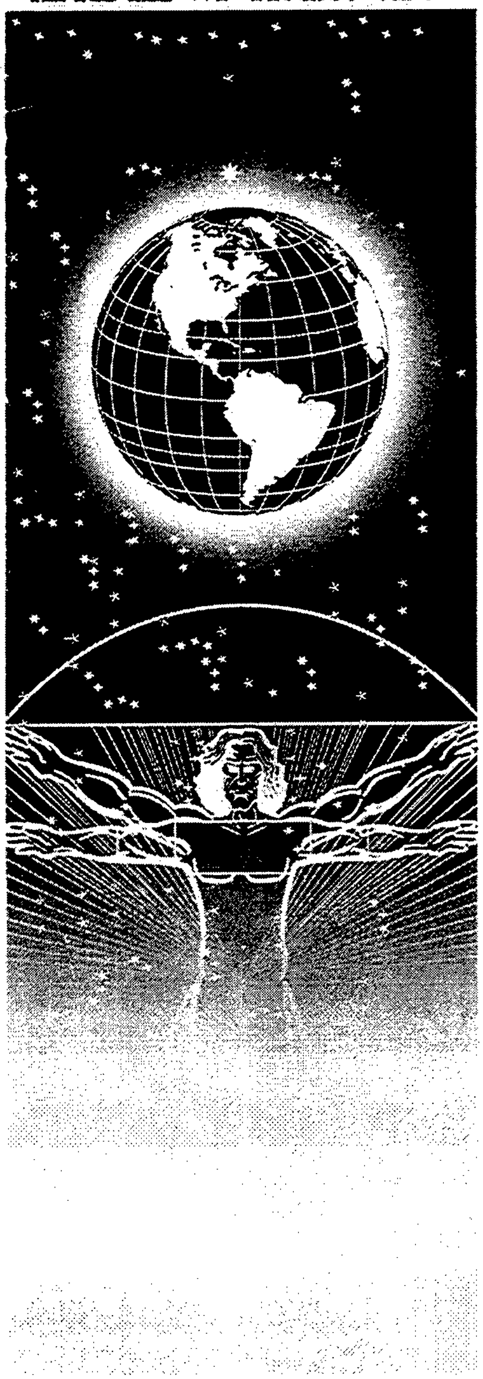
AFRL-HE-WP-TR-1999-0195



UNITED STATES AIR FORCE RESEARCH LABORATORY

CAN WE EVER ESCAPE FROM DATA OVERLOAD? A COGNITIVE SYSTEMS DIAGNOSIS

David D. Woods
Emily S. Patterson

COGNITIVE SYSTEMS ENGINEERING LABORATORY
INSTITUTE FOR ERGONOMICS
THE OHIO STATE UNIVERSITY
COLUMBUS OH 43210

Emilie M. Roth

ROTH COGNITIVE ENGINEERING
89 RAWSON ROAD
BROOKLINE MA 02445-4509

MARCH 1998

INTERIM REPORT FOR THE PERIOD 10 FEBRUARY 1997 TO 31 MARCH 1998

19991207 039

Human Effectiveness Directorate
Crew System Interface Division
2255 H Street
Wright-Patterson AFB OH 45433-7022

NOTICES

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Please do not request copies of this report from the Air Force Research Laboratory. Additional copies may be purchased from:

National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22161

Federal Government agencies and their contractors registered with the Defense Technical Information Center should direct requests for copies of this report to:

Defense Technical Information Center
8725 John J. Kingman Road, Suite 0944
Ft. Belvoir, Virginia 22060-6218

TECHNICAL REVIEW AND APPROVAL

AFRL-HE-WP-TR-1999-0195

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

MARIS M. VIKMANIS
Chief, Crew System Interface Division
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	March 1998	Interim - 10 February 1997 to 31 March 1998

4. TITLE AND SUBTITLE
Can We Ever Escape from Data Overload?
A Cognitive Systems Diagnosis

5. FUNDING NUMBERS
C: F41624-94-D-6000
PE: 62202F
PR: 7184
TA: 10
WU: 46

6. AUTHOR(S)
* David D. Woods
* Emily S. Patterson
** Emilie M. Roth

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
* Cognitive Systems Engineering Laboratory, Institute for Ergonomics
The Ohio State University, Columbus OH 43210
** Roth Cognitive Engineering
89 Rawson Road
Brookline MA 02445-4509

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory, Human Effectiveness Directorate
Crew System Interface Division
Air Force Materiel Command
Wright-Patterson AFB OH 45433-7022

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-HE-WP-TR-1999-0195

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
Data overload is a generic and tremendously difficult problem. In this report, we diagnose why this is the case and how intelligence analysis presents a particularly difficult version of data overload. We examine three different characterizations that have been offered to capture the nature of the data overload problem and how they lead to different proposed solutions. The first characterization is the clutter problem where there is "too much stuff," which leads to proposals to reduce the number of data bits that are displayed. The second characterization is a workload bottleneck where there is too much data to analyze in the time available. Data overload as a workload bottleneck shifts the view to particular activities rather than elemental data and leads to proposals to use automation to perform activities for the practitioner or cooperating automation to assist the practitioner. The third characterization is a problem in finding the significance of data when it is not known a priori what data will be informative. This characterization leads to model-based abstractions and representation design techniques as potential solutions. By focusing attention on the root issues that make data overload a difficult problem, we have identified a set of challenges that all potential solutions must meet. Most notably, all techniques must deal with the importance of context sensitivity in interpreting data. In order to place data in context, designers need to display data in a conceptual space that depicts the relationships, events, and contrasts that are informative in a field of practice. The diagnosis also reveals how data overload is manifested in intelligence analysis-like situations.

14. SUBJECT TERMS	15. NUMBER OF PAGES
alarm, cognitive engineering, data overload, inferential analysis, information retrieval, information visualization, intelligence analysis, intelligent agent	51
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNL

This page intentionally left blank.

PREFACE

This effort was accomplished under Contract F41624-94-D-6000, Delivery Order 0007 for the Air Force Research Laboratory's Human Effectiveness Directorate, under the direction of the Crew System Interface Division, Information Analysis and Exploitation Branch (AFRL/HECA). It was completed for the prime contractor, Logicon Technical Services, Inc. (LTSI), Dayton Ohio, under Work Unit No. 71841046: "Crew Systems for Information Warfare." Mr. Don Monk was the Contract Monitor.

This report was made possible by the contributions and support of many people, of which several deserve special mention. First, we thank Mr. Gil Kuperman for his many forms of support throughout this project, and particularly for his aid during initial interviews with intelligence analysts. Similarly, we thank Mr. Jay Finley for exceptional support on several levels, including but not limited to providing insight into the current intelligence analysis process and electronic tools. Interactions with members of the National Air Intelligence Center were greatly facilitated by Ms. Janet Reynolds. Tools under development for future use by intelligence analysts were demonstrated by Mr. James Homer and Mr. David Senders. Our understanding of the nature of the data overload problem faced in intelligence analysis-like situations, particularly as instantiated in the Zairean civil war scenario, was clarified during many stimulating discussions with Dr. Wayne Redenbarger. And, of course, we thank the intelligence analysts who donated their valuable time and expertise during a series of interviews, including Mr. James Homer, Mr. Shawn Bucy, Mr. Keith Anthony and Mr. Paul Kladitis.

TABLE OF CONTENTS

LIST OF FIGURES

This page intentionally left blank.

PART I. INTRODUCTION

> If we truly understand cognitive systems, then we must be able to develop designs that enhance the performance of operational systems; if we are to enhance the performance of operational systems, we need conceptual looking glasses that enable us to see past the unending variety of technology and particular domains.
>
> Woods and Sarter, 1993

Data overload is the problem of our age—generic yet surprisingly resistant to different avenues of attack. In order to make progress on innovating solutions to data overload in a particular setting such as intelligence analysis, we need to identify the root issues that make data overload a challenging problem everywhere and to understand why proposed solutions have broken down or produced limited success in operational settings.

Our overall guiding assumption is that to make progress on data overload in any setting requires a kind of complementarity. We need to advance our understanding of data overload in general by synthesizing results from past studies of data overload problems in control centers for engineered and physiological processes. Then we need to use this understanding to develop techniques to cope with data overload in the specific case of intelligence analysis, which has additional challenges beyond those encountered in other domains. Making progress in this specific case will serve simultaneously as a way to test for more generic concepts providing feedback to the general research base on this issue and as a place where the research base can stimulate practical innovations about what would be useful to build.

In order to achieve these objectives, a cognitive engineering team from the Institute for Ergonomics at the Ohio State University (David Woods, Principal Investigator, Emily Patterson, Emilie Roth, and Wayne Redenbarger) has

- completed a Cognitive Systems diagnosis of the sources of data overload problems in general,
- identified how the data overload problem is expressed in intelligence analysis-like situations,
- designed scenarios that instantiate aspects of data overload as experienced by intelligence analysts (e.g., the Ariane 501 launch failure, see Patterson, Roth, and Woods, in preparation; the Zairean civil war, see Woods, Patterson, Roth, and Redenbarger, 1998), and

- begun to use these scenarios to explore how analysts' strategies work when they are confronted with massive amounts of data in the electronic medium (Patterson, Roth, and Woods, in preparation).

In this report, we provide a "diagnosis" of what makes data overload a difficult problem based on past studies where we have examined how new computerized devices can help overcome or can exacerbate data overload related problems in control centers such as mission control for space shuttle operations, highly automated aviation flight decks, computerized emergency operations control centers in nuclear power plants, and surgical anesthetic management systems in operating rooms. Then we describe how intelligence analysis instantiates these issues and the additional challenges that are presented when monitoring human or organizational processes as opposed to engineered or physiological processes.

1. Data Overload is a Difficult and Generic Problem

Information is not a scarce resource. Attention is.
Herbert Simon[1]

Each round of technical advances, whether in artificial intelligence, computer graphics, or electronic connectivity promises to help people better understand and manage a whole host of activities from managing businesses to space missions to the national air space. Certainly, this ubiquitous computerization of the modern world has tremendously advanced our ability to collect, transmit, and transform data producing unprecedented levels of access to data.

However, our ability to interpret this avalanche of data (i.e., to extract meaning from artificial fields of data) has expanded much more slowly, if at all. In studies across multiple settings, we find that practitioners are bombarded with computer processed data, especially when anomalies occur. We find users lost in massive networks of computer based displays, options, and modes. Such difficulties help spur technologists to new rounds of development, but after each round, we continue to find beleaguered practitioners in virtually all areas of work and activity trying to cope with data overload in one form or another.

---

[11] In written publications, Simon has made this point several times:
"The information-processing systems of our contemporary world swim in an exceedingly rich soup of information, of symbols. In a world of this kind, the scarce resource is not information; it is the processing capacity to attend to information. Attention is the chief bottleneck in organizational activity ..." (Simon, 1976, p. 294).
"A design representation suitable to a world in which the scarce factor is information may be exactly the wrong one for a world in which the scarce factor is attention." (Simon, 1981, p. 167).

Why is data overload such a generic and persistent problem? Why has it been so tremendously difficult to make progress on this hallmark of the "information age?" In this paper we provide a diagnosis for why data overload is a difficult problem that has resisted progress. Focusing attention on root issues reveals paths for innovation.

1.1 An Explosion in Data Availability

> "The whole place just lit up. I mean, all the [alarm] lights came on. So instead of being able to tell you what went wrong, the lights were absolutely no help at all."
>
> Comment by one space controller in mission control after the Apollo 12 spacecraft was struck by lightning (Murray and Cox, 1989).

> "I would have liked to have thrown away the alarm panel. It wasn't giving us any useful information."
>
> Comment by one operator following the Three Mile Island nuclear power plant accident (Kemeny, 1979).

One level of technology change has created or exacerbated data overload problems: as computerization increasingly penetrates a field of activity, the power to collect and transmit data outstrips our ability to interpret the massive field of data available. Our problem is rarely getting the needed data, instead the problem is finding what is informative given our interests and needs in a very large field of available data. For example, one can find a version of the following statement in most accident investigation reports:

> "although all of the necessary data was physically available, it was not operationally effective. No one could assemble the separate bits of data to see what was going on" (Joyce and Lapinski, 1983).

This problem has expanded beyond technical fields of activity (an airplane cockpit or power plant control room) to everyday areas of activity as access to and the capabilities of the Internet have grown explosively. People have access to huge quantities of data in principle. However, they don't have the tools to cope with email overload or the thousands of "hits" returned by a web query.

Let us refer to this level of impact of technology change on data overload as the *data availability paradox:* more and more data is available in principle, but our ability to interpret what is available has not increased. This seems paradoxical because all participants in a field of activity recognize that having greater access to data is a benefit in principle. On the other hand, these same participants recognize how the flood of available data challenges their ability to find what is informative or meaningful for their goals and tasks.

The data availability paradox is an example of a paradox of simultaneous success and vulnerability. Technological change grows our ability to make data readily and more directly accessible – the success, and, at the same time and for the same reasons, the change increasingly and dramatically challenges our ability to make sense of the data available – the vulnerability.

1.2 People Can Find the Significance of Data: The "Wow!" Signal

The irony of the data availability paradox is that people in general are very good at finding the significance of data in many conditions. For example, Figure 1 is a printout of numbers and letters in a structure of columns and rows. An observer highlighted some of these data elements, writing the note "**Wow!**" in the margin. Clearly, these data elements were highly significant to this observer.[2]

---

[2] Kraus, J. (1979). We Wait and Wonder. Cosmic Search Vol. 1 No. 3 (http://www.bigear.org/vol1no3/wonder.htm)

Excerpts:
"Are we alone or are there other beings out there across the immense reaches of space who might be sending out radio signals we could hear?" Radio observatories, such as the Ohio State - Ohio Wesleyan radio observatory, try to answer this question by searching for signals which might indicate an extraterrestrial intelligent origin.

"In mid-August (1977) Jerry Ehman showed Bob Dixon, Dick Arnold and me a section of new computer print-out with all of the characteristics that one might expect from an extraterrestrial beacon signal. Jerry's amazement was reflected by the words "Wow!" which he had written on the margin of the print-out (Figure 1). Bob, Jerry, Dick and I had urgent discussions about its significance. We soon were referring to it as the "Wow!" signal.

The print-out format, which Bob Dixon had designed, consisted of 50 columns, one for each channel, with a single digit printed every 12 seconds indicating the signal level in that channel in units above the background level (the technical term for the unit used is one "standard deviation" or one "sigma"). A blank signified that the level was at zero. Any number above 4 or 5 might be considered as significant and probably not due to some random fluctuation. In order to accommodate levels above 9 with a single character, Bob arranged that the computer run through the alphabet with A for 10 through Z for 35.

What Jerry had noted was a sequence of characters in Channel 2 running: 6, E, Q, U, J, 5. When plotted up they produced a pattern which matched exactly (within measurement error) the telescope antenna pattern. This told us that the source was very probably celestial, that is, fixed with respect to the star background and that it passed through the telescope beam with the earth's rotation. It was strong (30 sigmas or 30 times the background) and because it appeared

Several things should strike us as we consider this example. To us, the data elements look like a meaningless mass of numbers and letters, since we

- lack the knowledge of this observer (a radio astronomer),
- lack any knowledge about what and how the elements symbolize (e.g., they represent radio telescope signals coded as the number of standard deviation units above background level),
- have no particular expectations about what is background, typical, or recent (the norm for years has been random, low level signals),
- do not know the goals of the observer (searching space for patterns of signals that might indicate an extraterrestrial intelligent origin), and
- do not know how patterns of signals that might indicate an extraterrestrial intelligent origin would be expressed in the representation.

But the data elements are not meaningless for the experienced, knowledgeable observer. While the representation looks quite crude, it does provide some support. The data is selected, pre-processed, and organized to enable experienced observers to scan for patterns. The observers are looking for an unknown, new signal, yet they can determine some properties or relationships to look for —departures from background, patterns associated with signals coming from different kinds of sources and, particularly, sources of extraterrestrial intelligent origin. The data is laid out in parallel. Given their knowledge in the field of practice and experience at scanning this representation, observers can recognize interesting patterns that stand out against the background. For example, radio astronomers at one point noticed an unusual pattern which further investigation revealed as a new natural phenomenon—pulsars.

While knowledgeable, experienced observers can find significant patterns in this data field, we are also struck by the fragility of this process given representations and tools like this printout. First, to succeed at all requires great investment in human expertise – people knowledgeable in the field of practice and practiced at observing through this representation. Second, even though people can succeed, we often find cases where the people involved miss significant aspects of the data field. Third, the kinds of representations developed in this case and others (such as status boards, annunciator panels, logs, and trend plots in traditional control rooms) are technologically crude. It seems so obvious that applying more sophisticated computer

---

in only one channel it was narrow-band (width 10 kilohertz or less). But even more significant, it was intermittent. A steady signal would have appeared two times on the record a few minutes apart as our telescope with its twin-beam scanned the sky. (The possibility that only one horn was functioning at the time can be ruled out because the two horns are balanced and, if one were out, the system would have been inoperative.) So it was an "on and off" signal! Was it intended for us? We decided that we should continue to scan the same region of sky on the chance that the signal might reappear. But it never did, and after weeks of patient listening we moved on with our survey to other parts of the sky.

processing and graphics capabilities should lead to more effective representations and tools with respect to data overload.

This example illustrates that people can find the significance in a field of data (i.e., people possess the competence to find the significance in a field of data though they may not always exhibit this competence in practice for specific cases). The questions for us to consider here become: How are people able to do this at all? How does computer technology affect people's ability to do this? How can we design visualizations to help people do this?

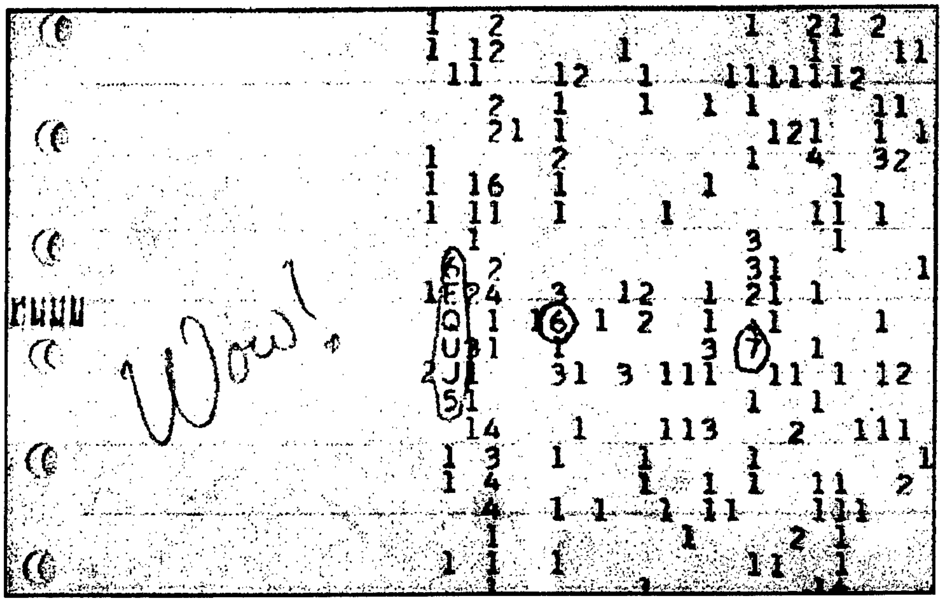


Figure 1. The "Wow!" signal.

1.3 "A Little More Technology Will Be Enough": The Technology-Centered Response to Data Overload

> Criando difficuldades para vender facilidades (creating difficulties to sell solutions).
>
> common Brazilian saying

Technologists are aware of the data availability paradox, at least implicitly. This occurs because the systems they develop almost always have surprising effects and sometimes have disappointing effects such as new operational problems (e.g., the case of cockpit automation: Billings, 1996; Sarter, Woods, and Billings, 1997; Woods and

Sarter, in press). However, these automation surprises only spur new levels of technology-centered work in the hope that the next technological advance will prove to advance our performance in interpreting the ever-larger fields of data previous technological advances have created.[3]

As the powers of technology explode around us, developers imagine potential benefits and charge ahead in pursuit of the next technological advance. The claim is that data overload and other problems will be solved by significant advances in machine 'information' processing (i.e., the technology for creating sophisticated graphics, for connecting distant people together, and for creating intelligent software agents). However, developers typically proceed in a technology-centered way (Winograd and Woods, 1997):

1. Technologists imagine how technological advances or new technological systems have promise to influence human cognition, collaboration, and activity.
2. They justify investment in the new developments by claiming that the new technology will increase productivity and reduce errors and costs, while they calm fears about development risks by saying they are just providing a supporting tool for practitioners who can choose to use it as a backup or as another option.
3. The proponents assume that, if you could build it, the imagined impact would come to pass. Research and development activity is focused exclusively on demonstrating advances towards such systems.
4. Studies of the field of practice occur only as impasses in technological development occur.
5. Eventually, interfaces are built which connect the technology to users. These interfaces typically undergo some usability testing and usability engineering to make the technology accessible to potential users.

Occasionally, useful systems emerge from the pursuit of technological advances without reference to human activities (though such advances tend to be used in ways quite different from the intentions of the developers). However, empirical studies on the impact of new technology on actual practitioner cognition, collaboration, and performance have revealed that new systems almost always have surprising consequences or even fail (e.g., Norman, 1990a; Woods, 1993; Sarter, Woods and Billings, 1997). Often the message from users, a message carried in their voices, their performance, their errors, and their adaptations, is one of technology-induced complexity. In these cases, technological possibilities are used clumsily relative to the conditions in the field of practice so that systems intended to serve the user turn out to

[3] Data overload difficulties obviously pre-date the recent rapid changes to different computer technologies (e.g., the quotes referring to data overload in the traditional control center in operation during the Three Mile Island nuclear power accident). The transition to the computer medium as a basic mechanism for data access and display has provided the opportunity for considerable tightening of the data availability paradox across a broader range of settings.

add new burdens often at the busiest times or during the most critical phases of the task and create new types of error traps (Woods et al., 1994, chapter 5).

For example, users can be surprised by new autonomous technologies that are strong but silent (Billings, 1996; Woods and Sarter, in press), asking each other questions like:

- What is it doing now?
- What will it do next?
- Why did it do this?

In other words, new technology transforms what it means to carry out activities within a field of practice—changing what knowledge is required and how it is brought to bear to handle different situations, changing the roles of people within the overall system, changing the strategies they employ, changing how people collaborate to accomplish goals.

This is a fundamental finding repeated in many fields of practice and with many kinds of technology. Systems which are developed putatively to aid users, when viewed in context, often turn out to create new workload burdens when practitioners are busiest, new attentional demands when practitioners are plagued by multiple voices competing for their attention, new sources of data when practitioners are overwhelmed by too many channels spewing out too much "raw" data. *Ironically, in practice such technology-centered systems become yet another voice in the data cacophony around us.*

The conclusion from research on the impact of technology change is that expanding the powers of technology is a *necessary but not sufficient* activity for supporting human cognition, collaboration, and performance (Winograd and Woods, 1997). What is paradoxical about this (Woods, 1984) is that human-centered solutions will almost always make use of technological powers for

- creating new kinds of visualization that reveal how a system, process, device, or activity normally functions and how it functions in anomalous ways,
- connecting people in ways that support collaboration and coordinated activity, and
- creating semi-autonomous machines that function as team players in support of their human supervisor or manager.

Ultimately, solving data overload problems requires both new technology and an understanding of how systems of people supported by various tools extract meaning from data. Our design problem is less can we build a visualization or an autonomous machine, but rather—what would be useful to visualize and how to make automated and intelligent systems team players. A little more technology, by itself, is not enough to solve generic and difficult problems like data overload, problems that exist at the intersections of cognition, collaboration, and technology.

PART II. A COGNITIVE DIAGNOSIS OF DATA OVERLOAD

2. Characterizations of Data Overload

There are three basic ways that data overload has been characterized:

1. As a clutter problem where there is *too much data*: therefore, we can solve data overload by reducing the number of data bits that are displayed,
2. As a *workload bottleneck* where there is too much to analyze in the time available: therefore, we can solve data overload by using automation and other technologies to perform activities for the user or to cooperate with the user during these activities,
3. As a problem in *finding the significance in data* when it is not known a priori what data from a large data field will be informative: therefore, we can solve data overload through model-based abstractions and representation design (Woods, 1984; Vicente and Rasmussen, 1992; Zhang and Norman, 1994)—better organizing the data to help people extract meaning despite the fact that what is informative depends on context.

2.1 Clutter

> "Clutter and confusion are failures of design, not attributes of information."
>
> Tufte, 1990, p. 51

The first way that people have characterized data overload is simply that there is "too much stuff." Such a diagnosis leads designers to try to reduce the available data. The filtering is applied at the level of the base data elements for that application.

This approach arose in the early 1980's as a "solution" to the problem of "clutter" in the design of individual computer-based displays. The approach led developers to ask: how much is too much data for people to perceive at one time or what is the maximum rate of data people can process? Developers proposed guidelines for display design that limited the number of pixels that could be lit on the screen (given technological advances this measure of screen density is obsolete, but other ways to define what are too many screen elements can and have been proposed).

This has not proven to be a successful or fruitful direction in solving data overload and has faded in large part. As we will examine in section 4.2, this approach has failed because:

- it misrepresents the design problem—see for example Tufte (1990) and Zhang and Norman (1994); one specific thematic example is that reducing data elements on one display does not reduce the available data, but rather shifts where and how data is accessed in the larger system, it increases people's need to navigate across multiple displays (Woods and Watts, 1997),

- it is based on erroneous assumptions about how human perception and cognition work; for example, the questions about maximum human data processing rates are meaningless and misleading because among other things people re-represent problems, re-distribute cognitive work, and develop new strategies and expertise as they confront complexity,
- it is utterly incapable of dealing with the context sensitivity problem—in some contexts, some of what is removed will be the relevant data.

Systems that reduce or filter available data are brittle in the face of context sensitivity. First, some of the usually unimportant data may turn out to be critically informative in a particular situation. For example, one nuclear power plant accident scenario is difficult precisely because the critical piece of data is usually unimportant (Roth et al., 1992). Second, some data that seems minor now may turn out to be important later after new events have changed the context. For example, in geopolitical affairs, in the 1997 Zaire civil war, one opposition figure, Kabila, surprisingly emerged as the leader of the rebel forces. Previous data about Kabila would have been considered minor, but it took on new significance after he emerged as a major figure in the events of 1997 (e.g., Kabila's ties to Ugandan and Rwandan leaders forged during their time as rebels is one key to Kabila's rise from obscurity and the later course of events in the civil war).

2.2 Workload Bottleneck

The second characterization of data overload has emerged in settings where access to data has grown quickly and explosively. In these contexts, such as web-based activities and intelligence analysis, participants use the words "data overload" in an everyday way that means they are experiencing what Human Factors professionals call a workload bottleneck--there are simply too many individual data units to examine them all manually in the time available.

In search of mechanisms to ease the bottleneck, people propose techniques such as:

- aids to search a database of reports, messages, etc.
  - ~ indexing the data base
  - ~ search aids
  - ~ visualization of the data base
- automation
  - ~ software agents as sentinels, notifiers, etc.

Workload bottlenecks may be a potentially useful way to think about data overload. It certainly describes the phenomena experienced by some users as their field of activity changes as a function of the reverberation of technological and organizational changes. In such settings, user tasks previously involved manually examining each report, source, or message they found as potentially relevant and synthesizing an assessment from these base data. Technological changes provide the users with access to so many reports, sources, messages, etc., that the resulting bottleneck forces users to decide what subset of reports they should examine or read.

Seeing data overload as a workload bottleneck leads developers to propose autonomous machine information processing – syntactic "relevance" metrics to prioritize reports for the user, intelligent software agents to notify users when particular types of data become available, push technologies—that identify what has been defined as the most relevant stuff for the user and modify the availability or accessibility of that material relative to other material. Development and deployment of such systems are moving forward rapidly in many areas.

Interestingly, the developments underway with software agents parallel previous cases where technologists have developed and deployed automation ostensibly to reduce workload bottlenecks – most notably in cockpit automation, but also in other settings such as automated systems in anesthetic management during surgery. The effects produced by these natural experiments have been examined (cf., e.g., summaries in Norman, 1990b; Billings, 1996; Sarter, Woods, and Billings, 1997; Woods and Sarter, in press).

The findings clearly show that technology for autonomous software agents is necessary but not sufficient to create useful systems to cope with data overload seen as a workload bottleneck. Introducing autonomous machine agents changes the cooperative structure creating new roles, new knowledge requirements, new judgments, new demands for attention, and new coordinative activities. Failing to address or support these requirements in design leads to patterns of breakdowns in coordination such as clumsy automation and automation surprises (Patterson et al., 1998).

Rather than pursue how to make intelligent and automated systems team players (summaries of research results are available based on investigations in aviation; e.g., Sarter et al., 1997), we will pursue another interpretation of the cognitive factors that underlie data overload.

2.3 The Significance of Data

> It is of the highest importance in the art of detection to be able to recognize, out of a number of facts, which are incidental and which are vital.
>
> Sherlock Holmes[4]

A cognitive systems view can provide a framework for understanding why data overload has been so resistant to technology-centered developments and why such developments seem to exacerbate rather ameliorate the data availability paradox. By providing a better diagnosis of why it has been so difficult to escape from or cope with data overload, these concepts will help frame the design problem in more productive channels.

The starting point for this approach is recognizing that large amounts of potentially available data stress one kind of cognitive activity—focusing in on the relevant or interesting subset of data for the current problem context. When people are unable to assemble or integrate all of the relevant data, this cognitive activity has broken down.

People are a competence model for this cognitive activity because people are the *only* cognitive system that we know of that is able to focus in on interesting material in natural perceptual fields, *even though what is interesting depends on context.* When people work in artificial perceptual fields, their ability to carry out this cognitive activity depends on the design of artifacts, representations, and supporting systems.

The ability to orient focal attention to "interesting" parts of the natural perceptual field is a fundamental competency of human perceptual systems (Rabbitt 1984; Wolfe 1992).

> "The ability to look, listen, smell, taste, or feel requires an animal capable of orienting its body so that its eyes, ears, nose, mouth, or hands can be directed toward objects and relevant stimulation from objects. Lack of orientation to the ground or to the medium surrounding one, or to the earth below and the sky above, means inability to direct perceptual exploration in an adequate way (Reed, 1988, p. 227 on Gibson and perceptual exploration in Gibson, 1966)."

---

[4] From A. Conan Doyle's "The Reigate Squire," first published in both *Strand* and *Harper's* in June 1893 (Hardwick, 1986).

Both visual search studies and reading comprehension studies show that people are highly skilled at directing attention to aspects of the perceptual field that are of high potential relevance given the properties of the data field and the expectations and interests of the observer. Reviewing visual search studies, Woods (1984) commented, "When observers scan a visual scene or display, they tend to look at 'informative' areas . . . informativeness, defined as *some relation between the viewer and scene,* is an important determinant of eye movement patterns" (p. 231, italics in original). Reviewing reading comprehension studies, Bower and Morrow (1990) wrote, "The principle . . . is that readers direct their attention to places where significant events are likely to occur. The significant events . . . are usually those that facilitate or block the goals and plans of the protagonist."

In the absence of this ability, for example in a newborn, as William James put it over a hundred years ago, "The baby assailed by eye, ear, nose, skin, and entrails at once, feels it all as one great blooming, buzzing confusion" (James, 1890, I 488). The explosion in available data and the limits of computer-based displays have left us often in the position of that baby—seeing a "great blooming, buzzing confusion" in the virtual data fields that technology makes so easy to create.

If we understand the mechanisms that support the ability which people possess to find the significance of data when acting in natural perceptual fields, we will be able to identify constraints, criteria, and techniques that will help people exhibit this ability when they work in the virtual perceptual fields created by modern technology.

3. Cognitive Factors and Data Overload

> Given an enormous amount of stuff, and some task to be done using some of the stuff, what is the *relevant stuff* for the task? (italics in original)
> Glymour 1987, p. 65

3.1 Why Is Focusing In On What Is Interesting Difficult? The Problem of Context Sensitivity

> "1202." Astronaut announcing that an alarm buzzer and light had gone off and the code 1202 was indicated on the computer display. Followed by these replies from mission controllers:
> "What's a 1202?"
> "1202, what's that?"
> "12...1202 alarm."
> Dialog as the LEM descended to the moon during Apollo 11
> (Murray and Cox, 1990).

The cognitive activity of focusing in on the relevant or interesting subset of the available data is a difficult task because what is interesting depends on context. What is informative is context sensitive when the meaning or interpretation of any change (or even the absence of change) is quite sensitive to some but not all the details of the current situation or past situations.

Consider this example:

> A [computer] program alarm could be triggered by trivial problems that could be ignored altogether. Or it could be triggered by problems that called for an immediate abort [of the lunar landing]. How to decide which was which? It wasn't enough to memorize what the program alarm numbers stood for, because even within a single number the alarm might signify many different things. "We wrote ourselves little rules like 'If this alarm happens and it only happens once, don't worry about it. If it happens repeatedly, but other indicators are okay, don't worry about it.'" And of course, if some alarms happen even once, or if other alarms happen repeatedly and the other indicators are not okay, then they should get the LEM [lunar module] the hell out of there.
>
> Response to discovery of a set of computer alarms linked to the astronauts displays shortly before the Apollo 11 mission (Murray and Cox 1990).

In this example, the alarm codes mean different things depending on the context in which they occur. This and other examples reveal that the meaning of a particular piece of data depends on

- what else is going on,
- what else could be going on,
- what has gone on, and
- what the observer expects or intends to happen.

To take another example from the Zaire civil war during 1997, many reports contained the same basic fact—a town fell to the rebels. However, the significance of that fact could be quite different depending on other data

- when Lubutu fell, the significance was that the rebel soldiers were pursuing refugees from their rival ethnic group;
- when Lubumbashi fell, the issue was control of mineral resources;
- when Kasindi fell, the significance was that allies (Uganda) had crossed national borders;
- when Kisangani fell, it showed the rebels were effective fighters in their showdown with government forces.

Formally, information is a relation between the data, the world the data refers to, and the observer's expectations, intentions, and interests (cf., Woods, 1991).

Understanding this is critically important to making progress on data overload. To repeat, the significance of a piece of data depends on

- other related data,
- how the set of related data can vary with larger context,
- the goals and expectations of the observer, and
- the state of the problem solving process and stance of others.

There is a widespread myth that information is something in the world that does not depend on the point of view of the observers and that it is (or is often) independent of the context in which it occurs. This is simply not the case. There are no facts of fixed significance. The available data are raw materials. A particular datum gains significance or meaning only from its relationship to the context in which it occurs or could occur including the perspective of observers. As a result, informativeness is not a property of the data field alone, but is a relationship between observers and the data field.

Take the case of a message about a thermodynamic system which states that valve X is closed. Most simply, the message signals a component status. If the operator knows (or the message also states) that valve X should be opened in the current mode of operation, the message signals a misaligned component. Or the message could signify that with valve X closed, the capability to supply material to reservoir H via path A is compromised. Or given still additional knowledge (or data search), it could signify that with valve X closed, the process that is currently active to supply material to reservoir H is disturbed (e.g., data such as actual flow less than target flow, or no flow, or reservoir H inventory low). Furthermore, the significance of the unavailability or the disturbance in the material flow process depends on the state of other processes (such as, is an alternative flow process available or is reservoir H inventory important in the current operating context). Each interpretation is built around what an object affords the operator or supervisor of the thermodynamic system, including an implicit response: correctly align component, ensure capability to supply material (or take into account the consequences of the inability to do so), repair the disturbance in the material flow process (or cope with the consequences of the disturbance), or discount these messages based on other current objectives of greater importance for the context.

In this example, the significance of a datum depends on, first, a set of contextual data. Second, which pieces of data fall into this relevance set can change both with system state and with the state of the problem solving process. The latter is particularly important—what data are relevant depend on where one is in the problem solving process. Examples of how the supervisor's situation assessment or mindset affects the interpretation of an alarm include:

- If the background situation assessment is "normal system function," then the alarm is informative, in part, because it signals that conditions are moving into abnormal or emergency operations.

- If the background line of reasoning is "trying to diagnose an unexpected finding," then the alarm may be informative because it supports or contra-indicates one or more hypotheses under consideration.
- If the background line of reasoning is "trying to diagnose an unexpected finding," then the alarm may be informative because it functions as a cue to generate more (or to broaden the set of) candidate hypotheses that might explain the anomalous process behavior.
- If the background line of reasoning is "executing an action plan based on a diagnosis," then the alarm may be informative because it functions as a cue that the current working hypothesis may be wrong or incomplete since the monitored process is not responding to the interventions as would be expected based on the current working hypothesis.

Given hindsight or the position of an omniscient observer, one can specify exactly what data are needed for the ultimate solution. However, this misses the cognitive task of focusing in on that relevant subset that is critical from the point of view of the person in the problem solving situation. Hindsight bias obscures the critical cognitive activity. This is why technology-centered approaches have been unsuccessful in coping with data overload. They miss what makes the problem difficult, and they miss the opportunity to learn from how people are able to extract meaning in natural fields despite being bombarded with sensory stimulation at an elemental level of analysis.

All techniques to cope with data overload must specify how they deal with context sensitivity. Particular techniques may try to finesse the context sensitivity problem, that is, they avoid confronting the problem directly, remaining content to nibble away at it through indirect means. Some techniques may be brittle, others robust. Brittle techniques cope with some sources of context sensitivity but break down quickly when they encounter more difficult cases. Some techniques may attempt to make machine reasoning more sensitive to context as an autonomous agent, while others are aimed at restructuring virtual worlds to enable the basic human competence to operate as it does in natural perceptual fields. But in the end, no substantial progress is possible on data overload without coping in one way or another with the context sensitivity of what is informative.

3.2 How Are People Able To Focus In On What Is Interesting?

Since people have the ability to cope with the context sensitivity of what is informative, they become the model for how to be competent at this task, a model that we need to understand in order to make fundamental progress. It is important to note that people are the only extant competence model.

Mechanisms of human perception and cognition that enable people to focus on the relevant subset of the available data, even though what is interesting depends on context, include:

- processes of perceptual organization, e.g.,
  ~ pre-attentive processing that organizes the perceptual field into meaningful units and relationships,
  ~ the fact that there exist nested layers of structure in natural perceptual fields.
- processes of attentional control, e.g.,
  ~ a mix of goal-directed and stimulus-driven processing,
  ~ the center-surround structure of vision,
  ~ the relationship between focal attention and orienting perceptual functions,
- anomaly-based processing, e.g.,
  ~ contrast-based computations that pick out and focus on anomalies (departures from typicality) and that depend on relative differences (difference in a background).

3.2.1 Perceptual Organization.

> I am standing at the window and see a house, trees, sky. And now, for theoretical purposes, I could try to count and say: there are...327 nuances of brightness [and hue]. Do I see "327"? No; I see sky, house, trees.
>
> Wertheimer, 1923/1950[5];

The quote from Wertheimer captures a fundamental aspect of human perception and cognition that relates to data overload in virtual environments. If I count elements in the perceptual field, there are an overwhelming number of basic elements varying in hue, saturation, and brightness across the visual field. But this avalanche of data does not overwhelm us because the processes of perception structure the scene into a few objects, events, and relationships between those objects (sky, house, trees). As one commentator on perception put it, "The process of organization reduces the stimulus data ... it groups large number of picture elements into a small number of *seen objects* and their parts" (Goldmeier, 1982, p. 5).

Meaning attaches to the end product of the grouping. The parts of the scene exist not as simply components of a larger whole, rather they act as carriers of their function within the whole. "What is perceived ... are the units and subunits, figures on a background, which result from perceptual grouping." The observer sees a field "... composed of objects, things, their form, their parts and subparts, *rather than of an enormous list of stimulus elements*" (Goldmeier, 1982, p. 5, emphasis added). The parts

[5] Translated from the original by N. Sarter.

and elements define higher levels of structure—objects and events in the world (Flach et al., 1995).

The ubiquitous computerization of the workplace provides the designer with the freedom to create a virtual perceptual field. The designer can (and must) manipulate the perceptual attributes of the virtual field (Wertheimer's 327 nuances of hue, brightness, saturation and shape, motion, etc. relative to other parts of the visual scene) that would automatically specify objects and their relationships in a natural scene. This results in a need to understand how perceptual attributes and features can be used as resources whose joint effect produces an organized and coherent virtual perceptual field.

One approach (data overload results from too much stuff) suggests that the answer to cluttered computer displays and data overload is to reduce or filter out data. Only use a few color categories. Reduce the number of pixels. Indicate less on the display. In contrast, studying how the perceptual system works in natural fields, as summarized above, leads us to a different approach. What matters in avoiding clutter and confusion is *perceptual organization*. More marks in the medium for representation, if they are used to better organize the virtual field, will reduce clutter. Tufte (1990) illustrates this approach admirably.

For example, some interface design guidelines have suggested that limits be set for optimal or maximum density, where density was defined as the number of graphical elements (pixels) versus the maximum number of locations available for graphical elements (the total pixels available in the display)—"18% is the optimal number of CRT pixels which should be lit." However, as Wertheimer indicates, the raw density of points of luminance is not an appropriate unit of analysis from a human perception point of view (nor are they the effective stimuli). Rather, one should count in units based on what is perceived. As Hanson (1958, p. 13) put it, "the plot is not another detail in the story, nor is the tune another note."

Tufte discusses how adding more marks need not result in a more cluttered display; the added marks may serve to organize the data. "It is not how much empty space there is, but rather how it is used. It is not how much information [read, data] there is, but rather how effectively it is arranged" (Tufte, 1990, p. 50). Clutter results from a failure to design the elements into a coherent perceptual organization or from a failure to manipulate the elements so that the resulting perceived organization captures a meaningful organization in the referent domain. Clutter occurs when people can perceive only the perceptual attributes themselves instead of a small number of objects, their parts, and their inter-relationships in the scene. Perceptual organization (perceptual grouping and figure/ground relationships) is one critical factor in avoiding clutter and confusion. We perceive objects and events rather than elemental physical parameters of the stimuli themselves. For human perception, attributes cohere to form objects and events, and we always experience all of the perceptual attributes associated with an object.

3.2.2 Control of Attention.

> Everyone knows what attention is. It is the taking possession by the mind, in a clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought.
>
> William James, 1890 I, pp. 403-404

We are able to focus, temporarily, on some objects, events, actions in the world or on some of our goals, expectations or trains of thought *while remaining sensitive to new objects or new events that may occur.*

Focus of attention is not fixed, but shifts to explore the world and to track relevant changes in the world. On flight decks, in operating rooms, and everyday work activities, attention must flow from object to object and topic to topic. In other words, one re-orients attentional focus to a newly relevant object or event from a previous state where attention was focused on other objects or on other cognitive activities (such as diagnostic search, response planning, and communication to other agents). New stimuli are occurring constantly. Sometimes such new stimuli are distractions. But other times, any of these could serve as a signal we should interrupt ongoing lines of thought and re-orient attention. This re-orientation involves disengagement from a previous focus and movement of attention to a new focus. Interestingly, this control of attentional focus can be seen as a skillful activity that can be developed through training or supported (or undermined) by the design of artifacts and intelligent machine agents.

Thus, a basic challenge for any cognitive agent at work is where to focus attention next in a changing world. Which object, event, goal, or line of thought we focus on depends on the interaction of two sets of activity. One of these is goal or knowledge directed, endogenous processes that depend on the observer's current knowledge, goals, and expectations about the task at hand. The other set of processes are stimulus- or data-driven where attributes of the stimulus world (unique features, transients, new objects) elicit attentional capture or shifts of the observer's focus. These salient changes in the world help guide shifts in focus of attention or mindset to relevant new events, objects, or tasks.

The ability to notice potentially interesting events and know where to look next (where to focus attention next) in natural perceptual fields depends on the *coordination* between orienting perceptual systems (i.e., the auditory system and peripheral vision) and focal perception and attention (e.g., foveal vision). The coordination between these mechanisms allows us to achieve a "balance between the rigidity necessary to ensure that potentially important environmental events do not go unprocessed and the flexibility to adapt to changing behavioral goals and circumstances" (Folk et al. 1992, p. 1043).

The orienting perceptual systems function to pick up changes or conditions that are potentially interesting and play a critical role in supporting how we know where to look next. To intuitively grasp the power of orienting perceptual functions, try this thought experiment mentioned by Woods and Watts (1997): put on goggles that block peripheral vision, allowing a view of only a few degrees of visual angle. Now think of what it would be like to function and move about in your physical environment with this handicap. Perceptual scientists have tried this experimentally through a movable aperture that limits the observer's view of a scene (e.g., Hochberg, 1986). Although these experiments were done for other purposes, the difficulty in performing various visual tasks under these conditions is indicative of the power of the perceptual orienting mechanisms.

3.2.3 Anomaly-Based Processing.

> ... readiness to mark the unusual and to leave the usual unmarked—to concentrate attention and information processing on the offbeat.
>
> J. Bruner, 1990, p. 78

Another hallmark of human cognitive processing is that we tend to focus on departures from typicality (this is demonstrated at all levels of processing). We do not respond to absolute levels but rather to contrasts and change. Meaning lies in contrasts—*some departure from a reference or expected course.*

Our attention flows to unexpected events. An event may be expected in one context and therefore go apparently unnoticed, but the same event will be focused on when it is anomalous relative to another context. An event may be an expected part or consequence of a quite abnormal situation, and therefore draw little attention. But in another context, the absence of change may be quite unexpected and capture attention because reference conditions are changing.

Our processing is tuned to contrasts--how behavior departs or conforms to the contrasting case. We process how the actual course of behavior follows or departs from reference or expected sequences of behavior given the relevant context.

4. Understanding Cognitive Processes Points to Human-Centered Techniques

4.1 Constraints on Effective Solutions to Data Overload

This diagnosis has led us to identify a number of constraints on effective solutions to data overload.

1. All approaches to data overload involve some sense of *selectivity*.

However, there are different forms of selectivity: facilitation or inhibition of processing. In the former, selectivity *facilitates* or enhances processing of the selected portion of the whole. In this form of selectivity, we use positive metaphors such as a spotlight of attention or a peaked distribution of resources across the field.

In the latter, selectivity *inhibits* processing of non-selected areas, for example stimuli in the selected portion can pass through and go on for further processing, whereas stimuli in the non-selected portion do not go on for processing. In this form of selectivity, we use negative metaphors such as a filter or a gatekeeping function.

Current research on cognitive solutions to data overload suggests that we need to develop positive forms of selectivity and develop techniques that support thorough exploration of the available data. This is the case in part because observers *need to remain sensitive to non-selected parts in order to shift focus fluently as circumstances change or to recover from missteps.*

2. *Organization* precedes selectivity.

Selectivity presumes a structured field on which attention can operate, focusing on potentially interesting areas depending on context. Designers of computer technology need to define the groups/objects/events and relationships attention can select.

The default in computer systems has been to organize around elemental data units or on the units of data appropriate for computer collection, transmission, and manipulation (Flach et al., 1995). These are either too elemental, as if we saw the world in "327" variations in hue, saturation, and brightness, or too removed from the meaningful objects, events, and relationships for the user's field of practice.

This finding means that effective systems for coping with data overload

- will have elaborate indexing schemes that map onto models of the structure of the content being explored
- will need to provide multiple perspectives to users and allow them to shift perspectives easily

3. All techniques to cope with data overload must deal with *context sensitivity*.

Data are informative based on *relationships* to other data, relationships to larger frames of reference, and relationships to the interests and expectations of the observer. Making data meaningful always requires cognitive work to put the datum of interest into the context of related data and issues.

This finding means that solutions to data overload will help practitioners put data into context. Presenting data in context shifts part of this burden to the external display rather than requiring the observer to carry out all of this cognitive work "in the head."

This can be done in many ways. When we display a given datum, we can show it in the context of related values. Rather than organizing displays around pieces of data, we can organize data around meaningful issues and questions--model based displays. These are models of how data relationships map onto meaningful objects, events, and processes in the referent field of activity (Flach et al., 1995).

We can use the power of the computer to help extract events from the flow of elemental data. Events are temporally extended behaviors of the device or process involving some type of change in an object or set of objects. The computer could also help observers recognize anomalies and contrasts by showing how the data departs from or conforms to the contrasting case (a departure from what is expected, from what is the plan or doctrine, from what has been typical). Since there are usually many possible contrasting cases, each defines a kind of perspective around which one views the elemental data available.

There is a prerequisite for the designer to be able to put data into context: they need to know what relationships, events, and contrasts are informative in what contexts in the field of practice.

4. Observability is more than mere data availability.

> The greatest value of a picture is when it <u>forces</u> us to notice what we never expected to see.
>
> Tukey, 1977, p. vi

There are significant differences between the available data and the meaning or information that a person extracts from that data. *Observability* is the technical term that refers to the cognitive work needed to extract meaning from available data. This term captures the relationship among data, observer, and context of observation that is fundamental to effective feedback. Observability is distinct from data availability, which refers to the mere presence of data in some form in some location. For human perception, "it is not sufficient to have something in front of your eyes to see it" (O'Regan, 1992, p.475).

One example of displays with very low observability occurs on the current generation of flight decks. The flight mode annunciations are a primary indication of how automated systems are configured to fly the aircraft. These crude indications of automation activities contribute to automation surprises where the automation flies the aircraft in a way that the pilots did not anticipate. As one pilot put it, "changes can always sneak in unless you stare at it" (see Woods and Sarter, in press for more on this example).

Observability refers to processes involved in extracting useful information. It results from the interplay between a human user knowing when to look for what information at what point in time and a system that structures data to support attentional guidance (see Rasmussen, 1985; Sarter, Woods and Billings, 1997). *The critical test of observability is when the display suite helps practitioners notice more than what they were specifically looking for or expecting* (Sarter and Woods, 1997). If a display only shows us what we expect to see or ask for, then it is merely making data available.

5. To cope with data overload, ultimately, will require the design of *conceptual spaces.* One builds a conceptual space by depicting relationships in a frame of reference (Woods, 1995; Rasmussen et al., 1994).

The search to solve data overload begins with the search for *frames of reference* that capture meaningful relationships for that field of practice. A frame of reference is a fundamental property of a space and what makes a space or map special from the point of view of representation. With a frame of reference comes the potential for concepts of neighborhood, near/far, sense of place, and a frame for structuring relations between entities. A frame of reference is a prerequisite for depicting relations rather than simply making data available.

Almost always there are multiple frames of reference that apply. Each frame of reference is like one perspective from which one views or extracts meaning from data. Part of designing a conceptual space is discovering the multiple potentially relevant frames of references and finding ways to integrate and couple these multiple frames.

4.2 Typical Finesses to Avoid the Context Sensitivity Problem

Standard approaches to data overload generally try to finesse the context sensitivity problem, either avoiding or hiding how context affects what is informative. For example, all of the following finesses have been tried with minimal success in coping with data overload in alarm systems (Woods, 1995b).

Calling these techniques finesses points to a contrast. In one sense, a finesse is a positive pragmatic adaptation to difficulty. All of these finesses are used to try to reduce data overload problems to manageable dimensions to allow experienced people to exhibit the fundamental human competence at extracting significance from data. However, a finesse is a limited adaptation because it represents a workaround rather than directly addressing the factors that make it difficult for people to extract meaning from data. Technology-centered approaches to data overload generally adopt strategies based on one or more of these finesses because of inaccurate or oversimplified models of why data overload is a generic and difficult issue.

Typical finesses that attempt to skirt the cognitive challenges that underlie data overload include:

(a) the scale reduction finesse—reduce available data.

Scaling back the available data is an attempt to reduce the amount of stuff people have to sort through to find what is significant. The belief is that if we can keep the scale of the problem manageable, then human abilities to find the critical data as the context changes will function adequately. Often scale reduction attempts are manifested as shifting some of the available data to more "distant" secondary displays with the assumption that these items can be called up when necessary.

This approach breaks down because of the *context catch*—in some contexts some of what is removed will be relevant. Data elements that appear to be less important on average can become a critical piece of evidence in a particular situation. But recognizing their relevance, finding them and integrating them in to the assessment of the situation becomes impossible if they have been excluded or pushed into the background of a virtual data world.

This finesse also breaks down because of the *keyhole catch*—it creates navigation burdens by proliferating more displays hidden behind the keyhole of the CRT screen (Woods and Watts, 1997). This occurs when scale reduction is applied to individual displays. Reducing the data available on individual displays pushes data onto more displays increasing demands for across display search and integration.

Ultimately, the irony of scale reduction as a finesse is that it runs counter to the technological trend – if one of the benefits of technology is more access to data, it is ironic that people have to throw away some of that access to cope with the complexity of trying to work with the available data.

(b) the global, static prioritization finesse—only show what is "important."

A related finesse is to select only the "important" subset of the available data. Often, the world of data is divided into two or three "levels of importance." Domain knowledge is used to assign individual data items to one of the two or three levels. All data items identified in the highest level of "importance" would be displayed in a more salient way to users. Data elements that fall into the second or third class of less important items would be successively less salient or more distant in the virtual world of the display system and user interface.

This approach also breaks down because of the *context catch*—how do we know what is important without taking context into account? Context sensitivity means that it is quite difficult to assign individual elements to a place along a single, static, global priority, or importance dimension. Inevitably, one is forced to make comparisons between quite disparate kinds of data and to focus on some kinds of situations and downplay others. Again, data items that are not important based on some overall criteria can be critical in particular situations.

This finesse, like the first, uses inhibitory selectivity, that is, they both, in effect, throw away data. In this case, developers will object saying that users can always call up data assigned to lower levels of importance if they feel they are relevant in a particular situation. But the problem is to help people recognize or explore what might be relevant to examine without already knowing that it is relevant. To aid this process requires one to consider perceptual organization, control of attention, and anomaly recognition as discussed earlier.

(c) the intelligent agent finesse—the machine will compute what is important for you.

Another version of the *context catch* plagues this approach—how does the machine know what is important without being able to take context into account?

However, this finesse also breaks down in the face of a new catch—the *clumsy automation catch.* The observer now has another data source/team member to deal with when they can least afford any new tasks or any more data (Sarter et al., 1997).

All intelligent agent algorithms, from agents programmed by practitioners specifically to flag data items to agents that "learn" rules from observing practitioners, are unable to escape the need to take context into account. The irony here is that developers believe that shifting the task to a computer somehow makes the cognitive challenges of focusing in on the relevant subset disappear. In fact, all finite cognitive processors face this challenge, whether they are an individual, a machine agent, a human-machine ensemble, or a team of people. It always takes cognitive work to find the significance in data.

For example, attempts in the mid-80's to make machine diagnostic systems handle dynamic processes ran into a data overload problem (these diagnostic systems monitored the actual data stream from multiple sensors). The diagnostic agents deployed their full diagnostic reasoning power in pursuit of every change in the input data streams (see Woods, Pople, and Roth, 1990; Roth, Woods and Pople, 1992; Woods, 1994). As a result, they immediately bogged down, dramatically failing to handle the massive amounts of data now available (previously, people mediated for the computer by selecting "significant" findings for the computer to process). To get the diagnostic systems to cope with data overload required creating a front end layer of processing that extracted, out of all of the changes, which events were "significant" findings that required initiating a line of diagnostic reasoning. In this case, determining what were significant events for diagnosis required determining what were unexpected changes (or an unexpected absence of a change) based on a model of what influences were thought to be acting on the underlying process.

(d) the syntactic finesse—use syntactic or statistical properties of text (e.g., word frequency counts) as cues to semantic content.

This finesse is relied on heavily in keyword search systems, web search engines, and information visualization algorithms that utilize "similarity" metrics based on statistical properties of the text (e.g., frequency counts of different content words) to place documents in a visual space (e.g., Morse & Lewis, 1997; Wise, Thomas, Pennock, Lantrip, Pottier, Schur, & Crow, 1996). The primary limitation of this approach is that syntactic and statistical properties of text provide a weak correlate to semantics and domain content. There is rarely a simple one to one relationship between terms and concepts. It is frequently the case that one term can have multiple meanings (e.g., Ariane is both a rocket launcher and a proper name; ESA stands for the European Space Agency, Environmental Services Association, and the Executive Suite Association) and that multiple terms can refer to the same concept (e.g., the terms "failed," "exploded," "was destroyed" can be used interchangeably).

The problem is compounded by the fact that the "relevance" metrics employed (e.g., the weighting schemes used by web search engines) are often opaque to the user. This is the *lack of observability* catch. The user sees the list of documents retrieved based on the query and the relevance weighting generated by the search engine. However, in many cases how the relevance weighting was generated is unclear, and the resulting document ordering does not accord well with how the user would have prioritized the documents (i.e., documents that come up early with a high weighting can be less relevant than documents that come up later.) This forces the user to resort to attempting to browse through the entire list. Since the generated list is often prohibitively long, it can leave the user unsure about whether important documents might be missed. Users will often prefer to browse documents ordered by metrics that do not attempt or claim to capture "relevance," such as date or source, rather than by syntactic relevance weighting because the organizing principle is observable and they know how to interpret values along those dimensions.

Attempts to place documents in a visual space based on syntactic properties are also subject to the *over-interpretation* catch. The spatial cues and relationships that are visible to the observer will be interpreted as meaningful even if they are incidental and not intended to be information bearing by the designer. For example, visualizations that attempt to represent multi-dimensional spaces (4 or more dimensions) on a two dimensional display can create ambiguities with respect to the position of a document relative to each of the dimensions. Users may assume that two documents that are located close to each other on the display reflect a similar degree of relationship to each of the dimensions represented in the space, when in fact they are not in the same position in the multi-dimensional space – even though it looks that way on the display. Similarly, information visualizations that attempt to reveal thematic relationships between documents through visual patterns are subject to over-interpretation. The visualizations can be dominated by patterns that are unimportant, such as missing data, and the underlying relationships may be distorted in the mapping to the perceptual field.

4.3 How Do People Find the Significance in Data Now?

How do people find the significance of data even though they are confronting an expanding field of data?

In many environments, there are artifacts, often quite traditional artifacts, that assist people. We find that artifacts that represent abstract data in a physical, spatially dedicated space help people cope with data overload to some degree. One example of this is traditional control centers. Ironically, the move to computerized control centers has created a massive data overload problem as the mechanisms, albeit crude ones, that supported the cognitive activities involved in finding the significance in data were removed and while the computerization led to an explosion in the number of displays that could be called up on one of a few CRT screens (see Woods and Watts, 1997 for a summary of this change). In other cases, we have found that users will tailor highly flexible devices to try to create a physically distributed workspace where individual types of data or data sources occur in one fixed position (Woods et al., 1994, chapter 5; Woods and Watts, 1997).

Another kind of artifact that seems to help people cope with data overload to some degree is event capture mechanisms. These are typically very crude mechanisms that indicate state changes or limit crossing such as annunciators in traditional control centers. Interestingly, work on data overload on the web has re-discovered the need for and re-created such basic event capture mechanisms as part of software agents that notify users when such events have occurred.

A different kind of coping strategy is found in the distribution of work. When confronted with the potential for data overload, organizations sometimes adopt or shift to a watch organization. In this case, people are assigned to monitor a portion of the overall data field (a subsystem or subfunction) reporting to supervisors who integrate reports from focused individuals or sub-teams. The most notable successful example of this is the structure of mission control for the Space Shuttle at Johnson Space Center (Patterson, Watts-Perotti, and Woods, in press, Watts et al., 1996). Coordination across people is an important component of how such work organizations are able to extract the significance from elemental data (e.g., Hutchins, 1990).

Different ways to organize work are more than mechanisms to cope with workload bottlenecks. The success of work organizations depends on and facilitates a build up of human experience and practice which provides people with the expertise to find the significance of data (usually) themselves with limited external support. This points to the most general coping strategy for data overload—human expertise and experience (what Norman, 1990a calls knowledge in the head). Ironically, the organizational changes underway today challenge coordination in the distribution of work as economic pressures reduce the investment in human expertise often while demanding more coordinated assessments and activities to deliver "just in time expertise."

4.4 Context-Sensitive Approaches

The promise of new technology is more than making data available. New technology does provide the power to develop external support for the cognitive activities involved in extracting the significance from data. The question is how to use that power. Ironically, this power can be used (and has been) to exacerbate data overload as well as to support people's ability to interpret large fields of data.

The diagnosis presented here points to criteria and constraints, in particular the central role of context sensitivity, that need to drive an innovation process. Using the basic human competence for finding what is informative in natural perceptual fields despite context sensitivity is another guide for innovation.

We will not attempt to lay out a complete set of techniques here. That work lies ahead of us in this project. However, one can already see the outlines of some of these techniques. For example, one family of techniques can be termed *sharpening* where local outposts of contextual data are used to compute aspects of what is relevant. These are areas where how context affects what is informative is well understood and robust.

Another family of techniques is concerned with *re-organization/re-representation*—more marks can reduce "clutter" if they produce a larger organization. One example of this is what Tufte (1990) terms micro/macro displays. These displays graphically combine two levels of analysis of the data field (for example, individual patients and the status of the sector in aeromedical evacuation planning; see Potter et al., 1996). A more elemental level is represented by individual graphic elements (patients in this example) which combine to produce an emergent structure that captures higher order properties (the state of the overall schedule and bottlenecks), while preserving access to the lower level elements (ability to trace an individual's status and itinerary).

Model-based representations are another direction. In this family, the semantics of the underlying processes or field of activity are used to help define the relationships that give data meaning (Vicente and Rasmussen, 1992). Related techniques would develop expectation based displays that highlight when events depart from expected or typical behavior and event based displays that capture the flow of events in the world at different levels of abstraction or in comparison to the expected flow of events (Potter and Woods, 1991).

Such techniques become the basis for developing pattern based displays and conceptual spaces that support people's abilities to explore spatially structured environments and recognize patterns across elements. For many kinds of data overload problems, there will be multiple organizing themes each of which defines a perspective on the field of data. Mechanisms to help users coordinate across a set of these perspectives will be needed.

However, all of these families of techniques have their own "catches," just as the finesses we explored earlier. Sharpening methods can fall prey to a completeness catch. New representations are subject to the catch of custom innovation--each is a unique creation tailored to a specific setting. Model-based methods to depict more than the base data are subject to an uncertainty catch--given high uncertainty in the data and significant consequences in possible outcomes, experts tend to revert to raw data, and the "right" model catch—how do you know the model that specifies how data is informative is appropriate for the task or situation? Expectation based displays are limited by the fact that it can be difficult to track/compute expectations about a process or about another agent.

All of these approaches hold promise for going beyond data availability to aid how a person extracts meaning from data. They demonstrate the need to develop ways to use the power of technology

- to enhance observability,
- to take into account context sensitivity, and
- to build conceptual spaces

These are some of the areas to make progress in if we are to escape the flood of data that technology has made available to all of us in so many settings.

PART III. THE INTELLIGENCE ANALYSIS VERSION OF DATA OVERLOAD

5. Intelligence Analysis Is a Challenging Version of Data Overload

Intelligence analysis in the United States is undergoing simultaneous organizational and technological change. Resulting from a shift in emphasis from the Cold War paradigm of monitoring a small number of countries for their ability to directly attack the United States to monitoring many more countries for a more diverse set of reasons (e.g., peacekeeping and humanitarian interventions), analysts are now being asked to cover a more diverse set of countries and technologies. At the same time, there have been reductions in both staffing and average years of experience. The net result is that intelligence analysts are increasingly asked to analyze situations that are outside their immediate base of expertise on shorter time horizons, increasing workload constraints and vulnerability to superficial or erroneous assessments.

In addition to this organizational backdrop, there are technological changes continuously underway. There has been a significant increase in the available amount of electronic data, particularly data that has not been generated specifically for intelligence analysis (e.g., information on the World Wide Web). This increase in data has generated interest in systems that are attempting to help analysts cope with the data, such as keyword search and browsing/filtering aids. These systems are also impacting cognitive workload—reducing some aspects while creating new types of cognitive work (e.g., managing lists of keywords that select incoming messages on a daily basis).

Given this situation, two characterizations of data overload and their associated solutions are relevant to intelligence analysis. The characterization of data overload as a workload bottleneck is a useful characterization. The increased workload demands have created bottlenecks that could potentially be alleviated by automation designed to coordinate its activities with the practitioner. For example, intelligent agents could organize available data for the analyst (e.g., by report quality). In addition, intelligent agents could critique the ongoing analysis process, for example, by tracking the "breadth" of the sampling of reports in relation to the available databases, detecting when sampling might be narrow, and suggesting broadening strategies.

Secondly, the characterization of data overload as a problem in finding the significance of data is clearly relevant to intelligence analysis. Given a reduction in the base of human expertise and reduced time to respond to analysis questions, it is important that the data be organized based on a model of domain semantics (e.g., ethnic group dynamics in sub-Saharan Africa or the structure of rocket launch programs for commercial and military satellite launch). This will allow analysts to better recognize unexpected, informative patterns and determine how an event is embedded in a contextual flow of other related events. Similarly, representation aids are needed that allow the practitioners to more quickly target the most relevant and profitable information and display the relationship of that information to other information that would corroborate it or conflict with it.

These kinds of displays have been termed model-based because they organize the suite of data displays on models of domain content and semantics rather than on properties of data per se (e.g., Woods, 1991; Vicente and Rasmussen, 1992).[6] It is important to remember that organizing data around models of the domain or field of practice is not the same as examining data with a pre-conceived solution in mind. Organizing data around fundamental relationships inherent in the process in question, rather than around data elements, is necessary, not optional, in the cognitive work needed to extract significance from data. While needed, the process can be flawed. People can carry out this cognitive process in their head without significant external assistance, or we can provide artifacts to assist in this process while remaining sensitive to how it can break down. One aspect of research in this area is concerned with how to provide robust functional models to use as the basis for display design (e.g., Flach et al., 1995).

While it is important to recognize the elements in data overload that are common across diverse operational settings in order to build upon existing research bases and design ideas, it is also important to identify the characteristics that are unique to intelligence analysis. There are several factors that complicate the practitioner's cognitive activities in intelligence analysis, as compared to those of practitioners in more heavily studied worlds. These include

- the kind of processes being monitored,
- the nature of the data available about the state of those processes, and
- the capabilities of the tools available to support analysis.

5.1 Monitoring Human/Organizational Processes

Cognitive engineering studies and designs generally have been addressed at practitioners who monitor and control engineered or sometimes physiological processes. In intelligence analysis, the underlying process that is monitored (what we refer to as the monitored process) is sometimes a technical process (e.g., communications network technology or the technology of specific weapons systems), but often consists of various kinds of human/organizational processes. For example, in analyzing events in one region of the world, an analyst may need to understand current and past ethnic group processes, alternative kinds of political processes—such as those

[6] They are also referred to as pattern-based, integrated, emergent property, or object displays in the literature. The term model based refers to the use of an organizing principle based on the structure, function, and activities in the underlying process.

of a theocracy—economic processes, geopolitical processes, and the development and implementation of military doctrine, to name just a few.

Adding such human/organizational processes to the mix leads us to consider the differences between different kinds of monitored processes. Figure 2 indicates that monitored processes can be loosely ordered on a dimension that describes how "definitive" we can be both in understanding and in modeling how the processes work. Figure 2 illustrates this dimension by ordering three classes of monitored processes: engineered, physiological, and human/organizational processes.

Engineered processes are physical systems that are designed and implemented by people, and are exemplified by such systems as the space shuttle, nuclear power plants, and military and commercial aircraft. These processes obey well understood physical laws. Physiological processes are self-tuning processes that exist naturally in the environment but can be altered by human intervention, as is the case in cardiovascular systems during open heart surgery. Human/organizational processes involve situations or activities in which groups of people interact, such as situations of low-intensity regional conflicts or activities involving supply logistics, economic behavior, or development and application of military doctrine. These processes may be defined or described by sets of rules, but these rules provide only a partial description of the actual behavior of people or organizations (e.g., for various reasons a military unit may deploy in a way inconsistent with standard doctrine).

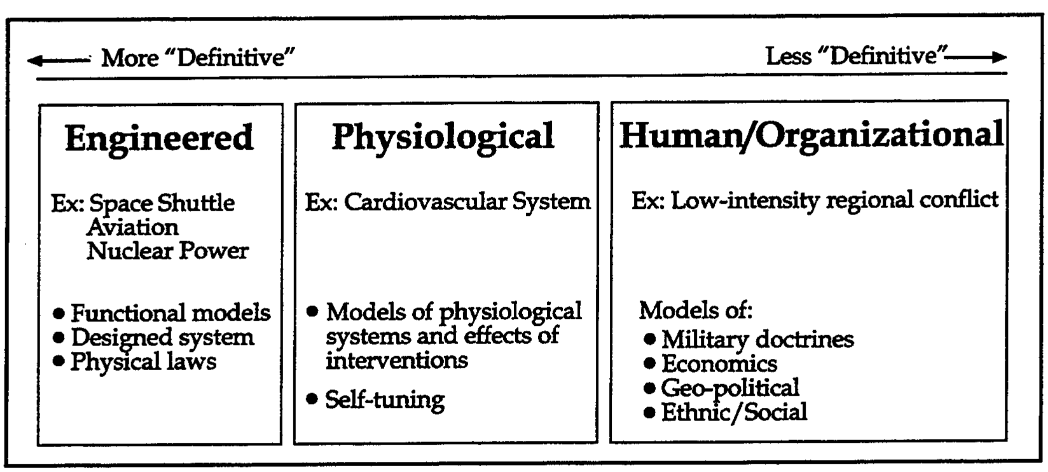


Figure 2. Different kinds of monitored processes can be ordered on a dimension of how "definitive" we can be in understanding, modeling, and predicting how that process works.

Highly "definitive" models, such as models of physical systems that were designed by people to accomplish certain goals, provide comparatively strong analytical frameworks because their component parts obey and are constrained by physical laws (e.g., heat exchangers always work a certain way functionally). Note that for all monitored processes, uncertainty and variability exist, but that the degree of uncertainty and variability changes as we move from less to more "definitive."

Many kinds of monitored processes can be relatively well-modeled at a functional level but are complex enough that many situations arise that are not predicted in advance. For example, regarding physiological systems, we know a great deal about the laws that govern such processes. However, we find that

- the models of physiological systems are not as detailed and accurate as those of the typical engineered process,
- the individual differences in physiological systems are larger between people than they are within analogous components of an engineered process (such as the variations found within examples of a particular model of aircraft),
- physiological processes have built in interactions and self-tuning control loops that are difficult to model completely.

In intelligence analysis, the models that are available to analysts are less "definitive" than the models available in engineered and physiological processes. Rather than a functional model, the frameworks available to analysts tend to be collections of heuristics and knowledge, such as how the military doctrine of a particular country's armed forces would influence behavior in a particular situation. These "models" are inherently less precise and support weaker predictions about actual behavior in specific situations. Yet these models are still very important, because the skilled use and application of these models is what is responsible for the recognizable differences in performance between more and less experienced analysts.

An additional complication in modeling human/organizational processes is that the division is less clear-cut between the "supervisory controller" and the "monitored process" given that the processes being monitored by intelligence analysts involve people. In engineered processes, for example, people are clearly outside of the processes to be monitored. Even in engineered processes, the roles of different people in the operational system can become quite complex in terms of scope of authority, supervisory control, and field of view. However, in discussing engineered processes, usually the confusion we try to guard against is ambiguity about the different roles different machines can play. The monitored process is technological, but we also now create machines that help us observe, evaluate, diagnose, and act on the monitored process. These support systems and automation are usually better seen as a part of the operational team along with the human monitors and supervisors (Billings, 1996). Similarly, with physiological processes the role of technology can be ambiguous: is it part of the process, (e.g., a programmable pacemaker), or is it part of the treatment team, (e.g., an infusion device)? But another potential complication emerges with physiological processes since the people are both the process being controlled and the controllers, (e.g., the patient, the physiological processes in question, can be part of the treatment process; see the case in Obradovich and Woods, 1996).

In the case of human/organizational processes, people, groups of people, or human organizations are active in every role. In an attempt to reduce the potential for confusion, Figure 3 provides a very rough schematic of the interacting roles when the monitored process is human/organizational. The figure contains three global roles (represented as the columns):

1. People in other parts of the world in various roles as part of economic, political, religious, ethnic, and military processes.
2. People in U.S. organizations in various roles as monitors of those processes (intelligence analysts) and as policy makers who decide about U.S. policies and actions in response to events in those parts of the world.
3. Investigators who try to understand the role of intelligence analysts and help shape new supporting tools to cope with issues like the potential for data overload.

The figure is tremendously oversimplified. There are other groups (e.g., humanitarian) and governments monitoring events in one part of the world that influence or shape the interactions. Governments may be watching and predicting how their people will behave (e.g., polls) or how different subgroups (e.g., constituencies) will react to different events, while outsiders may be monitoring how one group is anticipating how other groups will behave.

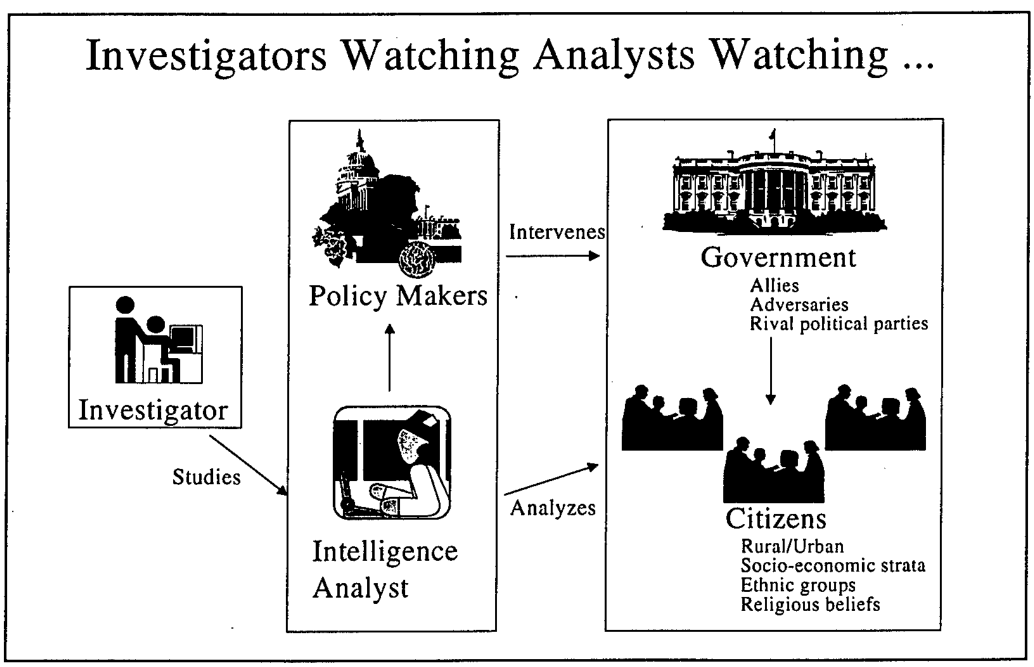


Figure 3. Analysts monitoring a human/organizational process.

5.2 The Nature of Data Available to Intelligence Analysts

In all of these different domains, the processes are monitored with data that is captured through "sensors." The nature of the data that is available is dependent on how the sensor information is processed, packaged, and displayed. In engineered and physiological processes, there are physical sensors placed at various points that monitor certain variables continuously. In general, the sensors always monitor the same thing in the same way and are displayed as single parameter sensor readings in dedicated locations (although there has been movement away from this one-sensor, one display organization with displays that integrate parameter values based on functional models of how the process works). In engineered processes, it is possible, though complicated, to define "nominal ranges" and signal an alarm when a parameter goes out of the nominal range. In physiological processes, it is also possible to point to possible limit values, but they function much more as landmarks or very general guidance because "significant" values depend so much on the patient and context. For example, what is too much or too little of some parameter for an individual may vary tremendously based on the stage of the surgical procedure, previous disease history, or relative to a baseline established for that particular individual at that particular time. In intelligence analysis the situation is even more difficult. It is not easy to flag abnormal data; indeed that may be part of the analysis process itself. It is often contentious what is an abnormal state, and even when it is not, there are currently no systems that can reliably recognize and flag textual descriptions of abnormal states.

When monitoring less definitive human/organizational processes, the "sensors" are more diffuse, with data about the process gathered remotely, indirectly, or by human observers on the scene. In human/organizational processes, when humans serve as the "sensors," the situation is actually better, in a sense. People can use their intelligence in terms of what variables to sample and what format is best to use to describe their observations. On the other hand, the data becomes more difficult to find and interpret because there is less consistency about what is sampled, how it is sampled, and where the information is displayed. In addition, there is the qualitative difference created by the fact that human/organizational processes are intentional systems. They can realize that they are being monitored and change their behavior or actively attempt to deceive observers. The observational sub-processes may, in fact, be specifically targeted for destruction, disruption, degradation, or denial.

Note that sensor data is not the only form of data available in any of these processes. Direct observation of the process, either by the supervisory controller or other agents in the distributed system, plays a role. In engineered processes, for example, controllers can directly touch a pipe to determine if it is hot. In physiological processes, anesthesiologists can look directly at the surgical field or check the color of the skin (e.g., if one notices the patient turning blue, then it is clear something is preventing adequate oxygenation of tissues). In intelligence analysis, agents can directly perceive information from satellite pictures or receive reports from agents who are dispersed to the area of interest to opportunistically perceive and report information.

In all of these domains, the reliability of the data is a critical concern. Physical sensors in engineered and physiological processes are uncertain indicators because they are placed in only a few locations; they are, in fact, model-based: the parameter of interest is often measured indirectly through other more tractable data, and they can fail. Data that is obtained through direct perception could also be unreliable: the observation relies upon the expertise and perceptual ability of the observer to identify subtle cues. In intelligence analysis, data comes in the form of reports created by humans who serve as the "sensors." The reports integrate a selection of data based on an interpretation and therefore need to be "unpacked" in order to identify the elemental data, which is used to generate an analysis product with a potentially different interpretation frame. People may bring a new set of reporting biases that create new forms of uncertainty. In addition, the difference between a normal state and an abnormal observation is contentious, and there is the added complication that the adversary in human/organizational processes may deliberately attempt to deceive the "sensors." As a result of the potential for unreliable data, similar strategies are observed in all of these different domains where data is cross-checked from independent sources in order to determine if the sensor is providing "invalid" data.

In intelligence analysis, data conflicts can be more subtle than in other domains due to the nature of the data. With engineered and physiological sensor data, there are concerns about effects being masked and sensors failing, but often the practitioners have the ability to check sensors on similar systems that are measuring the same information the same way and see if they agree. Intelligence analysts also employ a variation of this strategy, but it is more difficult to determine if information agrees because the information is often not measured the same way or at the same time and is not always identical in content. Analysts need to break down textual reports to a more elemental data level and then interpret the reports in order to determine the relationship of the data elements. When two or more independent sources give the same description of the same event, the information is more likely to be accurate. Schum (1987) refers to this as corroborative redundancy. When two or more sources provide information that inferentially favor the same hypothesis, this is referred to as convergent evidence which makes a particular hypothesis more likely. If information from two or more sources appear to corroborate or converge but stem from the same information source (e.g., a press release), there is no inferential value.

Conversely, items can be conflicting by saying logically opposing things or favoring different hypotheses. When information is discrepant, judgments of source quality are often important to decide what information to incorporate in the analysis. Factors that are considered in the credibility of a source include competency of the source to understand the issue at hand (e.g., Financial Times is a good source for financial information), predictable biases (e.g., self-reports by individuals or companies tend to be overly optimistic and less judgmental), and even attempts to actively deceive in the past (e.g., reports from countries with controlled media such as China might be publishing inaccurate accounts for political reasons).

Nevertheless, global judgments of source quality, such as "X is a trustworthy source," are under-specified, oversimplifications of how variations across sources play a role in the analysis process. Although certain sources are weighted as more credible than others based on past experience with a source, these judgments need to be tempered by other cues. Reports that are published immediately after the occurrence of an event often contain inaccuracies, although they tend to contain more detail than later reports. These reports are missing details that are provided in later updates – in other words, these reports contain "stale" information in relation to later reports. At the early stages, they are forced to speculate on causes without having all of the available information yet, and so there is a larger "hypothesis set" across reports as compared to later reports when there is more of a convergence on a small number of hypotheses. In addition, reports that are "distanced" from the original data are suspect. Having direct access to eyewitnesses, recorded data such as video, and telemetry data improves the quality of the analysis. In addition, having direct access to people who have interpreted the data in depth, such as the inquiry board after an accident investigation, is important. Reports of other reports suffer from the problems evidenced in the game "Telephone," where the story changes with each telling. This is exacerbated when the reports are translated from foreign languages. Finally, reports that are making predictions about future events are inherently uncertain, regardless of the competency of the person providing the prediction.

5.3 The Nature of Tools Available to Intelligence Analysts

As previously described and as we have observed in other domains, ongoing technological and organizational changes are fundamentally changing the task of intelligence analysis. As a result of data being available more in electronic media, shorter timelines, and a broader range of analytical responsibilities, it is becoming increasingly difficult, if not impossible, for analysts to read all of the potentially pertinent individual messages and potentially relevant reports necessary to do an analysis. In this new situation, the analyst now needs to search through an electronic data-base/document-base in order to identify relevant information. This is the "new world of data" that has begun to emerge for analysts, and therefore the nature of the tools available to intelligence analysts need to be somewhat different in nature than tools designed for real-time monitoring of sensor data in engineered and physiological processes.

The main complication introduced by this new situation is the relationship between events in the world, database(s) of electronic information about events in the world, and sampled information about events in the world (Figure 4). The intelligence analyst rarely directly observes events in the world. Rather, other humans generate reports about events in the world. These reports make up a set of databases whose characteristics are often opaque to the analyst, particularly since the available information is constantly being updated and the information is generally not indexed. Information is "sampled" from these databases, first by keyword search queries and then by browsing dates and titles through the computer "keyhole," a small CRT screen. The relationship of the sample to the database is generally not available to the analyst (although some ways to characterize the database are being developed that could be used to determine the relationship, e.g., Wise et al., 1996). How does an analyst know if (s)he has read all of the available relevant information or if the information that is retrieved by a keyword search is high quality in comparison to what is available? How does an analyst know what information in the database is contradicted or corroborated by other information in the database?

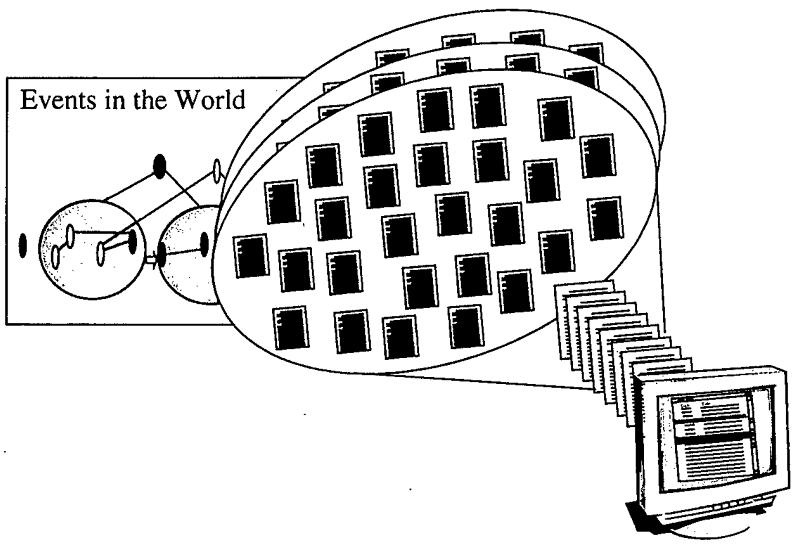


Figure 4. The analyst's new world as information sampling through a computer "keyhole."

A complicating factor in the search for information is that the report is not an elemental data unit. Intelligence analysts do not make judgments of how information is related at the level of the report. Instead, those judgments occur about selected descriptions taken from reports (Figure 5). The search and retrieval tools available to analysts return "bundles" at the report level, not at the level of selected descriptions within reports. There is no easy way for analysts to search for information that will corroborate a selected description at that level. Analysts would need to look for the selected description in all of the returned reports manually because the date and title information is unlikely to provide clues about the information at the level of a selected description. This process makes it particularly difficult for analysts to know when information about a topic has been updated or changed without reading all of the available documents.

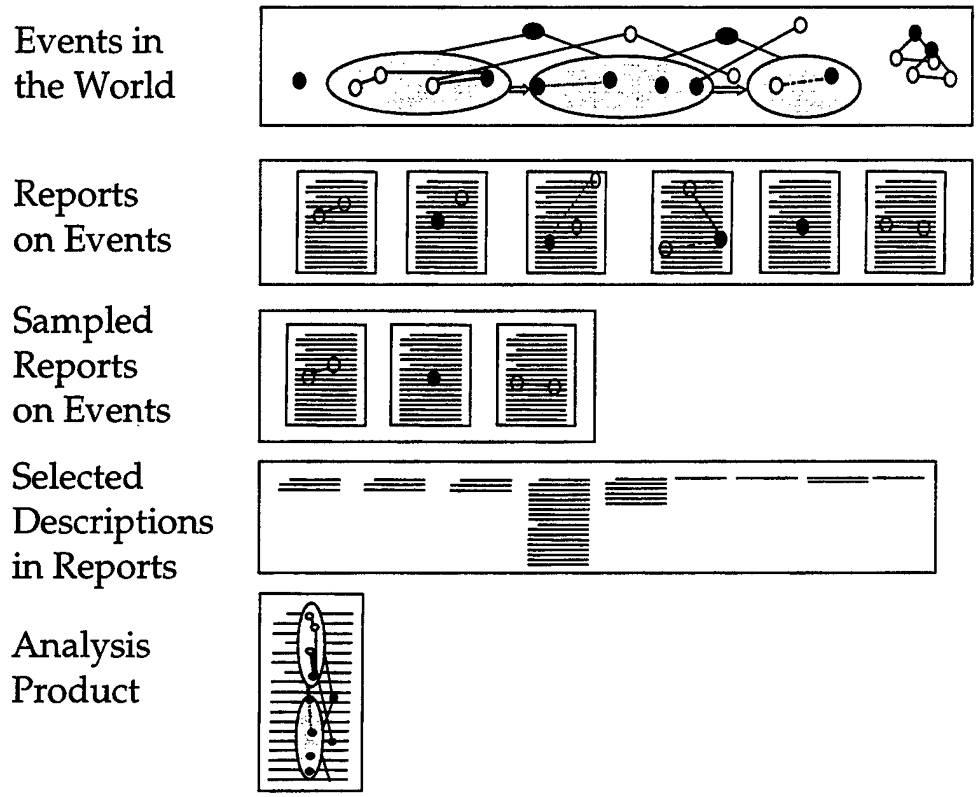


Figure 5. Sequence of information "bundles" in the analytical process.

Figure 5 gives an abstract view of how data is manipulated during the analytical process. Events occurring in the world are represented as textual descriptions within reports. These reports partially overlap and are distorted by the interpretation of the reporters on what the event was in relation to past, present, and future contexts. An analyst samples a subset of the available reports using keyword search and browsing mechanisms. The analyst then must break down the report into smaller units in order to compare whether descriptions in different reports are corroborating or discrepant along various dimensions. The corroborated descriptions are then incorporated into a coherent story, the analysis product, based on an interpretive frame provided by the analyst.

One can imagine a variety of tools which could better support these levels of data manipulations, such as:

- model-based information visualization tools to characterize the database (e.g., event-based displays)
- display of indications of report/source quality factors (e.g., "distance" from the primary data sources, temporal relation to event landmarks, report length)
- targeted support/critiquing systems to broaden data sampling
- targeted support/critiquing systems to aid the construction of a coherent story
- targeted support/critiquing systems for tracking conflicts in the data

We have observed in other contexts that the technology for visualization and software agents is necessary but not sufficient to create useful systems for practitioners in a work setting. Our and others' research on the use of technological powers have shown that computer technology for supporting data base search through visualization and autonomous software agents can be deployed skillfully or clumsily. The new world of data could overwhelm analysts with options for searching and viewing data in reports in the data/document base that their attention is focused more on the interface capabilities and less on the analysis task. On the other hand, the new world of data offers more possibilities for aiding the analysts as they build a picture of events in some area of the world or on some issue of concern.

Based on our diagnosis of what makes data overload hard, we see that both the characterizations of data overload as a workload bottleneck and finding the significance of data are relevant to the intelligence analysis setting. In designing solutions for data overload in intelligence analysis, complications stemming from the kinds of processes that are being monitored, the nature of the available data, and the capabilities of the available tools need to be considered. Research efforts are underway to extract the relationships, events, and contrasts that are informative to intelligence analysts in an identified scenario (e.g., Ariane 501 launcher failure) based on the results of a simulation study with experienced analysts (Patterson, Roth, and Woods, in preparation). This case provides a forum to demonstrate more tangibly what some of the concepts for addressing data overload might look like for a realistic topic and event. The relationships, events, and contrasts that are informative in this case illustrate the importance of context in finding the significance of data and illustrate techniques, such as depicting relationships in a conceptual space, which support the cognitive system process of analysis.

REFERENCES

Billings, C. E. (1996). *Aviation automation: The search for a human-centered approach.* Hillsdale, NJ: Lawrence Erlbaum.

Bower G. & Morrow, D. G. (1990). Mental models in narrative comprehension. *Science, 24,* 44-48.

Bruner, J. (1990). *Acts of meaning.* Cambridge MA: Harvard University Press.

Flach, J., Hancock, P., Caird, J. & Vicente, K. (Eds.) (1995). *An ecological approach to human machine systems I: A global perspective.* Hillsdale, NJ: Lawrence Erlbaum.

Folk, C. L., Remington, R. W. & Johnston, J. J. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance, 18*(4), 1030-1044.

Glymour, C. (1987). *Discovering causal structure: Artificial intelligence, philosophy of science, and statistical modeling.* Orlando: Academic Press.

Goldmeier, E. (1982). *The memory trace: Its formation and its fate.* Hillsdale, N.J.: Lawrence Erlbaum.

Hanson, N. R. (1958). *Patterns of discovery.* Cambridge England: Cambridge University Press.

Hardwick, M. (1986). *The complete guide to Sherlock Holmes,* pp. 86-87. New York: St. Martin's Press.

Hochberg, J. (1986). Representation of motion and space in video and cinematic displays. In K. Boff, L. Kaufman, and J. Thomas (Eds.), *Handbook of perception and human performance* (pp. 22-1 to 22-64). New York: John Wiley & Sons.

Hutchins, E. (1990). The technology of team navigation. In J. Galegher, R. Kraut and C. Egido (Eds.), *Intellectual teamwork: Social and technological foundations of cooperative work.* Hillsdale, NJ: Lawrence Erlbaum.

James, W. (1890/1981). *The principles of psychology.* New York: H. Holt & Company.

Joyce, J. P. & Lapinski, G. W. (1983). A history and overview of the safety parameter display system concept. *IEEE Transactions on Nuclear Science,* NS-30.

Kemeny, J. G. (1979). *Report of the President's Commission on the Accident at Three Mile Island, The Need for Change: The Legacy of TMI.* New York: Pergamon Press.

Morse, E & Lewis, M. (1997). Why information retrieval visualizations sometimes fail. In *Proceedings of the 1997 IEEE International Conference on Systems, Man, and Cybernetics*, Oct. 12-15, 1997, pg. 1680 - 1685.

Murray, C. & Cox, C. B. (1989). *Apollo: The race to the moon*. New York: Simon & Schuster.

Norman, D.A. (1990a). *The design of everyday things*. New York: Doubleday

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

Obradovich, J. H. & Woods, D. D. (1996). Users as designers: How people cope with poor HCI design in computer-based medical devices. *Human Factors, 38*(4), 574-592.

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology, 46*, 461-488.

Patterson, E. S., Roth, E. M., & Woods, D. D. (in preparation). A simulation study of computer-supported inferential analysis under data overload.

Patterson, E. S., Watts-Perotti, J., & Woods, D. D. (in press). Voice loops as coordination aids in space shuttle mission control. *Computer Supported Cooperative Work.*

Patterson, E. S., Woods, D. D., Sarter, N. B., & Watts-Perotti, J. (1998). Patterns in cooperative cognition. *Proceedings of COOP '98, Third International Conference on the Design of Cooperative Systems*. Cannes, France, 26-29 May.

Potter, S.S., Bell, R.W., & Elm, W.C. (1996). Supporting Aeromedical Evacuation Planning Through Information Visualization. *Proceedings of Human Interaction with Complex Systems*. Los Alamitos, CA: IEEE Computer Society Press.

Potter, S. S. and Woods, D. D. (1991). Event-driven timeline displays: Beyond message lists in human-intelligent system interaction. In *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*. IEEE.

Rabbitt, P. (1984). The control of attention in visual search. In R. Parasuraman and D. R. Davies (Eds.), *Varieties of Attention.* New York: Academic Press.

Rasmussen, J. (1985). Trends in human reliability analysis. *Ergonomics, 28*(8), 1185-1196.

Rasmussen, J., Pejtersen, A. M. & Goldstein, L. P. (1994). *Cognitive systems engineering.* New York: John Wiley & Sons.

Reed, E. S. (1988). *James J. Gibson and the psychology of perception.* New Haven, CT: Yale University Press.

Roth, E. M., Woods D. D., & Pople, H. E. Jr. (1992). Cognitive simulation as a tool for cognitive task analysis. *Ergonomics, 35,* 1163–1198.

Sarter, N. & Woods, D. D. (in press). Teamplay with a powerful and independent agent: A corpus of operational experiences and automation surprises on the Airbus A-320. *Human Factors, 39,* 553—569.

Sarter, N., Woods, D. D. & Billings, C. (1997). Automation surprises. In G. Salvendy, (Ed.), *Handbook of human factors/ergonomics* (2nd ed.). New York: John Wiley & Sons.

Schum, D. A. (1987). *Evidence and inference for the intelligence analyst, Volume I.* Lanham, MD: University Press of America.

Simon, H. A. (1976). *Administrative behavior* (3rd ed.). New York: Free Press.

Simon, H. A. (1981). *Sciences of the artificial* (2nd ed.). Cambridge, MA: MIT Press.

Tufte, E. R. (1990). *Envisioning information.* Chesire, CT: Graphics Press.

Tukey, J.W. (1977). *Exploring data analysis.* Reading MA: Addison-Wesley.

Vicente, K. J. & Rasmussen, J. (1992). Ecological interface design: Theoretical foundations. *IEEE Transactions on Systems, Man, and Cybernetics, 22*(4), 589-606.

Watts, J. C., Woods, D. D., Corban, J. M., Patterson, E. S., Kerr, R., & Hicks, L. (1996). Voice loops as cooperative aids in space shuttle mission control. In *Proceedings of computer-supported cooperative work.* Boston MA.

Wertheimer, M. (1923/1950). Laws of organization in perceptual forms. In W. D. Ellis (Ed.), *A Source book of gestalt psychology.* New York: Humanities Press.

Winograd, T. & Woods, D. D. (July 1997). Challenges for human-centered design. In J. Flanagan, T. Huang, P. Jones, and S. Kasif, (Eds.), *Human-centered systems: Information, interactivity, and intelligence.* Washington, DC: National Science Foundation.

Wise, J. A., Thomas, J. J., Pennock, K., Lantrip, D., Pottier, M., Schur, A., & Crow, V. (1996) Visualizing the non-visual: spatial analysis and interaction with information from text documents. *Proceedings of Info Viz 96.*

Wolfe, J. M. (1992). The parallel guidance of visual attention. *Current directions in psychological science, 1,* 124-128.

Woods, D. D. (1984). Visual momentum: A concept to improve the cognitive coupling of person and computer. *International Journal of Man-Machine Studies, 21,* 229-244.

Woods, D. D. (1991). The cognitive engineering of problem representations. In G. R. S. Weir and J. L. Alty, (Eds.), *Human-computer interaction and complex systems.* London: Academic Press.

Woods, D. D. (1993). The price of flexibility in intelligent interfaces. *Knowledge-Based Systems, 6,* 1-8.

Woods, D. D. (1994). Cognitive demands and activities in dynamic fault management: abduction and disturbance management. In N. Stanton (Ed.), *Human factors of alarm design.* London: Taylor & Francis.

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

Woods, D. D. (1995b). The alarm problem and directed attention in dynamic fault management. *Ergonomics, 38*(11), 2371-2393.

Woods, D. D., Johannesen, L., Cook, R. I. & Sarter, N. B. (1994). *Behind Human Error: Cognitive Systems, Computers and Hindsight.* Wright-Patterson AFB, OH: Crew Systems Ergonomic Information and Analysis Center (CSERIAC).

Woods, D. D., Patterson, E. S., Roth, E. M., & Redenbarger, W. J. (1998). *Aiding the Intelligence Analyst in Situations of Data Overload: The Zairean Civil War Scenario.* (Institute for Ergonomics/Cognitive Systems Engineering Laboratory Report, ERGO-CSEL 98-TR-01). Columbus OH: The Ohio State University, January 1998. (Prepared for Air Force Research Laboratory Crew System Integration Branch [AFRL/HECA], WPAFB, OH.)

Woods, D. D., Pople, H.E. Jr. & Roth, E. M. (1990). *The Cognitive Environment Simulation as a Tool for Modeling Human Performance and Reliability,Volumes I & II* (Technical Report NUREG-CR-5213). Washington D.C.: U.S. Nuclear Regulatory Commission.

Woods, D. D. & Sarter, N. (1993). Evaluating the impact of new technology on human-machine cooperation. In J. Wise, V. D. Hopkin, and P. Stager, (Eds.), *Verification and validation of complex systems: Human factors issues.* Berlin: Springer-Verlag.

Woods, D. D. & Sarter, N. (in press). Learning from automation surprises and going sour accidents. In N. Sarter and R. Amalberti (Eds.), *Cognitive engineering in the aviation domain.* Hillsdale NJ: Lawrence Erlbaum

Woods, D. D. & Watts, J. C. (1997). How not to have to navigate through too many displays. In M. G. Helander, T. K. Landauer, and P. Prabhu (Eds.) *Handbook of human-computer interaction,* (2nd ed.). Amsterdam: Elsevier Science.

Zhang, J. & Norman, D.A. (1994). Representations in distributed cognitive tasks. *Cognitive Science, 18,* 87–122.