, J.L., Rumelhart, D.E., and the PDP Research Group, eds.
1986 *Parallel Distributed Processing, Volume 2*. Cambridge, MA: MIT Press, A Bradford Book.
- Newell, A.
in press *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.
- Newell, A., Rosenbloom, P.S., and Laird, J.E.
in press Symbolic architectures for cognition. In A. Newell, P.S. Rosenbloom, and J.E. Laird, eds., *Symbolic Architectures for Cognition*. Cambridge, MA: MIT Press.
- Rumelhart, D.E., McClelland, J.L., and the PDP Research Group, eds.
1986 *Parallel Distributed Processing, Volume 1*. Cambridge, MA: MIT Press, A Bradford Book.
- Schneider, W., and Detweiler, M.
1987 A connectionist/control architecture for working memory. Pp. 53-119 in G.H. Bower, ed., *The Psychology of Learning and Motivation*. San Diego, CA: Academic Press.

Schneider, W., and Mumme, D.

- 1987 *A Connectionist/Control Architecture for Attention, Automaticity, and the Capture of Knowledge*. Unpublished manuscript.

Smolensky, P.

- 1988 On the proper treatment of connectionism. *Brain and Behavioral Sciences* 11:1-74.

VanLehn, K.

- 1983 Felicity Conditions for Human Skill Acquisition: Validating an AI-Based Theory. Ph.D. Thesis, Massachusetts Institute of Technology.
- 1989 *Architectures for Intelligence*. Hillsdale, NJ: Lawrence Erlbaum.

Resource Management and Time-Sharing

CHRISTOPHER D. WICKENS

OVERVIEW

The domain of this chapter is models that predict the loss of quality of processing information from multiple channels, or multiple tasks, which occurs as a direct result of that multiplicity. It is often assumed, therefore, that this multiplicity induces competition for some scarce commodity called "resources." At issue is whether one can predict the loss in quality, given characteristics of (1) the processing on each channel (or task) in isolation and (2) the relation between channels (tasks).

There are a number of psychological models of the resource allocation process. Unfortunately, it appears that those models which have the most precise quantitative formulation and have received the strongest empirical validation, have been derived in domains that may be furthest removed from the complex, heterogeneous task environment of the rotorcraft cockpit; whereas those that have addressed task environments of greatest complexity are furthest removed from a quantitative formulation (or alternatively are models that have yet to be validated). This disparity is unfortunate because it is clear that the objective should be one of obtaining quantitative models in which levels of performance in heterogeneous environments can be predicted from quantitative specification of task parameters.

Two general characteristics of the resource process have been addressed by models: (1) the allocation of resources—the selective aspects of attention—and (2) the sources of variance in competition between channels or tasks—the "scarcity" commodity of resources. Within each category, further discrimination may be made between two classes: (1) those models that assume, for convenience, a sequential mode of processing and address the logic of switching, or

TABLE 15-1 The Domain Of Models

		<u>Resource</u>	
		Allocation	Competition
Serial Aspects	Visual Scanning:	Moray Senders Harris and Spady Allen et al.	Switching: Gopher LaBerge et al. Kristofferson
	Task Selection:	Chu and Rouse Tulga and Sheridan Zacharias (PROCRU) Siegel and Wolf Wortman et al. (SAINT) Wherry (HOS)	
Time-sharing (Parallel Aspects)	Multichannel Detection:	Shaw Swets	Multiple Resources: Wickens Friedman et al. Navon and Gopher North Laughery et al.
	Manual Control:	Levison and Tanner Levison	
	Performance Operating Characteristics:	Sperling and Doshier	Confusion: Carswell Integration: Wickens

the serial aspects of performance. These models are not concerned directly with the level of performance, but rather with the timing of when tasks will be initiated. They may be contrasted with (2) those that do not make the serial assumption, that address domains of time-sharing and concurrent processing, and that make specific predictions about performance levels.

A 2 by 2 structure of the domains of models is presented in Table 15-1 in which the dichotomy of serial-parallel processing is crossed with that of allocation-selection versus competition. Within each cell key terms are identified that characterize the phenomena associated with the modeling efforts, along with key references or sources.

It is important to note that a number of elegant efforts have discussed models of the strategic or microscopic processes by which performance is produced—e.g., whether processing is serial or parallel (Kantowitz, 1986; Townsend, 1974; Townsend & Ashby, 1983) or whether information integration and selection are early or late (Kantowitz, 1986; Norman, 1968; Pashler, 1984; Shaw, 1982). Therefore, a distinction must be drawn between models of how performance is produced and models that predict how performance will vary as a function of task characteristics. The latter are clearly relevant to

design environments. The former will be relevant only if the modeled mechanism has robust and important implications for the level of performance obtained.

This chapter reviews the models that exist in terms of the four cells of Table 15-1 and concludes by describing in more detail the hybrid model that is considered to be most appropriate for the current applications.

SERIAL ALLOCATION

In the upper left cell of Table 15-1 are those models that have dealt unambiguously with the serial allocation of some processing resource. Such as the availability of foveal vision between saccades or the complete allocation of cognitive effort to one task rather than another. From the standpoint of these models, the issue of whether processing may be parallel is simply not relevant. They focus on those aspects of processing that are distinctly and unambiguously serial (i.e., that require decision of where to allocate over time).

Visual Sampling

There are two critical aspects of the visual sampling process to be modeled: the scanning process that assesses the transitions of visual fixations from one display to another and the fixation itself, which is characterized by a location, a useful field of view (diameter around the central location from which information is extracted), and a dwell time. Visual sampling has been examined both in supervisory control tasks employing fixed displays such as an aircraft instrument panel, where potential information sources are known in advance to the operator, and free field search, in which an area is searched for a target of unknown location, and sometimes uncertain form (e.g., the radiologist examining an x-ray plate, or the airborne observer engaged in search for a crash site or ground installation). Some successful quantitative models have been applied to the first domain, and a number of useful principles have emerged from the second, which can provide the foundation for effective model development.

The foundation for models of display scanning in supervisory control were provided by Fitts, Jones, and Milton (1949), who analyzed the frequency of fixations and transitions between cockpit instruments during flight. Senders (1966, 1983) provided a quantitative basis for an instrument sampling model which was based on

information theory. He reasoned that instruments which varied according to their bandwidth or autocorrelation (information delivered per unit time) would be sampled with a frequency in direct proportion to that information. Empirical validation revealed that this sampling model accounted for instrument scanning data reasonably well, although operators tended to sample more frequently (relative to optimal) from sources with low information content and sample less frequently from those sources with high content.

Senders' sampling model has been elaborated by Carbonell (1966) to include elements related to the subjective uncertainty of a given information source. According to this elaboration, the source is assumed to have zero uncertainty immediately after it is fixated. Uncertainty then grows over time as a result of two factors: the bandwidth or autocorrelation of the signal, and the decay properties of working memory (see Chapter 16). The next fixation of a given instrument will occur when the level of uncertainty reaches an internal criterion whose level is based upon the expected cost of *not* sampling the instrument, and therefore missing a critical event. Carbonell, Ward, and Senders (1968) obtained reasonably good validation of the fixation model using experienced pilots flying an aircraft simulator. Moray (1986) described validation of many of the model's characteristics in describing the radar scanning behavior of fighter aircraft controllers. To date, however, eye fixation models of supervisory control in aviation have addressed issues of instrument scanning and have not considered sampling of motion gradients in visual contact flight.

Scanning behavior in free field search has been less successfully modeled than in supervisory control, in part because search patterns and inspection performance tends to be quite heavily influenced by individual differences. Drury (1975), however, reports the validation of a combined search and decision model of sheet metal inspection in which operators search for flaws of unknown location. Williams (1966) describes a generic target search model that focuses on the role of target conspicuity in search tasks. An important characteristic of both of these models is the non-linear (logarithmic) function relating detection probability to search time allowed.

In addition to all of the models described above which can be applied to helicopter cockpit instrument scanning, or out of the cockpit search, there are a number of general principles of visual scanning that have emerged from the considerable body of research

in the area (see Abernethy, 1988; Moray, 1986; Sheridan and Ferrell, 1974, for good reviews). These are

1) The operator's search strategy is driven in large part by his expectations, or a "mental model" of where information is likely to occur on a display. (This model is formalized and quantified in the bandwidth characteristics of signals in the supervisory control models.) Differences in the mental model account for differences in search behavior between novices and experts (the latter having a better formed set of expectancies), and for the fact that higher information areas on a display (greater element density, contours) tend to be fixated more frequently. Expert-novice differences in scanning vehicular environments have been examined by Harris and Spady (1985) in aviation and by Mouront and Rockwell (1972) in automobile driving.

2) The fact that scanning behavior is internally driven by cognitive factors, rather than externally driven by display factors, is apparently responsible for substantial individual differences in scan patterns, particularly in search tasks. Unfortunately, these strategic differences impose difficulties in developing models that capture a higher degree of variance. Nevertheless, certain additional generalizations of search behavior across individuals can be made.

3) There is a tendency to avoid searching near the edges of a display even when targets may be likely to be located there (Parasuraman, 1986).

4) Fixation dwell times vary in their duration between 200 msec and approximately 1 second. Within this range there is no systematic evidence that longer dwells lead to more efficient search in free field tasks. However, in information extraction tasks, dwells are typically longer on displays that are less legible and from which more information is extracted. For example, Harris and Christliff (1980) noted that primary flight instruments necessary for control (e.g., the attitude display indicator) are sampled with longer dwells than those employed for check reading.

5) In search tasks, each fixation is characterized by a useful field of view (UFOV, Mackworth, 1976) which may vary in its diameter, depending on the density of material to be searched. Successive scans will not overlap UFOVs. The UFOV may range between approximately 2 and 4 degrees of visual angle. Combinations of the UFOV and the maximum fixation rate (2 to 3 saccades per second) constrain the amount of area that can be searched per unit time. However, even with sufficient time to search, it appears that operators do not cover

an entire area with UFOVs, and targets may be fixated (sometimes frequently) and yet not detected.

6) In many scanning tasks, some use may be made of peripheral vision, not necessarily to detect targets, but to guide the destination of the next fixation.

The extent to which these general principles may be incorporated into the quantitative predictive models of free-field scanning and target search formulated by Williams (1966) and by Drury (1975) remains unclear. The identification of these factors remains an important step, but the ultimate degree of success of the predictive models will clearly depend upon the ability to characterize the operator's internal model of an environment (for search task) or of a system (for supervisory control tasks) that guides sampling behavior via cognitive factors.

Task Selection

The characteristics of task selection on the basis of expected utilities and costs also lie at the core of the concurrent performance assumptions made by many of the predictive models of complex task performance (Pew, Baron, Feehrer, and Miller, 1977), such as the human operator simulator (HOS) (Harris, Iavecchia, Ross, and Shaffer, 1987; Strieb, Lane, Glenn, and Wherry, 1981; Wherry, 1976), SAINT¹ (Laughery, Drews, and Archer, 1986; Wortman, Duket, Seifert, Hann, and Chubb, 1978;), PROCURU (Zacharias, Baron, and Muralidharan, 1981), STALL (saturation of tactical aviator load limits; Chubb, Stodolsky, Fleming, and Hassoun, 1987), and those models developed by Siegel and Wolf (1969), Corker, Davis, Papazian, and Pew (1986), Chu and Rouse (1979), and Tulga and Sheridan (1980). Essentially these models assume that when two (or more) tasks compete for attention (call for completion at the same time), an algorithm assesses the order in which the tasks are to be performed. This algorithm is based on user-defined priorities (HOS; Harris et al., 1987), on computation of expected costs of ignoring those activities not immediately performed and expected benefits of undertaking the action that is highest in the priority sequence (PROCRU; Zacharias et al., 1981), or on the application of strategy-driven decision rules

¹SAINT is not actually a model of complex task performance but rather a structured programming language that allows user-defined task sequences to be played out.

and the differing degrees of competition fostered by greater or lesser similarity between tasks. In this regard they represent more complex elaborations of single-channel models of attention that were developed in psychology (Welford, 1967). However, recent elaborations of some of the models have begun to address the issue of concurrent performance, as described later in this chapter.

PARALLEL ALLOCATION

The emphasis of models in the lower left cell of Table 15-1, is on the loss of information-processing quality that results from concurrence and from shifts in resource allocation, rather than the forces (such as expectancy and utility) that predict when a sequential shift will take place. Furthermore, in contrast to models in the first quadrant, these models assume that parallel processing between tasks is ongoing and, hence, that interference effects result from competition for something more than time (or, at least, from more than time at a relatively low sampling frequency). Basically, these models have taken two generic approaches. One approach is to model performance on two perceptual (detection or recognition) tasks of equal priorities, as a function of such variables as signal strength, signal uncertainty, and signal differences (Shaw, 1982; Swets, 1984; Taylor, Lindsay, and Forbes, 1967). Several examples of this approach have been based upon the theory of signal detection. The empirical data to validate these models have been collected under fairly carefully defined conditions (near-threshold stimuli in constrained display locations), and these factors may constrain their relevance to the helicopter environment.

The second approach focuses on the differential allocation of resources to different channels or tasks, modeling this allocation from the standpoint of economic theory as a utility-based decision problem. Sperling (1984; Sperling and Doshier, 1986) provides an elegant integrative treatment of the factors underlying this modeling approach. This approach has its origins in the assumption that resources are continuously allocatable commodities that facilitate performance through a function referred to as the "performance resource function" (Norman and Bobrow, 1975). Performance is seen to improve or degrade on the basis of the allocation of something other than or in addition to time. Here again, reported data do not extend far beyond simple detection and recognition tasks.

One important quantitative modeling approach to time-sharing, however, that is applicable to a more diverse set of complex tasks is found in the multitask extension of the optimal control model of manual control (Levison, 1982; Levison, Elkind, and Ward, 1971). Fundamental to this model is a parameter of "observation noise" that is assumed to perturb the internal representation of analog signals used for tracking and monitoring. Observation noise is typically expressed as a ratio to relevant signal amplitude; that is, as an "observation noise ratio." On the one hand, the effects of changes in this observation noise ratio on tracking error may be predicted quantitatively within the model (Levison, 1982). On the other hand, the causes of change in noise level are incorporated in an attention sharing model by the formula $P_I = P_o/F_i$, in which P_o is the single task observation noise ratio, P_I is the observation noise ratio under multitask conditions, and F_i is the fraction of attention allocated to the task.

The quantitative aspects of Levison's approach have been validated (e.g., Stein and Wewerwinke, 1983), but the constraints are clear as well. The observation noise ratio is applicable only to tasks whose inputs are linear spatial quantities (position, velocity) and not qualitative or configurational feature-defined patterns, such as those used in symbolic or verbal processing or in object recognition.

While the model of attention modulation of the observation noise ratio was originally developed in the context of multi-axis tracking tasks, it is important to realize that the model is applicable to any task in which actions are taken on the basis of signals of ranging magnitudes. Thus, it may be applied to monitoring and decision tasks as well as to tracking, as has been done by Levison and Tanner (1971) and in the PROCURU model of Corker et al. (1986) to be described in the following section. It should be noted that this quantification of visual resolution from time-sharing is an important component of the model integration effort under this project. It stands as a parameter that can be passed to the visual models.

SERIAL COMPETITION

On the right side of Table 15-1 are models that focus on the nature of the competition between channels, as a consequence of structural similarities and differences between tasks or channels. When such processing is serial, as in the top right cell, any competition

must then be the result of a discrete attention switch, whose properties have been modeled by Sperling and Doshier (1986), LaBerge (1973), and Kristofferson (1967). These switching costs, however, are sufficiently small that the time actually involved in the switch itself will have a minimal effect on operational performance. In contrast, whether a switch does or does not take place is, of course, critical to operational performance. This issue is dealt with in the section on serial allocation.

PARALLEL COMPETITION

More relevant are the efforts to account for the competition between heterogeneous tasks carried out in parallel (or at least in such a way that long intervals of neglect do not characterize the performance of one task or the other).

Computer Simulation Models

Three of the computer simulation network models described earlier have recently taken a step toward acknowledging that not all performance is serial and that task demands vary in intensity as well as in time. The HOS model is currently being revised (Harris et al., 1987), and the revision, which will be available in a user-friendly microcomputer form, explicitly allows parallel processing of activities. Thus, for example, the model allows the operator to reach while scanning or to encode while controlling. However, parallel processing is assumed to be perfect processing. There is no mechanism for specifying interaction between tasks. The activities that are processed in parallel are user defined, as is a preemption mechanism that terminates a particular activity when one of higher priority is imposed. In addition, the software is designed to be flexible enough so that the user's own model may be substituted.

A recent elaboration of SAINT has also spawned a microcomputer version known as MICROSAINTE. Laughery et al. (1986) have used the programming capabilities of MICROSAINTE language to expand upon previous developments in two important respects:

- They accommodate demand specifications of tasks (or mental operations) that are not defined only in terms of time. Rather, the model employs a set of tabled demand values for different tasks, ranging from 0 to 7. These values were generated by expert pilots and compiled by McCracken and Aldrich (1984) and by Aldrich,

Szabo, and Bierbaum (1988). For example, the activities "monitor, scan, survey" have a demand level of one. "Trace, follow, track" have a demand level of 3. "Read, decipher text, decode" have a demand level of 7.

- They acknowledge the multiplicity of processing resources by assuming that task demands will interfere on particular combinations of channels, but not on other combinations. Four "channels" are defined: visual, auditory, cognitive, and psychomotor (VACP) (McCracken and Aldrich, 1984). Within each channel, simultaneous demands are summed, and values of greater than 5 on the visual channel are assumed to exceed a threshold that requires the abandonment of monitoring to support situational awareness. The Aircrew-Aircraft Integration (A³I) model developed by Corker et al. (1986) makes similar assumptions about the association of tasks and task demands to the four channels. An assumption made in this model is that demands greater than 7 in any channel will lead to a temporary postponement of the last task added, which caused demand to exceed the threshold.

Although the developments reported by Laughery et al. (1986) and by Corker et al. (1986) are a marked advance over previous efforts, they still suffer from a number of limitations. First, the demand level codings of activities within a channel do not appear to acknowledge the degree of difficulty of tasks within a level. Thus, for example, detecting a change in size (coded demand 2), if it is a subtle change in a dynamic environment, could be far more difficult than reading a simple one-word message (which is coded demand level 7).

Second, the assumption of parallel processing between channels (demand levels do not add across channels) appears to be unwarranted. For example, there is clear experimental evidence that auditory and visual tasks interfere, as do perceptual (both auditory and visual) and cognitive ones (Wickens, 1984). However, no assumptions are made regarding this sort of interference.

Finally, a concern directed toward all of the modeling efforts, echoing a lament voiced by Meister (1985) in his comprehensive review of these simulation models, is the lack of validation data. In the absence of empirical data necessary to determine if the predictions of the models are accurate, no firm evaluation can be offered.

It should be noted that there are at least two reports of validation of the four-channel (VACP) approach to complex task prediction in complex aviation simulations (Bateman and Thompson, 1986;

Laughery et al., 1986). Unfortunately, both used as criteria pilot-generated subjective ratings of task workload, rather than actual performance. Because subjective ratings and performance may differ from each other in important ways (Yeh and Wickens, 1988), some caution must be taken in accepting these as full validations.

Psychological Models

A contrast can be offered by the models of time-sharing that have grown directly out of the psychological laboratories. Here, validation data exist, but the direct applicability to systems design issues remains less well developed. The model in this domain that has received the greatest degree of validation and is also most appropriately tuned to the current application is probably the multiple resource model (North, 1985; Tsang and Wickens, 1988; Wickens, 1984, 1987, 1988; Wickens and Liu, 1988). Because the model can be used to improve upon existing simulation models, it is described here in some detail.

According to the multiple resource model, two tasks will suffer interference to the extent that the component tasks are more difficult (demand more resources) and that they compete for overlapping resources.

These resources are described at a more general level (e.g., spatial-verbal) than are the processing mechanisms of the tasks themselves. The current version of the multiple resource model proposed by Wickens (1987; Wickens and Liu, 1988) defines three dichotomous dimensions, each of which defines two resources. These are processing codes (spatial-analog versus verbal-linguistic), processing modalities (auditory-speech versus visual-manual) and processing stages (perceptual-cognitive versus response). However, it is possible, particularly in the helicopter environment, that a dimension of ambient-focal vision, which contrasts orientation judgment with object recognition, postulated by Leibowitz and Dichgans (1980) and by Christensen, O'Donnell, Shingledecker, Kraft, and Williams (1985) might well be relevant. Validation of the model in basic laboratory experiments has been carried out by a number of studies (e.g., Tsang and Wickens, 1988; Wickens, 1980; Wickens and Liu, 1988; Wickens and Weingartner, 1985). Validation in a more complex aviation simulator environment has been carried out by Wickens, Sandry, and Vidulich (1983) and by Wickens, Harwood, Segal, Tczkavic, and Sherman (1988).

North (1985, North and Riley, 1988) incorporated many of the assumptions of the multiple resource model into a predictive workload index algorithm known as WINDEX. Applicable to cockpit design modifications, WINDEX assigns resource demand levels (rated 1-5) to different channels or processing systems (e.g., window, helmet-mounted display, cathode-ray tube (CRT), auditory, stick, keypress, speech, and cognitive activity). Critical to the operation of WINDEX is a conflict matrix by which concurrent activities in different channels will interfere more or less, depending on their similarity in the multiple-resource space. This feature was absent from the Laughery et al. (1986) version of the MICROSAINTE simulation and from the A³I application developed by Corker et al. (1986). Thus, for example, in the WINDEX conflict matrix, large penalties are assigned to tasks that impose concurrent demands on two visual channels (e.g., window and helmet-mounted display). Reduced, but still substantial, conflicts may apply to simultaneous use of the window and auditory channel (both involving perceptual encoding), to the speech and key press channel (both involving responses), or to speech output and verbal rehearsal (both involving verbal processing). Minimum penalties would be assigned to concurrent use of the auditory and stick channel, which lie "far apart" in the multiple resource space. Although the model has been applied to the design of the light helicopter family (LHX) prototypes by McDonnell Douglas, the results of this application (and resulting validation of the model) unfortunately remain proprietary.

More recently, the WINDEX-type model has been applied in a competitive validation effort to data collected in a helicopter flight simulation (Wickens et al., 1988). Algorithms involving the complexity of multiple resource competitions were compared with simpler ones based on adding task demands and on pure time line analysis. The multiple resource algorithms were found to provide significantly (and substantially) better predictions of the performance data.

Three limitations of the multiple resource model make it difficult to move from a qualitative to a quantitative domain. These limitations are inherent in the model's efforts to address interference between heterogeneous tasks, but they are limitations for which potential solutions exist.

- The amount of resource overlap between tasks depends on careful definition of what constitutes a resource. Wickens' (1984) heuristic specification of resources defined by three dichotomous dimensions, allows for some quantification to be accomplished at four

levels of resolution, according to a "shared features" approach. For example, two tasks may compete for resources on zero, one, two, or three dimensions. Using this approach, Derrick and Wickens (1984) and Gopher and Braune (1984) have obtained reasonably good predictions of the degree of interference between a collection of heterogeneous tasks.

- Even more serious is the lack of a single metric that can be used to quantify the demand for resources (i.e., task difficulty) applicable across different component tasks. However, four possibilities exist. First, single-task performance differences imposed by a change in demand can be used to predict dual-task interference. Second, a relatively generic task analytic metric such as information rate or working memory load can be employed to quantify demands. Third, subjective ratings or estimates of single-task difficulty levels can be used. Fourth, it is possible to depend on expert opinion ratings to code demands (i.e., 0, 1, or 2). This technique is used by Gopher and Braune (1984), and advocated by North (1985; North and Riley, 1988), Laughery et al. (1986) and Corker et al. (1986) in their applications of WINDEX or MICROSAINTE, respectively. All three rely on the tabled values proposed by McCracken and Aldrich (1984; Aldrich et al., 1988) for coding these demand levels.

- There is yet no invariant metric for scaling the decrement or interference between tasks that may involve different performance measures (for an informative debate on this point, see Kantowitz and Weldon, 1985; Wickens and Yeh, 1985).

SYNTHESIS OF THE OPTIMAL MODEL

Table 15-2 presents a general review and comparison of several of the performance models described. It focuses on the assumptions made by the models about attention with regard to their serial or queuing characteristics (the logic by which tasks are selected to be performed), their parallel or resource assumptions (how many channels or resources, and whether tasks are defined in terms of their demand levels), and any assumptions the models make regarding the effects of workload on performance. As can be seen from the bottom of the table, an unfortunate facet of all models is the absence of available performance data necessary to validate them.

From the larger set shown in Table 15-2, 3-1/2 plausible models can be identified for potential application to the current design problem. These models are shown in Table 15-3. Each model has

TABLE 15-2 Comparison Of Key Features Of Some Selected Models

Features	SAINT/Siegel and Wolf (Laughery et al., 1986)	HOS	WINDEX (North, 1985)	PROCRU
<u>Attention Assumptions</u>				
Queuing	Production system Single-channel delay (Serial input) Memory decay of neglected tasks	Production system Preemption, Priorities	Under development Currently user-defined	Production system (Expected value) Single channel
<u>Resources</u>				
Demand Level	Yes (Aldrich and McCracken) Resource demands add within a channel	No (but allows user- defined models to be input)	Yes (Aldrich and McCracken)	No
Multiplicity	Yes (Four multiple parallel channels)	Allows perfect parallel processing	Yes, Includes conflict matrix Resource demands multiplied by interference level in conflict matrix	No
<u>Workload/Stress effects on performance</u>				
	Performance time is a function of time demand/ time available	No	No	No
<u>Inputs</u>				
Task identity	Task time and distribution Triggering conditions Following tasks Priority Task demand and channel identification	Task identities Mental functions System states Control/Display locations Information absorption time	Task demands per channel	
<u>Performance Outputs</u>				
Time	Yes (P(Complete))			
Errors	Yes (P(Error))	Yes	Yes	Yes
Time line analysis	Yes (Lead/Channel)			
<u>Model Running</u>				
	Monte Carlo simulation	Analytical models	Analytic	
<u>Software availability</u>				
	PC (MICROSAINT)	PC (Available only to contractor customer)	No	No
<u>Validation</u>				
	No	No	Yes (but not available)	No

TABLE 15-3 Models With Potential Application To The Current Design Problem

	WINDEX	A ³ I	LAUGHERY	HOS
Task Selection Decision	NA	Strategy-based skill-rule-knowledge Cost to rule implementation	SAINT: User-specified if-then selection rules	User-specified if-then selection rules
Workload and Time Sharing	Conflict matrix Partial parallel processing McCracken and Aldrich demands	Parallel between VACP channels McCracken and Aldrich demands	Parallel between VACP channels McCracken and Aldrich demands	User-specified parallel or serial processing
Subtask Models	Time and resource demand	Time and resource demand	Time and resource demand	Resident in HOS

its strength and weaknesses; provided below is some specification of the attributes in which they differ, so that rational selection of the optimal model can be facilitated.

The 3-1/2 models listed in Table 15-3 are described in terms of three relevant attributes: their task selection logic, their workload and task interference assumptions, and their mechanism for specifying performance of the component tasks. As is immediately apparent from the table, WINDEX in its current form contains no decision mechanism for the selection of tasks in a serial mode of processing. However, because it is the model that goes farthest toward making plausible theory-based assumptions regarding the interference between concurrent tasks, it is recommended that the logic of the conflict matrix underlying WINDEX be incorporated into whichever of the other three models is ultimately chosen.

The remaining three model approaches may be contrasted first in terms of the task selection algorithms that they adopt. All three involve user-specified rules for task selection. For example, a rule might be "if a target is not visible, then continue to navigate to bring it within visual range. If it is visible, then activate aiming device." All models allow for some specification of the priorities of actions when there is a choice, and HOS allows preemption of ongoing activities of lesser importance by those of greater importance. The A³I model, however, differs from the other three in terms of the sophistication of its assumed decision mechanism. The model allows for action choices to be made at three levels of Rasmussen's (1983)

decision continuum of skill-, rule-, and knowledge-based behavior. This continuum describes the number of contingent conditions that must be considered before arriving at a decision. Increasing levels of contingency yield decisions that take longer and are more demanding of cognitive effort, factors which feed directly into the predicted workload.

The second attribute in Table 15-3 concerns the workload or resource model adopted. Here, it would appear that the A³I model and the SAINT/Siegel and Wolf adaptation of Laughery et al. provide some advantage over the HOS model. This is primarily because the former have incorporated the shell, if not the appropriate details, of a multiple resource approach through the inclusion of the visual, auditory, cognitive, and psychomotor (VACP) channels, and the specification of task demand coding. Furthermore, both approaches appear to allow the number of these channels and the degree of interaction between channels (the latter nonexistent in the current versions) to be modified easily according to user preference. Hence, it would be feasible to modify workload computation algorithms to incorporate the multiple resource assumptions and conflict matrix inherent in WINDEX (North, 1985).

Although the HOS model appears to be less sophisticated (and modifiable) in terms of the dual-task assumptions, it appears to have a greater degree of sophistication built into the operator performance models, which are specified at levels of detail related to retaining information in memory, absorbing information, performing mental computations, and so forth. However, the assumptions lying behind these models do not appear to be documented in the open literature, nor is the most recent version of HOS IV available at this time for public distribution.

Hence, a final recommendation would appear to lie in the choice between the Bolt Beranek and Newman A³I model of Corker et al. and the Laughery et al. SAINT/Siegel and Wolf simulation. Factors favoring the former are (1) the greater sophistication of the task selection decision logic, a logic which is based on plausible assumptions and empirical data, and (2) the fact that the simulation was explicitly developed for an A³I helicopter simulation environment and, therefore, is directly compatible with the goals of the current project. Factors favoring the model of Laughery et al. are the relatively long history of development and application of the SAINT/Siegel and Wolf approach, as well as the commercially available documentation

and user friendliness of the MICROSAINTE software. A final recommendation is that both of these approaches be examined seriously and compared with regard to (1) their feasibility for incorporating WINDEX multiple resource assumptions and (2) their compatibility with other human performance models to be used in the simulation.

CONCLUSION

In conclusion, there is a trade-off between the degree of quantifiable prediction achieved (and perhaps possible) by models of interference and interaction, and the level of environmental complexity and heterogeneity at which those models are suited to operate. Three approaches are possible to extend quantitative prediction to the level of complexity existing in the helicopter cockpit: (1) Attempt to build quantitative elements into a multiple resource/element similarity model. (2) Attempt to extend the more quantitatively precise models of multichannel detection and recognition (e.g., Shaw, 1982; Sperling and Doshier, 1986) to heterogeneous task performance. (3) Establish how accurately complex performance can be accounted for by serial queuing models with assumptions of single-task neglect.

Each approach has its own costs and benefits. The first approach is bound to fall short of precise prediction because of the complexity and heterogeneity of the task environments that its goal is to predict. Yet, clearly, the helicopter pilot will often have to time-share different tasks or mental activities that are heterogeneous in their demand. The second alternative awaits verification: to establish whether, for example, the prediction of performance on a detection task when time-shared with a second simultaneous detection task will generalize to instances when the synchrony in timing is less precise or the concurrent task is of a different qualitative sort (i.e., tracking). The Optimal Control model is a good step in this direction. The third alternative already offers promise as far as it goes, but it is not designed to handle those aspects of time-shared performance that are truly parallel (e.g., flying while communicating). As a final note, whatever combination of approaches is chosen, researchers must increase their tolerance for models that less than perfectly account for the data and allow for adequate, rather than precise, fits.

REFERENCES

- Abernethy, B.
1988 Visual search in sport and ergonomics: Its relationship to selective attention and performer expertise. *Human Performance* 1(4):205-235.
- Aldrich, T.B., Szabo, S.M., and Bierbaum, C.R.
1988 *The Development and Application of Models to Predict Operator Workload During System Design*. Unpublished paper. Fort Rucker, Alabama: Anacapa Sciences, Inc.
- Bateman, R.P., and Thompson, M.W.
1986 Correlation of predicted workload with actual workload measured using the subjective workload assessment technique. *SAE Aerotech 86 Conference Society for Automotive Engineers*.
- Carbonnell, J.R.
1966 A queuing model for many-instrument visual sampling. *IEEE Transactions on Human Factors in Electronics* HFE-7:157-164.
- Carbonnell, J.R., Ward, J.L., and Senders, J.W.
1968 A queuing model of visual sampling: Experimental validation. *IEEE Transactions on Man-Machine Systems* MMS-9:82-87.
- Card, S.K., Moran, T.P., and Newell, A.
1986 The model human processor: An engineering model of human performance. In K.R. Boff, L. Kaufman, and J.P. Thomas, eds., *Handbook of Perception and Human Performance*. New York: John Wiley and Sons.
- Christensen, J.M., O'Donnell, R.D., Shingledecker, C.A., Kraft, C.L., and Williamson, G.
1985 *Optimization of Peripheral Vision*. USAFSAM Technical Report TR-85-96. Brooks AFB, Texas.
- Chu, Y.Y., and Rouse, W.B.
1979 Adaptive allocation of decision making responsibility between human and computer in multi-task situations. *IEEE Transactions on Systems, Man, and Cybernetics* SMC-9:769-778.
- Chubb, G.P., Stodolsky, N., Fleming, W.D., and Hassoun, J.A.
1987 STALL: A simple model for workload analysis in early system development. *Proceedings of the 31st Annual Meeting of the Human Factors Society*. Santa Monica, CA: Human Factors.
- Corker, K., Davis, L., Papazian, B., and Pew, R.
1986 *Development of an Advanced Task Analysis Methodology and Demonstration for Army-NASA Aircrew/Aircraft Integration*. Report No. 6124. Cambridge, MA: Bolt Beranek and Newman, Inc.
- Derrick, W.L., and Wickens, C.D.
1984 *A Multiple Processing Resource Explanation of the Subjective Dimensions of Operator Workload*. Technical Report EPL-842/ONR-84-1. University of Illinois Engineering Psychology Laboratory.
- Drury, C.
1975 Inspecting sheet metal: Model and data. *Human Factors* 17:257-265.
- Fitts, P.M., Jones, R.E., and Milton, J.L.
1949 Eye movements of aircraft pilots during instrument landing approaches. *Aeronautical Engineering Review* 9:1-5.

- Gopher, D., and Braune, R.
1984 On the psychophysics of workload: Why bother with subjective measures? *Human Factors* 2:519-532.
- Harris, R.L., and Christliff, D.M.
1980 What do pilots see in displays? In G. Corrick, E. Hazeltine, and R. Durst, eds., *Proceedings of the 24th Annual Human Factors Society*. Santa Monica, CA: Human Factors.
- Harris, R.M., Iavecchia, H.P., Ross, L.V., and Shaffer, S.C.
1987 Microcomputer human operator simulator (HOS-IV). *Proceedings of the 31st Annual Meeting of the Human Factors Society*. Santa Monica, CA: Human Factors.
- Harris, R., and Spady, A.
1985 Visual scanning behavior. *Proceedings of NAECON 85*. Dayton, OH.
- Kantowitz, B.
1986 Channels and stages in human information processing: A limited analysis of theory and methodology. *Journal of Mathematical Psychology* 29:135-174.
- Kantowitz, B., and Weldon, M.
1985 On scaling performance operating characteristics: Caveat emptor. *Human Factors* 27:531-547.
- Kristofferson, J.
1967 Attention and psychological time. *Acta Psychologica* 27:93-100.
- LaBerge, D.
1973 Identification of two components of the time to switch attention. In S. Kornblum, ed., *Attention and Performance IV*. New York: Academic Press.
- Laughery, R., Drews, C., and Archer, R.
1986 A micro SAINT simulation analyzing operator workload in a future attack helicopter. *Proceedings NAECON*. New York: IEEE.
- Leibowitz, H.W., and Dichgans, J.
1980 The ambient visual systems and spatial orientation. *Spatial disorientation in flight: Current Problems*. NATO AGARD Conference Proceedings.
- Levison, W.H.
1982 The optimal control model for the human operator: Theory, validation, and applications. AFFTCOTR-82-5. *Proceedings, Workshop on Flight Testing to Identify Pilot Workload*. Edwards Air Force Base, Air Force Flight Test Center.
- Levison, W.H., Elkind, J.I., and Ward, J.L.
1971 *Studies of Multi-variable Manual Control Systems: A Model for Task Interference*. NASA CR-1746. Washington, DC: NASA.
- Levison, W.H., and Tanner, R.B.
1971 *A Control-Theory Model for Human Decision Making*. NASA CR-1953. Washington, D.C.: NASA.
- Mackworth, N.H.
1976 Useful field of view. Pp. 173-178 in R.A. Monty and J.W. Senders, eds., *Eye Movements and Psychological Processes*. Hillsdale, NJ: Lawrence Erlbaum Associates.

- McCracken, J., and Aldrich, T.B.
1984 *Analysis of Selected LHX Mission Functions*. Technical Note ASI 479-024-84(B). CA: Anacapa Sciences, Inc.
- Meister, D.
1985 *Behavioral Analysis and Measurement Methods*. New York: John Wiley and Sons.
- Moray, N.
1986 Monitoring behavior and supervisory control. In K. Boff, L. Kaufman, and J.P. Thomas, eds., *Handbook of Perception and Human Performance Volume 2*. New York: John Wiley and Sons.
- Mourant, R.R., and Rockwell, T.H.
1972 Strategies of visual search by novice and experienced drivers. *Human Factors* 14:325-335.
- Norman, D.
1968 Toward a theory of memory and attention. *Psychological Review* 75:522-536.
- Norman, D., and Bobrow, D.J.
1975 On data limited and resource limited processes. *Cognitive Psychology* 7:44-64.
- North, R.
1985 WINDEX: A workload index for interactive crew station evaluation. *Proceedings NAECON*. New York: IEEE.
- North, R. and Riley, V.
1988 WINDEX: A predictive model of operator workload. In G. MacMillan, ed., *Human Performance Models*. NATO AGARD Symposium, Orlando, FL.
- Parasuraman, R.
1986 Vigilance, monitoring, and search. Pp. 43.1-43.39 in K. Boff, L. Kaufman, and J.P. Thomas, eds., *Handbook of Perception and Human Performance*. New York: John Wiley and Sons.
- Pashler, H.
1984 Evidence against late selection: Stimulus quality effects in pre-viewed displays. *Journal of Experimental Psychology: Human Perception and Performance* 10:413.
- Pew, R.W., Baron, S., Fehrer, C.E., and Miller, D.C.
1977 *Critical Review and Analysis of Performance Models Applicable to Man-Machine Systems Evaluation*. BBN Report No. 3446. Cambridge, MA: Bolt Beranek and Newman, Inc.
- Rasmussen, J.
1983 Skills, rules, and knowledge: Signals, signs, and symbols, and other distinction in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics* SMC-13:257-266.
- Rouse, W.
1980 *Systems Engineering Models of Human Machine Interaction*. New York: North Holland.
- Senders, J.W.
1966 A re-analysis of the pilot eye-movement data. *IEEE Transactions on Human Factors in Electronics* HFE-7:103-106.
1983 *Visual Scanning Processes*. Netherlands: University of Tilburg Press.

- Shaw, M. L.
1982 Attending to multiple sources of information: The integration of information in decision making. *Cognitive Psychology* 14:353-409.
- Sheridan, T.B., and Ferrell, W.R., eds.
1974 *Man-Machine Systems*. Cambridge, MA: MIT Press.
- Siegel, A.I., and Wolf, J.J.
1969 *Man-Machine Simulation Models*. New York: John Wiley and Sons.
- Sperling, G.
1984 A unified theory of attention and signal detection. Pp. 103-177 in R. Parasuraman and R. Davies, eds., *Varieties of Attention*. New York: Academic Press.
- Sperling, G., and Doshier, B.A.
1986 Strategy and optimization in human information processing. In K. Boff, L. Kaufman, and J.P. Thomas, eds., *Handbook of Perception and Performance Volume 2*. New York: John Wiley and Sons.
- Stein, W., and Wewerwinke, P.
1983 Human display monitoring and failure detection: Control theoretic models and experiments. *Automatic* 19:711-718.
- Strieb, M., Lane, N., Glenn, F., and Wherry, R.J., Jr.
1981 The human operator simulator: An overview. Pp. 121-152 in J. Moraal & K.F. Kraiss, eds., *Manned System Design: Method, Equipment, and Applications*. New York: Plenum Press.
- Swets, J.A.
1984 Mathematical models of attention. In R. Parasuraman and D.R. Davies, eds., *Varieties of Attention*. New York: Academic Press.
- Taylor, M.M., Lindsay, P.H., and Forbes, S.M.
1967 Quantification of shared capacity processing in auditory and visual discrimination. In A.F. Sanders, ed., *Attention and Performance I*. Amsterdam: North-Holland Publishing Company.
- Townsend, J.
1974 Issues and models concerning the processing of a finite number of inputs. In B. Kantowitz, ed., *Human Information Processing*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Townsend, J.T., and Ashby, F.G.
1983 *Stochastic Modeling of Elementary Psychological Processes*. New York: Cambridge University Press.
- Tsang, P.S., and Wickens, C.D.
1988 The structural constraints and strategic control of resource allocation. *Human Performance* 1:45-72.
- Tulga, M.K., and Sheridan, T.B.
1980 Dynamic decisions and workload in multitask supervisory control. *IEEE Transactions on Systems, Man, and Cybernetics* SMC-10:217-232.
- Welford, A.T.
1967 Single-channel operation in the brain. In A.F. Sanders, ed., *Attention and Performance I*. Amsterdam: North-Holland Publishing Company.

- Wherry, R.J.
1976 The human operator simulator-HOS. In T.B. Sheridan and G. Johanssen, eds., *Monitoring Behavior and Supervisory Control*. New York: Plenum Press.
- Wickens, C.D.
1980 The structure of attentional resources. Pp. 239-257 in R. Nickerson, ed., *Attention and Performance VIII*. Hillsdale, NJ: Lawrence Erlbaum Associates.
1984 Multiple resources in attention. In R. Parasuraman and R. Davies, eds., *Varieties of Attention*. New York: Academic Press.
1987 Attention. In P. Hancock, ed., *Human factors in Psychology*. Amsterdam: North-Holland Publishing Company.
1988 Attention and skilled performance. In D.H. Holding, ed., *Human Skills, Second Edition*. New York: John Wiley and Sons.
- Wickens, C.D., Harwood, K., Segel, L., Teakevic, I., and Sherman, W.
1988 TASKILLAN: a simulation to predict the validity of multiple resource models of aviation workload. *Proceedings of the Human Factors Society, 30th Annual Meeting*. Santa Monica, CA: Human Factors.
- Wickens, C.D., and Liu, Y.
1988 Codes and input modalities in multiple resources: A success and a qualification. *Human Factors* 30.
- Wickens, C.D., Sandry, D., and Vidulich, M.
1983 Compatibility and resource competition between modalities of input, output, and central processing. *Human Factors* 25:227-248.
- Wickens, C.D., and Weingartner, A.
1985 Process control monitoring: The effects of spatial and verbal ability and current task demand. In R. Eberts and C. Eberts, eds., *Trends in Ergonomics and Human Factors*. Amsterdam: North-Holland Publishing Company.
- Wickens, C.D., and Yeh, Y.Y.
1985 POC's and performance decrements: A reply to Kantowitz and Weldon. *Human Factors* 27:549-554.
- Williams, L.
1966 Target conspicuity and visual search. *Human Factors* 8:80-92.
- Wortman, D.B., Duket, S.D., Seifert, D.J., Hann, R.L., and Chubb, G.P.
1978 *Simulation Using SAINT: A User-Oriented Instruction Manual*. AMRL-TR-7761. West Lafayette, IN: Pritsker and Associates.
- Yeh, Y.Y., and Wickens, C.D.
1988 Dissociation of performance and subjective measures of workload. *Human Factors* 30:111-120.
- Zacharias, G.L., Baron, S., and Muralidharan, R.
1981 *A Supervisory Control Model of the AAA Crew*. BBN Report No. 4802. Cambridge, MA: Bolt Beranek and Newman, Inc.

16

Models of Working Memory

STUART K. CARD

Working memory refers to a functional part of human memory that accomplishes "the temporary holding and manipulation of information during the performance of a range of cognitive tasks such as comprehension, learning, and reasoning" (Baddeley, 1986, p. 34). At least three different functions performed by working memory, as expressed by models of cognitive processing, can be described in computational terms. Working memory functions as (1) a place to hold operands, things to be operated on by the operations of cognitive processing; (2) a cache to hold in a rapidly accessible state recently input or used information; and (3) a buffer between processes that happen at incommensurate rates.

In addition to its functions, working memory has also been characterized from two other points of view: time and structure. From a temporal point of view, working memory is the memory people have for information that lasts a few seconds. In this case it is called short-term memory, as distinguished from long-term memory which lasts hours or years. It is also distinguished from very short-term memory which lasts for only a fraction of a second. From a structural point of view, working memory is described in terms of a fixed number of slots, a set of activated nodes, or some other mechanism. In this case, it is usually given a name such as the short-term store (STS) and distinguished on the one hand from a long-term store (LTS) and on the other from sensory buffers, such as visual image store (VIS) or auditory image store (AIS). In structurally oriented descriptions, working memory is sometimes described not as a separate structure but as part of the state of a single, unified memory. For example, working memory may be described as the set of all nodes in a semantic memory that are activated.

Although distinctions among the several kinds of memory and between the two viewpoints for describing them are clear in principle, and are of the sort found in all systems of information storage (such as modern computers), the several memories function together in an integrated way so as to make explication of their interrelationships a difficult problem. The observed behavior of people is the result of the combined mechanisms at work.

Modeling working memory is important because working memory is limited. These limits produce errors or require the use of methods that function within the limits of memory. In cockpit design, the limits of working memory are manifested in pilot errors, especially those induced by high workload and are a strong constraint on the design of cockpit procedures.

PHENOMENA OF WORKING MEMORY

While a number of partial models of working memory exist, they do not yet embrace in a computational framework all the phenomena related to it. This is not surprising when one considers the close coupling of working memory with other cognitive functions. Comprehensive models for working memory may need to co-evolve with comprehensive models of human cognitive architecture, rather than being developed as isolated pieces of that architecture.

Nevertheless, a fair amount of knowledge has developed about the functioning of working memory, at least in the handling of verbal tasks (and, more recently, for certain visual tasks). Some of this information may be used in the design of cockpits. Models exist that account for some of these empirically derived phenomena and constrain the properties that comprehensive cognitive architectural models would have to exhibit. These phenomena, and references in the literature discussing them, are listed below. Some 32 phenomena can be classified into (1) the size and decay of verbal working memory, (2) contextual effects, (3) representational effects, (4) chunking, (5) skilled memory, (6) spatial working memory, and (7) phenomena related to long-term memory.

Size and Decay of Verbal Working Memory

The phenomena of size and decay are more or less directly related to limits imposed by working memory on the processing of verbal information.

1. **Short-term memory (STM) decay:** When people are given a list to recall (and prevented from rehearsing), the amount they can recall decays exponentially with the time elapsed before recall (Baddeley, 1986; Peterson and Peterson, 1959).

2. **Immediate memory span:** When people are given a list to recall, the number of items they can recall is about five to nine (Miller, 1956), or three to four reliably and seven to nine probabilistically (50 percent of the time) (Broadbent, 1975).

3. **Buffer span (or running span):** When people are given an information-processing task that prevents the use of long-term memory, the number of things they seem to be able to keep track of is approximately two to four items (Card, Moran, and Newell, 1983; Crowder, 1976).

4. **Effect of item type on span:** The working memory span depends on the type of material being memorized (Cavanaugh, 1972).

5. **Effect of word length:** People asked to repeat sequences of words are much more likely to do so correctly if the words are short than if they are long (Baddeley, Thomson, and Buchanan, 1975).

6. **Temporal span:** People remember the number of words they can read in approximately 1.6 seconds or the number of words they can speak in 1.3 seconds (Baddeley, 1986; Vellar and Baddeley, 1982).

7. **Articulation rate effect:** People who can articulate more rapidly tend to have a longer working memory span (Baddeley, 1986).

8. **Performance despite loading:** People required to keep in memory as many items as their memory span can hold nevertheless perform many other tasks (Baddeley, 1986).

9. **Suffix effect:** An irrelevant item at the end of an auditorily presented list reduces recall of the last few items on the list (Crowder and Morton, 1969).

Context Effects

Context phenomena concern the effects of earlier or later items in working memory on each other.

10. **Recency effect:** The last members of a list of items are recalled better than the others (except for those near the beginning). The closer they are to the end, the better these items are recalled (Postman and Phillips, 1965).

11. **Primacy effect:** The first members of a list are recalled better than the others (except for the ones near the very end). The

closer they are to the beginning, the better these items are recalled (Glanzer, 1972).

12. Release from proactive interference: When people are given a list of similar words to recall and rehearsal is prevented, recall is decreased with each sequential item. However, if an unrelated item occurs on the list, recall for that item is nearly as good as for the first item (Loess, 1968; Wickens, 1970).

13. Episodic memory: A task that is interrupted by another task which consumes the full immediate memory span does not have to be restarted from scratch, but can be resumed after some effort (Tulving, 1972, 1983, 1984).

Working Memory Representation

Representational phenomena concern the way in which items in working memory are actually coded or represented.

14. Phonological similarity effect: When people are given a list to recall immediately, they tend to confuse items that sound the same, reducing the number they can remember. This is true even if the list is presented visually (Baddeley, 1986; Conrad, 1964).

15. Unattended speech effect: When people are given a visual digit to remember in the presence of background noise consisting of spoken digits, recall is reduced and reduced much more than if the unattended audio input had been simply white noise (Salame and Baddeley, 1982).

16. Sequential output bias: When people are given a list to recall, it can be recalled forward much more easily than in reverse (Anders and Lillyquist, 1971).

17. Independence of item order information: When people remember lists, order information is lost more rapidly than content (Healy, 1982).

Chunking

The next set of phenomena arises because items in working memory comprise links to elements in long-term memory, rather than the elements themselves.

18. Chunking of recall: When people are given a list to recall, they naturally group the items in time into groups of three to four elements (Johnson, 1970, 1972).

19. **Between-chunk pauses:** When recalling information, people pause longer between, than within, chunks. Pauses between chunks tend to be around 2 seconds (Broadbent, 1975; Chase and Simon, 1973; McLean and Gregg, 1967; Reitman and Reuter, 1980).

20. **Opaqueness of chunks:** Retrieving a chunk at one level does not give one direct access to the content of the chunk at the next lower level (Johnson, 1970, 1972).

Skilled Memory

Phenomena of skilled memory relate to a few ways in which humans can optionally control processes in working memory so as to improve recall.

21. **Efficacy of rehearsal:** Items can be retained in immediate memory indefinitely if rehearsal is allowed (Baddeley, 1986).

22. **Efficacy of mnemonics:** The use of peg words (e.g., one is a bun, two is a shoe) or the method of loci can improve recall (Bellazza, 1981, 1982; Bower, 1970).

23. **Efficacy of elaboration:** Elaboration of associations improves storage and hence recall of information (Craik and Lockhart, 1975).

Spatial Working Memory

The following phenomena reflect working memory for nonverbal information.

24. **Multiple buffers:** The number of items people can remember is larger if they can simultaneously make use of several modalities (visual, motor, auditory) (Baddeley and Hitch, 1974).

25. **Spatial memory disruption:** Tasks involving spatial memory disrupt the simultaneous performance of other spatial tasks (Baddeley, 1986).

26. **Spatial imagery interference:** A concurrent spatial task disrupts the attempt to use an imagery-based mnemonic technique (Baddeley, 1986; Baddeley and Lieberman, 1980).

Long-Term Memory Effects

The following phenomena relate to operations with working memory that give rise to effects in long-term memory.

27. Total time hypothesis: The amount learned is proportional to the amount of time spent learning (Cooper and Pantle, 1967).

28. Elaborative versus maintenance rehearsal: The longer an item spends in working memory under elaborative rehearsal (in which its associations are elaborated), the greater is the probability that it will be recalled. However, maintenance rehearsal, in which an item is rehearsed without thinking about it, does not improve the chances of later recall (Craik and Lockhart, 1975).

29. Long-term recency effect: People recall more recent items better than earlier items, even extending over lengthy periods, provided the events concerned constitute a sufficiently separable category (Baddeley, 1986; Baddeley and Hitch, 1977).

30. Simultaneous long-term recency effect: Long-term recency effects can occur separately for separate categories of events remembered (Watkins and Peynircioglu, 1983).

31. Learning despite impaired working memory: Some neurological patients with impaired working memory appear to have normal long-term learning (Baddeley, 1986).

32. Weber's law time discriminability: The probability of recalling an item is proportional to $\log(DT/T)$, where DT is the time interval between the presentation of items and T is the total elapsed time at recall (Baddeley, 1986; Glenberg, Bradley, Stevenson, Kraus, Tkachuk, Gretz, Fish, and Turpin, 1980).

MODELS OF WORKING MEMORY

A number of models have been devised to handle these memory phenomena. Five models cover the major types: (1) Waugh and Norman (1965), (2) Atkinson and Shiffrin (1968), (3) Baddeley and Hitch (1974; Baddeley, 1986), (4) Anderson's ACT* model (1983), and (5) Schneider and Detweiler's connectionist/control model (1988).

Waugh and Norman's (1965) model includes a short-term store (their version of working memory) and a long-term store. The short-term store is a limited memory with a small number of fixed slots. Items enter the short term store and can get lost either by decay over time or by being displaced by new items. They can be retained through rehearsal. The rehearsal process also allows items to be transferred to the long-term store.

Atkinson and Shiffrin's (1968) model is similar but more differentiated. In addition to the short- and long-term stores it has a sensory

store that is presumed to hold information from one sense modality. The sensory store feeds information into the short-term store that acts as a working memory for various cognitive processes. The longer information is retained in the short-term store, the higher its probability of being transferred to the long-term store. This model also distinguishes between processing structure (the architectural, involuntary structures through which information is processed) and control processes (retrieval strategies, problem-solving techniques, etc.). This model handles many of the basic effects but has difficulty explaining some types of neurological disorders, the lack of certain kinds of incidental learning, long-term storage-based recency effects, and the fact that codes other than phonological codes can be used in the short-term store (see Baddeley, 1986, for a review).

Baddeley and Hitch's (1974; Baddeley, 1986) model of working memory assumes a central executive and two "slave" processors, an "articulatory loop" and a "visual-spatial sketch pad." The articulatory loop consists of a phonological store and an articulatory refreshing process. The visual-spatial sketch pad consists of a spatial memory and an eye-movement-like process. The articulatory loop stores basically verbal information; the visual-spatial sketch pad is specialized to maintain and manipulate visual-spatial images. A central executive coordinates information from the two, allocates attention, and is the medium for what Atkinson and Shiffrin called control processes. This model is broader in its coverage than the others and, in particular, addresses some problems of working memory for images. Although the model gives insights into an impressive number of experimental results, it has unfortunately not been reduced to computational or mathematical form.

Anderson's (1983) ACT* model contains three memories: working memory, declarative memory, and production memory. Declarative memory contains knowledge in the form of chunks (called "cognitive units" in this model). Cognitive units are such things as propositions, strings, or spatial images. Each cognitive unit in declarative memory can have associated with it a certain level of activation. Activation of a cognitive unit spontaneously decays at a certain rate. Chunks have links of different strengths to other cognitive units, and activation spreads along these links depending on their strength. Working memory is simply the set of all cognitive units in declarative memory activated at some particular time.

WORKING MEMORY PHENOMENA		MODELS				
		WN	AS	A	BH	SD
Size and Decay of Verbal Working Memory	1. STM Decay					
	2. Immediate Memory Span					
	3. Buffer Span					
	4. Item-type Effect on Span					
	5. Word Length Effect					
	6. Temporal Span					
	7. Articulation Rate Effect					
	8. Performance Despite Working Memory Loading					
	9. Suffix Effect					
Context Effects	10. Recency Effect					
	11. Primacy Effect					
	12. Release from Proactive Interference					
	13. Episodic Memory					
Working Memory Representation	14. Phonological Similarity					
	15. Unattended Speech Effect					
	16. Sequential Output Bias					
	17. Independence of Order Information					
Chunking	18. Chunking of Recall					
	19. Between-Chunk Pauses					
	20. Opaqueness of Chunks					
Skilled Memory	21. Efficacy of Rehearsal					
	22. Efficacy of Mnemonics					
	23. Efficacy of Elaboration					
Spatial Working Memory	24. Multiple Buffers					
	25. Spatial Memory Disruption					
	26. Spatial Imagery Interference					
Long-Term Memory Effects	27. Total Time Hypothesis					
	28. Elaborative Versus Maintenance Rehearsal					
	29. Long-Term Recency Effect					
	30. Simultaneous Long-Term Recency					
	31. Learning Despite Impaired Working Memory					
	32. Weber's Law Discrimination					

Model Abbreviations: WN = Waugh and Norman (1965); AS = Atkinson and Shiffrin (1968); BH = Baddeley and Hitch (1974), Baddeley (1986); A = Anderson (1983); SD = Schneider and Detweiler (1988).

FIGURE 16-1 Approximate coverage of working memory phenomena by models.

Schneider and Detweiler's (1988) connectionist/control model represents the contents of working memory as weights of arcs connecting neural-like units. Individual units of knowledge (e.g., the letter A) are represented as vectors of activation, such as 0 1 1 1 1 (where the zeros and ones represent the absence and presence of features such as vertical lines, horizontal lines, etc.). The model is described at three levels of detail: a microlevel neural-like network that can produce associative processing and attentional phenomena, a macrolevel that describes attentional control and communication within the system (e.g., how memory scanning works), and a system level that represents interactions between major parts of the system (e.g., the coordination of visual and auditory signals). Simulations have been run with this model to explain a number of the working memory effects listed earlier.

These five models can be divided into two groups: those that are largely models of the working memory component itself (Waugh and Norman, Atkinson and Shiffrin, Baddeley and Hitch) and those in which the working memory model is part of a larger human cognitive architecture (Anderson's ACT* and Schneider and Detweiler). In addition, the second group of models is more computationally oriented than the first. Figure 16-1 shows the approximate coverage of the working memory phenomena listed for the five models considered. Although it has not been possible to assign individual entries with complete certainty, because there is room for controversy on exactly what certain models predict for certain phenomena, it seemed desirable to give some indications of coverage of the various models. A solid square indicates coverage of the phenomenon by the model (although not necessarily computational coverage). A white square indicates lack of coverage. A square shaded gray indicates partial coverage. The figure is intended to suggest which models might be considered depending on what phenomena are important in design. Baddeley's model is a development of, and dominates, the other two traditional psychological models in terms of coverage. Its main problem, in the current context, is that it is not computationally expressed or part of a cognitive architecture. Anderson's ACT* model is attractive because of its integration with such an architecture. Its main drawback is its lower coverage of phenomena. Schneider and Detweiler's model appears to have the most detailed computational coverage of working memory phenomena, although it is not yet part of a comprehensive cognitive architecture.

For a review of the current state of the literature on working memory, the reader is directed to Baddeley (1986), Murdock (1974), and Crowder (1976). For a review of a computational model of working memory, the reader is directed to Schneider and Detweiler (1988).

REFERENCES

- Anders, T.R., and Lillyquist, T.D.
1971 Retrieval time in forward and backward recall. *Psychonomic Science* 22:205-206.
- Anderson, J.R.
1983 *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.
- Atkinson, R.C., and Shiffrin, R.M.
1968 Human memory: A proposed system and its control processes. Pp. 89-198 in K.W. Spence, ed., *The Psychology of Learning and Motivation: Advances in Research and Theory, Volume 2*. New York: Academic Press.
- Baddeley, A.D.
1986 *Working Memory*. Oxford: Clarendon Press.
- Baddeley, A.D., and Hitch, G.J.
1974 Working memory. In G.H. Bower, ed., *The Psychology of Learning and Motivation, Volume 8*. New York: Academic Press.
1977 Recency re-examined. Pp. 205-217 in S. Dornic, ed., *Attention and Performance VI*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Baddeley, A.D., and Lieberman, K.
1980 Spatial working memory. Pp. 521-539 in R. Nickerson, ed., *Attention and Performance VIII*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Baddeley, A.D., Thomson, N., and Buchanan, M.
1975 Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior* 14:157-589.
- Bellazza, F.S.
1981 Mnemonic devices: Classification, characteristics, and criteria. *Review of Educational Research* 51:247-275.
1982 *Improve Your Memory Skills*. Englewood Cliffs, NJ: Prentice-Hall.
- Bower, G.H.
1970 Analysis of a mnemonic device. *American Scientist* 58:496-510.
- Broadbent, D.E.
1975 The magic number seven after fifteen years. In A. Kennedy and A. Wilkes, eds., *Studies in Long-Term Memory*. New York: John Wiley and Sons.
- Card, S.K., Moran, T.P., and Newell, A.N.
1983 *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cavanaugh, J.P.
1972 Relation between the immediate memory span and the memory search rate. *Psychological Review* 79:525-530.

- Chase, W.G., and Simon, H.A.
1973 Perception in chess. *Cognitive Psychology* 4:55-81.
- Conrad, R.
1964 Acoustic confusion in immediate memory. *British Journal of Psychology* 55:75-84.
- Cooper, E.C., and Pantle, A.J.
1967 The total time hypothesis in verbal learning. *Psychological Bulletin* 68:221-234.
- Craik, F.I.M., and Lockhart, R.S.
1975 Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General* 104:268-294.
- Crowder, R.G.
1976 *Principles of Learning and Memory*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Crowder, R.G., and Morton, J.
1969 Precategorical acoustic storage (PAS). *Perception and Psychophysics* 5:365-373. Also in G.H. Bower, ed., *The Psychology of Learning and Motivation: Advances in Research and Theory, Volume 5*. New York: Academic Press.
- Glanzer, M.
1972 Storage mechanisms in recall. In G.H. Bower, ed., *The Psychology of Learning and Motivation: Advances in Research and Theory, Volume 5*. New York: Academic Press.
- Glenberg, A.M., Bradley, M.M., Stevenson, J.A., Kraus, T.A., Tkachuk, M.J., Gretz, A.L., Fish, J.H., and Turpin, B.A.M.
1980 A two-process account of long-term serial position effects. *Journal of Experimental Psychology: Human Learning and Memory* 6:355-369.
- Healy, A.F.
1982 Short-term memory for order information. In G.H. Bower, ed., *The Psychology of Learning and Motivation, Volume 16*. New York: Academic Press.
- Johnson, N.F.
1970 The role of chunking and organization in the process of recall. In G.H. Bower, ed., *The Psychology of Learning and Motivation, Volume 4*. New York: Academic Press.
1972 Organization and the concept of a memory code. In A.W. Melton and E. Martin, eds., *Coding Processes in Human Memory*. Washington, D.C.: Winston.
- Loess, H.
1968 Short-term memory and item similarity. *Journal of Verbal Learning and Verbal Behavior* 7:87-92.
- McLean, R.S., and Gregg, L.W.
1967 Effects of induced chunking on temporal aspects of serial recitation. *Journal of Experimental Psychology* 74:455-459.
- Murdock, B.B., Jr.
1974 *Human Memory: Theory and Data*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Peterson, L.R., and Peterson, M.J.
1959 Short-term retention of individual verbal items. *Journal of Experimental Psychology* 58:283-290.

- Postman, L., and Phillips, L.W.
1965 Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology* 17:132-138.
- Reitman, J.S., and Reuter, H.H.
1980 Organization revealed by recall orders and confirmed by pauses. *Cognitive Psychology* 12:554-581.
- Salame, P., and Baddeley, A.D.
1982 Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning Behavior* 21:150-164.
- Schneider, W., and Detweiler, M.
1988 A connectionist/control architecture for working memory. In G. Bower, ed., *The Psychology of Learning and Motivation: Advances in Research and Theory*. San Diego: Academic Press.
- Tulving, E.
1972 Episodic and semantic memory. In E. Tulving and W. Donaldson, eds., *Organization of Memory*. New York: Academic Press.
1983 *Elements of Episodic Memory*. Oxford: Clarendon.
1984 Precise Elements of Episodic Memory. *The Behavioral and Brain Sciences* 7:223-268.
- Vellar, G., and Baddeley, A.D.
1982 Short-term forgetting and the articulatory loop. *Quarterly Journal of Experimental Psychology* 34:53-60.
- Watkins, M.J., and Peynircioglu, Z.F.
1983 Three recency effects at the same time. *Journal of Verbal Learning and Verbal Behavior* 22:375-384.
- Waugh, N.C., and Norman, D.A.
1965 Primary memory. *Psychological Review* 72:89-104.
- Wickens, D.D.
1970 Encoding categories of words: An empirical approach to meaning. *Psychological Review* 77:1-15.

Training Models to Estimate Training Costs for New Systems

WALTER SCHNEIDER

OVERVIEW

The current theory used to project learning time for systems does not allow detailed projection of training times for systems based on theoretical analysis alone. Some modeling techniques may provide ballpark estimates of learning time that are likely to correlate with true learning times. Learning functions can be reasonably extrapolated from pilot training data. Such estimates could greatly improve the accuracy of projected training times.

Learning time is very dependent on the criterion for performance and the combination of tasks. The time needed to acquire a component skill at a level sufficient for correctly choosing the correct answer on a multiple-choice exam may represent only a small percentage of the time needed to perform the task quickly under high workload; for example, Simon (1986) has estimated that eight seconds is required to learn a new production in long-term memory versus 300 trials if the task is to be performed under high workload (Schneider, 1985). Training for a single task may not transfer well to performing the same task in combination with other tasks (Schneider and Detweiler, 1988).

The problem of accessing skill maintenance is critical to predicting human performance. Some skills decline markedly without practice (Annett, 1979; Farr 1986). Many critical combat skills (e.g., launching a missile) are practiced rarely, with long periods between the training and the critical execution of the skill. Maintenance training is expensive and may require redesign of the equipment (e.g., embedded training).

Cognitive psychology offers a variety of basic research models that provide an interpretation of how practice changes performance. However, these models cannot usefully predict learning time estimates for tasks performed in a virtual cockpit by a virtual human.

Several engineering approximation models can be used to estimate or extrapolate learning times for tasks. The basic approach to these modeling techniques is to estimate the number of components that must be learned, assess the learning time of a subset of the components for a few subjects developing a modest skill level, and then project the total training time for the average subject learning all the components to the desired skill level.

It is important to note that there are few if any "constants" in human learning time. Human learning time depends on the similarity of the new material to previously learned material: the compatibility of the material; and the speed, reliability, and resource (attention) requirements of the task. Human learning needs to be characterized in a high-dimensional space with all the dimensions interacting. Hence one must be very cautious when making a projection of learning time based on a small sample of the learning space. It is important to identify any boundary conditions and the expected error of any projected learning time.

SKILL DEVELOPMENT

In general, skills are developed via execution of the skill in the target task or in a task very similar to the target. Many researchers (e.g., Anderson, 1983; James, 1980; LaBerge, 1976; Posner and Snyder, 1975; Shiffrin and Schneider, 1977) conclude that skills can be characterized by at least two stages. Some models have broken down skills to as many as five stages (i.e., Schneider and Detweiler, 1987). The two major stages will be referred to here as controlled and automatic processing (Anderson, 1983, uses the terms interpretive and compiled processing of productions). Controlled processing is characterized as the slow, serial, effortful form of processing typical of a novice performer. For example, dialing a novel telephone number requires control processing to rehearse the number and enter the random string of digits. Tasks requiring variable responding of the processing of degraded stimuli are likely to require attentional resources even after extended training (e.g., inconsistent arming sequences for different weapon systems required to identify a target in camouflage). Automatic processing develops after extended training

and is characterized as fast, parallel, reliable, low effort, and somewhat difficult to inhibit. Dialing a well learned telephone number is an example of an automatic process. Dialing can be fast, requires little effort, can be done while performing other tasks, and may occur when not intended (e.g., dialing your home number when meaning to dial a related number).

The training requirement and resource demands of performing a task vary greatly depending on whether the task is performed in a controlled or automatic mode. Most simple rule tasks (e.g., a 10-step procedure for setting a radio to receive messages) can be acquired in a few trials as long as the subject can attend fully to the task and not be distracted by having to perform other tasks. However, if the subject must perform the task after months of delay while engaged in a concurrent high workload task, hundreds or even thousands of trials may be needed to learn to develop the task reliably. For example, Schneider and Fisk (1984) trained subjects to perform a category search task (e.g., respond to animal names). When subjects were allowed to attend to the task, they could perform the task accurately after a single trial. However, if subjects had to concurrently perform a digit search task, they required eight hours of training before category detection was high while performing a concurrent category/digit search task. Depending on the criterion (e.g., good performance in an attended state versus heavy dual-task load), the required number of learning trials can vary by a factor of 100. This large variability makes it difficult to estimate learning time without precise specification of the performance criteria (response time and accuracy), task environment (concurrent tasks), and similarity of the task to other tasks.

Engineering design decisions can have a large effect on whether automatic processing is possible for a task and on the amount of training necessary to make the task automatic. For automatic processing to develop there must be a consistent relationship between the internal (e.g., operator's goal state) and external states (e.g., press the "Esc" key to exit the current function); if the exit goal processing can be learned quickly, it could transfer to all other situations, and processing could become automatic. Unfortunately, all too often, different programs use different sequences. Thus, when users need to perform a function they must remember to recall what program they are in and what the exit function is for that program. If distracted, they will tend to enter the keys for the most frequently

exited program. In cockpit design, it is essential to maintain compatibility across controls and between control and real-world consistency (e.g., some locks lock by turning the key clockwise; others, counterclockwise). Operators can work with systems for years and still have to consciously recall tasks before each execution (or perform multiple tasks such as turning the key in the lock both ways).

Providing the user with a consistent response set can have order of magnitude effects on performance speed, resource load, effort, reliability, and retention. Schneider and Fisk (1984) have studied the development of automatic processing using a consistent and varied mapping paradigm. They typically use a search task in which subjects respond when stimuli of a particular class match (e.g., respond if an animal word appears). In a consistent mapping, subjects always respond to the stimuli in the same way (e.g., always respond to animal words and not to color words). In a varied mapping, the assignment is altered across trials (e.g., on one trial search for animals, ignore colors, on the next trial do the opposite). After several hundred trials of consistent mapping, automatic processing usually develops. In contrast, practice in a varied mapping task remains controlled even after months of training (see Schneider and Shiffrin, 1977).

There are large qualitative differences between controlled and automatic processing. In memory comparison (Fisk and Schneider, 1983, searching for semantic categories) controlled processing was 100 times slower: 202 milliseconds for controlled versus 2 milliseconds for automatic (Figure 17-1A). In dual-task memory comparison and digit search, control processing was 25 times more sensitive to the additional workload of the dual task (61 percent decrement for the controlled versus 2 percent for the automatic (Figure 17-1B)). The rated subjective workload category was much higher for controlled processing (Vidulich and Pandit, 1985) than automatic. Automatic processing is more reliable: resistant to the effects of heat stress, alcohol intoxication, and fatigue (see Hancock, 1984; Hancock and Pierce, 1984). In an inconsistent response search task, a 0.1 percent blood level alcohol caused a relative deficit of 37 percent in a controlled processing task and zero percent in an automatic processing task (Fisk and Schneider, 1982). Recent research has shown that automatic processing is retained well after long periods of inactivity. For example, Healy, Fendrich, and Proctor (1988) found no loss in an automatic detection skill after fifteen months with a single refresher session at 6 months. Bahrack (1984) has shown that well learned

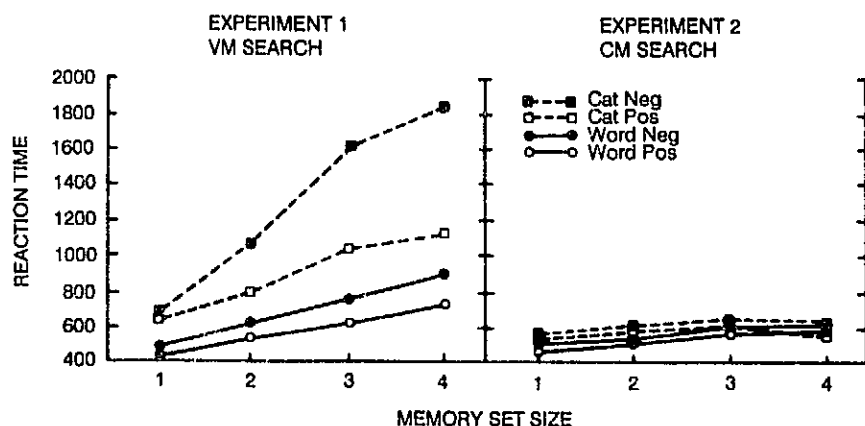


FIGURE 17-1A Search reaction time in Experiments 1 (VM) and 2 (CM) as a function of the memory set size with a display size of two items for the last ten blocks. (VM = varied mapping; CM = consistent mapping). SOURCE: Fisk and Schneider (1983).

material (terms remembered from a high school Spanish language course) can be maintained after 49 years with little loss (see Ericsson and Crutcher, 1988, for a review). In contrast, tasks that are practiced only until they can be performed at a control process level can show rapid decay (e.g., learning the programming commands to implement an averaging algorithm) and often not be retained from the previous night's cramming session for an exam the next day.

Extensive consistent practice can make complex tasks easy. Colle and DeMaio (1978) found that highly trained pilots could perform complex supersonic aircraft formation maneuvers (in a simulator) with no measurable deficit resulting from performing a concurrent digit canceling task. Allport, Antonis, and Reynolds (1972) found experts could sight-read music without deficit while repeating auditory information. Hirst, Spelke, Reaves, Caharack, and Neiser (1980) found that some subjects could read one passage, while simultaneously taking dictation on an unrelated passage, as well as they could perform each task individually. The importance of these results for predicting pilot performance is that any count of the number of components needed to perform a task which does not deal with the nature of the consistency of the task will provide a poor, and probably useless, prediction of actual performance.

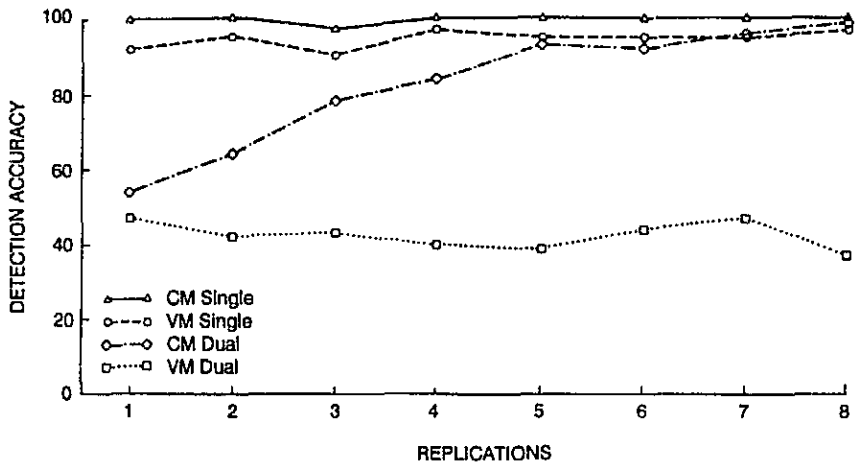


FIGURE 17-1B Single- and dual-task category detection. (For the first four replications, the category-search conditions varied between trials; the last four search conditions varied between blocks. CM = consistently mapped semantic search; VM = variably mapped semantic search.) SOURCE: Schneider and Fisk (1984).

MODELS FOR PREDICTING HUMAN PERFORMANCE

Although there are no global models for accurately simulating the virtual human, a variety of functional relationships can be used to estimate and project performance, given some sample data from the domain. These allow extrapolations of performance to be made, and may provide estimates of training time and performance levels from data developed on a virtual design. A common practice in engineering is to fit some approximation function (e.g., a Taylor series) to predict the behavior of a system that is not characterized precisely in terms of underlying functional relationships. In psychology, a variety of modeling approaches have demonstrated their effectiveness at characterizing performance.

Basic research models of human learning and performance are generally computer simulation models that perform the target task and predict human performance and learning data. None of these models has been developed on a scale that could be applied to the task of flying an aircraft. However, the techniques could model component tasks (e.g., setting up a radio).

Curve-Fitting Techniques

The major techniques of modeling learning have been based on fitting acquisition and decay functions. The major modeling classes currently available are random process models, learning curve models, and identical component models.

Random process models assume human learning to be a random process in which the learner goes from an unlearned to a learned state (see Atkinson, Bower and Crothers, 1965; Coombs, Dawes, and Tversky, 1970; Wickens, 1982). A typical example is to fit human performance to a Markov model with a number of knowledge components; for each learning trial, there is a certain probability that the knowledge state will change to a learned state. The transition probability must be derived empirically for a given problem area. However, once this has been derived it can be used to project the number of trainees that will have a given knowledge level as a function of the number of knowledge components to be trained and the number of trials to be learned. This has been successful in estimating training time (e.g., Rigg, Gray, Tillman, and Pryor, 1982) and in determining how to change practice sets in computerized training systems (e.g., see Suppes and Ginsberg, 1963).

The second curve-fitting technique involves modeling learning and retention functions as a negatively accelerated function. Learning is typically modeled as a power, exponential, hyperbolic, or logarithmic function (for a review, see Lane, 1986) of the number of training trials. Depending on the specific data, these all fit approximately equally (in terms of variance accounted for), generally accounting for more than 90 percent of the practice variance. The power law (Figure 17-2) and negative exponential fit equally well (almost always within 1 percent of variance accounted for; see Lane, 1986). For purposes of projecting training time, either function could be used. Recently, the power law has been the most popular representation of performance. Plotting the log of reaction time as a log function of trials produces a straight line for a power law. The remaining discussion focuses on the power law, but the same comments apply to the other functions. Newell and Rosenbloom (1981) reviewed dozens of studies ranging from cigar rolling to playing bridge and showed that all the data were well fit by a power law.

It is important to note that the parameters for the power law must be determined by empirical data. There are at least two parameters in the power law: (1) the time to perform the trial the first time and (2) the learning rate, the amount of reduction in learning time.

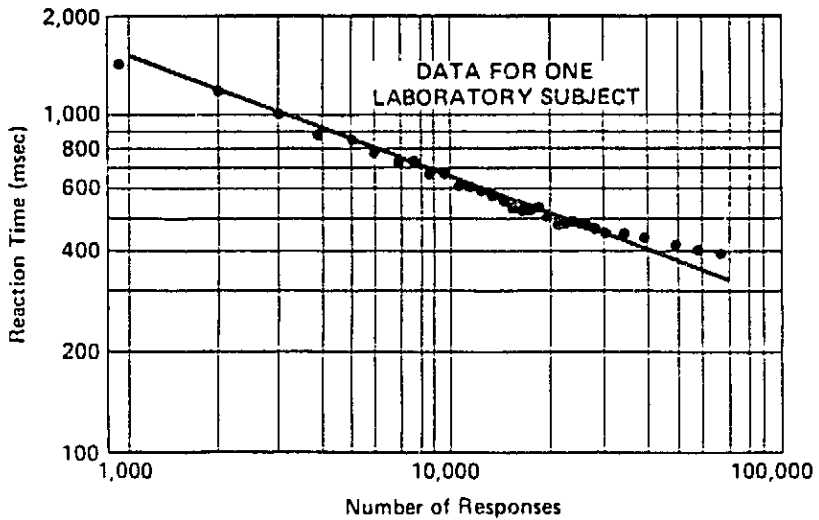


FIGURE 17-2 An example of the Power Law of Practice. Improvement of reaction time with practice on a 1023-choice task. Subjects pressed keys on a ten-finger chordset according to pattern of lights directly above the keys. After Klemmer (1962). SOURCE: Card, Moran, and Newell (1983).

In many situations one must estimate two additional parameters for the number of pretraining trials and the asymptotic performance level of the task. In the Newell and Rosenbloom (1981) review, the initial response time parameter ranged from 0.68 to 1,763 seconds. Such a wide range of variability illustrates the need for empirical data to estimate the learning rate for a given task component. One can get a reasonable approximation of these parameters by measuring the behavior of only a few individuals performing a modest number of executions (e.g., 100) of the task. This provides data that enable predicting performance improvement as a function of extended training.

In addition to predicting response processing, one must be able to predict error rates. A power law can be used to predict the log of the error rate as a function of the log of the number of trials (e.g., Anderson, Conrad, and Corbett, in press). The predictive validity of fits to the accuracy data has not been studied extensively. Accuracy is difficult to predict in situations of high workload because single-task accuracy is often a poor predictor of task performance under high workload (see Schneider and Detweiler, 1988).

Simulation Models

Computer simulation models have been developed to understand and predict human learning and performance. These models generally involve developing a cognitive architecture to accomplish a task and then fitting parameters of the model to the human data to predict performance on a variety of tasks and practice levels.

The most active learning effect models can be divided into production system, connectionist, and hybrid models. Each of these developed in some branch of cognitive science to simulate human learning. The models are generally developed as an existence proof to show that the assumptions of the model are sufficient to perform the task.

Production System Models

Production system models model human performance in terms of a series of "if-then" rules that operate in a working memory to perform tasks. Operations involve changes in memory, goal states, and actions. The process of modeling involves specifying the productions necessary to perform a task, the resources available to store intermediate results, and the learning and decay rates of various operations of the system. Developing a model involves building a program to perform the task. The models are similar to expert system models of performance.

The range of phenomena that can be modeled is limited in the same sense that expert system modeling is limited. If one could build a complete expert system for a pilot, one could simply replace the pilot, rather than having to develop a model to predict learning time. Given the current limitations of modeling, the full task cannot be modeled. However, models can provide estimates of learning time and performance of the procedural tasks (e.g., how long it would take to learn the engine start-up procedure of a variety of configurations).

A variety of models are production system oriented. The model most directly oriented to solving engineering models is the GOMS model of Card, Moran, and Newell (1983). This model has been applied to evaluating human computer interfaces to determine the relative merits of editor command sequences. Building the model requires identifying the set of goals, operators, methods of achieving the goals, and selection rules for choosing among competing goals (hence, the name GOMS). To model a series of computer word-processing tasks required a model with 20 goals, 13 operators, 6

methods, and 4 selection rules. Detailed second-by-second protocols were collected on 2 operators performing the tasks. The parameters were estimated based on the protocols. The coding process is very time-consuming, typically requiring hundreds of hours of coding time for a single study. The time for each of the component operations was estimated from the protocols. The duration of operators varied over two orders of magnitude (e.g., from 0.13 to 9.72 seconds). This illustrates the critical need to estimate the parameters. No global operator constant would produce a useful prediction. The model was tested by having it predict new unit tasks not originally used to estimate the parameters. The model was able to predict new unit task performance time within 35 percent and total time to perform a 20 minute editing task within 4 percent. Developing and validating the model is a time-consuming process.

Once the model has been developed, simulations can be performed to predict behavior on new configurations and at various skill levels. This involves specifying what operators are needed to perform the tasks with different designs (e.g., how can you replace a word in different editors) and then running the simulation. The relative merits of different designs on a variety of work tasks can then be estimated without further empirical study. One can also run sensitivity analyses on the model to determine the potential gain from changes in the engineering design. The GOMS model illustrates the potential gains and the large front-end costs to develop the model and estimate the parameters from protocols required for this class of modeling.

A variety of cognitive learning models can be used to estimate learning time. For example ACT* (Anderson, 1983), SOAR (Laird, Rosenbloom, and Newell, 1989), and SIERRA (VanLehn, 1983) all model human learning. ACT*, for example, has been applied to learning ranging from LISP programming to basic addition. These models predict how humans develop new productions during problem solving behavior; they might allow the benefits of practice in developing the skill to be predicted. Developing models for specific tasks is time consuming (e.g., requiring five man years for the LISP learning model) but allows estimation of the practice functions and can often be the basis for developing an intelligent tutoring system (e.g., Anderson et al., in press).

Polson and Kieras (1985) have used a production system model to predict learning times for various editor commands. The model

and empirical validation show that tasks which share similar productions exhibit a large degree of transfer. This technique may provide an estimate of learning time for tasks without requiring empirical data on every component.

Connectionist Modeling

Recently, connectionist modeling has generated a large amount of interest in cognitive science (see Rummelhart and McClelland, 1987; Schneider, 1986). These models represent learning as a process of changing connection weights between simple neuron-like units and might be applied in two ways in future engineering modeling. First, the gradient descent learning algorithms might be used as a nonlinear curve-fitting technique to predict learning time. Such models are currently being used in diverse areas (e.g., to predict chemical properties of new molecules or loan qualifications based on simple features). Perhaps such techniques could be used to predict the learning times of new tasks. However, in order to fit the many parameters of such models, very large data bases are required with clear measures of performance. Note that in many real-world tasks, it is difficult to obtain clear quantitative measures of performance.

The second use of connectionist models is to model human cognitive functioning. In sharp contrast to production system models, all the information in connectionist models interacts. All the knowledge is stored in a small number of connection matrices. For example, all associations between the acoustic and semantic representation of a task would be stored in one matrix. The implication of this is that all knowledge interacts. These models clearly show a wide variability of learning times for new components as a function of similarity to previous material (e.g., in NET-TALK, Sejenowski and Rosenberg, 1987). New words with similar phonemic relationships can be learned with little or no training (e.g., three trials or less) whereas new words with dissimilar patterns may require hundreds of trials. Basic research understanding of these models may provide a useful prediction of new learning as a function of its similarity to previous learning. The inability to predict the effects of similarity is probably the greatest hindrance to predicting human performance. It is important to note, however, that it may be years before such models can deal with similarity effects in real-world learning environments.

Hybrid Architectures

Recent, hybrid models have combined elements common to both production system and connectionist models. These hybrid models promise a better understanding of the stages of skill acquisition. Initial learning and performance appear to be based on rules. As practice continues, connectionist associative retrieval is substituted for rule-based execution. The Hunt and Lansman (1986) model has a production rule interpreter that directly associates the input to output across processing stages (e.g., a visual cue evoking a cognitive process). Schneider and colleagues (Oliver and Schneider, 1988; Schneider, 1985; Schneider and Detweiler, 1987, 1988) have developed a connectionist/control architecture that models controlled and automatic processing. Initially, performance is rule governed. However, as practice occurs, performance passes through five phases as automatic processing develops. This approach may allow interpretation of why single-task training is such a poor predictor of high workload performance (see Schneider and Detweiler, 1988). As with connectionist models, models in this area must be developed substantially before they can be applied directly to estimating human learning.

ENGINEERING GUIDANCE WITHOUT AN ALL-INCLUSIVE MODEL

There are, at present, no complete models of cognitive processing that can predict total task performance in tasks having the complexity of flying an aircraft. However, there is substantial knowledge about the impact of engineering decisions on training time. This knowledge can provide guidelines to better develop skill learning. Traditional workload analysis has proceeded without an all-inclusive model to identify points of unreasonable workload in a design (e.g., having to perform two different movements at a given point in time). For projecting training time, one can analyze the static parameters of the design, determining the number of component tasks to be performed and using an approximation model to estimate learning time. One can determine which component tasks must be done with concurrent workload and which are compatible with previous responses. It is important to remember that training time is determined by many dimensions of the task, most of which have strong interactions (e.g., compatibility between tasks is more important than the raw

number of tasks, see Polson, 1988). To keep the user of such information aware of the limited variance accounted for, it is important to provide standard error estimates, as well as mean training times, and to validate any model with human data.

Trade-offs in design must be addressed. For example, in current cockpit designs information appears on virtual displays. Complex systems may have a few to dozens of display modes. Information that is in the computer but not attended to (either due to operator inattention or to the operator's not displaying the appropriate screen) results in poor performance. Data on the learning time and operator requests for screens might be used to limit the number and types of virtual displays employed during critical segments of missions.

An initial workup of a design should include a number of factors. For example, how many new component steps must be learned to perform the task (e.g., firing a gun requires a given number of steps)? How many of those steps are new relative to previously learned tasks? How many are incompatible with other operations? Will these steps be performed under heavy workload or in degraded stimulus conditions? What information must be maintained in working memory, and how rapidly must the operations be performed? What is the frequency of the operations in normal training, operations, and time-critical combat situations? What is the cost of errors of the system?

USE OF RAPID PROTOTYPING AND QUICK EMPIRICAL EVALUATIONS

The inability to make accurate projections of training time emphasizes the need to obtain empirical data early in the design process. Rapid prototyping of design systems with quick empirical tests of loaded pilot performance would allow evaluation testing of designs. It is important to note that most critical combat-related tasks must be performed under conditions of high workload. In combat, the aircrew is always engaged in navigation, threat avoidance, and flight control, which severely limits the resources available for other tasks. Evaluation tests should simulate such a load either in a simulated environment or in a calibrated secondary task load situation. Training tests should include initial acquisition, reliability under high workload conditions, and skill maintenance assessment.

NEEDED RESEARCH

To more accurately project training costs and human performance in systems, more research is needed to develop approximation models of training performance, and detailed empirical and theoretical understanding of skill acquisition and retention are required.

Currently available techniques allow extrapolation of training time only after extensive collection of empirical data on either real or simulated systems. Training performance is determined by the interaction of the number of components to be trained, component consistency-compatibility, workload, similarity to other tasks, and retention periods. These dimensions are highly interactive, and no validated modeling technique can currently relate all of them.

Attempts should be made to develop and evaluate projection models of training time. An example of the beginning of such an attempt is the Knerr, Nadler, Dowell, and Trifano (1983) army project. The modeling might be either in the factor analytic tradition or in nonlinear factor analysis (e.g., connectionist modeling) techniques. Attempts to predict software development costs and time (e.g., Brooks, 1975; Putman, 1983) might provide an example of analogous prediction problems. In all such cases, the development of an empirical data base to validate such a model is critical (Maitland, 1982; Neal, 1982). The current lack of training cost projection models leaves the system evaluator with no objective criterion for assessing the potentially most expensive aspect of a design.

Better basic research understanding and modeling of skill acquisition, particularly under high workload situations, is critical. If researchers cannot predict multitask performance based on single-task performance (see Schneider and Detweiler, 1988) or if adding a new task substantially alters the rank ordering of all previous tasks, the accuracy of prediction is severely limited. Simply collecting more data from empirical research on training is unlikely to help. There have been three decades of research on part-task training that provide only broad guidelines (Adams, 1987; Stammers, 1982). Research must focus on characterizing the learning space in quantifiable dimensions and predicting skill acquisition times as a function of training time and procedures. The understanding of cognitive architectures via computer simulation provides methods of testing the learning theories of skill acquisition. Models are required that can identify predictor variables of learning time with all the interacting variables present in real-world design trade-off situations.

REFERENCES

- Adams, J.A.
 1987 Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. *Psychological Bulletin* 101:41-74.
- Ailport, D.A., Antonis, B., and Reynolds, P.
 1972 On the division of attention: A disproof of the single channel hypotheses. *Quarterly Journal of Experimental Psychology* 24:225-235.
- Anderson, J.R.
 1983 *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.
- Anderson, J.R., Conrad, F.G., and Corbett, A.T.
 in press Skill Acquisition in the LISP Tutor. *Cognitive Science*.
- Annett, J.
 1979 Memory for skill. In M.M. Gruneberg and P.E. Morris, eds., *Applied Problems in Memory*. New York: Academic Press.
- Atkinson, R.C., Bower, G.H., and Crothers, E.J., eds.
 1965 *An Introduction to Mathematical Learning Theory*. New York: John Wiley and Sons.
- Bahrack, H.P.
 1984 Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General* 113:1-29.
- Brooks, F.
 1975 *The Mythical Man-Month*. Reading, MA: Addison-Wesley Publishing Company.
- Card, S.K., Moran, T.P., and Newell, A.
 1983 *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Colle, H.A., and DeMaio, J.
 1978 *Measurement of Attentional Capacity Load Using Dual-Task Performance Operating Curves*. Interim report AFHRL-TR-78-5. Air Force Systems Command, Brooks AFB, Texas.
- Coombs, C.H., Dawes, R.M., and Tversky, A., Eds.
 1970 *Mathematical Psychology: An Elementary Introduction*. Englewood Cliffs, NJ: Prentice-Hall.
- Ericsson, K.A., and Crutcher, R.J.
 1988 *Long-Term Retention of Sequentially Organized Information, Knowledge, and Skills: An Empirical Review*. Unpublished manuscript, University of Colorado, Boulder.
- Farr, M.J.
 1986 *The Long-Term Retention of Knowledge and Skills: A Cognitive and Instructional Perspective*. IDA Memorandum Report M-205. Institute for Defense Analyses, Alexandria, VA.
- Fisk, A.D., and Schneider, W.
 1983 Category and word search: Generalizing search principles to complex processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 9:177-195.

- Hancock, P.A.
1984 Environmental stressors. J.S. Warm, ed., *Sustained Attention in Human Performance*. NY: Wiley & Sons.
- Hancock, P.A., and Pierce, J.O.
1984 Toward an attentional theory of performance under stress: Evidence from studies of vigilance in heat and cold. In A. Mital, ed., *Trends in Ergonomics/Human Factors I*. Amsterdam: North-Holland.
- Healy, A.F., Fendrich, D.W., and Proctor, J.D.
1988 *The Effects of Training on Letter Detection*. Unpublished manuscript.
- Hirst, W., Spelke, E.S., Reaves, C.C., Caharack, G., and Neisser, U.
1980 Dividing attention without alternation or automaticity. *Journal of Experimental Psychology: General* 109:98-117.
- Hunt, E., and Lansman, M.
1986 Unified model of attention and problem solving. *Psychological Review* 93:446-461.
- James, W.
1890 *Principles of Psychology, Volume I*. New York: Holt.
- Knerr, C.M., Nadler, L.B., Dowell, S.K., and Trifano, D.R.
1983 Comparison of training transfer and effectiveness models. Pp. 906-910 in *Proceedings of the Human Factors Society 27th Annual Meeting*. Santa Monica, CA: Human Factors Society.
- LaBerge, D.
1976 Perceptual learning and attention. Pp. 237-273 in W.K. Estes, ed., *Handbook of Learning and Cognitive Processes, Volume 4*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Laird, J., Rosenbloom, P., and Newell, A., eds.
1986 *Universal Subgoalng and Chunking: The Automatic Generation and Learning of Goal Hierarchies*. Boston, MA: Kluwer Academic Publishers.
- Lane, M.E.
1986 *Skill Acquisition Curves and Military Training*. IDA Paper P-1945. Institute for Defense Analyses, Alexandria, VA.
- Maitland, A.J.
1982 Training effectiveness analysis: Where the operator meets the equipment. Pp. 255-257 in *Proceedings of the Human Factors Society*. Santa Monica, CA: Human Factors Society.
- Neal, G.L.
1982 Overview of training effectiveness. Pp. 244-248 in *Proceedings of the Human Factors Society*. Santa Monica, CA: Human Factors Society.
- Newell, A.
1989 *Unified Theories of Cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Newell, A., and Rosenbloom, P.S.
1981 Mechanisms of skill acquisition and the law of practice. In J.R. Anderson, ed., *Cognitive Skills and Their Acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Oliver, W.L., and Schneider, W.
1988 Using rules and task division to augment connectionist learning. *Proceedings of the Tenth Annual Conference of the Cognitive Science Society* 55-61.

- Polson, P.G.
1988 The consequences of consistent and inconsistent user interfaces. Pp. 59-108 in R. Guindon, ed., *Cognitive Science Over its Application for Human Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Polson, P.G., and Kieras, D.E.
1985 A qualitative model of the learning and performance of text editing knowledge. *Proceedings of the CHI 1985 Conference on Human Factors in Computing* 207-212.
- Posner, M.I., and Snyder, C.R.R.
1975 Attention and cognitive control. In R.L. Solso, ed., *Information Processing and Cognition: The Loyola Symposium*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Putnam, L.H.
1978 A general empirical solution to the macro software sizing and estimation problem. *IEEE Transactions on Software Engineering* SE-4:335-361.
- Rigg, K.E., Gray, B.B., Tillman, B.W., and Pryor, J.H.
1982 Estimating skill retention from initial training. Pp. 263-265 in *Proceedings of the Human Factors Society*. Santa Monica, CA: Human Factors Society.
- Rumelhart, D.E., and McClelland, J.L.
1986 On learning the past tense of English verbs. Pp. 216-271 in J.L. McClelland and D.E. Rumelhart, eds., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models*. Cambridge MA: MIT Press.
- Schneider, W.
1985 Training high-performance skills: Fallacies and guidelines. *Human Factors* 27:285-300.
1987 Connectionism: Is it a paradigm shift for psychology? *Behavioral Research Methods, Instruments, and Computers* 19:73-83.
- Schneider, W., and Detweiler, M.
1987 A connectionist/control architecture for working memory. Pp. 54-119 in G.H. Bower, ed., *The Psychology of Learning and Motivation, Volume 21*. New York: Academic Press.
1988 The role of practice in dual-task performance: Toward workload modeling in a connectionist-control architecture. *Human Factors* 30:539-566.
- Schneider, W., and Fisk, A.D.
1982 Concurrent automatic and controlled visual search: Can processing occur without resource cost? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 8:261-278.
1984 Automatic category search and its transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10:1-15.
- Schneider, W., and Shiffrin, R.M.
1977 Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review* 84:1-66.
- Sejnowski, T.S., and Rosenberg, C.R.
1987 Parallel networks that learn to pronounce English text. *Complex Systems* 1:145-168.

- Shiffrin, R.M., and Schneider, W.
1977 Controlled and automatic human information processing: II. Perceptual learning, automatic, attending, and a general theory. *Psychological Review* 84:127-190.
- Simon, H.K.A.
1986 The parameters of human memory. Pp. 299-309 in F. Klix and H. Hagendorf, eds., *Human Memory and Cognitive Capabilities*. New York: Elsevier Science Publishers.
- Stammers, R.B.
1982 Part and whole practice in training for procedural tasks. *Human Learning* 1:185-207.
- Suppes, P., and Ginsberg, R.
1963 A fundamental property of all-or-none models. *Psychological Review* 70:139-161.
- VanLehn, K.
1983 *Felicity Conditions for Human Skill Acquisition: Validating an AI-Based Theory*. Technical Report No. CIS-21. Palo Alto, CA: Xerox, Palo Alto Research Center, Cognitive and Instructional Sciences.
- Vidulich, M., and Pandit.
1985 *A Laboratory Study of Workload and Automaticity*. Unpublished manuscript.
- Wickens, C.D.
1982 *Models for Behavior: Stochastic Processes in Psychology*. San Francisco, CA: W.H. Freeman and Company.
- Wightman, D.C., and Lintern, G.
1985 Part-task training for tracking and manual control. *Human Factors* 27:267-283.

Modeling Scenarios for Action

STUART K. CARD

To bring human performance (or other) models to bear at design time, it is necessary to predict the interactions of pilot, aircraft, and environment. This can be done roughly, and at great time and expense, by humans in simulators. It would be more effective if a greater portion of the task could be done analytically and computationally. Computational methods have potentially great advantages over empirical human simulations: (1) they might be made vastly cheaper; (2) they might be much faster to use; (3) many more contingencies might be explored; and (4) the need for measurement, data reduction, and interpretation (e.g., by eye-movement cameras) might be replaced by simple data capture. The problem is that pilot behavior depends on higher-order perceptual and mental functions—just the part of the system that is most difficult to model. This chapter collects some of the modeling techniques potentially applicable to this problem.

FIXED SCENARIOS

The standard technique that has evolved to model pilot action is based on fixed scenarios. Each scenario consists of a list of actions, fixed by the analyst, that accomplishes some mission. The actions are then used as input for later analyses. For example, a scenario might involve all the perceptual, control, and communications actions required to take off, fly to a certain destination, and land. From the detailed scenario, the analyst could then pursue other time line construction, workload analysis, anthropometric analyses, or analysis of eye-scanning patterns.

Scenarios come in a number of formats ranging from tables of actions to graphical versions. Figure 18-1 is a summary of a scenario for an aircraft flying a logistical mission (Murphy, Pizzicara, Hamson, and Bernberg, 1967). Figure 18-2 gives a fragment of the scenario detail. The full scenario extends 40 pages, includes 626 named actions, and is one of 3 scenarios used for this cockpit analysis. The scenario includes perceptions (e.g., "assess fuel flow rates"), actions (e.g., "adjust rpm, egt, epr, and oil temperature, pressure, and quantity"), and communications (e.g., "report intelligence to CP"). Figure 18-3 shows control, display, and automation analyses that have been expanded around Task 56 ("adjust throttle"). One analysis (Figure 18-3A) considers what sort of display is needed, how frequently it will be read, and how critical it is; another analysis (Figure 18-3B) considers how the throttle will be controlled; and a third (Figure 18-3C) what kind of automation to provide for the control.

The use of fixed scenarios is a simple, but tedious, technique to model enough of the interaction between the pilot and his environment for other analytical methods to be applied. In fixed scenarios, the analyst transforms a general and brief plan of interaction, such as that in Figure 18-1, into detailed lists of actions by imagining what would happen if one were to interact in the specific situation. Some degree of variability is handled by using sets of different scenarios, strategically chosen so that interesting realms of interaction will be traversed. The scenario technique depends on the fact that the world and the behavior of interest are composed of skilled, routine tasks with designed methods (e.g., landing an aircraft) and that sampling a set of tasks from this world can help identify the major infelicities of the test cockpit.

There are several strong limitations to this approach, however:

1. An analyst might not expand the scenario correctly or might miss the use of items in the environment for memory.
2. Because no contingencies are permitted, even minor changes to the mission, such as flying over new terrain or the addition of other actors, require new analyses.
3. Some inputs, such as determining how high a helicopter would have to pop up to see over a hill, might be tedious to perform.
4. No contingent interactions, such as having a human pilot perform one of the roles of the mission or having the simulation respond to the terrain or to the actions of other actors, are possible.

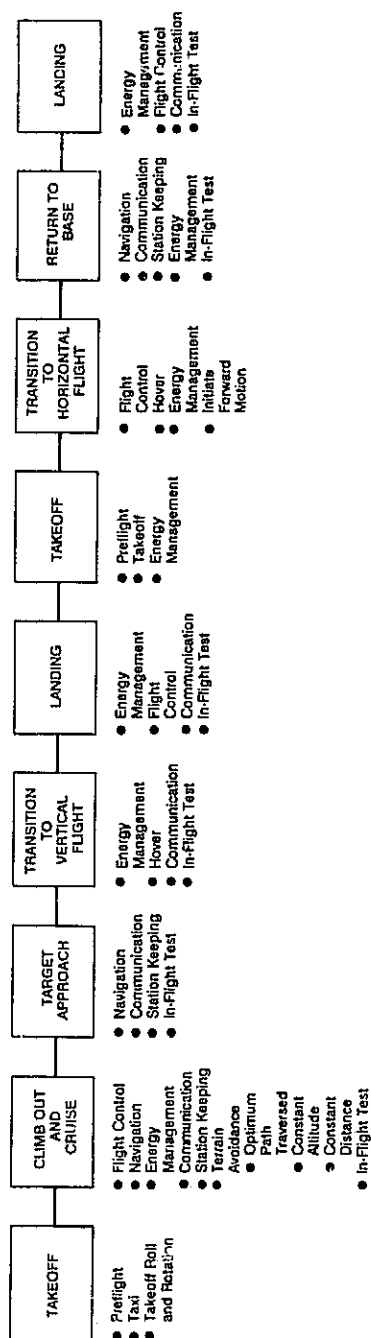


FIGURE 18-1 Summary of scenario for simple logistics mission. SOURCE: Murphy et al. (1967).

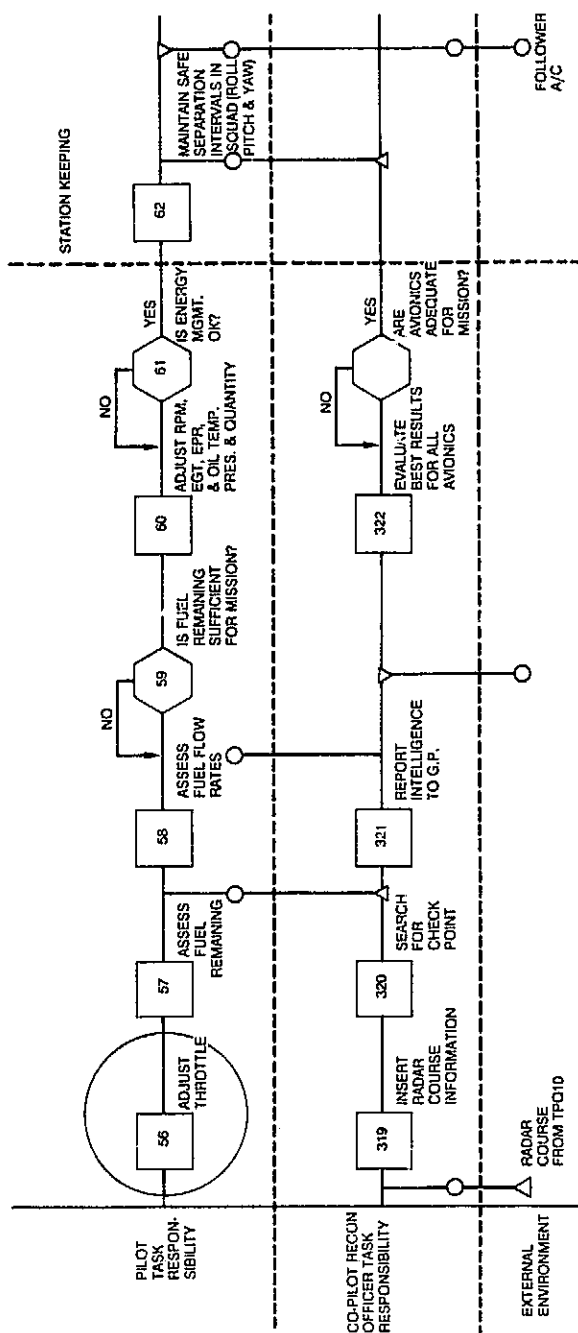


FIGURE 18-2 Fragment of scenario. SOURCE: Murphy et al. (1967).

Functions	Display Requirements	Items to be Displayed	Range of Interest	Accuracy Required	Display Frequency				Display Category		Display Criticality 1=high 2=med 3=low	Remarks
					Continuous	On Demand	Warning Displays	Pri-mary	A/N	Back-Up		
53	Cut gas generators	-----	-----	-----	•	•		•	•			Will be apparent from other displays.
54	Fan doors and bottom louvers	Status	-----	-----								Information is continuous in integrated form on the primary display and in specific numerical form on the A/N display.
55	Check Point Search	Geographical with Ownship and Check Point Indicated	Geographical Position Ownship Position Check Point Position	-----	•	•	•	•	•	•	3	Latitude-Longitude data available on demand.
56	Throttle	Status	0-100%	To be determined	•				•		3	Manual throttle in neutral position during autothrottle mode.
57	Fuel	Status	Main Tanks Auxiliary Tanks	0 - 100% 0 - 100%	•	•	Fuel Level Low	•	•		1	Can demand display of fuel levels or tank being used.
58	Fuel Flow Rate	Status	Fuel Flow Rate Fuel Quantity Headwind	0 - 100% 0 - 100%	•	•	Net Optimum	•	•	•	2	
59	Fuel Against Mission Requirements	Geographic Display with Ownship and Fuel Range Indicated	Ownship Position Geographical Map Range Perimeter	To be determined 500 sq. mi.	•	•	Out of Tolerance	•	•	•	1	Scale to be determined by tactical requirements.
60	Energy Management Components	Status	RPM EGT EPR Oil Temperature Oil Pressure Oil Quantity	± 5% ± 5% ± 5% To be determined ± 5% ± 5%	•	•	Out of Tolerance	•	•	•	3	
61	Energy Management System of Tolerance	Attention Signal Fault Location	-----	-----	•			•	•	•	1	Identify Malfunction area in A/N display.

FIGURE 18-3A Matrix display: Display information analysis (pilot responsibility). SOURCE: Murphy et al. (1967).

No.	Function	Auto- mated	Items to be Controlled	Control Requirements			Control Category		Remarks
				Control Methods	Crit- icality	Fre- quency of Use	Automatic from Program	Manual Program Enable	
51	Is Attitude OK?	Yes	-----	-----	---	---	•	---	Energy Management Control system.
52	Is Altitude and Speed OK?	Yes	-----	-----	---	---	•	---	
53	Cut Gas Generators	Yes	Gas Generators	Automatic - EMCS Back-up - Switch	---	---	•	---	
54	Fan Doors and Bottom Louvers	Yes	Lift Fan Doors and Louvers	Automatic AFCS Back-up - Doors - Switch Louvers - Continuous Rate Device Switch	1 1 1	Low Low Low	• --- ---	• • •	
55	Search for Checkpoint	Yes	Demand Display of Lower Position ILLTV, IR, Laser, SLAR, Radar, Visual Search Scan Sector	Sensor Mode Select Switch	1	12/Men	---	•	Sensors used in combinations appropriate to mission segment. Switches request demand display of search point designated by the Cursor.
				Sensor Azimuth and Elevation Switches	1	12/Men	---	•	
				Scan Pattern Switch	1	12/Men	---	•	
				Cursor Positioning Control and a Switch	1	4/Men	---	•	
				Automatic - AFCS and EMCS Back-up - Continuous Rate Device Switch	1	20/Men	• --- ---	• • •	
56	Adjust Throttle	Yes	Throttle	Automatic - AFCS and EMCS Back-up - Continuous Rate Device Switch	1	Low	•	---	Switch controls demand display.
			Demand Display of Throttle Setting		1	20/Men	---	•	

FIGURE 18-3B Matrix analysis: Control information analysis (pilot responsibility). SOURCE: Murphy et al. (1967).

No.	Functions	Auto- mated	Sensor Requirements	Information Requirements	Display Requirements	Data-Processing Requirements	Control Requirements	Remarks
56	Adjust throttle	Yes	Transducers on throttle subsystem	Indicated setting	-----	Energy management computation subroutine	AFCSS	-----
57	Assess remaining fuel	Yes	Transducer on fuel stores	Indicated setting	Status	Energy management computation subroutine	-----	-----
58	Assess fuel flow rate	Yes	Flow transducer on lines	Actual vs command setting	Status	Energy management computation subroutine	-----	-----
59	Is fuel remaining sufficient for mission?	Yes	-----	Computation of distance to target, fuel flow rate, fuel quantity, mission parameters, weather parameters	Geographic display with ownship and fuel range indicated	Energy management computation subroutine	-----	-----

FIGURE 18-3C Matrix analysis: Automation of functions (pilot responsibility). SOURCE: Murphy et al. (1967).

5. The entire analysis is tedious, slow, expensive, and impractical to update.

SCENARIOS WITH SIMPLE CONTINGENCIES

Especially in tasks that involve largely routine skill, it is possible to go beyond fixed scenarios to add simple contingencies. An early example was the SAINT system in which, whenever workload became sufficiently high, the simulation would compress the time for actions or even skip steps if necessary.

A more recent example is the GOMS (Card, Moran, and Newell, 1983) analysis in which a task is analyzed in terms of goals, operators, methods, and selection rules. Operators are actions that can be performed directly. Goals are actions that can be broken down further and often have alternative ways of being accomplished. Methods are procedures composed of goals and operators and simple control structures that can be used to achieve goals. Selection rules are rules for choosing among alternative methods for accomplishing goals. For example, the major contingencies in using a computer-based text editor to edit a manuscript might be described as follows:

GOAL: EDIT-MANUSCRIPT

GOAL: EDIT-UNIT-TASK

GOAL: ACQUIRE-UNIT-TASK

GET-NEXT-PAGE

GET-NEXT-TASK

GOAL: EXECUTE-UNIT-TASK

GOAL: LOCATE-LINE

[select USE-QS-METHOD
USE-LF-METHOD]

GOAL: MODIFY-TEXT

[select USE-S-COMMAND
USE-M-COMMAND]

- repeat until no more unit tasks
- if task not remembered
- if at end of manuscript page
- if an edit task was found
- if task not on current line

In this case, goals are explicitly indicated by the tag **GOAL:** and **GET-NEXT-PAGE** and **GET-NEXT-TASK** are operators. **USE-QS-METHOD**, **USE-LF-METHOD**, **USE-S-COMMAND**, and **USE-M-COMMAND** are methods. An example of a typical set of selection rules for **GOAL: MODIFY-TEXT** is

Rule 1: Use the **S-COMMAND** method as a default.

Rule 2: However, if the correction is at the very beginning or the very end of the line, then use the **M-COMMAND** method.

Similar methods have been used to describe other tasks (Kieras and Polson, 1985; Carroll and Olson, 1987; Singley and Anderson, 1985) and even routines in other cultures (Randall, 1987).

A similar representation was used to supply simple contingencies for early versions of the NASA Aircrew/Aircraft Integration (**A³I**) helicopter simulator (Corker, Davis, Papazian, and Pew, 1986). For example, the goal structure for a scenario fragment in which a helicopter pops up high enough for the pilot to see certain objects of interest is described:

```
POP-UP-AND-SCAN
  POP-UP-FOR-SCAN
    [in-parallel-do:
      LOOK-FOR
      POP-UP]
  STABILIZE-CRAFT
  HOVER-AND-SCAN
    [in-parallel-do:
      HOVER
      SCAN]
```

In this case, goals and operators, as in the GOMS analysis, are distinguished mainly by whether they are considered primitive or whether they can be expanded. **LOOK-FOR**, **POP-UP**, and **SCAN** are primitive operators. Alternative methods for the actions with selection rules are not given, but some actions are allowed to proceed in parallel. As in the GOMS analysis, each of the goals or operators can handle a set of arguments (possibly through an inheritance hierarchy). For example, the goal-like action **POP-UP-AND-SCAN** is implemented (slightly simplified) as:

```

(deffavor POP-UP-AND-SCAN
  (scan-list nil)(max-elevation 100)
  (sequential-forms
    '((POP-UP-FOR-SCAN
      :maximum-elevation max-elevation :agent agent
      :scan-list scan-list)
      (STABILIZE-CRAFT
        :elevation (send agent :z) :agent agent))))

```

The action for a primitive operator is not further expanded in terms of other modeled actions but is implemented directly in terms of internal system primitives. The primitive action POP-UP, for example, is given by:

```

(deffavor POP-UP
  (vacp '(5 0 5 3))(max-elevation 200)(pop-up-rate 3)
  (tick-procedure '(send agent :alter-vertical-
    velocity (min pop-up-rate (- max-
    elevation (send agent :z))))))
  (termination-conditions '(<= (abs (- (send
    agent :z) max-elevation))(send agent
    :vertical-acceleration))))

```

In this case, a fixed constant is used to estimate the visual ($v = 5$), auditory ($a = 0$), cognitive ($c = 5$), or perceptual ($p = 3$) loading of the action. These fixed constants could be replaced by modeled parameters supplied by computational human performance models.

As these examples show, a number of simple scenario contingencies, such as how high to pop up or actions contingent on being able to see other objects, can be handled. Some other simple decisions based on doctrine can also be handled by building the doctrine into the model. Behavior that depends on problem solving cannot be handled in this fashion, but stochastic elements can be added to the models (e.g., Card, Moran, and Newell, 1983, Chapter 6). Learning and transfer of training analyses can also be done from such an analysis (Kieras and Bovair, 1986). The approach has only limited application to the analysis of errors.

MODELING MORE COMPLEX SCENARIOS

Several techniques exist that are potentially applicable for setting more complex scenarios of high-level interactions between the

pilot and the environment. These look promising but are not in established use: opportunistic planning and blackboard architectures, modeling of informal procedures by agent commitment, and artificial intelligence (AI) planning models.

Opportunistic Planning and Blackboard Architectures

The GOMS sort of analysis models the activities of settled skill in more or less routine environments. By contrast, the planning of novel action has been modeled by what has come to be called "opportunistic planning" and is based on a "blackboard architecture" of control (Cohen and Feigenbaum, 1982; Hayes-Roth and Hayes-Roth, 1978, 1979; Hayes-Roth, 1980). This model is applicable when the agent is trying to combine multiple sources of knowledge that put constraints on one another. The idea is that the different sources of knowledge independently add information to a global data structure known as blackboard. These data are then independently available to, and serve as a constraint on, other processes that use the blackboard. The blackboard concept derives from the Hearsay-II speech understanding system (Hayes-Roth, 1985; Lesser, Fennell, Erman, and Reddy, 1975) where it was used to coordinate information sharing and control by semiautonomous parallel processes all simultaneously processing different aspects of an input sentence. However, it has also been used to model image understanding (Prager, Nagin, Kohler, Hanson, and Riseman, 1977), protein-crystallographic analysis (Nii and Feigenbaum, 1978), inductive inference (Soloway and Riseman, 1977), and interactions between the different knowledge processes active in a single person doing routine planning (in this case, planning Saturday errands).

In this model, planning processes are triggered bottom up by something the planner notices about the world. This causes the planner to introduce new steps into a plan opportunistically, whenever it is convenient to do so. For example, in planning errands a person might notice that two errands are near each other and decide to do them together. Alternatively, the person might decide abstractly to group errands into regions and look for clusters of errands near each other. The blackboard contains the same data at different levels of abstraction to model the complex way in which people shift back and forth among abstractions (e.g., in the example above, the detail of proximity between stores triggering a shift to a global strategy of trying to group all errands by region).

Hayes-Roth and Hayes-Roth (1979) found several characteristics of human planning in their studies, which they claim the blackboard technique models:

- **Opportunistic decision sequences:** Each decision was motivated by one or two immediately preceding decisions, rather than by some high-level executive program.
- **Multiple levels of abstraction:** In thinking-out-loud protocols, people switched back and forth among levels of abstraction in reasoning about decisions.
- **Multidirectional processing:** Decisions at a higher level of abstraction could influence decisions at a lower level of abstraction and vice versa.
- **Global tactics:** People could make global decisions that they were going to think of their planning problem as, for example, a scheduling problem or a traveling salesman problem. This would influence the processing strategy for the whole task.

Modeling of Informal Procedures by Agent Commitment

Recently, there has been interest in understanding the ways in which informal plans are refined by interaction with the external world and how the external world can be used as a participant in the information processing. This interest is based on social science research (for example, see Heritage, 1984; Suchman, 1987) showing that, in many human activities, the procedures people do are only partially defined, the consequences of actions are not very predictable, and manipulations of world objects are a potent way to overcome information-processing limitations.

Fikes (1982) has suggested modeling informal procedures in terms of making and fulfilling commitments to other agents. Whether or not a goal has been achieved in this model depends only on whether the client agent agrees it has been. Responsibilities for fulfilling commitments can be subcontracted to other agents. This model attempts to overcome two major difficulties in basing models of procedures on the usual computer science notion of procedure: (1) the variability in the way tasks are accomplished (e.g., the task may be accomplished by skipping part of it or renegotiating a deadline) and (2) the informality of task descriptions. Similar ideas are now being tested for coordinating the actions of multicomputer networks.

Artificial Intelligence (AI) Planning Models

Models of planning in artificial intelligence, really models of how to choose a sequence of actions that accomplishes given goals, have a long history. The models of human planning presented above stand in some contrast to models of activity planning that have been used in AI (see Chapman, 1987; Cohen and Feigenbaum, 1982; Vere, 1983a). This reflects, in part, different strengths of humans and of current machines. Humans have severe limitations on immediate memory, but good visual perception capabilities and abstraction abilities. Current machines have no difficulty in keeping track of large numbers of partial states, are very limited perceptually, and are much better at syntactically oriented processing.

AI planners are distinguished on a number of dimensions, but the most fundamental one is whether they work in the space of individual actions or abstractions of individual actions (like ABSTRIPS, Sacerdoti, 1974), or whether they work in the space of entire plans (like NONLIN, Tate, 1977). In the latter case, each step is an entire plan. As work proceeds, the plan gets more refined, is better sequenced, and has fewer errors. A review of AI planning models is beyond the scope of this chapter, other than to note that some AI planning systems have been put to use in applications related to scenario generation: KNOBS (Engelman, 1983) for Air Force tactical missions, DEVISER (Vere, 1983b) for planning spacecraft activities, and SIPE (Wilkins, 1984) for aircraft carrier deck operations.

REFERENCES

- Card, S.K. Moran, T.P. and Newell, A.
1983 *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Carroll, J.M., and Olson, J.R.
1987 *Mental Models in Human Computer Interaction*. Committee on Human Factors, National Research Council. Washington, D.C.: National Academy Press.
- Chapman, D.
1987 Planning for conjunctive goals. *Artificial Intelligence* 32:333-377.
- Cohen, F.R., and Feigenbaum, E.A., eds.
1982 *The Handbook of Artificial Intelligence, Volume III*. Los Altos, CA: William Kaufman, Inc.
- Corker, K., Davis, L., Papazian, B., and Pew, R.
1986 *Development of an Advanced Task Analysis Methodology and Demonstration for Army-NASA Aircrew/Aircraft Integration*. Report No. 6124. Cambridge, MA: Bolt Beranek and Newman, Inc.

- Engelman, C.
1983 KNOBS: An integrated AI interactive planning architecture. Pp. 450-462 in *Proceedings of the AIAA Computers in Aerospace Conference*. Hartford, CT: AIAA.
- Fikes, R.
1982 A commitment-based framework for describing informal cooperative work. *Cognitive Science* 6:331-347.
- Hayes-Roth, B.
1980 *Human Planning Processes*. Report No. R-2670-ONR. Santa Monica, CA: Rand Corporation.
1985 A blackboard architecture for control. *Artificial Intelligence Journal* 26:251-321.
- Hayes-Roth, B., and Hayes-Roth, F.
1978 *Cognitive Processes in Planning*. Report No. R-2366-ONR. Santa Monica, CA: Rand Corporation.
1979 Modeling planning as an incremental, opportunistic process. *Proceedings of the 1979 International Joint Conference on Artificial Intelligence* 375-383.
- Heritage, J.
1984 *Garfinkel and Ethnomethodology*. Cambridge, England: Polity Press.
- Kieras, D.E., and Bovair, S.
1986 The acquisition of procedures from text: A production-system analysis of transfer of training. *Journal of Memory and Language* 25:507-524.
- Kieras, D.E., and Polson, P.G.
1985 An approach to the formal analysis of user complexity. *International Journal of Man-Machine Studies* 22:365-394.
- Lesser, V.R., Fennell, R.D., Erman, L.D., and Reddy, D.R.
1975 Organization of the Hearsay-II Speech Understanding system. *IEEE Transactions on Acoustics, Speech, and Signal Processing* ASSP-23:11-23.
- Murphy, J.V., Pizzicara, D.J., Belcher, J.J., Hamson, R.L., and Bernberg, R.E.
1967 *Integrated Cockpit Research Program-Appendix I*. Report No. NR-213-041; DTIC Number AD-658753. Woodland Hills, CA: Litton Guidance and Control Systems Division, Advanced Programs Directorate.
- Nii, H.P., and Feigenbaum, E.A.
1978 Rule-based understanding of signals. In D.A. Waterman and F. Hayes-Roth, eds., *Pattern-Directed Inference Systems*. New York: Academic Press.
- Prager, J., Nagin, P., Kohler, R., Hanson, A., and Riseman, E.
1977 Segmentation processes in the VISIONS system. *Proceedings of IJCAI-77*.
- Randall, R.A.
1987 Plans in cross-cultural settings. Pp. 39-75 in S.L. Freidman, E.K. Scholnick, and R.R. Cocking, eds., *Blueprints for Thinking*. Cambridge, MA: Cambridge University Press.
- Sacerdoti, E.D.
1974 Planning in a hierarchy of abstraction spaces. *Artificial Intelligence* 5:115-135.

- Singley, M.K., and Anderson, J.R.
1985 The transfer of text-editing skill. *International Journal of Man-Machine Studies* 22:403-423.
- Soloway, E.M., and Riseman, E.M.
1977 *Proceedings of the Workshop on Pattern-Directed Inference Systems, SIGART Newsletter*. New York: ACM.
- Suchman, L.
1987 *Plans and Situated Action*. Cambridge, MA: Cambridge University Press.
- Tate, A.
1977 Generating project networks. *Proceedings of the Fifth IJCAI*. Cambridge: IJCAI.
- Vere, S.A.
1983a Planning. In M. Vidulich, Y. Yeh, and W. Schneider, eds., *Encyclopedia of Artificial Intelligence*. Also Time compressed components for air intercept control skills. Pp. 161-1654 in *Proceedings of the Human Factor Society 27th Annual Meeting*. Santa Monica, CA: Human Factors Society.
- 1983b Planning in time: Windows and durations for activities and goals. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-5(3):246-267.
- Wilkins, D.E.
1984 Domain-independent planning: Representation and plan generation. *Artificial Intelligence* 15(3):269-301.

Modeling and Predicting Human Error

DAVID D. WOODS

INTRODUCTION

One cannot survey and rate "error models" for inclusion in a computer-aided engineering and design (CAD/CAE) framework.¹ In part, this is because a model of error is also a model of processing mechanisms and, in part, because there are few models available that address the way in which errors occur at the scale of behavior relevant to pilot performance. In other words, models of processing mechanisms are not necessarily models of how processing can break down or lead to erroneous performance. The "error models" available are either descriptive taxonomies (e.g., Rasmussen, 1986; Reason, 1987b) or cognitive simulations that assist an analyst in discovering error-prone points in a person-machine system (e.g., Corker, Davis, Papazian, and Pew, 1986; Woods, Roth and Pople, 1987).

Definition of Error

If the end brings me out all right, what is set against me won't amount to anything. If the end brings me out all wrong, ten angels swearing I was right would make no difference.

Abraham Lincoln

There have been long and unresolved debates among researchers on human performance as to what human error is. Some of these

¹For an overview of research trends on the topic of human error, the best single source is Rasmussen, Duncan, and Leplat (1987); see also, Rasmussen (1986) and Senders and Moray (in press). Reason has conducted a large and far-reaching research program on human error (cf. Reason and Mycielska, 1982; various chapters in Rasmussen et al., 1987; Reason, 1987, in press).

discussions are reflected in the pages of Rasmussen et al. (1987) and Senders and Moray (in press). To guide further discussions in the context of this report, Figure 19-1 illustrates graphically one approach to establish a temporary and pragmatic truce among the differing camps.

The concept illustrated in Figure 19-1 is to separate performance failures from information-processing deficiencies. Performance failures are defined in terms of a categorical shift in consequences on some dimension related to performance in a particular domain. For the helicopter domain, examples include failure to fulfill the tactical mission goal, failure to survive the mission, failure to prevent a helicopter system failure, and failure to mitigate the consequences of a helicopter system failure. Performance failures can be defined in terms of some potentially observable external standard. Note that the definition is in terms of the language of the domain.

Information processing deficiencies involve some type of "defect" in cognitive function that, if uncorrected, could lead to a performance failure. This is the point at which attempts to characterize the nature of human error have floundered (cf. Rasmussen et al., 1987; Senders and Moray, in press). The problem, in short, is what criterion or standard to use to judge a defect (e.g., see Garbolino, 1987, for one discussion of this issue). One complicating factor is the possibility of "error" recovery. Thus, Figure 19-1 shows an initial information processing deficiency followed by a recovery interval. If error detection occurs before there are any shifts in negative consequences, then the problem solver has recovered; if not, then a performance failure has occurred. This way to call a truce in the debates on defining error illustrates that error modeling must be concerned with the processes of error detection and correction as well as error genesis (Allwood, 1984; Perkins and Martin, 1986; Rizzo, Bagnara, and de Visciola, 1987; Woods, 1984).

This definition of error also points to one of the difficulties in human performance modeling: the customer is interested in domain consequences or outcomes; the psychologist is capable of addressing the kinds of information processing that go on in the course of solving domain problems. However, a bridge is needed between the manner in which processing may unfold and the domain consequences of that processing.

Error

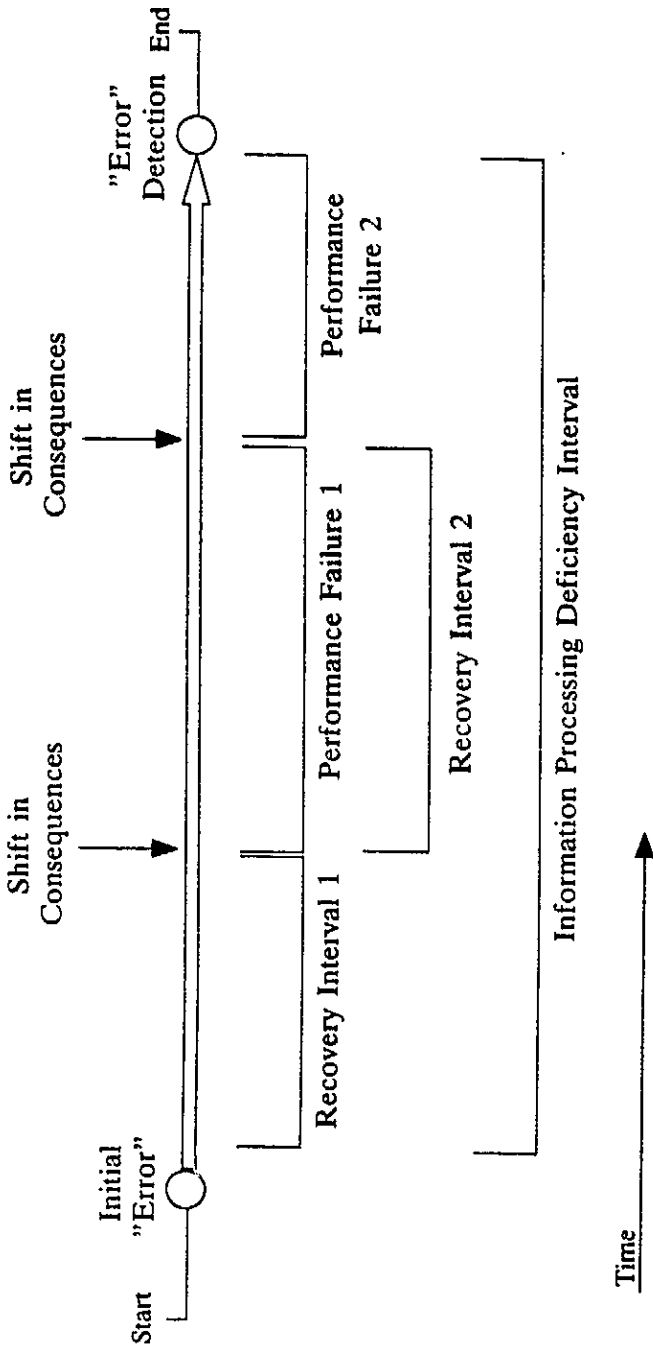


FIGURE 19-1 The nature of human error.

The Limited Rationality Approach

Most all of the research on human error today assumes that error is the result of limited rationality—people are doing reasonable things, given their knowledge, objectives, point of view, and limited resources, such as time or workload (Montmollin and De Keyser, 1986; Rasmussen et al., 1987; Woods et al., 1987). As a result, error analysis consists of tracing the problem-solving process to identify points at which limited knowledge and processing lead to breakdowns. This perspective implies that errors result from mismatches between problem demands and a person's knowledge and processing resources (e.g., Rasmussen, 1986). In this view, human error becomes person-machine system breakdown. Another implication of conceiving of error as produced by demand-resource mismatches is that one must consider what features of domain incidents and situations increase problem demands. The section on problem demand factors suggests some answers to this.

The limited rationality assumption also suggests a strategy for predicting human intention errors in complex systems via a simulation-based approach in which the investigator can vary the knowledge resources and processing characteristics of a limited-resource computer problem-solver and observe the behavior of the computer problem solver in different simulated domain scenarios. This cognitive simulation approach depends on mapping the cognitive demands imposed by the domain in question with which any intelligent but limited resource problem-solving agent would have to deal. This includes the nature of domain incidents, how they are manifest through observable data to the operational staff, and how they evolve over time. Then one can embody this model of the problem-solving environment as a limited resource, symbolic processing, problem-solving system. If the knowledge organization and processing characteristics of the symbolic processing system can be varied in psychologically meaningful ways (e.g., different mental models or diagnostic strategies that can be linked to those used by subsets of the practitioner population, as in Gitomer, 1988) and if the effects of external resources can be mapped into the program's resource settings (such as procedures, training, interface systems, aiding systems), then the errors committed by the computer problem solver are hypotheses about errors that people will commit given the same resource and demand conditions. Woods et al. (1987) have begun to develop a system based on this strategy and to apply it to identifying errors in nuclear power plant emergencies (cf. also Johnson, Moen, and

Thompson in press). Note that the cognitive simulation approach does not necessarily require strong theoretical assumptions about the detailed psychological processes underlying human behavior.

Errors in the Design of Person-Machine Problem-Solving Systems

The limited rationality approach emphasizes the role of knowledge resources in performance. These resources are established by training, experience, interface systems, and support systems. As a result, one can consider "human error" to be a symptom or manifestation of underlying flaws in the person-machine system (e.g., Hollnagel and Woods, 1983).

In this view, one objective of error modeling is to anticipate and correct designer errors in the development of interface and support systems—places where there are inadequate resources to meet the domain's demands or unanticipated negative consequences of interface/support system characteristics. There have been many cases in which the introduction of new technology to support or off-load the human has had unanticipated negative impacts in the form of changed human role, increased mental workload, and new error forms (cf. Adler, 1986; Elm and Woods, 1985; Mitchell and Foreen, 1987; Mitchell and Saisi, 1987; Wiener, in preparation). Roth, Bennett, and Woods (1987) and Suchman (1987) provide studies of specific cases of brittle machine problem solvers and communication breakdowns between person and machine, respectively. Other human performance problems created by interface/support system design that have been identified in the literature are mode errors, getting lost in large display systems, the alarm problem in alerting and monitoring systems, and tunnel vision due to keyhole effects in interface system design (see Wood and Roth, in press, for an overview).

Because problems in domains such as army helicopter scenarios are always solved with some external resources, the critical modeling question is what effects new resources or new configurations of resources have on performance. This question can be addressed via cognitive simulation, if known effects of interface/support systems on how states of the world are manifest and how they affect the human problem solver's knowledge activation can be represented within the settings of the computer-based problem solver.

ERROR MODELING

Sources of Error

Given the limited rationality approach, there are two basic sources of information-processing deficiencies which need to be modeled if one is to predict error-prone situations in domains such as army helicopter missions:

1. missing, incomplete, or erroneous (buggy) knowledge; and
2. inert knowledge (i.e., situation-relevant knowledge is not accessed under the conditions in which the task is performed).

Mental model and intelligent tutoring work has focused on the role of incomplete and erroneous knowledge of a domain in producing erroneous behavior (e.g., Brown and VanLehn, 1980). In general, modeling buggy knowledge depends on empirical studies to identify the kinds of missing and erroneous knowledge that characterize specific subsets of the practitioner population in specific domains. Gaps in knowledge may be related to Johnson's concept of chasm or missing bridge difficulties (Johnson and Thompson, 1981).

Another source of errors is inert knowledge. Does knowledge that is relevant in principle and available actually get called to mind in some problem solving context (e.g., Bransford, Sherwood, Vye, and Rieser, 1986; Gettys, Pliske, Manning and Casey, 1987; Hilton and Slugolski, 1986; Kahneman and Miller, 1986; Perkins and Martin, 1986). One tends to assume that if a person can be shown to possess a piece of knowledge in any circumstance, this knowledge should be accessible under all conditions in which it might be useful. In contrast, a variety of research has revealed dissociation effects, that is, knowledge accessed in one context remains inert in another (Bransford et al., 1986; Cheng, Holyoak, Nisbett, and Oliver, 1986; Gentner and Stevens, 1983). For example, Gick and Holyoak (1980) found that, unless explicitly prompted, people will fail to apply a recently learned problem-solving strategy to an isomorphic problem (cf. also Kotovsky, Hayes, and Simon, 1985). Thus, the fact that people possess relevant knowledge does not guarantee that this knowledge will be activated when needed. The critical factor is not whether the problem solver possesses domain knowledge but rather the more stringent criterion that situation relevant knowledge be accessible under the conditions in which the task is performed.

Monitoring/attentional strategies and cognitive processing strategies for coping with high workload, used by a limited-resource problem solver, have strong effects on the contexts in which knowledge is accessible. The concept of inert knowledge also shows how the representation of the domain can affect the quality of performance. The nature of the problem representation can help or hinder problem solvers in recognizing what information or strategies are relevant to the problem at hand. For example, Fischhoff, Slovik, and Lichtenstein (1978) and Kruglanski, Friedland, and Farkash (1984) found that judgmental biases (e.g., representativeness) were greatly reduced or eliminated when aspects of the situation cued the relevance of statistical information and reasoning. Thus, one dimension along which representations vary is their ability to provide prompts to the knowledge relevant in a given context. Inert knowledge is also important in modeling the hypothesis generation phase of diagnostic behavior under limited resources (Gettys et al., 1987). In dynamic limited resource problem-solving situations such as military helicopter missions, behavior depends on what hypotheses are called to mind and pursued first to explain the current pattern of findings (Johnson et al., *in press*; Woods et al., 1987). This means that incoming data also serves as retrieval cues, given the context of the current situation assessment and past experience.

Descriptive Error Forms

For a human performance model to address errors, it must be able to detect conditions that lead to known kinds of human error. This implies at least a partial taxonomy of descriptive error forms that people commit in domains with characteristics similar to those of military helicopter mission scenarios. What follows is a brief listing of some of the descriptive error forms that have been noted by various researchers. This list is not intended as a taxonomy of errors but only as a sample of the error forms that would make up such a taxonomy for helicopter mission scenarios. Note that the categories are based on psychological concepts and not on the language of the domain in which the error occurred or the physical form of the error. Only psychologically based taxonomies can provide the basis for more sophisticated modeling.

The errors noted are

- failures to revise situation assessment as new evidence comes in, also called fixation, mind set, or garden path (De Keyser et al., 1988; Johnson, Duran, Hassenbrock, Moller, Prietula, Feltovich, and Swanson, 1981; Woods, 1984);
- premature localization (Bechtel, 1982);
- vagabonding (Dorner, 1983);
- missing side effects in highly coupled systems (Dorner, 1983);
- availability/perfect knowledge in hypothesis generation (Gettys and Fisher, 1979; Johnson et al., 1981);
- representational errors (Evans, 1983) or failures of selective attention (Woods, 1986)—lack of attention to relevant data or paying attention to irrelevant data;
- confirmation bias in hypothesis evaluation;
- lapses or slips of action (Norman, 1981; Norman and Shallice, 1980; Reason and Mycielska, 1982);
- capture or substitution errors;
- mistake in choosing alternatives (Rasmussen, 1986);
- omitting or forgetting isolated acts (Rasmussen, 1986);
- mode errors (Monk, 1986; Norman, 1983);
- strong-but-wrong error forms based on matching bias, given variations in frequencies of encounter (Reason, 1987a); and
- over-reliance on familiar shortcuts (Rasmussen, 1986).

Each of these error forms could be examined in more detail. However, one error will be considered (and that one very briefly) which the author's own research and modeling experience suggests is particularly important for domains such as helicopter missions. The results of several studies (De Keyser et al., 1987; Johnson, et al., 1981; Johnson and Thompson, 1981; Woods, 1984) strongly suggest that a major source of human error in dynamic domains is a failure to revise situational assessment as new evidence comes in. Initial situation assessment tends to be accurate, in the sense of being consistent with the partial information available early in the event. Errors become manifest later, in the evolution of the event, as people fail to revise their assessments in response to new evidence which indicates a deviation of the event from the expected path (e.g., due to multiple failures). These results suggest that a major source of human error in dynamic domains is fixation or perseverance: a failure to revise situational assessment and planned actions when the situation changes (cf. De Keyser, Woods, Masson, and Van Daele, 1988a,b; Johnson et al., in press, for an in-depth discussion of this error form in complex dynamic worlds).

Knowledge of descriptive error forms alone can be applied to predict performance in person-machine systems in several ways. Consider, for example, flight management systems in fixed-wing aircraft. Current flight management systems are relatively dumb subordinates in that they require detailed, explicit instructions and input; they exhibit overly rigid patterns of behavior; and their performance is highly data bound.

Several predicted error forms result from this type of person-machine system. First, given human characteristics, input/instruction misentries will occur. Whether performance failures follow depends on the ability to detect the input error. This can occur at several levels of abstraction if feedback information is available at all. At the most concrete level, detection can occur through checks that the correct instructions were entered in the time interval immediately surrounding the input operation (e.g., data integrity checks). The design of the detailed human-computer interface and the presence of various kinds of data integrity checks affect the likelihood of error detection at this level.

Although instructions are concrete, they implicitly set up higher-level response strategies or goals. Therefore, error detection occurs at a more abstract level via feedback about the implications of the literal instructions. This type of feedback is not time synchronous with the input operations; rather, it is time linked to when information is available about performance envelope violations or intention violations. The likelihood of error detection at this level depends on the person's ability to maintain correct situational awareness. This, in turn, depends on a variety of processing factors (workload, fixation proneness) and interface/display factors (displays that characterize the state of the flight relative to the state of control of the flight and mission objectives). Actual cases of performance failures have resulted from failures to detect flight problems due to lack of situational awareness (Wiener, 1985a,b, 1988).

Expected future flight management systems will be more intelligent in the sense of being more flexible through the ability to fill in gaps in user instructions. Thus, the user will be able to specify instructions at higher levels of abstraction and the system will fill in the details. This shift in technology leads to predictions for the frequency and forms of errors that are likely to occur. Concrete input errors should be reduced. However, there will be a new type of error in which communication breakdowns like those documented in Suchman (1987) occur.

Error detection can be enhanced, in the case of literal communication errors (misentries), by displays that help to highlight intention/strategy mismatches. In the case of higher level communication breakdowns, error detection can be enhanced by displays that help highlight mismatches of machine intention and human intention.

Descriptive taxonomies are also the first step in more sophisticated approaches to modeling error. For example, Rouse, Geddes, and Curry (1987) used information about error forms to build a computer system that automatically assesses human performance. They began by defining a taxonomy of errors and then building a symbolic processing program, including user intent modeling, which checks for conditions that can lead to the error forms (cf. also Hollnagel, *in press*). Woods et al. (1987) and Johnson et al. (*in press*) illustrate another approach in which cognitive mechanisms hypothesized to underlie a set of descriptive error forms are set up in a cognitive simulation by adjusting or changing the processing mechanisms resident in the simulation. The simulation is then tested as a surrogate domain problem solver to see if it exhibits the error forms under the same or similar circumstances as the human practitioner.

Demand Factors

As part of error modeling, one must be able to vary demand factors as well as resource factors. Because helicopter missions are highly defined (i.e., a large amount of preplanned guidance about how to act in different situations is available in written form or in learned doctrine), problems increase in difficulty when some complicating factor goes beyond the rote implementation of preplanned routines and creates the need to adapt responses from the usual.

Complicating factors can take a variety of forms:

- underspecified or ambiguous instructions;
- special conditions or contexts (e.g., missing or failed means);
- errors in the plan;
- human execution errors; and
- impasses (where the plan's assumptions are not true).

In addition, multiple interacting factors in the scenario can produce situations that go beyond the preplanned routines, for example,

- a fault followed by additional failures;
- missing information;
- situations that remove or obscure the usual information;

- conflicting goals or responses;
- situations requiring actions that depart from the usual; and
- novel situations.

From this point of view, there are two error forms: (1) failures to recognize the need to adapt (behavior persists in one path in the face of changing circumstances that demand a shift in response) and (2) erroneous adaptation (the need for adaptation is recognized, but the attempted adaptation is inadequate due to incomplete knowledge). This view is important because it can provide a basis for why and when "violations" occur (Reason, 1988). Violations are responses other than the nominal response sequence specified in procedures or standard operating practice which hindsight suggests was most appropriate. From the viewpoint of adaptability, violations occur because of the need to adapt to circumstances that go beyond preplanned routines or because of plan breakdowns. If these circumstances are chronic, violations can then become habitual and occur in combination with circumstances that lead to disasters (e.g., the Zeebrugge and Chernobyl disasters).

Another example of violations occurs with increases in automation when, as is almost always the case, the designer has not taken into account all of the factors that are operative in the actual task world. When the designer does not provide pilots with explicit mechanisms to control or instruct the automatic systems, pilots will learn how to trick the automatic systems into doing what the pilot wants (see also Roth et al., 1987). For example, some commercial aviation pilots have learned that they can trick a flight control system into getting them down faster for landing by entering a fictitious tail wind. Circumstances often occur in which a landing must be carried out quickly. The problem is that there may be side effects of this action in terms of what the automatic systems will do under other circumstances. The result is that the trick or shortcut may work on many occasions but lead to unanticipated negative consequences when factors that turn the side effect virulent are present.

In this approach to defining problem demand, the difficulty of a problem depends both on the nature of the problem itself and on the resources (e.g., plans) available to solve it. Also note that the adaptability viewpoint emphasizes the ability of the skilled performer to compensate for environmental variability or disturbances. Error then becomes a breakdown in one's resistance to variability or disturbances. Hence the concept of "brittleness" in machine problem

solving, where performance breaks down when the problem is outside the design envelope (e.g., Roth et al., 1987).

Error Model or Processing Model

Knowledge and error flow from the same mental sources, only success can tell one from the other.

Ernst Mach (1905, p. 84).

Jens Rasmussen frequently quotes Mach on error and knowledge to make the point that a model of error is inherently a model of processing mechanisms. The question that a prospective error modeler must answer then is: What processing mechanisms and variants need to be included to be able to capture error forms? The following is a sampling of some processing mechanisms that must be included if a model is to address error in worlds such as that of military helicopter missions. Note that what follows is not a particular model, but some of the cognitive activities that must be modeled to predict error for this type of domain.

Coping with High Workload

A fundamental characteristic of the helicopter domain is the potential for problem solving under limited resources and high workload. Specific models of limited resource processing reflect two basic concepts for coping with high workload (e.g., Lane, 1982):

1. process fewer events, that is,
 - choose among competing activities,
 - defer activities,
 - monitor fewer channels,
 - eliminate gathering feedback on expected responses (substitute expectation for checking), or
 - consider fewer alternative hypotheses; and
2. process events less completely, that is,
 - monitor less often (sampling),
 - check less corroborating evidence,
 - gather partial feedback on expected responses (level of abstraction in checking),
 - narrow the field of attention,
 - limit anticipation of what might happen next or possible future trajectories in diagnostic search and planning,

- pursue possible explanations less thoroughly (shift response criterion).

In other words, information processing can be corrupted when demands are high relative to resources. When this occurs fewer events are processed or events are processed less completely. This is, of course, oversimplified since the level of skill, training and experience interacts with workload and limited-resource processing. The current approach is based on the concept of automatic versus effortful or controlled processing, where automatic processes are less vulnerable to excessive workload conditions (e.g., Fisk and Scerbo, 1987).

Other factors that affect workload/limited-resource processing are the nature of the interface to the domain and the level and philosophy of automation. The interface design can affect workload and cognitive strategies (e.g., Woods and Roth, 1988). For example, interface design can affect processing strategies by forcing a user to shift from highly automatic perceptual processes to effortful cognitive processes or vice versa (e.g., Woods, 1984b). Similarly, the design and organization of machine agents (either control or decision automation) can affect workload (e.g., Wiener, 1985b, in preparation). This means that in order to model errors in domains like helicopter missions, one must be able to specify—analytically, empirically, or theoretically—the effects of changes in the interface/automation on the cognitive processing involved in handling domain events.

Choosing among competing activities may have to occur at a strategic as well as at a tactical level. For example, early in a developing incident, one may focus on gathering evidence on the state of the world rather than pursuing one possible diagnosis. During hypothesis evaluation phase, one may focus on explanation driven search. During a plan execution phase, one may focus on plan monitoring. If the situation is changing rapidly, then one may focus on disturbance management.

Monitoring and Control of Attention

Another fundamental aspect of processing in military helicopter missions is monitoring strategy. This includes the way in which salient signals inside or outside the cockpit interrupt and capture processing resources. There are also knowledge-driven monitoring demands directed both by diagnostic activities and by the need to

monitor for expected responses from automatic systems, friendly forces, and opposing forces.

As a result of limited resources in an event-driven world, there is a need to control the focus of attention which may require capabilities such as context sensitive judgments of importance and discrimination of expected from unexpected events.

Control of attention may be a particularly critical element of a processing model for limited-resource problem solving in changing and uncertain situations. For example, March and Shapira (1987) in commenting on the state of decision theory note that "these observations suggest that choice behavior . . . is susceptible to an alternative interpretation in terms of attention. Theories that emphasize the sequential consideration of a small number of alternatives, . . ., or that highlight the significance of order of presentation and agenda effects are all reminders that understanding action in the face of incomplete information may depend more on ideas about attention than on ideas about decision" (see also Klein, in press).

The nature of the interface to the domain (e.g., problems in the design of alerting systems) and the level and philosophy of automation affect where attention is focused during unfolding scenarios.

Helicopter cockpits are highly automated and are likely to become more highly automated (at least with respect to weapon systems, navigation, and communication). As a result, the cockpit must be designed more and more for the human's supervisory control role. A critical part of this is determining how the attention of the monitor should be distributed in different contexts and states. One type of error in supervisory control is maldistribution of attentional resources, and one type of mistake in the design of the human interface to automation is introducing factors that force poor distributions of attention. A classic scenario that can be abstracted from several real cases in aviation is aircraft systems that require "heads down" to operate or instruct the systems at times in the flight where it is most important for the pilot to have "heads up" on the world. An example of this is focusing on control-display unit (CDU) data entry in order to reprogram the flight control system when a runway change has occurred during the landing approach phase of flight in commercial aviation. There are two basic variations on this scenario. In one, because attention should be focused outside the cockpit during this phase of flight, most pilots do not reprogram the flight control system and land the aircraft manually. In the other, the pilot tries to instruct or enter data into the automatic system. Error vulnerability

occurs if the pilot has trouble doing this and focuses more and more on getting the automatic system set up. As attention narrows on the interaction with the automatic system, new incoming signals that indicate a change in the situation and demand pilot attention are missed. This narrowing of attention has been cited as a factor in several industrial mishaps.

Diagnosis and Revision

Diagnosis and situation assessment in the domain of helicopter missions must address the possibility of multiple interacting failures and a situation that evolves and changes over time. This means that one must address the manner in which diagnosis and situation assessments are revised as evidence comes in over time (Klein, *in press*; Woods and Roth, 1986). Fixation on a hypothesis in the face of discrepant evidence (*i.e.*, revision failures) is a dominant descriptive error form in domains where multiple factors can account for the perceived pattern of evidence and where situations evolve over time (De Keyser, Woods, Massons, and Van Daele, 1988).

Limited resources and dynamic situations make hypothesis generation a critical part of diagnostic behavior—deciding what set of hypotheses is plausible or worth pursuing (Gettys and Fisher, 1979; Gettys, Mehle, and Fisher, 1986; Manning and Gettys, 1981). Hypothesis generation focuses on how knowledge is activated about plausible hypotheses which should be considered during hypothesis evaluation—the calling to mind of possible hypotheses. One kind of error in hypothesis generation is the failure to sample the space of potential hypotheses that could account for the currently perceived pattern of evidence. For example, Johnson (Johnson *et al.*, 1981; Johnson, *in press*) studied the performance of experienced medical diagnosticians on a problem prone to fixation. One class of errors occurred in hypothesis generation and included failures to call to mind the correct alternative. Note that modeling hypothesis generation requires gathering data on what hypotheses subsets of the population of practitioners call to mind in different contexts.

When problems unfold over time, hypothesis generation and hypothesis evaluation activities are not separate sequential stages but intermixed and interacting activities. This suggests another source of error where hypothesis generation is terminated prematurely leading to failures to revise. For example, a highly plausible hypothesis can block retrieval from other parts of the hypothesis space. Manning and

Gettys (1981) found this result in a study on the effects of providing an initial hypothesis as a retrieval cue on hypothesis generation performance (cf. also Arkes and Harkness, 1980). Johnson et al. (1981) also found revision errors in the performance of experienced diagnosticians when they were unable to shift from a highly plausible, but incorrect, initial hypothesis to the correct one.

Plan Selection, Monitoring, and Adaptation

Because the domain of military helicopter missions is highly proceduralized, formulating responses is initially a process of plan selection based on the current situation assessment and not one of plan generation. In cases of plan breakdowns or when the situation goes beyond the preplanned routines, plan adaptation is required. For example, choice under uncertainty and risk situations can arise when there are competing goals (cf., Woods and Roth, for two examples of nuclear power plant emergencies in which the problem crystallized into a classic dilemma of choice under certainty. Effective cognitive simulation must be able to capture the factors that lead pilots to "improvise" in adapting preplanned routines and doctrine to complicating factors.

Because the helicopter is an event-driven but highly doctrinal world, there is an interaction between whether and when processing is event driven (data driven) and when it is plan driven over the unfolding scenario. Errors occur when behavior is excessively plan driven, given situations that are incompatible with the preplanned routine (e.g., Woods, 1984a,b).

Multiagent Problem Solving

There are multiple human agents involved in missions (within and across helicopters), and there are (and will be more) machine agents involved in flying missions (both control and decision automation). The architecture of this multiagent problem solving system has consequences for knowledge activation and information processing, especially because different agents may have partial state information or knowledge (Fischhoff, Lanir, and Johnson, 1986). There is also the question of one agent controlling (supervisory control) or interacting with another (cooperative problem solving), which raises questions about mental models of how the other agent functions and qualitative reasoning (envisioning) on the expected behavior of the

other agents. The processing consequences can lead or contribute to performance breakdowns (e.g., descriptive error forms such as miscommunication, failures to communicate, working at cross-purposes due to different situational assessments, groupthink).

Other Processing Needs

Several other areas may be important in supporting the above-mentioned types of processing. One of these is qualitative reasoning about the way in which the behavior of the domain (engineered processes, friendly forces, opposing forces) will evolve, conditional on different actions. Initial analyses with the cognitive simulation developed by Woods et al. (1987) show that qualitative reasoning mechanisms are important for models (1) to capture diagnostic behavior when multiple interacting explanations can account for the current perceived state and (2) to capture one important characteristic of expertise in which experts are highly sensitive to domain behavior that departs from the expected, given the current situational assessment. The ability to discriminate expected from unexpected domain behavior as a function of context may be particularly important in modeling limited resource diagnosis in evolving scenarios (Rasmussen, 1986; Woods et al., 1987). Finally, qualitative reasoning may be an important element in simulating plan adaptation and repair. Qualitative reasoning for this domain needs to address both engineered systems and tactical processes (friendly forces, opponent forces).

Another aspect of processing is default reasoning as part of considering context sensitivities in human reasoning and problem solving. For example, there may be a typical relationship between two domain pieces of knowledge which only applies in certain contexts or which changes under exceptional circumstances. This addresses an error form in which a practitioner relies too much on a familiar shortcut in exceptional circumstances when it no longer applies (Rasmussen, 1986). Related to this is the need to capture different diagnostic search strategies such as symptomatic (search-based symptom-diagnostic category relations) or more explicit reasoning about intervening or abstract states (Abbot, 1988; Rasmussen, 1986).

Error as Gradual Breakdown in Processing

Finally, performance failures may not be traceable to specific knowledge bugs or corrupt processing strategies. Rather, in many

cases, performance failures may result from the cumulative effect of gradual breakdowns in the interaction of the kinds of processing mentioned above (this finesses the problem of defining criteria to judge defects in cognitive processing by focusing on and identifying cognitive processing which results in poor performance in certain classes of situations). For example, a signal may be missed (which could result from a variety of factors—attentional focus, low observability, high signal noise). By itself, the missed signal may have trivial performance consequences, in part because there are many opportunities for correction (in this example, sampling the channel later or observing other evidence for the state change). However, it can begin a chain of processing that leads to adverse performance consequences. For example, the missed signal could affect the set of hypotheses that are called to mind, leading to inadequate hypothesis evaluation and the formation of an incorrect situation assessment. In turn, this could lead to performance failures directly in terms of incorrect intentions to act or indirectly by affecting the interpretation of incoming evidence. This strategy depends on having a way to relate cognitive processing resources and external problem demands to outcomes over large numbers of situations, that is, a cognitive simulation that supports analytical experiments.

The gradual breakdown view illustrates the critical role of error detection and the factors that affect error detectability in the prediction of performance failures—the breakdown usually is corrected before negative consequences ensue. Unfortunately, only a small amount of research has investigated error detection and correction (Allwood, 1984; Perkins and Martin, 1986; Rizzo et al., 1987; Woods, 1984).

The gradual breakdown view also pinpoints the need to model human performance at the scale of behavior relevant to helicopter missions in order to capture the interactions among such fundamental cognitive processing categories as how information gathering interacts with diagnosis, hypothesis generation and evaluation interact as evidence comes in over time, and how plan adaptation and repair interact with diagnosis and monitoring activities. Unfortunately, little research in cognitive psychology has addressed these interactions.

Directions in Error Modeling

What follows is a broad menu of the potential strategies that a prospective error modeler currently can choose to begin to identify or

predict error-prone points. To use any of them for error modeling or prediction presupposes some psychologically based error taxonomy (cf. Rasmussen et al., 1987).

One possible approach is the use of computational technology to amplify a human error expert's search for error-prone points in a person-machine system. This can be done by using knowledge about the sources of, and contributing factors to, known categories of error as the basis for building one type of "error identification" system. Such a system would be directly analogous to systems that attempt to identify human errors on line as part of intelligent support systems (e.g., Hollnagel, *in press*; Rouse et al., 1987) except that it would not be monitoring the behavior of actual pilots. Building such a system is possible in principle, but a variety of major research hurdles exist. First, building systems that recognize errors made by a person during task performance has proven very difficult. Such systems tend to be based either on domain-specific criteria of what is good performance or on very simple error categories that can be defined through syntactic criteria (for example, simple execution errors in discrete tasks such as omissions, reversals, repetitions). Second, building an error identification system for the evaluation of person-machine system designs involves identifying error-prone points off-line, when there is no pilot actually performing the task. As a result, this is not a promising avenue.

Related to this is the idea of a doctrine tester, where one builds a computer simulation that attempts to respond to domain incidents based only on an implementation of the available doctrine, standard operating practices and procedures for handling different situations. In other words, the preplanned routines are actually programmed. By running the plans through a wide range of demand situations, one can identify gaps in the preplanned guidance and characterize the kinds of knowledge and processing necessary to span those gaps. The rule-based programming technology needed to make this strategy practical is available today, and this approach should be the minimum standard in design. The main investment required is the customization of the required computational technology to increase the productivity and reduce the cost of executing this approach (but see Corker et al., 1986).

Another possible approach is to extend techniques for evaluating the complexity or difficulty of a task. For example, Kieras and Polson have used production system based task simulations to develop a complexity/usability metric for some basic human-computer

interaction tasks (Kieras and Polson, 1985; Polson, 1987). A particular person-machine system is represented within the simulation and run through various situations. The outcome is a measure of cognitive complexity (e.g., the working memory load and the number of rules invoked to accomplish the task). This approach has been demonstrated successfully for evaluating simple human-computer interfaces, when error-free performance is assumed. One could try to extend it to errors if errors are assumed to be monotonically related to the complexity or difficulty of the task. Error prediction would be indirect: the more complex the human-computer interaction, the greater the error potential. However, an extension of this approach to errors is not without its difficulties. First, the assumption that models of error-free performance transfer to cases in which errors can occur is highly questionable. Second, building a complexity metric that would apply to the scale of tasks involved in helicopter missions goes well beyond what has been developed to date for simple human-computer interfaces. Finally, the difficulty approach provides no way of specifying the forms of errors to be expected in particular situations.

Another simulation-based approach to error modeling is to set up or constrain cognitive simulation by knowledge of how cognitive processing can contribute to errors (e.g., limited resources, control of attention, descriptive error forms) and by a model of the problem-solving environment that limited-resource problem solvers must confront (e.g., doctrine, incident evolution, complicating factors). The analyst then varies the knowledge resources and processing characteristics of the computer problem solver to represent the actual or hypothetical situation to be investigated, for example, different strategies for coping with high workload. The computer problem solver is used to see what specific performance failures (unsatisfactory mission outcomes) occur across a variety of domain scenarios. If the knowledge organization and processing characteristics of the computer problem solver can be varied in psychologically meaningful ways and if the effects of external resources can be mapped into the program's resource settings, then the computer problem solver's performance failures are hypotheses about errors people will commit given the same resource and demand conditions. In this way, a bridge can be built between psychological knowledge and domain consequences (cf. Woods et al., 1987).

Consider the cognitive simulation of experienced medical diagnosticians that Johnson and his colleagues (Johnson et al., 1981,

1983; Thompson, Johnson, and Moen, 1983) have built. The computer problem solver has processing mechanisms to call to mind a subset of potential hypotheses (hypothesis generation), to test for expectation violations, to evaluate plausible hypotheses, and to revise its diagnosis as more evidence is examined based on analyses of expert performance (the first version of this system was called *DIAGNOSER*; a more recent system is called *Galen*). Johnson and his colleagues compared the computer problem solver's performance and errors to those committed by experienced diagnosticians on problems designed to be fixation prone. They found that the computer problem solver and human diagnosticians committed many of the same descriptive error forms. One class of errors occurred in hypothesis generation and included failures to call to mind the correct alternative from the space of potential hypotheses that could account for the currently perceived pattern of evidence. Another shared error form was that of revision errors in which both the computer problem solver and experienced diagnosticians were unable to shift from a highly plausible, but incorrect, initial hypothesis to the actually correct hypothesis. The system also exhibited breakdowns in hypothesis evaluation that matched errors committed by experienced people. Johnson and his colleagues also changed the processing/knowledge resources of the computer problem solver and were able to control whether it was fixation prone or fixation resistant.

Woods et al. (1987) are in the process of building and testing a similar system explicitly designed to capture a wide range of descriptive error forms and the factors that affect limited resource problem solving. This system is called cognitive environment simulation and is based on Pople's work in medical problem solving (Pople, 1982, 1985). The system is designed specifically for error analysis in dynamic, highly doctrinal worlds (the initial application is human performance in nuclear power plant emergencies).

This approach to cognitive simulation is based on the limited rationality view of human error and its corollary that the source of errors is demand-resource mismatches (e.g., Rasmussen, 1986). The computer problem solver allows an analyst to examine how demands (the kind of problem, such as multiple interacting factors, goal competition, missing evidence) and resources (the dynamic flow of information from the domain, knowledge about the domain, processing strategies for coping with high workload, control of attention, and diagnostic search) interact as the problem evolves. The Johnson et

al. studies and the Woods et al. system show that this approach is viable given today's computational technology and state of knowledge of errors. Furthermore, it is the only approach that helps to predict the form of errors that may occur (and, therefore, to suggest ways of reducing error), which can hope to capture how person-machine system characteristics affect performance and can translate from the psychological world to consequences for domain behavior (and vice versa). Current experience with the cognitive simulation strategy also reveals large gaps in our knowledge of how different cognitive activities interact in complex task worlds and of the mechanisms that give rise to errors. The cognitive simulation approach, however, allows designers to take advantage of what is known at this time and provides a structure or framework model that can evolve as more is learned about human performance and error.

REFERENCES

- Abbott, K.H.
1988 Robust operative diagnosis as problem solving in a hypothesis space. In *Proceedings of the Seventh National Conference on Artificial Intelligence*, American Association for Artificial Intelligence.
- Adler, P.
1986 New technologies, new skills. *California Management Review* 29:9-28.
- Allwood, C.M.
1984 Error detection processes in statistical problem solving. *Cognitive Science* 8:413-437.
- Arkes, M.R., and Harkness, A.R.
1980 Effect of making a diagnosis on subsequent recognition of symptoms. *Journal of Experimental Psychology: Human, Learning and Memory* 6:568-575.
- Bechtel, W.
1982 Functional localization assumptions as impediments to clinical research. In L.R. Troncale, ed., *A General Survey of Systems Methodology, Volume 2*. Monterey, CA: Intersystems Publications.
- Bransford, J., Sherwood, R., Vye, N., and Rieser, J.
1986 Teaching and problem solving: Research foundations. *American Psychologist* 41:1078-1089.
- Brown, J.S., and VanLehn, K.
1980 Repair theory: A generative theory of bugs in procedural skills. *Cognitive Science* 4:379-426.
- Cheng, P., Holyoak, K., Nisbett, R., and Oliver, L.
1986 Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology* 18:293-328.
- Corker, K., Davis, L., Papazian, B., and Pew, R.
1986 *Development of an Advanced Task Analysis Methodology and Demonstration for Army Aircrew/Aircraft Integration*. Technical Report BBN 6124. Cambridge, MA: Bolt Beranek and Newman, Inc.

- De Keyser, V., Masson, M., Van Daele, A., and Woods, D.D.
1987 Fixation errors in dynamic and complex systems: Descriptive and explanatory approaches. In *Proceedings of the Fourth International Conference on Event Perception and Action*. Trieste, Italy.
- Dorner, D.
1983 Heuristics and cognition in complex systems. In R. Groner, M. Groner, and W.F. Bischof, eds., *Methods of Heuristics*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Elm, W.C., and Woods, D.D.
1985 Getting lost: A case study in interface design. In *Proceedings of the Human Factors Society 29th Annual Meeting*. Santa Monica, CA: Human Factors Society.
- Evans, J. St. B.T.
1983 Selective processing in reasoning. In J. St B.T. Evans, ed., *Thinking and Reasoning: Psychological Approaches*. Routledge & Kegan Paul.
- Fischhoff, B., Slovic, P., and Lichtenstein, S.
1978 Fault trees: Sensitivity of estimated failure probabilities to problem representation. *Journal of Experimental Psychology: Human Perception and Performance* 4:330-344.
- Fischhoff, B., Lanir, Z., and Johnson, S.
1986 *Military Risk Taking and Modern C³I*. Technical Report 86-2. Decision Research.
- Fisk, A.D., and Scerbo, M.W.
1987 Automatic and controlled processing approach to interpreting vigilance performance: A review and reevaluation. *Human Factors* 29:653-660.
- Garbolino, P.
1987 Bayesian theory and artificial intelligence: The quarrelsome marriage. *International Journal of Man-Machine Studies* 27. (Also in G. Mancini, D. Woods, and E. Hollnagel, eds. *Cognitive Engineering in Dynamic Worlds*. London: Academic Press.
- Gentner, D., and Stevens, A.L., eds.
1983 *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gettys, C.F., and Fisher, S.D.
1979 Hypothesis plausibility and hypothesis generation. *Organizational Behavior and Human Performance* 24:93-110.
- Gettys, C.F., Mehle, T., and Fisher, S.D.
1986 Plausibility assessments in hypothesis generation. *Organizational Behavior and Human Decision Processes* 37:14-33.
- Gettys, C.F., Pliske, R.M., Manning, C., and Casey, J.T.
1987 An evaluation of human act generation performance. *Organizational Behavior and Human Decision Processes* 39:23-51.
- Gick, M.L., and Holyoak, K.J.
1980 Analogical problem solving. *Cognitive Psychology* 12:306-365.
- Gitomer, D.H.
1988 Individual differences in technical troubleshooting. *Human Performance* 1:111-132.
- Hilton, D.J., and Slugoski, B.R.
1986 Knowledge-based causal attribution: The abnormal conditions focus model. *Psychological Review* 93:75-88.

- Hollnagel, E.
in press The genotype and phenotype of erroneous action: Implications for human computer interaction designs. In J. Alty and G. Weir, eds. *Human-Computer Interaction and Complex Systems*. Academic Press, London.
- Hollnagel, E., and Woods, D.D.
1983 Cognitive systems engineering: New wine in new bottles. *International Journal of Man-Machine Studies* 18:583-600.
- Johnson, P.E., Duran, F., Hassenbrock, A., Moller, J., Prietula, M., Feltovich, P., and Swanson, D.
1981 Expertise and error in diagnostic reasoning. *Cognitive Science* 5:235-283.
- Johnson, P.E., Moen, J.B., and Thompson, W.B.
in press Garden path errors in diagnostic reasoning. In M.J. Coombs and L. Bolec, eds., *Expert Systems*. New York: Springer-Verlag.
- Johnson, P.E., and Thompson, W.B.
1981 Strolling down the garden path: Detection and recovery from error in expert problem solving. In *Proceedings for the Seventh International Joint Conference on Artificial Intelligence*. Vancouver, British Columbia.
- Kahneman, D., and Miller, D.T.
1986 Norm theory: Comparing reality to its alternatives. *Psychological Review* 93:136-153.
- Kieras, D.E., and Polson, P.G.
1985 An approach to the formal analysis of user complexity. *International Journal of Man-Machine Studies* 22:365-394.
- Klein, G.A.
in press Recognition-primed decision. In W.B. Rouse, ed., *Advances in Man-Machine Research, Volume 5*. Greenwich, CT: JAI Press.
- Kotovsky, K., Hayes, J.R., and Simon, H.A.
1985 Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology* 17:248-294.
- Kruglanski, A., Friedland, N., and Farkash, E.
1984 Lay persons' sensitivity to statistical information: The case of high perceived applicability. *Journal of Personality and Social Psychology* 46:503-518.
- Lane, D.M.
1982 Limited capacity, attention allocation and productivity. In W.C. Howell and E.A. Fleishman, eds., *Human Performance and Productivity: Information processing and Decision Making*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Manning, C., and Gettys, C.F.
1981 The effect of a previously-generated hypothesis on hypothesis generating performance. *Catalog of Selected Documents in Psychology* 11(9).
- March, J.G., and Shapira, Z.
1987 Managerial perspectives on risk and risk taking. *Management Science* 33:1404-1418.

- Mitchell, C., and Forren, M.G.
1987 Multimodal user input to supervisory control system: Voice augmented keyboard. *IEEE Transactions on Systems, Man, and Cybernetics* SMC-17:594-607.
- Mitchell, C., and Saisi, D.
1987 Use of model-based qualitative icons and adaptive windows in workstations for supervisory control systems. *IEEE Transactions on Systems, Man, and Cybernetics* SMC-17:573-593.
- Monk, A.
1986 Mode errors: A user-centered analysis and some preventive measures using keying-contingent sound. *International Journal of Man-Machine Studies* 24:313-327.
- Montmollin, M., and De Keyser, V.
1985 Expert logic vs. operator logic. In G. Johannsen, G. Mancini, and L. Martensson, eds., *Analysis, Design, and Evaluation of Man-Machine Systems*. CEC-JRC Ispra, Italy: IFAC.
- Norman, D.A.
1981 Categorization of action slips. *Psychological Review* 88:1-15.
- Norman, D.A., and Shallice, T.
1980 *Attention to Action: Willed and Automatic Control of Behavior*. Technical Report 8006. San Diego: University of California.
- Perkins, D., and Martin, F.
1986 Fragile knowledge and neglected strategies in novice programmers. In E. Soloway and S. Iyengar, eds., *Empirical Studies of Programmers*. Norwood, NJ: Ablex.
- Polson, P.G.
1987 A quantitative theory of human-computer interaction. In J.M. Carroll, ed., *Interfacing Thought*. Cambridge MA: MIT Press.
- Pople, H. E.
1985 Evolution of an expert system: From internist to caduceus. In I. De Lotto and M. Stefanelli, eds., *Artificial Intelligence in Medicine*. New York: Elsevier Science Publishers B.V. (North-Holland).
1982 Heuristic methods for imposing structure on ill-structured problems: The structuring of medical diagnostics. In P. Szolovits, ed., *Artificial Intelligence in Medicine*. Boulder, CO: Westview Press.
- Rasmussen, J.
1986 *Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering*. New York: North-Holland.
- Rasmussen, J., Duncan, K., and Leplat, J.
1987 *New Technology and Human Error*. Chichester: John Wiley and Sons.
- Reason, J.
1987a Generic error-modelling system (GEMS): A cognitive framework for locating human error forms. In J. Rasmussen, K. Duncan, and J. Leplat, eds., *New Technology and Human Error*. Chichester: John Wiley and Sons.
1987b A preliminary classification of mistakes. In J. Rasmussen, K. Duncan and J. Leplat, eds., *New Technology and Human Error*. Chichester: John Wiley and Sons.
in press *Human Error*. Cambridge, England: Cambridge University Press.

- Reason, J., and Mycielska, K.
1982 *Absent Minded? The Psychology of Mental Lapses and Everyday Errors*. Englewood Cliffs, NJ: Prentice-Hall.
- Rizzo, A., Bagnara, S., and Visciola, M.
1987 Human error detection processes. *International Journal of Man-Machine Studies* 27. (Also in G. Mancini, D. Woods, and E. Hollnagel, eds. *Cognitive Engineering in Dynamic Worlds*. London: Academic Press.
- Roth, E., Bennett, K., and Woods, D.D.
1987 Human interaction with an 'intelligent' machine. *International Journal of Man-Machine Studies*, 27:479-525. (Also in G. Mancini, D. Woods, and E. Hollnagel, eds., *Cognitive Engineering in Dynamic Worlds*, London: Academic Press.
- Rouse, W.B., Geddes, N.D., and Curry, R.E.
1987 An architecture for intelligent interfaces: Outline of an approach to supporting operators of complex systems. *Human-Computer Interaction* 3(2).
- Senders, J. and Moray, N., eds.
in prep *Human Error*.
- Suchman, L.A.
1987 *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge, England: Cambridge University Press.
- Thompson, W.B., Johnson, P.E., and Moen, J.B.
1983 Recognition-based diagnostic reasoning. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*. Karlsruhe, West Germany.
- Wiener, E.
1985a Beyond the sterile cockpit. *Human Factors* 27:75-90.
1985b *Cockpit Automation: A Field Study of Flightcrew Transition*. Technical Report NASA CR-177333. National Aeronautics and Space Administration.
in prep A field study of pilot adaptation to highly automated transport aircraft.
- Woods, D.D.
1984a Some results on operator performance in emergency events. In D. Whitfield, ed., *Ergonomic Problems in Process Operations*. Institute of Chemical Engineering Symposium Ser. 90.
1984b Visual momentum: A concept to improve the cognitive coupling of person and computer. *International Journal of Man-Machine Studies* 21:229-244.
1986 Paradigms for intelligent decision support. In Hollnagel, E., Mancini, G. and Woods, D.D., eds., *Intelligent Decision Support in Process Environments*. New York: Springer-Verlag.
- Woods, D.D., and Roth, E.
1986 *Models of Cognitive Behavior in Nuclear Power Plant Personnel*. NUREG-CR-4532. Washington D.C.: U.S. Nuclear Regulatory Commission.
in press Cognitive systems engineering. In M. Helander, ed., *Handbook of Human-Computer Interaction*. New York: North-Holland.

Woods, D.D., Roth, E., and Pople, H.

1987 *Cognitive Environment Simulation: An Artificial Intelligence System for Human Performance Assessment.* Washington D.C.: U.S. Nuclear Regulatory Commission.

20

Modeling Decision Making for System Design

BARUCH FISCHHOFF

Judgment is needed to extract information from an uncertain environment. Decision making is needed to extract a course of action from those judgments in order to achieve some goals. The judgments that a person-machine system's design must support include: Where are we in the process? Is this instrument reading reliable? What does that display mean? Did I hear those instructions correctly? Do I remember my own plans? Is this an example of a Type Y contingency? Have others heard me? What will happen if I try to ride out this problem?

The decisions facing a system's operators include: What is the best plan for this task? Should I treat this as an emergency situation? If so, how? Should I trust the maintenance that has been done on the system? Should I override the on-line computer's recommendations? How should I describe my situation to others?

Systems used in military operations must support additional judgments and their associated decisions. For example, will my helicopter be seen if I pop up to survey the terrain, and is it worth that risk to establish my location? Also, what does doctrine say to do in this kind of situation, and how can I do something that makes more sense to me under these circumstances, while still being able to defend my actions?

Most organizations attempt to eliminate the guesswork from such judgments and decisions. They want to make life easier for their personnel, reducing their mental workload and allowing them to focus on leadership, innovation, and implementation. Organizations prefer to examine situations carefully at the planning stage, rather than hurriedly at the action stage. They want their operators' actions

to be predictable, for the sake of central control and for the sake of other operators who depend on them.

Unfortunately for these desires, there are limits to planning. Some contingencies cannot be anticipated at all. Particularly in military domains, those who prepare surprises must also be prepared for them. In other cases, while the general outlines of a contingency may be anticipated, there may be so many variants on that general theme that an operator cannot be expected to learn the precise response to every one of them. In still other cases planners and operators may disagree about what works. Even when planning is perfect, there is no guarantee that operators will diagnose an actual situation quickly and accurately enough to access and implement the appropriate plan. Whenever uncertainty remains, judgment is needed to interpret the situation and to convert that interpretation into an uncertain decision, or gamble.

How people make decisions under conditions of uncertainty has been an active area of research for about 35 years, with some longer historical roots (Edwards, 1954, 1961; Fischhoff, 1987; Levi and Abelson, 1983; von Winterfeldt and Edwards, 1986). That research offers a number of tools and perspectives on modeling operator performance. These include methods that might be incorporated in system design (including computer-aided design), methods that might be used to test the limits of person-machine systems, and methods that might make operator behavior more model-like and more optimal.

WHY DECISION MAKING SEEMS EASY TO MODEL—SOMETIMES

Research into behavioral aspects of decision making arose from the axiomatization of decision theory by von Neumann and Morgenstern (1947), Savage (1954), Wald (1948), and others. Psychologists (and also, for a time, economists, philosophers, and others) asked whether people actually conformed to the rules of behavior prescribed by decision theory as representing rationality. To a first approximation, people did, in many of the laboratory experiments conducted in the 1950s and 1960s. The decisions that subjects reached were close to those that followed from applying decision theory to the experimental tasks. This was encouraging evidence for mainstream U.S. economists who assume the descriptive validity of decision theory when modeling behavior in marketplace situations, where tests of people's decision-making processes are more difficult than in the lab.

Unfortunately, this "victory" proved to be a mixed blessing, for several reasons that emerged over time. One such reason is that the sort of model posited by decision theory is a powerful predictor of input-output (or stimulus-response) relations even for underlying processes that follow rather different rules (Dawes, 1979; Goldberg, 1968). As a result, predictive accuracy provides only weak assurance that people have, in fact, followed the rules of rational decision making in the experiment (and, hence, might be expected to follow those rules in less constrained situations). A second reason for concern is that many experiments are designed in a manner that makes it unlikely for sensible subjects to behave in a manner that deviates greatly from rationality, whatever rules they are actually using. As a result, behavior there says little about behavior in less constrained, real-world situations, where the opportunities for suboptimality (not to mention outright folly) are much greater.

A third reason for disappointment is that real-world situations are also a lot more complicated than laboratory situations, which involve, more or less, just the subject wrestling cognitively with a stylized set of considerations. In the real world, various other factors can encourage more or less optimal behavior (e.g., previous trial-and-error experience, social pressure, advice, advertising). This also means that it is much harder to identify the "effective stimulus," in the sense of the set of facts and values that an individual is combining in order to identify the best possible course of action.

It is hard to study simultaneously how someone construes a situation and how that is translated into action. Laboratory studies have typically resolved this methodological dilemma by creating tasks, such as choosing among gambles, in which the salient elements (i.e., the possible gains and losses, subjects' goals) are easily identified and require no interpretation. Such designs allow investigators to focus on the processes by which decisions are derived. In life, though, there are often many cues potentially commanding attention, each subject to multiple interpretations. Moreover, people may be choosing among a variety of alternative goals.

Faced with this richness, economic analysts have adopted the complementary research strategy. They assume the decision-making process, namely, that people follow the rules of rational inference, and then work backward to determine how people have interpreted the decision problem (i.e., what goals they have chosen to optimize). The difficulty with this strategy is that it affords little opportunity to test the underlying assumption of optimality. With some ingenuity,

it is possible to identify some set of goals that people have optimized, especially when there is also considerable freedom to guess at how they have interpreted the facts of the problem (Fischhoff and Cox, 1985).

Some constraints to this potentially tautological research strategy come from auxiliary assumptions that limit the set of possible interpretations of behavior. Thus, it is often assumed that decision makers (e.g., marketplace consumers) interpret the statistical information that they observe accurately and that their decisions are insensitive to any features of stimuli which have no representation in decision theory. Unfortunately, behavioral research in the last 20 years has documented many ways in which people misperceive statistical information or respond to seemingly irrelevant reformulations of problems (Fischhoff, Slovic, and Lichtenstein, 1980; Kahneman, Slovic, and Tversky, 1982; Turner and Martin, 1986). These findings have been theoretically productive, in the sense of stimulating research to account for particular patterns of suboptimal behavior. In some cases, they have been accompanied by "error theories," showing just how sensitive certain classes of decisions are to particular errors (von Winterfeldt and Edwards, 1982). They have yet to be complemented by procedures for predicting how people will construe decision problems for which varied interpretations are possible (Fischhoff, 1983).

IMPLICATIONS FOR MODELING OPERATOR PERFORMANCE

If one knows how operators have interpreted a situation, then it is often reasonable to assume optimality when predicting their behavior. Optimizing models have, therefore, an important place in the repertoire of system designers and modelers. In addition to providing possibly relevant predictions, such models force explicit consideration of the informational environment faced by operators attempting to make specific decisions: What cues are out there, where they are taken from, how easily are they accessed, how they should be interpreted, and how they should be combined. Designers committed to being user centered still need ways to focus their attempts to be sensitive. Formally modeling the decisions that operators should be making is one such way, even if those models are not held to be descriptively valid.

Adopting the operators' perspective in such a detailed way might reveal design problems that would escape less rigorous analyses. For example, it might show cases in which vital cues are ambiguous, needlessly redundant, poorly positioned, or scattered in ways that frustrate quick integration. Done properly, modeling should reveal the relative informational value of different cues (Raiffa, 1968), showing perhaps what cues bear watching in confusing situations, what cues might be relegated to subordinate displays, and what cues would benefit most from greater precision in how they are estimated or displayed.

The computational complexity of such a model might also provide a rough indication of the operators' mental workload, by assuming that the mental manipulations used by operators in achieving rationality are analogous to the formal calculations in the model. One striking feature of many models that assume optimality is the enormous complexity of the computations that people are supposed to do in their heads (not to mention their implied sophistication in knowing how to set up their work). Such modeling might help designers to estimate how much workload could be reduced by simplifying operators' decision-making tasks. For example, operators might be instructed to ignore particular cues in some situations or to combine them in less complex ways. These same models should also allow estimating the effects of simplifications on the optimality of operators' decisions. As a result, designers should have the basic inputs for identifying "best buys" among simplicity-optimality trade-offs (Johnson and Payne, 1985).

Using rational models to estimate mental workload assumes not only that operators identify the optimal course of action, but also that they do it by something similar to the process described by the models. However, even when people do the right thing, they may have followed other processes, which impose other workloads. Training may enable them to recognize a complex pattern of cues as calling for a particular response, with little deliberation at all. Conversely, they may come upon correct responses through a cumbersome rule-based process that circumvents the need for analytical thinking. In such cases, more behaviorally realistic models are needed for assessing the difficulty of decision-making processes (Bettman, Johnson, and Payne, 1986; Huber, 1980).

Whatever model is used for the process, it must focus on the concrete stimuli observed by operators and consider the real problems of observing, interpreting, and integrating them. Otherwise, modeling

becomes an exercise in operations research, rather than in human factors. Not only are the predictions likely to be inaccurate, there also will be little chance to identify design problems, which requires both admitting the possibility of problems and looking at the reality that operators actually face. Staying in the realm of the abstract will also obscure which real-world features cannot be expressed in formal terms, as well as the difficulty in finding the real-world equivalents of formal expressions. The ease with which formal models are generated and elaborated by those fluent in modeling may make flight from reality breathtakingly easy.

Those seeking fluency in modeling need familiarity with the kinds of models that are available and help in matching real situations to abstract models. Using the wrong kind of model dooms design efforts. Thus one of the most valuable things that an interactive computer-aided design (CAD) system could do is help a designer diagnose a situation as one in which the operator must perform a value-of-information analysis (for which the task is identifying the information source expected to contribute the most to decision making or determining whether there is any value to gathering additional information). The system could then lead the designer stepwise through the creation of such a model, perhaps even providing some reminders about its assumptions and limitations.

Decision making is not just a matter of interpreting informational cues. It also involves exploiting that information to achieve particular goals. Formal modeling requires explicit recognition of an operator's goals and of the trade-offs among them. That effort may reveal unclear or conflicting goals. The attempt to resolve these goals may prompt greater organizational self-awareness, or it might upset the entire modeling effort, when the organization cannot face or admit publicly to its values (e.g., the relative importance of life and property, or of the lives of different individuals).

Attempts to apply decision theory (e.g., in the form of cost-benefit analysis) to public decisions involving risks to life and safety frequently run afoul of charges that such cold calculations are immoral, taking the theory beyond its range of sensible application (Calabresi, 1970; Fischhoff, Lichtenstein, Slovic, Derby, and Keeney, 1981; Lowrance, 1976). The customary countercharge is that such trade-offs are implicit in any decision involving those stakes; as a result, it is best to face and make them deliberately. Even when the logic of this argument proves persuasive, there is no guarantee that system designers, operators, or senior organizational officials will be

able to make those hard trade-offs in a coherent fashion. Indeed, a growing literature shows people's evaluations for novel value questions, mixing barely commensurable consequences, to be unstable or labile, easily buffeted by nuances of how the question is posed (Fischhoff et al., 1980; Hogarth, 1982; Kahneman and Tversky, 1981; Turner and Martin, 1986). Such problems may strain the credibility of decision theory, even when they are within its formal range.

One final psychological limit to using formal decision theory, or any formal theory, to model behavior arises from experts' abilities to describe the environment they are attempting to model. Knowing a lot about an environment is no guarantee of being able to express that knowledge in the terms of a particular modeling language (Fischhoff, *in press*). A CAD facility ought to incorporate the best available techniques for eliciting information from technical experts (National Research Council, 1983).

MODELING WITHOUT OPTIMALITY

Economists, operations researchers, and others are fond of assuming optimality when they model behavior for several reasons. One is that it is hard for those who know how to solve a problem to empathize with those who do not; thus, optimality seems reasonable. A second reason is the observation that people clearly make sensible choices in many situations (e.g., when to cross the street, which grocery goods to purchase), although it must be admitted that such choices may reflect the exercise of specific learned habits, rather than the result of applying general decision-making principles. A third reason is that optimizing models are extraordinarily tractable analytically, allowing treatment of both individuals and collectives. For example, much macroeconomics would falter if microeconomics could not promise that individuals and firms are successful profit maximizers in their economic behavior.

Although it has been the topic of much debate, the question of whether people are optimal or suboptimal decision makers has no simple answer (Jungermann, 1985). Anecdotally, one can point to examples of either kind of behavior. Analytically, any observed behavior can be interpreted in quite different ways, showing different degrees of apparent wisdom. Presumably, performance varies with individuals and with tasks.

Those investigators willing to entertain the possibility of suboptimal behavior have adopted several research strategies, with different

implications for system design. Some have assumed that people use an optimizing decision rule but apply it to a suboptimal set of inputs. For example, people might ignore certain considerations (e.g., long-term consequences of their actions), estimate others imprecisely, or even show systematic biases (e.g., exaggerate some probabilities, underestimate the importance of some consequences). There is extensive literature documenting foibles of these types (e.g., Fiske and Taylor, 1984; Kahneman et al., 1982; Nisbett and Ross, 1980).

This approach preserves the analytical tractability of the optimizing models, but requires an empirical effort to establish what inputs decision makers are actually using. For decisions that are already being made, intensive concurrent or retrospective questioning might reveal what factors have been considered (Beach, Townes, Campbell, and Keating, 1976; Blackshaw and Fischhoff, 1988; Bowman, Frishkoff, and Frishkoff, 1987; Furby, Fischhoff, and Morgan, in press; Kunreuther, Ginsberg, Miller, Sagi, Slovic, Borkin, and Katz, 1978; Svenson, 1979). When systems are being designed, one must anticipate how operators will interpret novel situations. Those predictions might be based on performance with related systems already in operation, on tests with prototypes of the planned system, or on general behavioral principles (e.g., people tend to underestimate the time needed to execute plans). With rich situations, it may be much more difficult to determine what cues people attend to than how accurately they perceive those that they do notice.

A second approach to modeling suboptimal behavior assumes that people use an orderly decision rule, not just the one prescribed by decision theory. (The inputs to this rule may or may not be accurate and comprehensive.) That rule might be a rational one, not simply the one dictated by the situation. For example, there is a large field of study (Feather, 1982) devoted to fitting simple expected utility rules to various decisions having a small set of salient consequences that seem likely to be shared by most decision makers (e.g., decisions about careers, about health behavior). Although such rules are formally defensible, they seem overly simple for the problems to which they are applied.

Alternatively, the rule might be one with descriptive, but not normative, validity (Payne, 1982). That is, it is meant to represent how people do make decisions, but not necessarily how they should. Kahneman and Tversky's (1979) "prospect theory" is an example of this genre currently receiving considerable attention. Although it, too, computes an expectation, the decision rule of this theory uses

a different set of primitives than the comparable rule in decision theory. In addition, before the rule is applied, options are simplified through an "editing" process which transforms them in ways that have no representation in normative decision theory. As with the expected utility rule, applying prospect theory's rule in real-world situations requires a detailed and potentially difficult specification of how decision makers have interpreted their surroundings. Successful application of such models means being able to predict reliably behavior that is acknowledged to be inappropriate.

An approach to modeling suboptimality that is even less orderly views decision making as the result of applying deterministic rules, such as "do what we've always done," "do what others do," "do as we were told," "nothing ventured, nothing gained," "no price is too high for safety," "zero defects," "a bird in hand is worth two in the bush," or "ask Ed about these things." Certainly, people's explanations of their decisions are often summarized in such rules. The appeal of such rules as justifications to others presumably means that they have some appeal to decision makers as guides to action. Invoking such simple rules suggests that people have analyzed their decisions only cursorily. However, these rules could merely be handy, defensible summaries in situations where lengthier accounts are inappropriate. Alternatively, they may be invoked as a way of selecting among the options remaining after a more thoughtful analysis has eliminated clearly inferior ones.

Although these rules are simple, their application probably is not (for both those who use them and those who study their use). A great many rules might be invoked in a given situation. Each may have several interpretations. There is no obvious way to reconcile conflicts between them if more than one is evoked. Perhaps because of this messiness, there is relatively little systematic knowledge about such rules. A further complication for investigators is that the study of rules requires thinking about the substantive properties of concrete situations, rather than about the formal properties of abstract ones, the natural content of decision theory. The effect of deterministic rules on the optimality of decisions might be studied with techniques akin to those used to study the effects of simplifying heuristics in operations research. The question of when are they applied remains.

MAKING BEHAVIOR MORE MODEL-LIKE

Any model that presumes suboptimal operator behavior might make the designers of a system (and their employers) nervous, even if it could be shown that a suboptimal procedure often produces fairly good decisions at a modest expenditure of decision-making effort (Williamson, 1981). A natural response to problems is trying to fix them. One type of fix is to replace the fallible component, in this case the human decision maker. One type of replacement is automating the decision-making function. Unfortunately, not all decisions can be automated. Some cannot be anticipated. Others cannot be modeled by decision theory. Still others require a human hand (or mind) to generate the commitment needed to implement them (e.g., in military or sales campaigns). When many decisions are automated, one must still worry that the reduced role left to the operators' discretion will lead to deskilling or disengagement, reducing their ability to "get back in the loop" when distinctly human interventions are needed (Sheridan and Hennessey, 1985).

A second natural response to problems is changing the system's design so as to facilitate decision making and remove possible sources of error. As mentioned earlier, decision theory models can help identify the critical cues for decision making and the degree of precision required in each. Those analyses could, in turn, direct the positioning and design of displays, so that operators can spot the important cues most easily and get the right amount of detail on each. When decisions demand multiple cues, integrative displays could be designed, in order to reduce the mental workload required for their combination (at the possible expense of forcing operators to learn about novel composite cues). If cues tend to be misinterpreted, then care can be taken to avoid inadvertently misleading operators. For example, displays might instill too much or too little confidence in the information that they report (and, conversely, in the uncertainty surrounding those reports).

A less obvious part of the design involves the decisions themselves. Some decisions are just hard to make. These include ones with conflicting or ambiguous goals and ones with ill-defined option spaces. Looking closely (and sympathetically) at operators' tasks may reveal situations requiring clearer directives, which the organization might be able to supply and explain. A close look at a mass of options might reveal some way of reducing it to a set of more manageable ones (e.g., Fischhoff, Furby, and Morgan, *in press*), or it may be advisable to restrict the number of options that can be considered

to ensure that the remaining ones are considered thoughtfully. Simplifying tasks should improve performance and predictability, at the price of foregoing refinements.

A third response to human problems is human improvement. Training operators to identify and execute the prescribed responses in anticipated situations is part of the design for most engineered systems. The more comprehensive that training is, the fewer situations there will be that require decision making. The operators' job then becomes, first, to determine what situations have arisen and, then, to implement the appropriate solutions. Their success can be predicted, in part, by assessing the diagnostics of the cues available for identifying situations. Success might be improved somewhat by designs and training that helped them to match the concrete patterns of cues observed in the world with the abstract patterns described in plans.

Where decisions still must be made, operators need the raw intellectual skills for independent decision making. Studies of the limits to people's judgment and decision making have typically sought ways to reduce those limits. That literature provides a point of departure for training in decision making as a generalized skill (Beyth-Marom, Dekel, Gombo, and Shaked, 1985; Janis and Mann, 1977; Kahneman et al., 1982; Nisbett, Krantz, Jepson, and Kunde, 1983; von Winterfeldt and Edwards, 1986). Those skills include being able to discern the logical structure of decision-making situations, to access one's relevant knowledge (in memory), to assess the limits to that knowledge, to make unpleasant trade-offs, to control one's emotions, to evaluate past experiences fairly, and to balance such diverse considerations in one's head.

These general skills are needed if operators are to think their way through to situation-specific decisions, rather than just follow orders or apply known solutions. However, even though these skills are general, their application may be context dependent. Thus, for example, weather forecasters are outstanding at assessing the limits to their own knowledge with respect to precipitation, without necessarily showing equal facility at confidence assessment for other tasks (Murphy and Winkler, 1984). Thus, in assessing any skills, it is important to recognize their limits. That means, among other things, expecting a reduction in skills, at least initially, when operators are shifted to a new task or when their task changes under them. One common threat in highly engineered systems is the continuing introduction of changes in the system, in the attempt to remedy

imperfections. Although the operator's task may be unchanged formally, its quirks may have changed and, with them, the validity of the operator's predictions of system behavior (and of the behavior of other operators).

TESTING THE LIMITS OF DECISION MAKING

Although one hopes and designs for optimality in decision making, there are many reasons to doubt that it will be obtained. Often, the tasks are complex, the time for execution is short, and the optimal solution algorithms are unfamiliar (or even nonexistent). In such situations, it is incumbent on designers, operators, and organizations dependent on them to know the limits of decision making. Such knowledge can show designers where to design around operators. It can show operators where to mistrust their own intuitive thought processes and, instead, to seek guidance or rely on standard solutions. It can show system managers something about what problems to expect, allowing them to allocate resources and prepare for surprises.

Confronting the limits to decision making might express itself in a number of ways. One is reduced expectations regarding operators' performance, affecting both how their decisions are evaluated and what reign they are given in selecting options. It is a false freedom when operators are told to exercise discretion without being given adequate decision support. A second possible expression is greater attention to decision-making skills in training (e.g., simulator exercises focused on decision-making processes; coursework focused on intuitive decision-making processes, rather than on decision theory). A third possible expression is reduced faith in models that assume optimality, either as descriptions of operator performance or as prescriptions of preferred actions. Admitting to a limit in what can be modeled might change somewhat the balance between computation and improvisation within an organization, perhaps increasing the latitude afforded operators in deciding what to do and when to override model recommendations.

Although much progress has been made in studying and modeling decision-making behavior, that research has not yet been applied systematically to designing engineering systems in ways that are sensitive to the strengths and weaknesses of both human decision makers and optimal decision theory (Hollnagel, Mancini, and Woods, 1986; Woods and Roth, 1986). Its application would require individuals to

be knowledgeable about both the range of available optimal models and the research into suboptimal behavior (as well as about the systems being designed). It would also require additional research into topics such as how well (various) people can describe different kinds of operator behavior, how individual decision making is changed by being embedded in group or organizational settings, what price is paid (in terms of optimality) for relying on simplifying heuristics, what the mental workload associated with different rules is, how the benefits of modeling can be enjoyed without sacrificing off-model considerations, how the ability to generate decision options can be encouraged and modeled, and how general organizational goals can be made meaningful to operators faced by specific situations.

REFERENCES

- Beach, L.R., Townes, B. D., Campbell, F. L., and Keating, G. W.
1976 Developing and testing a decision aid for birth planning decisions. *Organizational Behavior and Human Performance* 15:99-116.
- Bettman, J. R., Johnson, E. J., and Payne, J. W.
1986 Cognitive Effort and Decision Making Strategies: A Componential Analysis of Choice. Unpublished manuscript.
- Beyth-Marom, R., Dekel, S., Gombo, R., and Shaked, M.
1985 *An Elementary Approach to Thinking Under Uncertainty*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Blackshaw, L., and Fischhoff, B.
1988 Decision making in online search. *Journal of the American Society for Information Science* 39(6):369-389.
- Bouwman, M. J., Frishkoff, P. A., and Frishkoff, P.
1987 How do financial analysts make decisions? A process model of the investment screening decision. *Accounting, Organizations and Society* 12(1):1-30.
- Calabresi, G.
1970 *The Costs of Accidents*. New Haven, CT: Yale University Press.
- Dawes, R. M.
1979 The robust beauty of improper linear models in decision making. *American Psychologist* 34:571-582.
- Edwards, W.
1954 The theory of decision making. *Psychological Bulletin* 51:380-417.
1973 Behavioral decision theory. *Annual Review of Psychology* 12:473-498.
- Feather, N. T., ed.
1982 *Expectations and Actions*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Fischhoff, B.
1983 Predicting frames. *Journal of Experimental Psychology: Learning, Memory and Cognition*:103-116.

- 1987 Judgment and decision making. In R.J. Sternberg and E. E. Smith, eds., *The Psychology of Human Thought*. New York: John Wiley and Sons.
- in press Eliciting information from experts. *IEEE Transactions on Systems, Man, and Cybernetics* (special issue on knowledge engineering).
- Fischhoff, B., and Cox, L. A. Jr.
1985 Conceptual framework for benefit assessment. In J. D. Bentkover, V. T. Covello and J. Mumpower, eds., *Benefits Assessment: The State of the Art*. Dordrecht, The Netherlands: D. Reidel.
- Fischhoff, B., Furby, L., and Morgan, M.
in press Rape prevention: A typology of strategies. *Journal of Interpersonal Violence*.
- Fischhoff, B., Lichtenstein, S., Slovic, P., Derby, S., and Keeney, R.
1981 *Acceptable Risk*. New York: Cambridge University Press.
- Fischhoff, B., Slovic, P., and Lichtenstein, S.
1980 Knowing what you want: Measuring labile values. Pp. 117-141 in T. Wallsten, ed., *Cognitive Processes in Choice and Decision Behavior*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Fiske, S., and Taylor, S.
1984 *Social Cognition*. Reading, MA: Addison-Wesley.
- Furby, L., Fischhoff, B., and Morgan, M.
in press Preventing rape: How people perceive the options. I: Assault prevention, victimology, cooperative work. *Cognitive Science* 6:331-347.
- Goldberg, L.
1968 Simple models or simple processes? Some research on clinical judgments. *American Psychologist* 23:483-496.
- Hogarth, R. M., ed.
1982 *New Directions for Methodology of Social and Behavioral Science: The Framing of Questions and the Consistency of Response*. San Francisco: Jossey-Bass.
- Hollnagel, E., Mancini, G., and Woods, D.D.
1986 *Intelligent Decision Support in Process Environments*. Berlin: Springer-Verlag.
- Huber, O.
1980 The influence of some task variables on cognitive operations in an information-processing decision model. *Acta Psychologica* 45:187-196.
- Janis, I., and Mann, L.
1977 *Decision Making*. New York: Free Press.
- Johnson, E., and Payne, J.
1985 Effort and accuracy in choice. *Management Science* 31:395-414.
- Jungermann, H.
1985 The two camps on rationality. In R. W. Scholz, ed., *Decision Making Under Uncertainty*. Amsterdam: Elsevier.
- Kahneman, D., Slovic, P., and Tversky, A., eds.
1982 *Judgment Under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.
- Kahneman, D., and Tversky, A.
1979 Prospect theory. *Econometrica* 47:263-292.

- 1981 The framing of decisions and the psychology of choice. *Science* 211(4481):453-458.
- Kunreuther, H., Ginsberg, R., Miller, L., Sagi, P., Slovic, P., Borkin, B., and Katz, N.
1978 *Disaster Insurance Protection: Public Policy Lessons*. New York: Wiley.
- Levi, A., and Abelson, R.
1983 Decision making. In *Handbook of Social Psychology*. Reading, MA: Addison-Wesley.
- Lowrance, W.
1976 *Of Acceptable Risk*. San Francisco: Freeman-MacKay.
- Murphy, A. W., and Winkler, R.
1984 Probability of precipitation forecasts. *Journal of the American Statistical Association*.
- National Research Council
1983 *Research Needs for Human Factors*. Committee on Human Factors. Washington, D.C.: National Academy Press.
- Nisbett, R. E., Krantz, D. H., Jepson, C., and Kunda, Z.
1983 The use of statistical heuristics in everyday inductive reasoning. *Psychological Review* 90:339-363.
- Nisbett, R.E., and Ross, L.
1980 *Human Inference: Strategies and Shortcomings*. Englewood Cliffs, NJ: Prentice-Hall.
- Payne, J. W.
1982 Contingent decision behavior. *Psychological Bulletin* 92:382-401.
- Raiffa, H.
1968 *Decision Analysis*. Reading, MA: Addison-Wesley.
- Savage, L. J.
1954 *The Foundations of Statistics*. New York: John Wiley and Sons.
- Sheridan, T., and Hennessey, R., eds.
1985 *Research and Modeling of Supervisory Control Behavior*. Committee on Human Factors. Washington, D.C.: National Research Council.
- Svenson, O.
1979 Process descriptions of decision making. *Organizational Behavior and Human Performance* 23:86-112.
- Turner, C., and Martin, E., eds.
1986 *Surveying Subjective Phenomena*. National Research Council. New York: Russell Sage.
- von Neumann, J., and Morgenstern, O.
1947 *Theory of Games and Economic Behavior* (2nd Ed.). Princeton, NJ: Princeton University Press.
- von Winterfeldt, D., and Edwards, W.
1982 Costs and payoffs in perceptual research. *Psychological Bulletin* 93:609-622.
- 1986 *Decision Analysis and Behavioral Research*. New York: Cambridge University Press.
- Wald, A.
1950 *Statistical Decision Functions*. New York: John Wiley and Sons.

Williamson, O. E.

- 1981 The Economics of Organization: The Transaction Cost Approach.
 American Journal of Sociology 87:548-573.

Woods, D.D., and Roth, E.

- 1986 *Models of Cognitive Behavior in Nuclear Power Plant Personnel.*
 NUREG/CR-4532. Washington, D.C.: U.S. Nuclear Regulatory
 Commission.

Knowledge Elicitation and Representation

DEBORAH A. BOEHM-DAVIS

Most pilots seek to maintain their aircraft within an operationally safe envelope. To accomplish this, they must successfully perform a number of tasks (Chambers and Nagel, 1985), such as

- executing flight procedures,
- planning and replanning flight mission goals,
- monitoring flight progress,
- planning and executing corrective actions,
- maintaining air-ground-air communication, and
- diagnosing system malfunctions.

These tasks all require that the pilot have some sort of internal representation of the environment. This representation must include both declarative knowledge, such as facts and characteristics associated with aircraft and flight, and procedural knowledge about how to use various systems and how to perform certain tasks (Roske-Hofstrand and Papp, 1986).

Several issues arise with regard to these representations. The first is how to determine the contents of this representation. The second is how to represent the information contained in these various cognitive structures. Finally, there is the impact of different design decisions on the form of these representations.

KNOWLEDGE ELICITATION

Knowledge elicitation is the term used to refer to any of the methods employed to gather data regarding what information people have about a particular system. This process generally elicits both procedural and declarative knowledge; furthermore, the knowledge is elicited from a person or people who are defined as expert in

the domain being studied. Thus, the type of information elicited is typically about how the system works, what the system components are, how they are related, what the internal processes of the system are, and how they affect the system components from an expert's point of view.

The techniques for eliciting this knowledge include both direct and indirect methods. The direct procedures, as their name indicates, involve asking experts to report directly their experiences in using a system. This can be done through interviews, questionnaires, or verbal protocols. For each of these techniques, the responses of the subject-matter (or domain) expert form the knowledge base of that domain.

In interviews and questionnaires, experts may be asked merely to describe their interactions with the system, or they may be asked structured questions such as cause and effect queries. Using verbal protocol techniques (see Learning Research and Development Center, 1985, for a guide to performing cognitive task analyses), the subject-matter expert "talks aloud" while either solving typical tasks or running through simulations designed to tap a variety of circumstances likely to be encountered in using the system. These protocols are then analyzed by the researcher, or knowledge engineer, and the data are translated into knowledge structures that capture the observed information-gathering and decision-making strategies.

Indirect techniques include traditional experiments, simulations, and observational studies that capture and analyze patterns of responses, such as errors or pauses. In traditional psychological experiments, the effects of different manipulations are used to infer the underlying cognitive structure. In simulation studies, simulations of the system are developed, and results of the simulation runs are then compared with what people do in using the actual system. In observational studies, the responses, errors, or pauses made by the users are collected and analyzed for consistent patterns. Many of these techniques rely heavily on statistical analyses, such as scaling, path analysis, and ordered trees, to discover the structure of the information in the domain (for example see, Reitman and Reuter, 1980; Schvaneveldt, Durso, Goldsmith, Breen, Cooke, Tucker, and DeMaio, 1985).

Recent research in this area has focused on building automated (or semiautomated) tools for acquiring this expertise (see, for example, the four-part series of special issues on knowledge acquisition for knowledge-based systems edited by Boose and Gaines, 1987).

Many of these tools, however, suffer from problems common to all the knowledge elicitation techniques discussed and to the whole approach for eliciting knowledge and representing it in a knowledge base.

Fischhoff outlined some of these concerns in a report by the National Research Council (1983). First, he points out the necessity of ensuring a common frame of reference between the researcher collecting the data and the subject-matter expert. Second, he notes the need to match the questions asked of the domain experts to their mental structures. Specifically, he stresses the assumption of most techniques that experts can answer any questions asked. Researchers, therefore, do not consider the possibility of getting misleading data. This may arise either because experts do not want to admit how they actually accomplish their tasks or because the specific question asked falls outside the particular person's expertise. Finally, he points out that the quality of the information elicited must be clarified, in terms both of how complete and accurate the expert's knowledge is and of how biased the reports are.

This raises the question of how to validate the knowledge gleaned from an elicitation procedure. Researchers have questioned the impact of reporting biases on the part of the expert (see, for example, Cleaves, 1987); the veridity of retrospections used by experts in developing answers to the questions posed, and the impact of the technique itself on the type of knowledge elicited and the organization of that information. Tied to this is the problem of knowing what an appropriate level of abstraction is in representing the knowledge collected from a subject-matter expert. These considerations make it difficult to determine whether the "correct" information has been elicited for any given system.

Another problem arises from the conceptions of the nature of novice-expert differences. All the techniques discussed so far are aimed at eliciting expert knowledge from experts. These techniques have buried in them the assumption that the differences between novices and experts are quantitative, not qualitative. In other words, the assumption is that what makes a person a novice is that he or she has not yet acquired as much information as the expert.

This assumption is not universally accepted. Rasmussen (1986) has suggested that the differences between novices and experts are qualitative, with expert models coming closer to what is true in the world. If this is the case, the emphasis on eliciting all the contents of

a user's mental model may be misplaced. Rather, one should concentrate on what the triggering conditions are for an expert to recognize (or diagnose) a particular situation. This would suggest that models as sophisticated as the ones described may not be necessary; rather, it may be preferable to get a first cut at people's understanding of the systems they use, which could be done with small, quick investigations. Some insights into a person's expertise could also be obtained by calibrating their general ability to use the information contained in a tool, rather than by trying to elicit all of their knowledge about a system.

KNOWLEDGE REPRESENTATION

Once the information that experts have about a system has been captured, a way must be found to characterize or represent that information. A number of cognitive structures have been proposed in the past few years to describe the content of people's declarative and procedural knowledge in a given domain.

A recent report (Carroll and Olson, 1987) proposes three basic representations to characterize what a user knows:

1. simple sequences,
2. methods, and
3. mental models.

Simple sequences refer to the sequence of actions that must be taken to perform a given task. These sequences are steps that allow users to get things done. They do not require that the user understand why the steps are being performed. Methods refers to the knowledge of which techniques or steps are necessary to achieve a specific goal. This characterization of knowledge, unlike simple sequences, incorporates the notion that people have general goals and subgoals, and can then apply methods purposefully to achieve them. Mental models refer to more general knowledge of the workings of a system. Specifically, mental models are defined as "a rich and elaborate structure, reflecting the user's understanding of what the system contains, how it works, and why it works that way" (Carroll and Olson, 1987, p. 12). Both sequences and methods are considered to be "task-oriented in that they contain no theory of how the system works or what the user's actions do internally to produce the results" (Carroll and Olson, 1987, p. 6).

Mental models, by incorporating the user's knowledge into the representation, provide a richer framework for study. Although mental models have been around for some time in the manual control area (see Rouse and Morris, 1986, for a discussion of this issue), the term has only recently been adopted by the cognitive psychology community.

A somewhat more general definition of mental models was proposed by Rouse and Morris (1986) in a review of research in this area. They define mental models as "the mechanisms whereby humans are able to generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future system states. This definition incorporates three purposes served by the models: description, explanation, and prediction.

Regardless of the specific definition used for a mental model, there are a number of characteristics that the concept shares across application domains. These characteristics have been summarized by Norman (1983, p. 8):

- Mental models are incomplete.
- People's abilities to "run" their models are severely limited.
- Mental models are unstable: People forget the details of the system they are using, especially when those details (or the whole system) have not been used for some period.
- Mental models do not have firm boundaries: similar devices and operations get confused with one another.
- Mental models are "unscientific": People maintain "superstitious" behavior patterns even when they know they are unneeded because they cost little in physical effort and save mental effort.
- Mental models are parsimonious: Often people do extra physical operations rather than the mental planning that would allow them to avoid those actions; they are willing to trade-off extra physical action for reduced mental complexity. This is especially true where the extra actions allow one simplified rule to apply to a variety of devices, thus minimizing the chances for confusion.

Huey (1986, pp. 6-7) has extended this list to include the following commonalities:

- They are fundamentally concerned with understanding human knowledge about the world.

- They are what people have in their minds and what guides their use of things—they reflect a user's beliefs about the system.
- They are not static entities having only a single form.
- They constitute an underlying understanding of how a system works.
- They are not directly observable—they must be inferred from overt behavior.
- They evolve naturally—through interaction with a target system, a person formulates a mental model of that system.
- They need not, and usually are not, technically accurate, but must be functional.
- They will continually be modified in order to get a workable result.
- They will be constrained by such variables as the user's technical background, previous experiences with similar systems, and the structure of the human information processing system.
- They may include contradictory, erroneous, and unnecessary concepts.
- They contain only partial descriptions of operations and large areas of uncertainties.

These commonalities raise a number of difficulties for someone trying to build a mental model of a particular system. The fact that mental models tend to be incomplete presents the first difficulty. Unless information is elicited from a number of experts, all the information needed to develop a complete mental model is unlikely to be available. Even if a number of experts are queried, the possibility remains that not all of the critical information needed to build a complete model will be elicited. Second, the fact that models tend to be dynamic and unstable suggests that it would be exceedingly difficult to generate a concrete, runnable representation of an actual mental model. If the instability observed in people's representations of systems is due to explicit, changing conditions in the external world, it may be possible to capture that information and represent it in the model. However, if the changes are a function of unobservable internal user states, it may be impossible to model the system, except perhaps as a random process. Third, because they are constrained by a user's technical background and current understanding of the system, mental models will be different for different users. This may make it difficult to construct an overall model

that is representative either of any given individual or of the range of people likely to use the system. Fourth, because mental models (even for an individual) include contradictory, erroneous, or unnecessary concepts, it will be difficult for the knowledge engineer to build an accurate model. Where elicited information is contradictory, it is not clear how one would choose which information to include in the system. Fifth, the fact that models are context sensitive suggests that even if models are built, they might be applicable only in narrowly defined situations. Context sensitivity also suggests that the models will have problems dealing with interactions among variables. That is, the experts may be unable to describe the impact of these same variables in combination with one another. Finally, the fact that mental models are not directly observable makes them difficult to validate. Thus, even the best knowledge elicitation procedure will be suspect because the knowledge elicited cannot be validated.

MENTAL MODELS AND DESIGN DECISIONS

The ultimate goal behind building a concrete, complete representation of mental models is to use them as input to other processes. The underlying assumption is that changes in mental models lead to changes in performance. Thus, if the impact of a design change on the mental representation can be captured and described, the impact of this change on performance could be predicted. As an example, consider a design decision to present altitude information by using a digital, rather than an analog, display. To the extent that this change increases or decreases the pilot's ability to access altitude information quickly and accurately while flying the aircraft, the design decision will have an impact on performance.

Full-blown, runnable systems based on this kind of analysis may not be possible immediately. Although, a number of techniques can be used to elicit knowledge from experts and build expert systems, few techniques are available for validating this knowledge (see Chignell and Peterson, *in press*, for a discussion of this issue). Eventually, validation will be needed both to assess the accuracy of the model and to determine whether information is being captured at an appropriate level of detail. Even if such models can be described, a problem remains. The visual models described in this report generally do not require mental models data to run; rather, they rely more heavily on biological data. On the other hand, the cognitive models of pilot performance discussed elsewhere in this report do not have input slots for this type of information.

REFERENCES

- Boose, J., and Gaines, B., eds.
1987 Special Issues: Knowledge acquisition for knowledge-based systems. Parts 1-4. *International Journal of Man-Machine Studies* 26(1, 2, 4) and 27(2).
- Carroll, J., and Olson, J.R., eds.
1987 *Mental Models in Human-Computer Interactions: Research Issues About What the User of Software Knows*. Committee on Human Factors. Washington, D.C.: National Academy Press.
- Chambers, A.B., and Nagel, D.C.
1985 Pilots of the future: Human or computer? *IEEE Computer* 74-87.
- Chignell, M.H., and Peterson, J.R.
in press Strategic human factors issues in knowledge engineering. *Human Factors*.
- Cleaves, D.A.
1987 Cognitive biases and corrective techniques: Proposals for improving elicitation procedures for knowledge-based systems. *International Journal of Man-Machine Studies* 27:155-156.
- Gentner, D., and Stevens, A.L., eds.
1983 *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Huey, B.
1987 *The Nature of Mental Models*. Unpublished manuscript, George Mason University.
- Learning Research and Development Center (LRDC).
1985 *Guide to Cognitive Task Analysis*. Pittsburgh, PA: University of Pittsburgh.
- National Research Council
1983 *Research needs for human factors*. Committee on Human Factors. Washington, D.C.: National Academy Press.
- Norman, D.A.
1983 Some observations on mental models. In D. Gentner and A.L. Stevens, eds., *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rasmussen, J.
1986 *Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering*. New York: North-Holland.
- Reitman, J.S., and Reuter, H.R.
1980 Organization revealed by recall orders and confirmed by pauses. *Cognitive Psychology* 12:554-581.
- Roske-Hofstrand, R.J., and Paap, K.R.
1986 Cognitive networks as a guide to menu organization: An application in the automated cockpit. *Ergonomics* 29:1301-1311.
- Rouse, W.B., and Morris, N.M.
1986 On looking into the black box: prospects and limits in the search for mental models. *Psychological Bulletin* 100:349-363.
- Schvaneveldt, R.W., Durso, F.T., Goldsmith, T.E., Breen, T.J., Cooke, N.M., Tucker, R.G., and DeMaio, J.C.
1985 Measuring the structure of expertise. *International Journal of Man-Machine Studies* 23:699-728.

22

Afterword

Do useful cognitive models exist for computer-aided design (CAD/CAE) facilities? The situation appears to be mixed. Developments in cognitive architecture are promising. Advances in cognitive architectures can be expected to lead to across-the-board improvements in the ability to use human performance models for engineering analysis and design work because they address directly one of the main limitations—the complex interactions among subsystems that occur for a human engaged in any macrolevel task. For researchers associated with one of the teams working on cognitive architectures, these models may be useful tools.

For time-sharing and workload, practical models are available. However, there is a trade-off between the degree of quantifiable prediction achieved by models of interference and interaction, and the level of environmental complexity and heterogeneity in which those models are suitable.

Much is known about human working memory, but traditional models in this area have not been developed in ways that lend themselves to inclusion in a computational workstation. Some recent developments show promise, however, if the issue of integration with models of tasks using working memory can be overcome.

There are taxonomies of errors and explanations for the existence of different classes of error. Certain approaches could be taken in predicting errors, such as combining a simulation with an error detector. These have, at least, a limited usefulness.

It is possible to extend current scenario techniques by adding some contingency to the scenario, such as Corker and colleagues (Corker, Davis, Papazian, and Pew, 1986) have done. This improvement leverages other models dependent on the scenario building.

Although much progress has been made in studying and modeling decision-making behavior, that research has not yet been applied systematically to designing engineering systems in ways that are sensitive to the strengths and weaknesses of both human decision makers and optimal decision theory.

It is probably too early to apply knowledge-based modeling to a CAD/CAE system, although work proceeding in artificial intelligence on developing a pilot's assistant could make this possible.

Three problems arise repeatedly in the engineering modeling of cognitive function:

- The central problem is to integrate models of diverse components into a coherent unity that works together. Real behavior involves a complex interaction between parts of the cognitive architecture that is difficult to address in isolated models of components.

- As with vision, there are numerous gaps among the functions that have been modeled successfully, so there is not yet a seamless repertory that can be drawn on automatically in any task.

- The role of perception as a part of cognition (e.g., as a form of external memory and a coinicator of procedural activity) has yet to be adequately attacked in technical models. The intimate interaction between cognition and perception is not well elucidated in the models reviewed.

In summary, if their strong limitations are taken into account, cognitive models can be useful for some practical CAD/CAE tasks. Furthermore, the attempt to create a CAD/CAE facility will put pressure on researchers to extend models in directions that are likely to lead to interesting theoretical (e.g., overcoming integration problems) and ultimately practical developments.

REFERENCE

- Corker, K., Davis, L., Papazian, B., and Pew, R.
1986 *Development of an Advanced Task Analysis Methodology and Demonstration for Army-NASA Aircrew/Aircraft Integration*. Technical Report No. 6124. Cambridge, MA: Bolt Beranek and Newman, Inc.

Part IV

23

Findings and Recommendations

This chapter contains the panel's findings on the adequacy of existing models to serve as the groundwork for a computation-based methodology and facility for aircraft cockpit design. It also presents the panel's recommendations for the research needed to provide a stronger base upon which such a methodology and related facilities can be built. The panel believes that a computationally based human factors design methodology is an important development that will have significant impact on many types of military, industrial and commercial human machine systems. The development of this methodology and related facilities should be encouraged. A stronger base of models more specifically directed toward the problems of design is required. The panel's recommendations are intended to define actions that NASA and other agencies might take to improve this base and to advance the development of computation-based design methodologies.

It is clear from the reviews in Parts II and III that the models available to us today do not support well the goal of providing a fairly complete simulation model of vision and related cognition. There are too many gaps in the models, linkages that are missing, validations that have not been performed, and, in the case of cognition, an overall architecture that is ill-defined. It is a disappointment, but not a surprise, to find that the cup is not completely filled, but neither is it entirely empty. It is also clear from these same reviews that there are some important design questions that can be addressed with the models that now exist. For example, questions in domains like mission analysis, workload, visual scanning, detectability, legibility and others can be addressed at least in part. The panel believes that a design facility which provides model-based tools for these addressable questions has significant potential for improving the design

process and resulting designs and would also serve to stimulate the development of more complete models and better tools.

DESIRABLE ATTRIBUTES AND TYPES OF MODELS

A model is of greater use in the human factors computer-aided engineering (HF/CAE) facility if it has certain attributes. First, it must be computational, either numerical or nonnumerical. Second, it must be explicit in its inputs and outputs. These are essential if the model is to connect to the physical reality of the environment at one end, and to compute and deliver the concrete performance of the human at the other. Third, a simulation model is preferable to a static analytical model of human performance for answering many design questions. This is because the simulation model necessarily incorporates the linkage between stimulus and response and can, therefore, illuminate the effect of cockpit design on that linkage and on performance in situations where the human's actions have an important effect on the operational environment. However, static analytical models are very effective in other situations and are preferable to an empirical description of some behavior derived from data collected in a limited set of experiments because they allow extrapolation beyond the measured conditions. Finally, simulation models themselves vary in the amount of available information about human behavior and its limitations that they exploit—ranging from normative models which represent ideal behavior, given human and situational limitations, to computer implementations developed to make a machine perform some human function. Because the panel is interested in human performance, a model that explicitly represents human performance is clearly preferable to one of equivalent functionality that represents some arbitrary machine performance.

Recommendations

- The Army-NASA Aircrew/Aircraft Integration (A³I) facility should focus on simulation models that are explicit in their inputs and outputs for use in the HF/CAE facility but should not ignore static analytic models.

- Where a normative model exists, it should be used. If none exists, a descriptive model should be used to complete a simulation.

- Where a human performance model exists, it should be used. If none exists, it is better to use a machine performance model or

even an arbitrary computer implementation to complete a simulation and allow some investigation of feasibility and sensitivity.

ADEQUACY OF MODELS FOR THE A³I DESIGN FACILITY

Many models exist, but they do not provide a complete description of human vision and associated cognitive performance. Models are missing in many important areas, and there are gaps in the models that do exist. The linkages among the models within the visual and cognitive domains are weak; the linkages between the two domains are weaker still. A satisfactory architecture is lacking for human information processing which would provide the integrative framework for these and similar models so that the needed linkages, omissions, and gaps could be illuminated and the models made to work together. Despite the lack of a completeness, there exist many models that would be useful for answering important design questions and which could provide the basis for a design facility that has the potential for significantly improving the design process.

Recommendations

- Efforts should be made to strengthen the research oriented infrastructure in government, academic, and industrial settings supporting computational human performance models for engineering design. This is critical to the long-term development of models needed for system design.

- The engineering design community in government and industry, which benefits from the development of models, should be encouraged to support the building of the academic infrastructure, perhaps through the vehicle of consortia.

- In developing models, emphasis should be placed on working both (1) from the top down by developing information processing architectures that specifies general interfaces, functionality, and data structures and (2) from the bottom up by focusing on models of specific complete subsystems (e.g., vision) that would force identification of needed model components, linkages, inputs, and outputs.

- The development of prototype HF/CAE facilities should be supported. Early versions of such facilities should focus on tools that address important design questions that are based on existing models. These prototype facilities should be tested in a design context to determine their utility and the validity of the assumptions upon which they were based.

VALIDATION

Validation against human performance of the models used in the A³I facility and of the integrated set of models is a critical and difficult problem. Many individual models and integrated sets of models have not been well validated against human performance data, thus casting doubt on the correctness of the analyses and designs based on the use of these models.

Recommendations

- A³I and all other programs developing human performance models should emphasize validation as part of their program and plan to conduct validation experiments that compare model data with human data.
- Validation must be a continuous effort. Validation techniques should be built into the A³I system, where possible, and into the processes controlling the use of the system so that a growing body of validation results is acquired.

NEED FOR ACCESS TO HUMAN FACTORS DATA BASE

The currently available models, although useful, are not sufficient to support the design process for a complex human-machine system without being supplemented by other human factors information such as experimental results, guidelines, and case histories. This situation will exist for a long time, if not indefinitely.

Recommendations

- If the A³I facility is to be a complete design facility, provision should be made to provide access to external data bases of information relevant to the design problem, such as experimental results, guidelines, and case histories.
- The research community should be alerted to the need for the types of human factors data to make effective use of models and should be encouraged to collect these data.
- Consideration should be given to techniques for applying such information effectively to design problems.

BROADER CONTEXT OF COMPUTATIONAL HUMAN FACTORS

The A³I program, if successful, will be an important contribution to the advancement of a computation-based design methodology. Such methodology can have an important impact on the design of many types of military, industrial, and commercial human-machine systems. For this impact to be significant, results of the A³I work must be made readily accessible to a larger community of other researchers and system designers, and these researchers and designers must be able to contribute to the development of future stages of the A³I system.

Recommendations

- The A³I program should lay the foundation for participation by the larger community of researchers and designers who are contributing to the development of computational human factors or who might become users of the methods and models developed by researchers in this field.

- Specific consideration should be given to making the architecture of the HF/CAE system modular, to making it be from many sources (a collection of models rather than one monolithic model), to writing the software so that A³I models can be distributed and used by other researchers and designers who are likely to have access to industry-standard professional workstations.

IMPORTANCE OF THE SYSTEMS DESIGN CONTEXT FOR RESEARCH ON MODELS

System design and analysis have special needs and require basic theoretical work on models that is aimed at supporting this kind of design and analysis. These requirements differ substantially from those that often motivate traditional academic research on models of human performance. The system design context is a powerful vehicle for exposing shortcomings in models and linkages and, as a result, imposes a valuable discipline on research and development of such models.

Recommendations

- More emphasis should be given by the research community to research on models to be used in systems design contexts. This will tend to force explicitness and completeness of the models that result.

- NASA should encourage the development of models by the academic community for use with the HF/CAE facility and should foster the integration and evaluation of these models in that context. All attempts to supplement and implement models of this type should be encouraged.

- NASA should stimulate research aimed at improving the understanding of design on a small and on a large scale, of how to apply models to design, and of how the use of models makes a difference in design.

FOCUSING THE A³I PROGRAM

The A³I program is potentially a very large effort at the forefront of a new and important design methodology. Initial efforts must be directed at understanding and proving the system design concepts underlying the A³I program and at demonstrating the benefits of computational methods of human factors designs.

Recommendations

- The A³I program should first focus on a well defined test case and attempt to determine the effectiveness of the HF/CAE concepts. A single or small set of important questions frequently encountered in aircraft or helicopter design and supportable by existing models should be identified. The goal should be to determine what is required to make the HF/CAE useful. Good candidates would be questions related to workload and visibility.

- Next, the program should assimilate a cohesive set of models appropriate for addressing these specific questions, build tools based on these models that are directed toward answering them, apply these tools to a specific problem of engineering analysis and run it to completion to develop an understanding of the difficulties of integration and use and to demonstrate the benefits of computation-based design, in order to prove the concepts underlying A³I.

- The problems chosen should be representative of important design issues and approached in a manner consistent with current

design practices. Analyses and results required by existing practices should be an essential output of the HF/CAE facility.

- These problems should be approached as publishable experiments with the goal of collecting reportable data (e.g., by keeping a journal) that can communicate the manner in which each tool and model is used and each design problem solved.

PROVIDING A FRAMEWORK AND A BOX OF TOOLS

A large collection of tools will be required in a successful HF/CAE facility. Some of these tools will provide general facilities useful in many parts of the design process; others will help the designer answer specific questions, synthesize specific elements of the design, run specific experiments, and analyze and produce specific outputs required of the design team. The collection of tools will grow and change over time. The HF/CAE facility should be designed as a framework within which a heterogeneous collection of tools can be integrated and used effectively by a design team. In addition, the HF/CAE facility must initially fit into existing design processes and help answer questions critical to and produce the outputs considered important by existing design processes.

Recommendations

- Current efforts to make the HF/CAE facility a flexible framework for a collection of tools should be continued and encouraged.
- The analytical tools incorporated into the HF/CAE facility should be developed to answer specific critical questions required by the existing cockpit design process.
- Studies of current cockpit design processes and problems should be undertaken to define the requirements for specific tools to be developed for the HF/CAE facility to perform specific steps required by the design process.