Human Factors in Transportation

Automation and Human Performance

Theory and Applications

Edited by
Raja Parasuraman
Mustapha Mouloua

Automation and Human Performance: Theory and Applications

HUMAN FACTORS IN TRANSPORTATION

A Series of Volumes Edited by
Barry H. Kantowitz

Parasuraman/Mouloua • *Automation and Human Performance: Theory and Applications*

Barfield/Dingus • *Human Factors in Intelligent Transportation Systems (In press)*

Billings • *Principles of Human-Centered Aviation Automation (In press)*

Noy • *Ergonomics and Safety of Intelligent Driver Interfaces (In press)*

Bainbridge • *Complex Cognition and the Implications for Design (In preparation)*

Garland/Wise/Hopkin • *Handbook of Human Factors in Aviation Systems (In preparation)*

Automation and Human Performance: Theory and Applications

Edited by

Raja Parasuraman
Catholic University of America

Mustapha Mouloua
University of Central Florida, Orlando

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an informa business

Reprinted 2009 by CRC Press

Library of Congress Cataloging-in-Publication Data

Automation and human performance : theory and applications / Raja
Parasuraman, Mustapha Mouloua, editors.
p. cm. -- (Series on human factors in transportation)
Includes bibliographical references and indexes.
ISBN 0-8058-1616-X (cloth : alk. paper)
1. Man-machine systems. 2. Automation--Human factors.
I. Parasuraman, Raja. II. Mouloua, Mustapha. III. Series.
TA167.A96 1996
620.8'2--dc20 96-10760
CIP

10 9 8 7 6 5 4 3 2 1

Contents

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Series Foreword

Barry H. Kantowitz
Battelle Human Factors Transportation Center
Seattle, Washington

The domain of transportation is important for both practical and theoretical reasons. All of us are users of transportation systems as operators, passengers, and consumers. From a scientific viewpoint, the transportation domain offers an opportunity to create and test sophisticated models of human behavior and cognition. This series covers both practical and theoretical aspects of human factors in transportation, with an emphasis on their interaction.

The series is intended as a forum for researchers and engineers interested in how people function within transportation systems. All modes of transportation are relevant, and all human factors and ergonomic efforts that have explicit implications for transportation systems fall within the series purview. Analytic efforts are important to link theory and data. The level of analysis can be as small as one person, or international in scope. Empirical data can be from a broad range of methodologies, including laboratory research, simulator studies, test tracks, operational tests, field work, design reviews, or surveys. This broad scope is intended to maximize the utility of the series for readers with diverse backgrounds.

I expect the series to be useful for professionals in the disciplines of human factors, ergonomics, transportation engineering, experimental psychology, cognitive science, sociology, and safety engineering. It is intended to appeal to the transportation specialist in industry, government, or academia, as well as the researcher in need of a testbed for new ideas about the interface between people and complex systems.

The present book is especially appropriate to launch this series because of the outstanding job of integrating theoretical and practical aspects of transportation human factors performed by the editors. I have always believed that the best practical tool is a good theory. The editors have honored this maxim by starting in Part I with four helpful theoretical perspectives. Part II explores human performance in the context of theory while Part III emphasizes practical applications. Part IV concludes with interesting speculations about the future that meld the theoretical and the practical. Forthcoming books in the series will continue this blend of practical and theoretical perspectives.

Foreword

Earl L. Wiener
University of Miami

> *"The question is no longer whether one or another function can be automated, but, rather, whether it should be."*
> (Wiener & Curry, 1980)

In 1979 Renwick Curry and I, at the request of NASA, embarked on a project to determine the influence of cockpit automation on flight safety. We found what we later called "promises and problems," good news and bad news. One of the questions we asked was whether cockpit automation had reached or passed a point of optimality, and was possibly doing more harm than good. Now, 15 years later, we still do not have an answer to that question.

The miracle of automated flight that the industry has witnessed over the last two decades was enabled by a device so tiny that you could hold hundreds of them in the palm of your hand. The development of the microprocessor chip ushered in a new generation of computer-based automatic devices in transport aircraft, and now in air traffic management. Automation has brought an era of fuel-efficient and potentially airspace-efficient flight. But certain incidents and accidents in recent years have led many to question the designers' claims for safety and even workload reduction in automatic flight. This is not due to shortcomings in the equipment itself, but to problems at the human interface. The equipment is highly reliable—as pilots like to say, it works "as advertised."

Still the problems persist. In early 1995, following a string of incidents and disastrous crashes of the most modern airliners, *Aviation Week and Space Technology* was moved to run a two-part series on human-computer interaction in the modern cockpit (Hughes and Dornheim, 1995). Emblazoned on the cover of the first issue was one of the questions that Curry and I had asked 15 years earlier, "Automated cockpits: who's in charge?"

Designers and operators of the modern equipment may have underestimated the changes brought by automation in the style of flying, the increase in mental workload, and the demands on the crews to monitor the equipment and the status of flight. Terms such as "situational awareness" suddenly appeared, and rapidly became part of the lexicon of human factors. The popular but elusive term "complacency" was seen as part of the automation problem. Authors began to demand "human-centered" automation. It seems ironic that after a half-century of human factors engineering, it was necessary to call for human-centered (rather than technology-centered) design.

In many ways the situation in aviation today resembles that which was first encountered in the World War II era: highly capable machines outstripping the ability of the human operators to manage them. The problem was attacked from every angle by the emerging field of human factors engineering, and was brought under control. By the time the jet airliners were introduced in the early '60s, cockpit human engineering was ready for the challenge.

With the microprocessor revolution upon us, we again find ourselves in the same position: an extremely sophisticated technology, which in many applications has not exploited its full potential, due to problems at the human interface. I once stated that when the history of cockpit design is written, the present age will be called the "era of clumsy automation."

As this volume indicates, the problem is ubiquitous. Aviation does not have exclusive rights to human factors problems, though it has generally been the case that other high-risk industries look in that direction for technological guidance. The papers assembled here paint a panorama of technology that has raced ahead of our understanding. They offer not despair, but directions for solution. Parasuraman and Mouloua are to be commended for creating a book with an understanding of the present and a vision for the future. The products of "the era of clumsy automation" will ultimately suffer the same fate as the World War II systems, largely at the hands of the authors of these chapters, their students and colleagues. History will describe the next generation of systems as "the era of effective, symbiotic, supportive automation." Operator complacency will dwell no more, and there will be a ringing reply to the question, "Who's in charge here?" The human operator, who else?

Finally, I reflect that when I was a graduate student, more years ago than I like to admit, something called "the systems approach" had emerged from post-war systems engineering, operations research, and human factors. Every author, in every paper it seemed, found it necessary, almost as if driven by some scientific or political doctrine, to state that what was needed was a systems approach (in contrast to a component-wise approach) to problems.

Today is it difficult to understand why this needed to be said at all, in fact if one were to write such a thing, it would be an immediate target for a reviewing editor's blue pencil. Why expound on the obvious, the editor would ask?

Let us hope that the time will not be distant before the same can be said of "human-centered automation."

Hughes D., and Dornheim, M. A. (Eds.) A series of articles by various authors on pilot-computer interfaces. *Aviation Week and Space Technology,* January 30 and February 6, 1995.

Wiener, E. I., and Curry, R. E. (1980). Flight-deck automation: Promises and problems. *Ergonomics, 23,* 995–1011.

Preface

There is perhaps no facet of modern society in which the influence of automation technology has not been felt. Whether at work or in the home, while traveling or while engaged in leisurely pursuits, human beings are becoming increasingly accustomed to using and interacting with sophisticated computer systems designed to assist them in their activities. Among these are diagnostic aids for physicians, automated teller machines for bank customers, flight management systems for pilots, navigational displays for drivers, and so on. As we approach the 21st century, it is clear that even more powerful computer tools will be developed, and that many of the more complex human–machine systems will not function without the added capabilities that these tools provide.

The benefits that have been reaped from this technological revolution have been many. At the same time, however, automation has not always worked as advertised; thus there is a real concern that the problems raised by automation could outweigh the benefits. Whatever the merits of any particular automation technology, however, it is clear that automation does not merely supplant human activity but also transforms the nature of human work. Understanding the characteristics of this transformation is vital for successful design of new automated systems. In general, the implementation of complex, "intelligent" automated devices in such domains as aviation, manufacturing, medicine, surface transportation, shipping, and nuclear power has brought in its wake significant challenges for human factors, cognitive science, and systems engineering.

The influence of automation technology on human performance has often been investigated in a fragmentary, isolated manner, with investigators conducting disconnected studies in different domains. Independent workshops and conferences have been held, for example, on automation in air traffic control or automation in anesthesiology. There has been little contact between these endeavors, although the principles gleaned from one domain may well have implications for the other. Also, with a few exceptions, the research has tended to be empirical and only weakly theory driven. In recent years, however, various groups of investigators have begun to examine human performance in automated systems in general, and have begun to develop theories of human interaction with automation technology. Our goal in this book is to present these theories and to assess the impact of automation on different aspects of human performance. The contributors to this volume were asked to examine automation and human performance from the dual perspective of theory and multiple domains of application. By presenting both basic and applied research, we hope to highlight the general principles of human–computer interaction in several domains in which automation technologies are widely implemented. The major premise of this approach is that a broad-based, theory-driven approach will have significant implications for the effective design of automation technology in specific work environments.

We have divided the book into four parts. Part I covers broad theoretical perspectives and concepts in automation research. The opening chapter by Woods provides a general theoretical framework for "decomposing" the complexity of human–automation interaction and poses the research challenges that must be met in the future. In chapter 2, Riley outlines an empirically supported theory of automation usage patterns by human operators. The theoretical and empirical bases of an alternative approach to implementation of automation, *adaptive* automation, are described by Scerbo in chapter 3. Flach and Bennett in chapter 4 present an original theoretical framework, based on the concept of *representation,* for the design of user interfaces to automated systems.

Part II consists of eight chapters devoted to assessing the impact of automation on different aspects of human performance. The domains of human performance covered include: monitoring (Parasuraman, Mouloua, Molloy, & Hilburn, chap. 5); mental workload (Kantowitz & Campbell, chap. 6; Kramer, Trejo, & Humphrey, chap. 7); situational awareness (Endsley, chap. 8); vigilance (Warm, Dember, & Hancock, chap. 9); decision making (Mosier & Skitka, chap. 10); and supervisory control (Coury & Semmel, chap. 11). The final chapter in this section, by Bowers, Oser, Salas, and Cannon-Bowers, discusses aspects of team performance in automated systems.

The eight chapters in Part III discuss issues related to human performance in different domains in which automation technologies have been introduced. Historically, air travel is one area in which many automation innovations were first introduced. Aviation is also the area where much human factors research on automation has been carried out. Accordingly, the first three chapters in Part III are devoted to aviation automation – both airborne, in the cockpit (Sarter, chap. 13; Rogers, Schutte, & Latorella, chap. 14), and on the ground, in air traffic management (Hopkin & Wise, chap. 15). Because the aim of this book was to examine human-automation interaction across domains of application, several systems other than those found in aviation are also discussed. The next two chapters describe automation in different modes of transportation: motor vehicles on the road (Hancock, Parasuraman, & Byrne, chap. 16), and maritime operations (Lee & Sanquist, chap. 17). The remaining three contributions discuss automation in medical systems (Guerlain & Smith, chap. 18), quality control and maintenance (Drury, chap. 19), and oil and gas pipeline operations (Meshkati, chap. 20).

Part IV consists of two chapters that look to the future. In chapter 21, Sheridan speculates on the future relationship between humans and automation. Finally, in chapter 22, Hancock also discusses this relationship in the context of his theme of understanding the "teleology," or grand purpose in design, of automation technology.

Many of the chapters in this book derive from papers presented at the First Automation Technology and Human Performance Conference, held in Washington, D.C. on April 7-9, 1994. This conference was conceived and organized by us, in collaboration with the following members of the Cognitive Science Laboratory, Catholic University of America: Charles A. Adams, Evan A. Byrne, Pamela Greenwood, David Hardy, Brian Hilburn, Robert Molloy, and Sangeeta Panicker. The assistance of the following members of the Scientific Committee is also gratefully acknowledged: Peter Hancock (University of Minnesota), Harold Hawkins (Office of Naval Research), James Howard, Jr. (Catholic University of America), Alan Pope (NASA Langley Research Center), Indramani L. Singh (Banaras Hindu University), and Joel S. Warm (University of Cincinnati). The conference received financial support from the following institutions and agencies: NASA Langley Research Center, Hampton, Virginia; Office of Naval Research, Arlington, Virginia; and the School of Arts and Sciences, The Catholic University of America. Additional support for preparation of this volume was received from NASA Langley Research Center, the National Institute on Aging, and the Office of Naval Research. Finally, we express our gratitude to the following persons for their assistance in the preparation of this book: William J. Bramble Jr., Guillermo Navarro, Carolyn Inzana, Eric Gruber, and Jacqueline A. Duley.

Raja Parasuraman
Washington, DC

Mustapha Mouloua
Orlando, FL

Contributors

Kevin B. Bennett 309 Oleman Hall, Psychology Department, Wright State University, Dayton, OH 45419.

Clint A. Bowers Team Performance Lab, Department of Psychology, University of Central Florida, Orlando, FL 32816-1390.

Evan A. Byrne Cognitive Science Laboratory, The Catholic University of America, Washington, DC 20064.

John L. Campbell Battelle HF Transportation Center, 4000 NE 41st Street, Seattle, WA 98105.

Janis Cannon-Bowers US Naval Air Warfare Center-TSD, 12350 Research Parkway, Code 4961, Orlando, FL 32826-3224.

Bruce G. Coury Johns Hopkins University, Applied Physics Laboratory, RSI/Rm. 2-246, Laurel, MD 20723-6099.

William N. Dember Department of Psychology, ML 376, University of Cincinnati, Cincinnati, OH 45221-0376.

Colin G. Drury State University of New York at Buffalo, Department of Industrial Engineering, 342 Belt Hall, Buffalo, NY 14260.

Mica R. Endsley Texas Technical University, Department of Industrial Engineering, Lubbock, TX 79409.

John M. Flach Wright State University, 309 Oelman Hall, Psychology Dept., Dayton, OH 45435.

Stephanie A. E. Guerlain 210 Baker Systems, Ohio State University, 1971 Neil Avenue, Columbus, OH 43210.

P. A. Hancock HF Research Lab, University of Minnesota, 141 Mariucci Arena-Ops., 1901 Fourth Street SE, Minneapolis, MN 55455.

Brian G. Hilburn National Aerospace Lab, Anthony Fokkerweg, Amsterdam, Netherlands.

V. David Hopkin Human Factors Consultant, Peatmoor, 78 Crookham Road, Fleet, Hants, GU13 0SA, Great Britain.

Darryl G. Humphrey Department of Psychology, Wichita State University, 1845 Fairmont Street, Wichita, KS 67208.

Barry H. Kantowitz Battelle HF Transportation Center, P.O. Box C-5395, Seattle, WA 98105.

Arthur F. Kramer University of Illinois, 603 E. Daniel Street, Psychology Department, Champaign, IL 61820.

Kara A. Latorella NASA Langley Research Center, Crew/Vehicle Integration Branch, M/S 152, Hampton, VA 23681-0001.

John D. Lee Battelle HF Transportation Center, P.O. Box 5395, 4000 NE 41st Street, Seattle, WA 98115.

Najmedin Meshkati Institute of Safety and Systems Management, University of Southern California, Los Angeles, CA 90089-0021.

Robert Molloy Cognitive Science Laboratory, Catholic University of America, Washington, DC 20064.

Kathleen Mosier NASA Ames Research Center, MS 262-4, Moffett Field, CA 94035.

Mustapha Mouloua Center for Applied Human Factors in Aviation, University of Central Florida, Orlando, FL 32816-1780.

Jodi Heintz Obradovich Cognitive Systems Engineering Laboratory, Department of Industrial, Welding and Systems Engineering, The Ohio State University, 210 Baker Systems, 1971 Neil Avenue, Columbus, OH 43210.

Randall J. Oser US Naval Air Warfare Center-TSD, 12350 Research Parkway, Code 4961, Orlando, FL 32826-3224.

Raja Parasuraman Cognitive Science Laboratory, Catholic University of America, Washington, DC 20064.

Victor Riley Honeywell, Incorporated, HTC, 3660 Technology Drive, MN65-2600, Minneapolis, MN 55418.

William H. Rogers NASA Langley Research Center, Crew/Vehicle Integration Branch, M/S 152, Hampton, VA 23681-0001.

Sally Rudmann School of Allied Medical Professionals, The Ohio State University, 535 School of Allied Medicine, 1583 Perry Street, Columbus, OH 43210.

Eduardo Salas US Naval Air Warfare Center-TSD, 12350 Research Parkway, Code 4961, Orlando, FL 32826-3224.

Thomas F. Sanquist Battelle Human Affairs Center, 4000 NE 41st Street, Seattle, WA 98105.

Nadine B. Sarter Ohio State University, 210 Baker Systems Building, 1971 Neil Avenue, Columbus, OH 43210-1210.

Ralph D. Semmel The Johns Hopkins University, Applied Physics Laboratory, RSI/Room 2-238, Laurel, MD 20723-6099.

Mark W. Scerbo Old Dominion University, Psychology Department, Norfolk, VA 23529.

Paul C. Schutte NASA Langley Research Center, Crew/Vehicle Integration Branch, M/S 152, Hampton, VA 23681-0001.

Thomas B. Sheridan Massachusetts Institute of Technology, 77 Massachusetts Avenue #3-346, Cambridge, MA 02139.

Linda J. Skitka Department of Psychology, M/C 285, University of Illinois at Chicago, Chicago, IL 60680.

Jack W. Smith Laboratory for Knowledge Based Medical Systems, Department of Computer Science, The Ohio State University, 395 Dreese Lab, 2015 Neil Avenue, Columbus, OH 43210.

Philip J. Smith Cognitive Systems Engineering Laboratory, Department of Industrial, Welding and Systems Engineering, The Ohio State University, 210 Baker Systems, 1971 Neil Avenue, Columbus, OH 43210.

Patricia Strohm Western Reserve Care System, Northside Medical Center Blood Bank, 500 Gypsy Lane, Youngstown, OH 44501.

Leonard J. Trejo Department of Psychology, University of Illinois, 603 East Daniel St., Champaign, IL 61820.

Joel S. Warm University of Cincinnati, Department of Psychology, Cincinnati, OH 45221.

John A. Wise Center for Aviation/Aerospace Research, Embry-Riddle Aeronautical University, 600 S. Clyde Morris Boulevard, Daytona Beach, FL 32114-5416.

David D. Woods Ohio State University, 1971 Neil Avenue, I & SE Department, Columbus, OH 43210.

I Theories and Major Concepts

1 Decomposing Automation: Apparent Simplicity, Real Complexity

David D. Woods
The Ohio State University

INTRODUCTION

We usually focus on the perceived benefits of new automated or computerized devices. Our fascination with the possibilities afforded by technology in general often obscures the fact that new computerized and automated devices also create new burdens and complexities for the individuals and teams of practitioners responsible for operating, troubleshooting, and managing high-consequence systems. First, the demands may involve new or changed tasks such as device setup and initialization, configuration control, or operating sequences. Second, cognitive demands change as well, creating new interface management tasks, new attentional demands, the need to track automated device state and performance, new communication or coordination tasks, and new knowledge requirements. Third, the role of people in the system changes as new technology is introduced. Practitioners may function more as supervisory controllers, monitoring and instructing lower-order automated systems. New forms of cooperation and coordination emerge when automated systems are capable of independent action. Fourth, new technology links together different parts that were formerly less connected. As more data flows into some parts of a system, the result is often data overload. Coupling a more extensive system more tightly together can produce new patterns of system failure. As technology change occurs we must not forget that the price of new benefits is often a significant increase in operational complexity. Fifth, the reverberations of technology change, especially the new burdens and complexities, are often underappreciated by the advocates of technology change. But their consequences determine when, where, and how technology change will succeed.

My colleagues and I have been studying the impact of technology change on practitioners—those people who do cognitive work to monitor, diagnoses and manage complex systems—pilots, anesthesiologists, process plant operators, space flight controllers (e.g., Woods, Johanessen, Cook, & Sarter, 1994). In these investigations we have seen that technology change produces a complex set of effects. In other words, automation is a wrapped package—a package that consists of changes on many different dimensions bundled together as a hardware/software system. When new automated systems are introduced into a field of practice, change is precipitated along multiple dimensions. In this chapter I examine the reverberations that technology change produces along several different dimensions:

- Automation seen as more autonomous machine *agents*. Introducing automated and intelligent agents into a larger system in effect changes the team composition. It changes how human supervisors *coordinate* their activities with those of the machine agents. Miscommunications and poor feedback about the activities of automated subsystems have been part of accident scenarios in highly automated domains.
- Automation seen as an increase in *flexibility*. As system developers, we can provide users with high degrees of flexibility through multiple options and modes. We also have the ability to place multiple virtual devices on one physical platform so that a single device will be used in many contexts that can differ substantially. But do these flexibilities create new burdens on practitioners, burdens that can lead to predictable forms of error?
- Automation seen as more *computerization*. Technology change often means that people shift to multifunction computer-based interfaces as the means for acquiring information and utilizing new resources. Poor design of the computer interface can force users to devote cognitive resources to the interface itself, asking questions such as: Where is the data I want? What does the interface allow me to do? How do I navigate to that display? What set of instructions will get the computer to understand my intention? Successful computer interfaces (e.g., visualization, direct manipulation) help users focus on their task without cognitive resources (attention, knowledge, workload) being devoted to the interface per se.
- Automation seen as an increase in *coupling* across diverse parts and agents of a system. Tighter coupling between parts propagates effects throughout the system more rapidly. This can produce efficiency benefits by reducing transfer costs, but it also means that problems have greater and more complex effects, effects that can propagate

quickly. But when automated partners are strong, silent, clumsy, and difficult to direct, then handling these demands becomes more difficult. The result is coordination failures and new forms of system failure.

- Much technology change is justified, at least in part, based on *claims* about the impact of technology on human performance—the new system will "reduce workload," "help practitioners focus on the important part of the job," "decrease errors," and so on. But these claims often go unexamined. A number of studies have examined the impact of automation on the cognition and behavior of human practitioners. These studies, many of which are discussed in other chapters of this book, have shown repeatedly that systems introduced to aid practitioners in fact created new complexities and new types of error traps.
- The success of new technology depends on how it affects the people in the field of practice. The dimensions addressed earlier represent some of the ways that technology change can have surprising impacts on human and system performance. By closely examining the reverberations of technology change we can better steer the possibilities of new technology into fruitful directions.

HOW TO MAKE AUTOMATED SYSTEMS TEAM PLAYERS

Heuristic and algorithmic technologies expand the range of subtasks and cognitive activities that can be automated. Automated resources can, in principle, offload practitioner tasks. Computerized systems can be developed that assess or diagnose the situation at hand, alerting practitioners to various concerns and advising practitioners on possible responses.

Our image of these new machine capabilities is that of a machine alone, rapt in thought and action. But the reality is that automated subtasks exist in a larger context of interconnected tasks and multiple actors. Introducing automated and intelligent agents into a larger system changes the composition of the distributed system of monitors and managers and shifts the human's role within that cooperative ensemble (Hutchins, 1994). In effect, these "intelligent" machines create joint cognitive systems that distribute cognitive work across multiple agents (Hutchins, 1990; Roth, Bennett, & Woods, 1987). It seems paradoxical, but studies of the impact of automation reveal that design of automated systems is really the design of a new human–machine cooperative system. The design of automated systems is really the design of a team and requires provisions for the coordination between machine agents and human practitioners (e.g., Layton, Smith, & McCoy, 1994).

However, research on human interaction with automation in many domains, including aviation and anesthesiology, has shown that automated systems often fail to function as team players (Billings, 1991; Malin et al., 1991; Sarter & Woods, 1994b). To summarize the data, automated systems that are strong, silent, clumsy, and difficult to direct are not team players. Automated systems are:

Strong when they can act autonomously.

Silent when they provide poor feedback about their activities and intentions.

Clumsy when they interrupt their human partners during high workload or high criticality periods, or when they add new mental burdens during these high-tempo periods.

Difficult to direct when it is costly for the human supervisor to instruct the automation about how to change as circumstances change.

Systems with these characteristics create new problems for their human partners and new forms of system failure.

"Strong" automation refers to two properties of machine agents. In simpler devices, each system activity was dependent on immediate operator input. As the power of automated systems increases, machine agents, once they are instructed and activated, are capable of carrying out long sequences of tasks without further user interventions. In other words, automated systems can differ in *degree of autonomy* (Woods, 1993). Automated systems also can differ in *degree of authority*. This means that the automated system is capable of taking over control of the monitored process from another agent if it decides that intervention is warranted based on its perception of the situation and its internal criteria (Sarter & Woods, 1994a).

Increasing autonomy and authority create new monitoring and coordination demands for humans in the system (Norman, 1990; Sarter & Woods, 1995; Wiener, 1989). Human supervisors have to keep track of the status and activities of their automated partners. For example, consider the diagnostic situation in a multi-agent environment, when one notices an anomaly in a process being monitored (Woods, 1994). Is the anomaly an indication of an underlying fault, or does the anomaly indicate some activity by another agent in the system, unexpected by this monitor? In fact, in a number of different settings, we observe that human practitioners respond to anomalies by first checking for what other agents have been or are doing to the process jointly managed (Johannesen, Cook, & Woods, 1994).

When machine agents have high autonomy, they will act in the absence of immediate user input. Human practitioners have to *anticipate* how the automated system will behave as circumstances change. Depending on the complexity of the system and the feedback about system activities, this may be difficult. As one commentator has put it, the most common questions people ask about their automated partners are: What is it doing? Why is it doing that? What will it do next? How in the world did we get into that mode? (Wiener, 1989). These questions are indications of coordination breakdowns—what has been termed "automation surprises." Automation surprises are situations where automated systems act in some way outside of the expectations of their human supervisors. Data from studies of these surprises in aviation and medicine (Moll van Charante, Cook, Woods, Yue, & Howie, 1993; Norman, 1990; Sarter & Woods, 1994b) indicate that poor feedback about the activities of automated systems to their human partners is an important contributor to these problems.

Autonomy and authority are properties that convey an agentlike status on the system from the point of view of human observers. This raises an important point. Automated systems have two kinds of interpretations. Based on knowledge of underlying mechanisms, an automated system is deterministic and predictable. However, those who monitor or interact with the system in context may perceive the system very differently. For example, with the benefit of knowledge of outcome and no time pressure, one can retrospectively show how a system's behavior was deterministic. But as system complexity increases, and depending on the feedback mechanisms available, predicting the system's behavior in context may be much more difficult.

A user's perception of the device depends on an interaction between its capabilities and the feedback mechanisms that influence what is observable about system behavior in relation to events in the environment. What feedback is available depends on the "image" the device presents to users (Norman, 1988). When a device is complex, has high autonomy and authority, and provides weak feedback about its activities (what has been termed "low observability"), it can create the image of an animate agent capable of independent perception and willful action. We refer to this as the *perceived animacy* of the automated system. In effect, the system, although determinate from one perspective, seems to behave as if it were an animate agent capable of activities independent of the operator (Sarter & Woods, 1994a).

Flightdeck automation on commercial transport jets illustrates how autonomy combined with low observability can create the perception of animacy (Sarter & Woods, 1994b). Pilots sometimes experience difficulties with tracking system behavior in situations that involve indirect mode transitions. In these situations, the system changes its behavior independent of any immediate pilot instructions. The system acts in response to reaching a preset target (e.g., leveling off at a target altitude) or because an envelope protection threshold is crossed. In other words, based on the programmed mechanisms, the system "realizes" the need for a mode change, carries it out without requesting pilot consent, and provides only weak feedback about the change or the implications of the change for future aircraft behavior. It is in this type of situation that pilots are known to ask questions such as: What is it doing? Why did it do this? What will it do next? (Wiener, 1989). These are questions one asks about another agent with an agenda of its own and an agent that does not communicate very well.

Much work in this area has noted that poor feedback on system status and behavior is at the heart of automation surprises. But what does it mean to say "poor feedback?" When we take a close look at the data provided to the operators of many advanced systems, it becomes quite clear that the amount of data available to the human is increasing. All of the necessary data to build a picture of their automated partner's activities is present in general. But the effectiveness of this data depends on the cognitive work needed to turn it into a coherent interpretation in context.

Effective feedback depends on more than display formats; it is a relation among the system's function, the image the system presents to outsiders, and the observer embedded in an evolving context (Woods, 1995). As a result, it is better to refer to interface and feedback issues in terms of *observability*. This term captures the fundamental relationship among thing observed, observer, and context of observation that is fundamental to effective feedback. Observability depends on the cognitive work needed to extract meaning from the data available. We, as researchers, need to make progress on better ways to measure this property of cognitive systems.

Because automated systems are deterministic, if one has complete knowledge of how a system works, complete recall of the past instructions given to the system, and total awareness of environmental conditions, then one can project accurately the behavior of the automated partner. However, as the system becomes more complex, projecting its behavior also becomes more cognitively challenging. One has to have an accurate model of how the system works, one has to call to mind the portions of this knowledge that are relevant for the current situation, one has to recall past instructions that may have occurred some time ago and may have been provided by someone else, one has to be aware of the current and projected state of various parameters that are inputs to the automation, one has to monitor the activities of the automated system, and one has to integrate all of this information and knowledge together in order to project how the automation will behave in the future. As a result, an automated system can look very different from the perspective of a user in context as compared to an analyst taking a bird's-eye view with knowledge of outcome. The latter will see how the system's behavior was a direct and natural result of previous instructions and current state; the former will see a system that appears to do surprising things on its own. This is the paradox of perceived animacy of automated systems that have high autonomy and authority but low observability (Fig. 1.1). This situation has strong implications for error analysis and incident reconstruction (Woods et al., 1994).

The trend in automation seems to be for greater increases in system autonomy and authority, whereas feedback mechanisms are stagnant at best. The result appears to be that "strong and silent" automation is on the increase (Norman, 1990). Yet the research to date has revealed that there are latent dangers of powerful yet silent automation (e.g., Cook, Woods, & Howie, 1992).

Designing automated systems is more than getting that machine to function autonomously, it is also making provisions for that automated agent to coordinate its activity with other agents. Or, perhaps more realistically, it is making provisions so that other human agents can see the assessments and activity of the automated agent so that these human practitioners can perform the coordination function by managing a set of partially autonomous subordinate agents (see Billings, 1991; Woods et al., 1994).

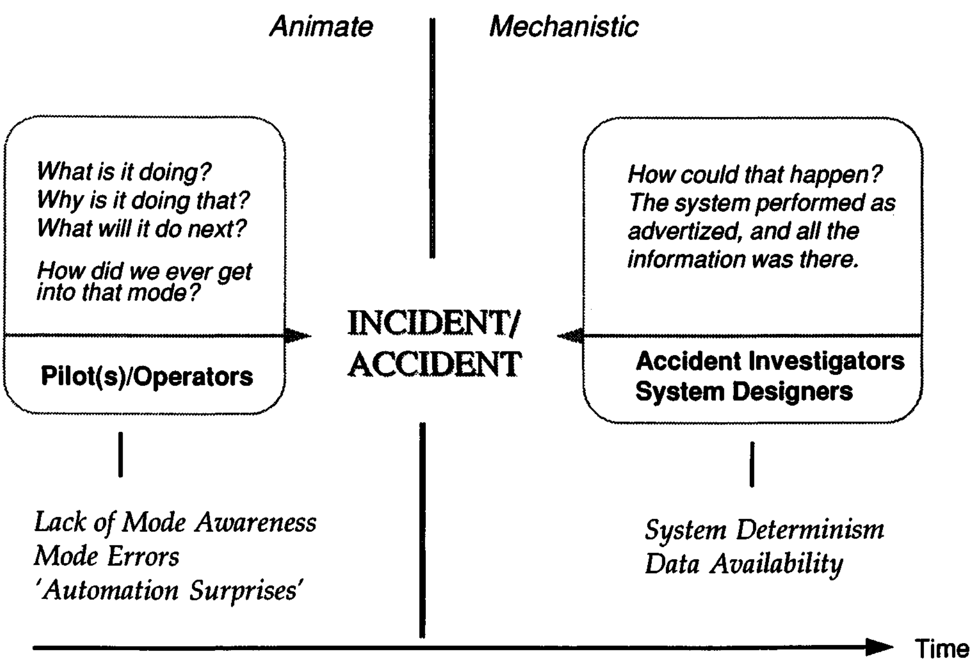

FIG. 1.1. A paradox associated with perceived animacy. Automated systems that have high autonomy and authority but low observability appear to behave as if they are animate agents capable of activities independent of the operator. However, such systems are deterministic, and their behavior is predictable if one has complete and available knowledge of how the system works, complete recall of the past instructions given to the system, and total awareness of the situation and environmental conditions.

FLEXIBILITY: BURDENSOME OR INSTRUMENTAL?

Flexibility and customizability are central to the perceived advantages of the growth in technological powers (Woods, 1993). New automated systems are often flexible in the sense that they provide a large number of functions and options for carrying out a given task under different circumstances. For example, the computers on commercial jet flightdecks provide at least five different mechanisms at different levels of automation for changing altitude. This customizability is construed normally as a benefit that allows practitioners to select the mode or option best suited to a particular situation. However, it also creates a variety of new demands.

To utilize highly flexible systems, the practitioner must learn about all of the available options, learn and remember how to deploy them across the variety of real operational contexts that can occur, and learn and remember the interface manipulations required to invoke different modes or features. Monitoring and attentional demands are also created as practitioners must keep track of which mode is active. All of this represents new burdens on the practitioner to set up and manage these capabilities and features.

If the new tasks and workload created by such flexible systems tend to congregate at high-workload and high-criticality periods, the result is a syndrome called *clumsy automation* by Wiener (1989). Clumsy automation is a form of poor coordination between the human and machine in the control of dynamic processes where the benefits of the new technology accrue during workload troughs and the costs or burdens imposed by the technology (i.e., additional tasks, new knowledge, forcing the user to adopt new cognitive strategies, new communication burdens, new attentional demands) occur during periods of peak workload, high-criticality, or high-tempo operations (Cook & Woods, 1995; Sarter & Woods, 1994b). Significantly, deficits like this can create opportunities for new kinds of human error and new paths to system breakdown that did not exist in simpler systems (Woods et al., 1994).

The result is that we need to understand the difference between two types of flexibility in cognitive artifacts: (a) flexibilities that serve to increase practitioners' range of adaptive response to the variability resident in the field of activity, and (b) flexibilities that simply create new burdens on practitioners, especially at high-tempo or high-criticality periods (Woods, 1993).

PROPERTIES OF THE COMPUTER SHAPE PRACTITIONER COGNITION AND BEHAVIOR

Today, technological change is transforming the workplace through the introduction and spread of new computer-based systems. Thus, automation can be seen in part as computerization. But there are a variety of properties of the computer as a medium that shape practitioner cognition and behavior in predictable but problematic ways.

Computer-based information technology allows designers to combine multiple features, options, and functions onto a single physical platform. The same physical device can be designed to operate in many different contexts, niches, and markets simply by taking the union of all the features, options, and functions that are needed in any of these settings. In a sense, the computer medium allows one to create multiple virtual devices concatenated onto a single physical device. After all, the computer medium is multifunction—software can make the same keys do different things in different combinations or modes, or provide soft keys, or add new options to a menu structure; the CRT or other visual display unit (VDU) allows one to add new displays that can be selected if needed to appear on the same physical viewport.

But to do this pushes the designer to proliferate modes and thereby create the potential for mode errors, to proliferate displays hidden behind the narrow viewport and create navigation problems, to assign multiple functions to controls so that users must remember complex and arbitrary sequences of operation. In other words, the modularity of the computer medium helps designers follow Norman's (1988) tongue-in-cheek advice on how to do things *wrong* in designing computer-based devices. Such systems appear on the surface to be simple because they lack large numbers of physical display devices and controls; however, underneath the placid surface of the CRT workstation a variety of clumsy features may exist that produce operational complexities.

Computerization also has tremendously advanced our ability to collect, transmit, and transform data. In all areas of human endeavor, we are bombarded with computer-processed data. But our ability to digest and interpret data has failed to keep pace with our abilities to generate and manipulate greater and greater amounts of data. Thus, we are plagued by data overload. User interface technology has allowed us to concentrate this expanding field of data into one physical platform, typically a single visual display unit (VDU). Users are provided with increased degrees of flexibility for data handling and presentation in the computer interface through window management and different ways to display data. The technology provides the capability to generate tremendous networks of computer displays as a kind of virtual perceptual field viewable through the narrow aperture of the VDU. These changes effect the cognitive demands and processes associated with extracting meaning from large fields of data (Woods, 1995).

We have demonstrated in several studies how characteristics of computer-based devices influence cognition and behavior in ways that increase the potential for erroneous actions and assessments. In one case (Cook & Woods, in press), a new operating-room patient-monitoring system was studied in the context of cardiac anesthesia. This and other similar systems integrate what was previously a set of individual devices, each of which displayed and controlled a single sensor system, into a single CRT display with multiple windows and a large space of menu-based options for maneuvering in the space of possible displays, options, and special features. The study consisted of observing how the physicians learned to use the new technology as it entered the workplace.

By integrating a diverse set of data- and patient-monitoring functions into one computer-based information system, designers could offer users a great deal of customizability and options for the display of data. Several different windows could be called, depending on how the users preferred to see the data. However, these flexibilities all created the need for the physician to interact with the information system—the physicians had to direct attention to the display and menu system and recall knowledge about that system. Furthermore, the computer keyhole created new interface management tasks by forcing serial access to highly interrelated data and by creating the need to periodically declutter displays to avoid obscuring data channels that should be monitored for possible new events.

The problem occurs because of a fundamental relationship, the escalation principle: the greater the trouble in the underlying system or the higher the tempo of operations, the greater the information processing activities required to cope with the trouble or pace of activities (Woods et al., 1994). For example, demands for monitoring, attentional control, information, and communication among team members (including human-machine communication) all tend to go up with the tempo and criticality of operations. This means that the burden of interacting with the display system tends to be concentrated at the very times when the practitioner can least afford new tasks, new memory demands, or diversions of his or her attention away from patient state to the interface per se.

The physicians tailored both the system and their own cognitive strategies to cope with this bottleneck. In particular, they were observed to constrain the display of data into a fixed, spatially dedicated default organization rather than exploit device flexibility. They forced scheduling of device interaction to low-criticality self-paced periods to try to minimize any need for interaction at high-workload periods. They developed stereotypical routines to avoid getting lost in the network of display possibilities and complex menu structures.

These are all standard tactics people use to cope with complexities created by the clumsy use of computer technology (Woods et al., 1994). We have observed that pilots, space flight controllers, as well as physicians cope with new burdens associated with clumsy technology by learning only a subset of stereotypical methods, underutilizing system functionality. We have observed these practitioners convert interface flexibility into fixed, spatially dedicated displays to avoid interacting with the interface system during busy periods. We have observed these practitioners escape from flexible but complex modes of automation and switch to less automated, more direct means of accomplishing their tasks when the pace of operations increases. This adaptation or tailoring process occurs because practitioners are responsible not just for the device operation, but also for the larger performance goals of the overall system. As a result, practitioners tailor their activities to insulate the larger system from device deficiencies (Cook & Woods, in press; Woods et al., 1994).

HIGHLY COUPLED SYSTEMS BREAK DOWN IN NEW WAYS

The scale of and degree of coupling within complex systems creates a different pattern for disaster where incidents develop or evolve through a conjunction of several small failures, both machine and human (Perrow, 1984; Reason, 1990). There are multiple contributors; all are necessary, but they are individually insufficient to produce a disaster. Some of the multiple contributors are latent; that is, conditions in a system that produce a negative effect but whose consequences are not revealed or activated until some other enabling condition is met. This pattern of breakdown is unique to highly coupled systems and has been labeled the *latent failure model of complex system breakdown* (Reason, 1990).

Computerization and automation couple more closely together different parts of the system. Increasing the coupling within a system has many effects on the kinds of cognitive demands to be met by the operational system. And increasing the coupling within a system changes the kinds of system failures one expects to see (Perrow, 1984; Reason, 1990). The latent failure model for disaster is derived from data on failures in highly coupled systems.

Automation and computerization increase the degree of coupling among parts of a system. Some of this coupling is direct, some is based on potential failures of the automation, and some is based on the effects of automation on the cognitive activities of the practitioners responsible for managing the system. For example, higher coupling produces more side effects to failures. A failure is more likely to produce a cascade of disturbances that spreads throughout the monitored process. Symptoms of faults may appear in what seem to be unrelated parts of the process (effects at a distance). These and other effects can make fault management and diagnosis much more complicated.

Highly coupled processes create or exacerbate a variety of demands on cognitive functions (Woods, 1988). For example, increased coupling creates:

- New knowledge demands (e.g., knowing how different parts of the system interact physically or functionally).
- New attentional demands (e.g., deciding whether or not to interrupt ongoing activities and lines of reasoning as new signals occur).
- More opportunities for situations to arise with conflicts between different goals. New strategic trade-offs can arise as well. Creating or exacerbating conflicts and dilemmas produces new forms of system breakdown (see Woods et al., 1994).

Automation may occur in the service of stretching capacity limits within a system. But these efficiency pressures may very well create or exacerbate double binds that practitioners must face and resolve. These pressures may also reduce margins, especially by reducing the error tolerance of the system and practitioners' ability to recover from error and failures. Characteristics such as error tolerance and the degree of observability through the computer interface can change the ability of practitioners to buffer the system in the face of contingencies and complications (Woods et al., 1994).

Technology change often facilitates greater participation by formerly remote individuals. People, who represent different but interacting goals and constraints, now can interact more directly in the decision-making process. Coordination across these diverse people and representatives and cooperation among their interacting interests must be supported.

Overall, changes in automation, through increased coupling, make systems more vulnerable to the latent failure type of system breakdown where multiple contributors come together in surprising ways (see also Hollnagel, 1993; Woods et al., 1994). Thus, increases in level of automation can change the kinds of incidents, their frequency, and their consequences in ways that can be very difficult to foresee. Interestingly, the signature of failure in tightly coupled systems is often misperceived and labeled as simply another case of human error.

TECHNOLOGY CHANGE TRANSFORMS OPERATIONAL AND COGNITIVE SYSTEMS

These effects of technology change run counter to conventional wisdom about automation. There are two broad themes that run throughout the previous discussion.

First, changes in level of automation *transform* systems. Technology change is an intervention into an ongoing field of activity (Flores, Graves, Hartfield, & Winograd, 1988; Winograd & Flores, 1987). When developing and introducing new technology, one should realize that technology change represents new ways of doing things; it does not preserve the old ways with the simple substitution of one medium for another (e.g., paper for computer-based; hardwired for digital; automatic for manual).

Marketing forces tout the universal benefits of all types of new technology without reservations. However, the difference between skillful and clumsy use of technological powers lies in understanding how automation can transform systems. For example, where and when does it create new burdens? How does the keyhole property of the computer shape practitioner cognition in ways that reduce error and failure recovery? What are the new patterns of system breakdown? What is the new cooperative or joint human–machine system created by more automation? How does this cooperative system function when complicating factors arise at and beyond the margins of normal routines?

Understanding the potential transformations allows one to identify and treat the many postconditions necessary for skillful use of technological possibilities. To do this one must unwrap the automation package. In doing so we must come to recognize that new technology is more than object in itself. When we design new automated and computerized systems we are concerned with more than just a hardware and software object. We are also designing:

- A dynamic *visualization* of what is happening and what may happen next that provides practitioners with feedback about success and failure and about activities and their effects.
- A *tool* that helps practitioners respond adaptively to the many different circumstances and problems that can arise in their field of activity.
- A *team* of people and machine agents that can coordinate their assessments and activities as a situation escalates in tempo, difficulty, and criticality.

APPARENT SIMPLICITY, REAL COMPLEXITY

Conventional wisdom about automation makes technology change seem simple. Automation is just changing out one agent (a human) for another. Automation provides more options and methods. Automation frees up people for other more important jobs. Automation provides new computer graphics and interfaces. However, the reality of technology change, as revealed through close examination of device use in context, is that technological possibilities often are used clumsily, resulting in strong, silent, difficult-to-direct systems that are not team players.

The discussion of how technology change transforms systems points out the irony present in conventional claims about the effects of automation. The very characteristics of computer-based devices that have been shown empirically to complicate practitioners' cognitive activities and contribute to errors and failures are generally justified and marketed on the grounds that they reduce human workload and improve human performance. Technologists assume that automation will automatically reduce skill requirements, reduce training needs, produce greater attention to the job, and reduce errors.

New technology can be used skillfully to increase skilled practice and to produce more reliable human-machine systems, but not through wishful thinking or superficial claims about the impact of new technology on human-machine systems. Understanding or predicting the effects of technology change requires close examination of the cognitive factors at work in the operational system. Studying and modeling joint human-machine cognitive systems in context is the basis for skillful as opposed to clumsy use of the powers of technology (Woods et al., 1994).

ACKNOWLEDGMENTS

The preparation of this manuscript was supported, in part, under a Cooperative Research Agreement with NASA-Ames Research Center (NCC 2-592; Technical Monitor: Dr. Everett Palmer) and under a grant with the FAA Technical Center (Grant No. 93-G-036; Technical Monitor: John Zalenchak).

REFERENCES

Billings, C. E. (1991). *Human-centered aircraft automation philosophy* (NASA Technical Memorandum 103885). Moffett Field, CA: NASA-Ames Research Center.

Cook, R. I., & Woods, D. D. (1995). Adapting to new technology in the operating room. *Human Factors,* in press.

Cook, R. I., Woods, D. D., & Howie, M. B. (1992). Unintentional delivery of vasoactive drugs with an electromechanical infusion device. *Journal of Cardiothoracic and Vascular Anesthesia, 6,* 1-7.

Flores, F., Graves, M., Hartfield, B., & Winograd, T. (1988). Computer systems and the design of organizational interaction. *ACM Transactions on Office Information Systems, 6,* 153-172.

Hollnagel, E. (1993). *Human reliability analysis: Context and control.* London: Academic Press.

Hutchins, E. (1990). The technology of team navigation. In J. Galegher, R. Kraut, & C. Egido (Eds.), *Intellectual teamwork: Social and technical bases of cooperative work* (pp. 191-220). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hutchins, E. (1994). *Cognition in the wild*. Cambridge, MA: MIT Press.

Johannesen, L., Cook, R. I., & Woods, D. D. (1994). Cooperative communications in dynamic fault management. In *Proceedings of the 38th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 225–229). Santa Monica, CA: Human factors and Ergonomics Society.

Layton, C., Smith, P. J., & McCoy, E. (1994). Design of a cooperative problem-solving system for enroute flight planning: An empirical evaluation. *Human Factors, 36*(1), 94–119.

Malin, J., Schreckenghost, D., Woods, D., Potter, S., Johannesen, L., Holloway, M., & Forbus, K. (1991). *Making intelligent systems team players: Case studies and design issues* (NASA Technical Report, TM-104738). Houston, TX: NASA Johnson Space Center.

Moll van Charante, E., Cook, R. I., Woods, D. D., Yue, Y., & Howie, M. B. (1993). Human–computer interaction in context: Physician interaction with automated intravenous controllers in the heart room. In H. G. Stassen (Ed.), *Analysis, design and evaluation of man-machine systems 1992* (pp. 000–000). New York: Pergamon Press.

Norman, D.A. (1988). *The psychology of everyday things*. New York: Basic Books.

Norman, D.A. (1990). The "problem" of automation: Inappropriate feedback and interaction, not "over-automation." *Philosophical Transactions of the Royal Society of London, B 327*, 585–593.

Perrow, C. (1984). *Normal accidents. Living with high-risk technologies*. New York: Basic Books.

Reason, J. (1990). *Human error*. Cambridge, England: Cambridge University Press.

Roth, E. M., Bennett, K. B., & Woods, D. D. (1987). Human interaction with an "intelligent" machine. *International Journal of Man-Machine Studies, 27*, 479–525.

Sarter N. B., & Woods D. D. (1994a). Decomposing automation: Autonomy, authority, observability and perceived animacy. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 22–27). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sarter N. B., & Woods D. D. (1994b). Pilot interaction with cockpit automation II: An experimental study of pilots' mental model and awareness of the Flight Management System (FMS). *International Journal of Aviation Psychology, 4*, 1–28.

Sarter, N. B., & Woods, D. D. (1995). "How in the world did we get into that mode?" Mode error and awareness in supervisory control. *Human Factors 37*(1), 5–19.

Wiener, E. L. (1989). *Human factors of advanced technology ("glass cockpit") transport aircraft* (Technical Report 117528). Moffett Field, CA: NASA-Ames Research Center.

Winograd, T., & Flores, F. (1987). *Understanding computers and cognition*. Reading, MA: Addison-Wesley.

Woods, D. D. (1988). Coping with complexity: The psychology of human behavior in complex systems. In L. P. Goodstein, H. B. Andersen, and S. E. Olsen (Eds.), *Tasks, errors, and mental models* (pp. 128–148). New York: Taylor and Francis.

Woods, D. D. (1993). Price of flexibility in intelligent interfaces. *Knowledge-Based Systems, 6*(4), 189–196.

Woods, D. D. (1994). Cognitive demands and activities in dynamic fault management: Abductive reasoning and disturbance management. In Stanton, N. (Ed.), *The human factors of alarm design* (pp. 63–92). London: Taylor and Francis.

Woods, D. D. (1995). Towards a theoretical base for representation design in the computer medium: Ecological perception and aiding human cognition. In J. Flach, P. Hancock, J. Caird, & K. Vicente (Eds.), *An ecological approach to human-machine systems I: A global perspective* (pp. 157–188). Hillsdale NJ: Lawrence Erlbaum Associates.

Woods, D. D., Johannesen, L. J., Cook, R. I., & Sarter, N. B. (1994). *Behind human error: Cognitive systems, computers, and hindsight*. Dayton, OH: Crew Systems Ergonomic Information and Analysis Center (CSERIAC).

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

2 Operator Reliance on Automation: Theory and Data

Victor Riley
Honeywell Technology Center

INTRODUCTION

On June 30, 1994, an Airbus A330 crashed during a test flight, killing all seven on board. The test was being performed to evaluate how well the aircraft's autopilot system performed with an engine out, simulated hydraulic failure, and rear center of gravity just after takeoff. According to the investigating committee, the crew appeared overconfident and did not intervene in time to prevent the accident. They calculated that if the test pilot had retaken manual control four seconds earlier than he actually did, the accident would have been avoided (Sparaco, 1994).

As illustrated by this accident, the decision to rely or not rely on automation can be one of the most important decisions an operator of a complex system can make. Indeed, the decision has been a critical link in the chains of events that have led to many incidents and accidents in aircraft, railroad, ship, process control, medical, and power plant operations. In these cases, the operators may rely too much on the automation and fail to check up on it or monitor its performance, or they may defeat the automation because they have high, possibly misplaced, confidence in their own abilities to perform the tasks manually. Several aircraft accidents illustrate the former problem; the Chernobyl nuclear power plant accident is one example of the latter.

Until recently, little has been known about what factors influence the decision to use or not use automation and what types of bias to which this decision may be subject. A better understanding of these factors and biases may help system developers anticipate the conditions under which operators may underrely or overrely on automation and guide the development of training methods and user interfaces to encourage rational automation use.

At a time when automation was being given increasing authority in complex systems, Sheridan and Farrell (1974) expressed concern about the changing roles of human operators and automation, and included operator trust in the automation as one of the fundamental elements of supervisory control tasks. Several investigators have since investigated human trust in automated systems. Muir (1987) performed a literature review on trust and developed a theory of trust in automation that combined two previous theories of trust between humans. Muir (1989) went on to perform two experiments using a process control simulation to demonstrate that operator trust in automation could be measured using a subjective scale, that operators could distinguish between faulty and reliable components of a system, and that a subjective measure of trust in a component correlated with the operator's use of that component.

Lerch and Prietula (1989) examined how attributions of qualities to agents affected operator trust in the agent. They also found evidence that it was harder to recover trust in an agent after a failure than to build trust in it initially, in support of one of the hypotheses proposed by Muir (1987). Will (1991) performed a study with petroleum engineers to determine the extent to which they relied on an expert system to perform well analysis. The expert system was intended to examine the well data and recommend a solution, but the experimenters had introduced a fault that would cause the system to make the wrong recommendation. Initially, all subjects used the system and expressed confidence in the result. One expert subject later redid the analysis by hand and discovered the error. The only significant difference between the expert and novice subjects was their expressed opinion of the utility of the system, with experts saying they would have done just as well without the system (without knowing that the answer they had reached was wrong).

Riley (1989) suggested that an operator's decision to rely on automation may not depend only on the operator's level of trust in the system, but rather on a more complex relationship among trust, self-confidence, and a number of other factors. The major thrust of this argument was that if the operator had more confidence in his or her own ability to do the task than trust in the automation, the operator was likely to do the task manually, whereas if the operator's trust in the automation was higher than the operator's self-confidence, the operator was likely to rely on the automation. However, this relationship was mediated by other factors, including the operator's level of workload and the level of risk associated with the situation. For example, a higher level of risk may exaggerate the effects of trust and self-confidence, or it might bias an operator toward manual control because of the operator's fiduciary responsibility for the process.

The theory is shown in Fig. 2.1, where arrows represent hypothesized influences between factors. The "reliance" factor represents the probability that an operator will use automation and is influenced by the operator's self-confidence and level of trust in the automation. Trust, in turn, is influenced by the actual reliability of the automation and a "duration" factor, which is meant to account for increasing stability of the operator's opinion of the automation as the operator gains experience with it.

Muir's (1989) results provide support for the proposed relationship among automation accuracy, trust in automation, and reliance. In addition, Lee (1992) (also Lee & Moray, 1992) performed an extensive series of studies to investigate these relationships and the relationship among trust, self-confidence, and automation use. Using an extension of the process control simulation used by Muir (1989), Lee provided evidence that the combination of trust in automation and self-confidence can influence automation use. He also found a very high level of variance due to individual differences in automation use and task performance strategy. However, the relationships proposed in Fig. 2.1 among workload, self-confidence, and automation use have not received as much support because little workload-related research has looked at the question of automation use. Parasuraman, Molloy, and Singh (1993) demonstrated that an element of workload is necessary for the development of automation related "complacency." Harris, and Arthur (1993) attempted to determine whether giving subjects advance notice of workload increases would prompt them to turn automation on as a workload management strategy, but were unable to

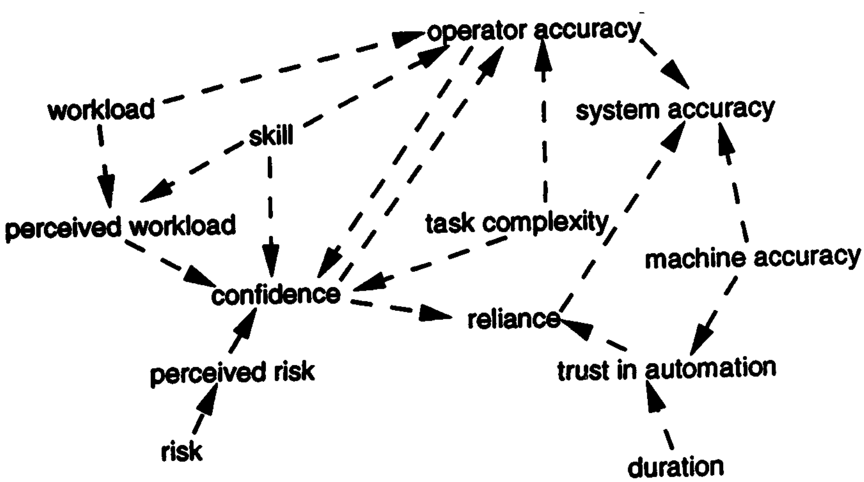

FIG. 2.1. An initial theory of operator reliance on automation. Arrows indicate the hypothesized directions of influence.

demonstrate a significant effect, partly due to a high degree of between-subject variance. Taken with the previously discussed findings, this evidence provides support for some of the relationships proposed in Fig. 2.1 while leaving others in question. The rest of this chapter summarizes the results of a series of studies to investigate these and further relationships.

AN INVESTIGATION OF FACTORS THAT INFLUENCE AUTOMATION RELIANCE USING A SIMPLE COMPUTER GAME

The experiments summarized here made use of a simple computer-based testbed (Riley, 1994). The primary purpose of this testbed was to enable independent manipulations of subject workload, the difficulty or uncertainty associated with the task that could be automated, automation reliability, and the riskiness of decisions. The ability to investigate the independent effects of these factors was a critical requirement for the testbed. In real operations, such as flight and process control, it is often difficult or impossible to separate factors. For example, when the level of risk in a situation rises, workload and task uncertainty often rise at the same time. This makes it difficult to attribute automation use decisions observed in real operations to specific factors. Isolating and testing these factors in the laboratory might shed light on how they may operate in the real world.

Another objective was to minimize the possibility that subjects' automation use decisions would depend on, or be influenced by, their different strategies in performing the task. In complex systems, many different system management strategies can be taken, and automation use decisions may be a fundamental part of these strategies. This was illustrated by Lee (1992). By using a very simple task with few possible strategies, automation use differences due to subject differences can be separated from task strategy differences.

The testbed used two tasks, one of which could be turned over to automation at the subject's discretion. The first task required the subject to categorize a character as either a letter or a number. The level of uncertainty of this task could be controlled by introducing characters that were neither letter nor number at some known rate, and scoring these characters randomly as if they were one or the other. This task could be turned over to an automated "aid" that performed the classifications for the subject. A distraction task required the subject to correct random disturbances of a marker from a target location. Workload could be controlled by varying the probability that the marker would move in a given trial. This task could not be automated, and scoring on the categorization task was contingent on the placement of the marker over the target at the end of each trial. Each trial lasted 1.75 seconds and trials were forced paced: If the subject did not respond to a trial, it was counted as wrong and the next trial began on schedule. All experiments summarized here ran for 2,050 trials (about an hour). A depiction of the screen used in the experiments is shown in Fig. 2.2.

Automation reliability was controlled by setting the probability that the automation would make a correct classification. Pretesting showed that subjects were approximately 85% accurate in performing the task manually, so automation was set at 90% accuracy in three of the experiments for the case when it was working and 50% (chance performance) when it failed; this would make the automation's performance resemble that of a good human operator, yet make it difficult to discriminate between own reliability at manual control and automation reliability. The approximation of expected subject accuracy and the remaining uncertainty due to random wrong answers was thought to reduce the influence of rational comparisons of own and automation accuracy in automation use decisions, making these decisions more open to influence from the factors of interest (workload, automation failures, uncertainty, and risk). Three of the experiments contained two automation failures in isolation and one failure in the

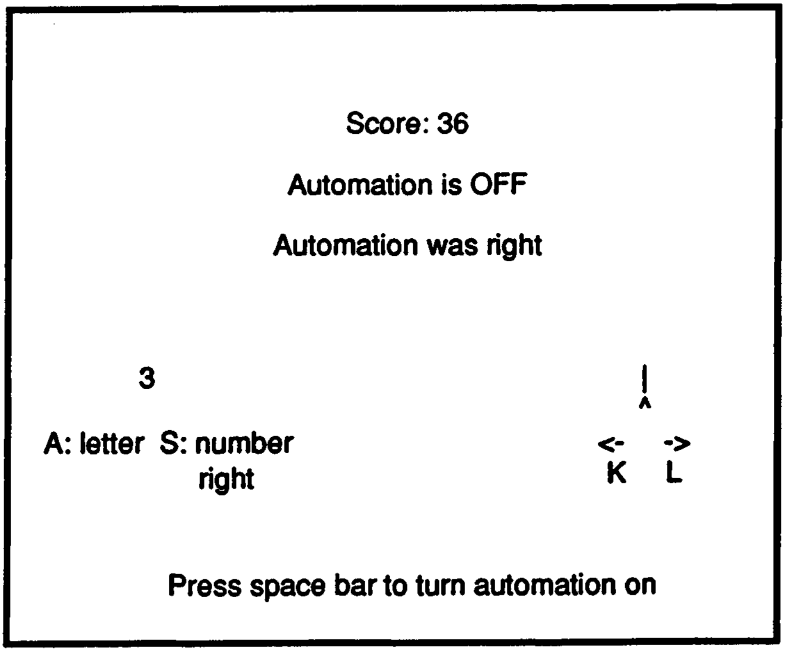

FIG. 2.2. The screen layout used by subjects in the computer game.

presence of higher workload and higher uncertainty, and workload and uncertainty also varied, both two times over the hour-long timeline. The primary dependent variable in three of the experiments was the proportion of subjects who used the automation during each trial. Because of the wide confidence limits for a proportion score, at least 30 subjects took part in each experiment.

Thirty University of Minnesota students enrolled in undergraduate psychology courses took part in Experiment One. The primary purpose of this experiment was to estimate the magnitudes of influence from automation reliability, task uncertainty, and workload on automation use decisions. A $25 award was offered to the subject who posted the highest score in the game, to provide an incentive and something of value that might be lost due to error, a necessary element of the risk manipulation to be explored later.

Fig. 2.3 shows the profiles of automation accuracy, workload, and task uncertainty over the course of Experiments One, Three, and Four. Experiment Two used a similar automation profile which will be described later. As shown, the experiment started with automation accuracy at 90%, workload at 40%, and task uncertainty at 10%. Because the experiment returned to this combination of conditions periodically, this will be referred to as the "normal" condition. After about 6 minutes, the workload

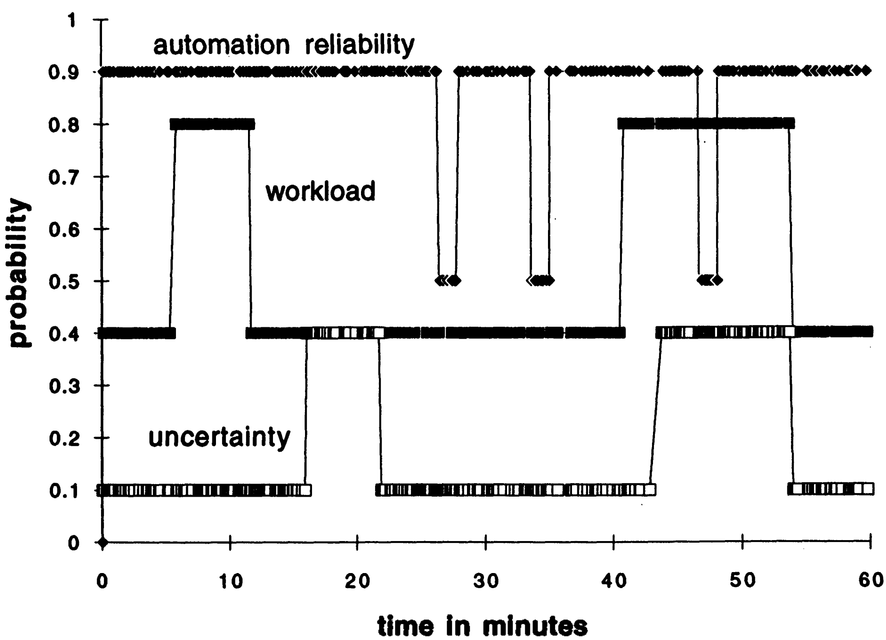

FIG. 2.3. The profiles of independent variables used in the computer game.

manipulation occurred, with workload rising to 80% for about 6 minutes. The normal condition then returned for about 4½ minutes, followed by the uncertainty manipulation where a variety of new uncategorizable characters appeared. In this condition, there was a 40% chance that the subject would be given a nonnumber/nonletter character, such as an asterisk or pound sign, reducing the expected level of manual task performance by 15%. The uncertainty manipulation lasted for about 6 minutes, followed by a return to the normal condition for 4½ more minutes before the first automation failure. This failure lasted for about 1¾ minutes, then the normal condition returned for about 6 more minutes before the automation failed again for 1¾ minutes. Following this failure, the normal condition returned for about 5½ minutes, then a combined manipulation occurred, started by a workload increase that was augmented, after about 3 minutes, by an increase in uncertainty. After about 3 minutes of the workload and uncertainty combination, the automation failed again for 1¾ minutes, then recovered. After about 6 more minutes, both workload and uncertainty returned to the normal level, which was maintained for the last 6 minutes of the game phase.

Fig. 2.4 shows the proportion of students who used the automation over the course of Experiment One in relation to the manipulation profiles. One

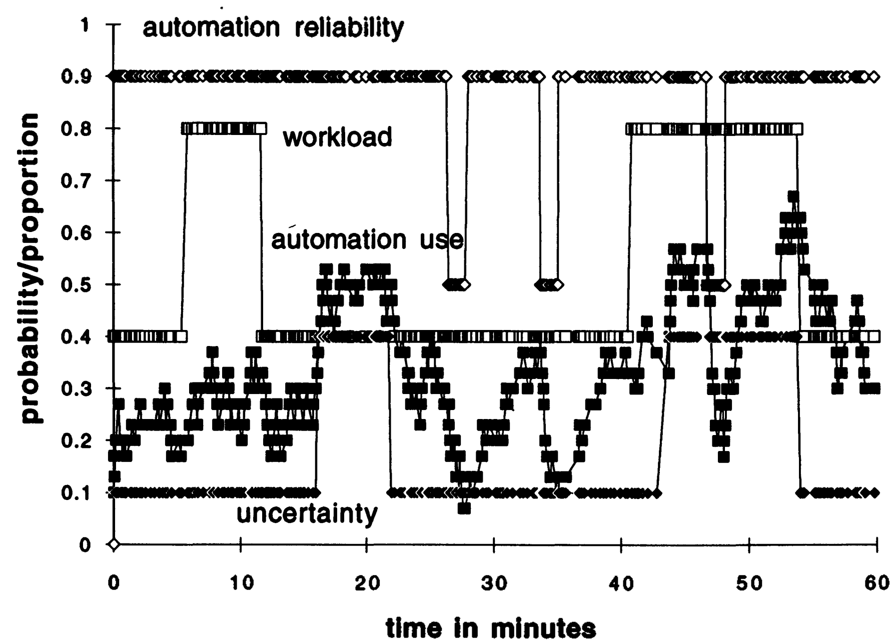

FIG. 2.4. The profile of the proportion of subjects who used the automation during each trial of Experiment One.

unexpected characteristic of the student automation use profile was its low overall level. In the normal condition, automation use tended to return to about a 35% level, although there was a gradually rising trend over the course of the experiment. Even during the uncertainty manipulation, when even the most skilled subject would expect to make a large number of errors, only about half the students relied on the automation. Although counter to an optimal strategy, this apparent bias toward manual control is consistent with observations by Lee (1992). Another surprising finding was that there was no apparent reluctance to use the automation after failures. Contrary to predictions based on Muir's (1987) investigation of human trust, subjects did not delay turning the automation back on when it appeared to recover after failure events, and automation use after each failure was not less than before the failure. However, the amount of dithering in the profile prevented precise comparisons across the profile, so the dynamics of trust in automation use were explored in more detail in the second experiment.

The independent parameters in Experiment One were very highly autocorrelated. To illustrate, consider that with three automation failures, the expected reliability of the automation changed only six times over 2,050 trials, and there were only 14 state changes (trials during which one of the independent variables changed value) overall. For this reason, a full regression analysis could not be performed to estimate parameter significance levels. Instead, the timeline was divided into segments corresponding to system states (combinations of automation reliability, workload, and task uncertainty), and the proportion of subjects using the automation in each segment was estimated by averaging across the last 10 data points in each segment. This allowed the profile to stabilize after each change of conditions and provided a good estimate of the stable value of automation use in the segment. A regression analysis of this estimator against workload, uncertainty, automation reliability, and the number of automation failures showed that uncertainty and reliability were both significant ($p = .0008$ and $.00004$, respectively), but that workload was not ($p = .96$). The regression produced a good fit, with $R = 0.94$ ($F = 17.45$, $p = .0003$). The Durbin-Watson statistic of this regression was 2.6, indicating that there was no serial correlation present.

One of the interesting results of Experiment One was the list of reasons given by subjects for choosing whether or not to rely on the automation. A comparison of total automation use and average lateness of automation use by reason given for using automation revealed differences in the strategies used by subjects. For example, subjects who cited fatigue tended to use the automation less and later than did other subjects, suggesting that they followed a manual control strategy for much of the experiment but turned the automation on late. Subjects who cited a high level of self-confidence in doing the task themselves used the automation very little, and those who cited the uncertainty manipulation used it differently from those who cited workload or errors. These results suggest that large individual differences exist in automation use strategies, and that the theory of automation use shown in Fig. 2.1 may not apply to individuals, but rather across a group; individuals may use much simpler strategies influenced by small numbers of factors.

The amount of dithering in the automation use profile prevented reliable comparisons that would have shed light on some interesting questions, such as whether it was more difficult to recover trust in the automation after failures than to gain trust in it initially. Experiment Two was intended to better understand the influences and dynamics of trust in automation and to separate the element of trust from possible uncertainty about automation states. Unlike the other three experiments, the workload and uncertainty levels remained constant and automation was 100% accurate when it worked and 0% when it failed. This was intended to reduce the subject's uncertainty about whether the automation or subject was better at the task. The dependent measure was the response time to state changes: how long it took for subjects to turn the automation on initially, off in response to the first failure, on in response to the first recovery, and so on. Three conditions were run with 17 subjects in each condition. As in Experiment One, subjects were students in undergraduate psychology courses at the University of Minnesota.

The three conditions differed only in the amount of information that subjects were given about the automation prior to playing the game. In the Trust/State condition, subjects had no prior information about the automation, so when an automation failure was encountered, subjects would be uncertain both about whether the automation had entered a partially or fully unreliable state and how long the state change would last; thus, both trust and state uncertainty would influence automation use decisions in this condition. In the Trust condition, subjects were told that the automation could only get all the answers right or wrong at a time. This was intended to eliminate the subjects' suspicion that the automation may have entered a partially reliable state, leaving trust (the subjects' projection of the automation's accuracy into the future) as the sole remaining influence. In the None condition, subjects were also told how long the automation would stay failed and recovered in each state transition, so neither trust nor state uncertainty would influence decisions. Differences among the three conditions would reveal the contributions of each element (state uncertainty and trust) to subject automation use decisions.

The response time data exhibited a large amount of skew, so a log transformation was applied to normalize the data for illustration. However, other characteristics of the data prevented the use of analysis of variance: because automation was fully under subject control, there were many cases in which a subject did not turn automation on or off prior to a state change, so no response was required for that state change. This meant that some of the cells in the data matrix were empty. For these reasons, nonparametric techniques were applied to estimate the significance of response time differences. Fig. 2.5 shows the normalized response times for reference; the normalized data correspond better to the significance levels of the differences diagnosed by the nonparametric measures than do the raw data.

The results of Experiment Two demonstrated that both state uncertainty and trust affect automation use decisions, but only early in the subjects' experience with the automation. Prior to the first failure, subjects in the None condition turned the automation on significantly faster than did those in the other two conditions, suggesting that low trust in the automation delayed subject use of the automation. However, trust was developed by flawless automation behavior over the first 20 minutes of the timeline, so high trust delayed subject responses to the first failure. Contrary to popular theories of trust, however, subjects did not show any reluctance to use the automation again after the first failure; in the Trust condition, subjects turned the automation on following the failure significantly faster than they turned it on initially, which opposes the hypothesis that trust would be harder to recover following failures than to gain initially, and there was no

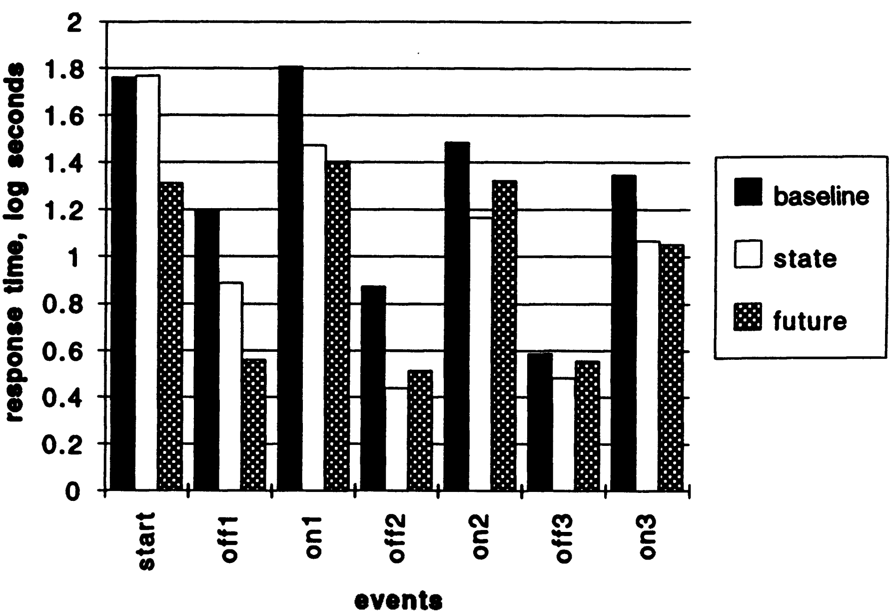

FIG. 2.5. Normalized response times to automation failure and recovery events in Experiment Two.

difference between on responses from subjects in the Trust and None conditions, suggesting that trust did not play a role in this response. After the first failure and recovery, differences between the conditions disappeared, suggesting that subjects learned to recognize automation states and anticipate its behaviors. This demonstrates that system-specific information can replace other factors as influences on automation use decisions.

One interesting feature of the response times was the large amount of skew observed. Histograms of all response times showed a cluster of responses early, then a scattering of responses late. This suggested that some of the subjects were responding directly to the manipulations whereas others turned the automation on or off in response to other factors, such as fatigue, boredom, or distraction. This was confirmed by the reasons given by subjects for making automation use decisions following the completion of the experiment. The large range of individual differences demonstrated here agrees with Lee (1992) and Harris et al. (1993) and suggests that large subject pools are required for automation studies.

Experiments One and Two provided important data regarding the fundamental influences on automation use decisions, but the use of university students and the artificial nature of the experiment environment prevent generalizing these conclusions to real systems. The first step toward drawing such generalizations is to determine whether the operators of real systems, with high levels of training and experience with advanced automation, exhibit the same automation use characteristics as did the students. After understanding the behaviors of real system operators in this highly controlled environment, we should be better able to interpret their behaviors in real systems.

Experiment Three was intended to carry out this first step. Experiment Three replicated Experiment One, but commercial airline pilots, who are trained and highly experienced with advanced automation, served as subjects. Although there was no evidence to suggest that pilots would use the automation any differently from the students, the fact that pilots use automation so extensively in real systems and the importance of their automation use decisions made this a topic of great interest. Thirty-four pilots from a major airline took part, and the conditions of Experiment One were replicated as closely as possible, except for two differences: First, the pilots were offered a $100 award for best performance due to their higher expected income than the students'; and second, the pilots were run using a laptop computer whereas the students were run on a desktop machine. The types of stimuli used and a comparison of the visual characteristics of the two displays suggested that any display differences would not produce task performance differences.

The automation use profile produced by the pilots is shown along with that produced by the students in Fig. 2.6. The pilot profile virtually

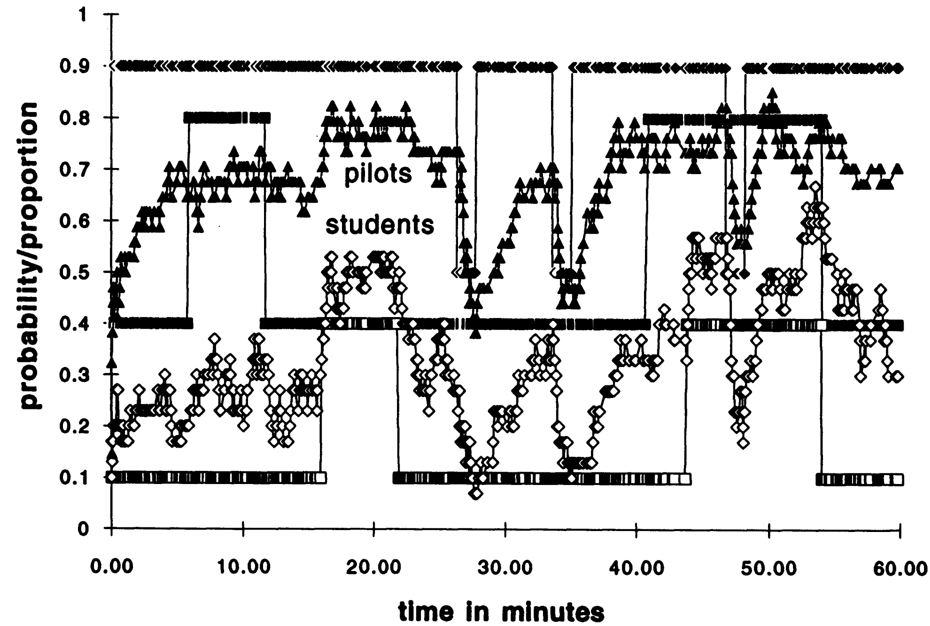

FIG. 2.6. The profile of the proportion of pilots who used the automation during each trial of Experiment Three, compared with student use of automation in Experiment One. The pilots used the automation much more than the students did, but the dynamic characteristics of automation use were very similar.

duplicated the student profile in its dynamic behaviors, but the pilots' overall level of automation use was much higher and showed a substantial bias in favor of automated control. The proportions were separated by an average of 34% ($p < .01$), and the greater number of trials over which automation was used by the pilots was highly significant ($p < .00005$, two-tailed test). Secondary tests supported no explanations for this difference due to age, willingness to accept risk, attitudes toward automation, or other differences between the groups. This result suggests that some aspect of pilot experience or training biases them in favor of automation use. However, the dynamic characteristics of pilot automation use almost replicated those of student use; workload was insignificant ($p = .90$), uncertainty significant ($p = .01$), and automation reliability highly significant ($p = 8.14E\text{-}7$), using the same method of taking the last 10 data points in each segment as point estimators for regression analysis.

One of the surprising aspects of the pilot automation use profile was its high level during automation failures. Fully one third of the pilots continued to use the automation throughout the failure periods. However, those pilots who did turn the automation off in response to failures did so somewhat faster than the students did, arguing against the possibility that those pilots who did not turn the automation off simply did not notice the failure. These results, however, do not imply that pilots will overrely on automation on the flightdeck; this particular experiment used a highly artificial environment, and the potential loss of a $100 award does not compare with the consequences of errors committed on the flightdeck. To investigate whether differences in risk, defined as the likelihood and consequences of error, might influence automation use, the final experiment was developed with an additional penalty for categorization errors. The expectation going into this experiment was that when the consequences of error increased, subjects would want to assume manual control, preferring errors of commission over errors of omission (letting the automation have control and make the costly errors).

Experiment Four replicated Experiment Three, but each categorization error produced a 5% probability of the loss of 10% of the subject's current point total. Over the course of 2,050 trials, each subject was expected to lose 10% of his or her current points between 10 and 15 times. As in Experiment Three, subjects in Experiment Four were offered a $100 award for the highest score posted on the game. Experiment Four subjects were pilots from the same airline as those in Experiment Three. Thirty-one pilots took part.

The automation use proportion produced by subjects in Experiment Four is shown in Fig. 2.7 along with the profile from Experiment Three. The Experiment Four profile matched that produced in Experiment Three until after the second automation failure. Subjects in the higher risk condition then took much longer to turn the automation back on following the failure and ended up using the automation at a much lower rate after the third failure. The ending difference between the proportions was about 20%, which was significant (p = .05). Otherwise, the close match between the two profiles prior to the second failure reinforced the results of Experiment Three, with the same high reliance on automation throughout failures.

Other questions of interest were also explored in the four experiments. It was thought that subjects may use automation in relation to their own level of manual proficiency; if a subject were able to accurately assess his or her own level of accuracy doing the task manually and compare that with the automation's apparent accuracy, that would influence the decision to rely on automation. As stated earlier, the levels of automation accuracy were selected to make this comparison difficult, to make the reliance decision more open to influence from the other factors of interest. To determine whether subject proficiency still played a part in the decision, the correlation between manual accuracy and automation use was examined, and no relationship was found. In addition, a measure of subject self-confidence was constructed by giving subjects the opportunity to perform additional trials under manual control and receive double the additional number of

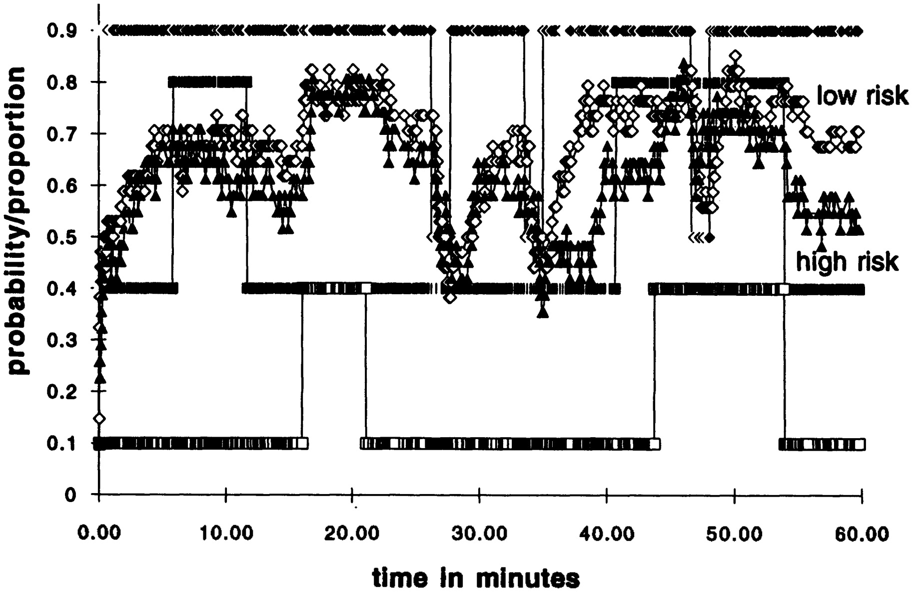

FIG. 2.7. The profile of the proportion of pilots who used the automation during each trial of Experiment Four compared with the automation use profile from Experiment Three. Pilots in the higher-risk condition tended to use the automation less than did those in the lower-risk condition after the second automation failure.

points if they performed those trials perfectly. There was no correlation between actual manual performance accuracy during the game phase and the number of additional trials selected, indicating that subjects were not good at estimating their own proficiency.

However, behavior in this additional trials task could also have been influenced by risk-taking attributes, and these attributes might also account for other differences of interest, such as between the student and pilot populations and between the low- and high-risk conditions in the pilot experiments. To examine this possibility, a subjective and an objective measure of risk taking were incorporated into all experiments. No differences in risk-taking attitudes were found between any of the groups using either measure.

Finally, it was thought that pilots and students might have different attitudes toward automation, and that this difference would account for behavior differences. To examine this possibility, the Complacency Potential Rating Scale developed by Singh, Molloy, and Parasuraman (1993) was administered to all subjects. No differences were found between groups using this scale.

CONCLUSIONS

Taken together, these results provide firm support for more of the relationships proposed in Fig. 2.1 and for some new ones. Fig. 2.8 shows a revised theory of automation use, incorporating these results. Dashed lines indicate those relationships that were originally hypothesized but not supported by the evidence (they were examined in this series of studies but the results are not reported here; see Riley, 1994, for a complete treatment), whereas solid lines indicate relationships for which evidence was gained in these and other existing studies. Fatigue and learning about system states now replace the duration factor in the first theory, and learning is shown as influencing trust and reliance directly. The influence of learning on trust is necessary to account for the effect that experience with the automation has on trust (as demonstrated in Experiment Two), but the direct influence of learning on reliance replaces the influence of trust on reliance as subjects are better able to recognize system states and anticipate system behaviors rationally. The results also reaffirm the wide range of individual differences in automation use. Therefore, the theory shown in Fig. 2.8 should not be considered as applying to an individual's automation use decisions, but rather to represent automation use behaviors produced across a group. It is likely that an individual uses a much simpler strategy than this theory suggests, and that individual strategies are influenced by a few factors or small subsets of this overall theory.

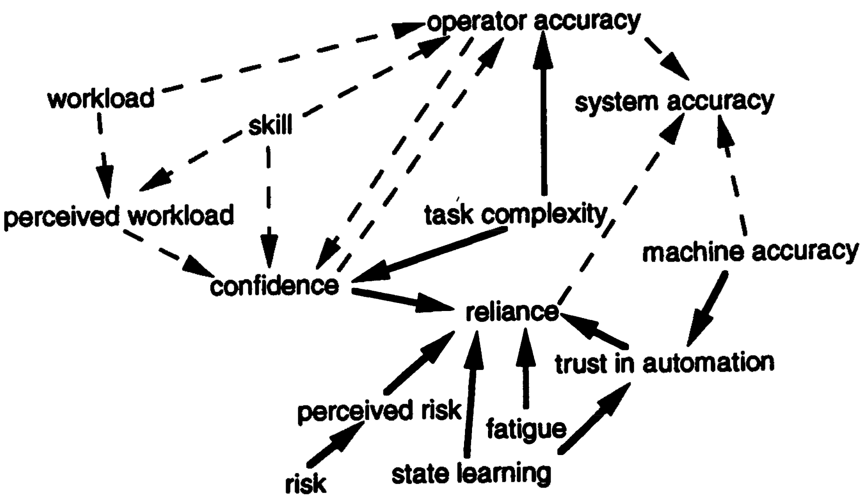

FIG. 2.8. The revised theory of automation use. Dotted arrows show hypothesized relationships that have not been confirmed by experimental evidence, whereas solid lines represent those relationships supported by evidence from these studies.