

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

David D. Woods
Emily S. Patterson
Institute for Ergonomics
Cognitive Systems Engineering Laboratory
The Ohio State University
Columbus, OH

Emilie M. Roth
Roth Cognitive Engineering
Boston, MA

Klaus Christoffersen
Institute for Ergonomics
Cognitive Systems Engineering Laboratory
The Ohio State University
Columbus, OH

Data overload is a generic and tremendously difficult problem. We examine three characterizations of the data overload problem and how they lead to different proposed solutions. The first characterization is a clutter problem where there is "too much stuff," which leads to proposals to reduce the number of data units that are displayed. The second characterization is a workload bottleneck where there is too much data to process in the time available. Data overload as a workload bottleneck shifts the view to practitioner activities rather than elemental data and leads to proposals to use automation to perform activities for the practitioner. The third characterization is a problem in finding the significance of data when it not known a priori what data will be informative. This characterization leads to model-based abstractions and representation design techniques as potential solutions. Many of the existing approaches to coping with data overload avoid directly confronting the problem that what is significant depends on context. We advocate an alternative approach that depends on model-based organization of the data in a conceptual space that depicts the relationships, events, and contrasts that are informative in a field of practice and uses active machine intelligence in circumscribed, cooperative roles to aid human observers in organizing, selecting, managing, and interpreting data.

CHARACTERIZATIONS OF DATA OVERLOAD

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, 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.

There are three basic ways that the data overload problem has been characterized (Woods, Patterson, and Roth, 1998):

1. As a clutter problem where there is *too much data*: therefore, we can solve data overload by reducing the number of data units that are displayed. This has not proven to be a fruitful direction in solving data overload because it misrepresents the design problem, is based on erroneous assumptions about how human perception and cognition work, and is incapable of dealing with the context sensitivity problem -- in some contexts, some of what is removed will be the relevant data.
2. As a *workload bottleneck* where there is too much to process 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. This is a potentially useful way to think about data overload, but the findings clearly show that automation support is necessary but not sufficient to create useful systems. Introducing autonomous machine agents changes the cooperative structure creating new roles, new knowledge requirements, new judgments, new demands for attention, and new coordinative activities. The automation must be directable and observable in order to avoid patterns of coordination breakdowns such as clumsy automation and automation surprises (Patterson, Woods, Sarter, and Watts-Perotti, 1998).

3. As a problem in *finding the significance of 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). Machine intelligence is used to better organize the data to help people extract meaning despite that fact that what is informative depends on context. 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*. Characterizing the data overload problem in this way will allow us to directly address the difficult challenges in creating truly useful solutions.

THE CONTEXT-SENSITIVITY PROBLEM

The key reason that the task of finding the significance of data is so difficult is what we refer to as the context-sensitivity problem. That is, the information carried by data is not absolute – it depends on the context in which it occurs, including:

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

Appreciating the effects of context-sensitivity is critical to a deep understanding of data overload. If people already know which data are relevant to the current context and how they relate, the problem reduces to navigating or searching within the virtual data field to find the required values. Supporting this process is an important and potentially difficult design challenge (see Woods and Watts, 1997), and failing to do so can significantly increase the operational impacts of

data overload. However, the central problem of data overload is not how to enable people to find specific data. Rather, the key question is how to help observers to understand which portions of the data space need to be examined given a dynamic operational context.

HOW CURRENT METHODS FINESSE CONTEXT SENSITIVITY

Many of the existing approaches to coping with data overload have applied various “finesses” to avoid directly confronting the context-sensitivity problem (Woods, Patterson, and Roth, 1998). In one sense, these finesses represent positive pragmatic adaptations to difficulty – they reduce data overload problems to dimensions where experienced people can often continue to function effectively. However, finesses are limited because they represent workarounds rather than directly dealing with the factors that make it difficult for people to extract meaning from data. All of these approaches involve a form of pre-defining which data within the data field are most important to present to observers. For example, techniques which attempt to filter the available data or present it in prioritized lists rely on a model specifying how to filter or alter the relative salience of data. These models tend to be static, based on global considerations of priority or importance. They are not designed to be responsive to subtle or novel variations in the problem environment which may significantly alter the actual importance of different data.

Another class of techniques finesse the context sensitivity problem by organizing data along syntactically generated dimensions which purport to correlate with “similarity” or “relevance” (e.g., as in most search engines for the World Wide Web). Generally, these cues are statistical properties of text that are not linked to domain- or task-specific models. The primary limitation of this approach is that syntactic and statistical properties of text provide at best a weak correlate to semantics. In addition, the dimensions are often algorithmic combinations of a multitude of factors which are not observable to the practitioner and do not accord well with how they would have organized the data themselves.

Contrary to what appears to be an implicit belief on the part of many designers, introducing intelligent machine agents to monitor and process data does not make the context-sensitivity problem go away. All such designs rely on a model of the relationship between the problem context and the significance of data in order to decide where to focus processing resources. These models can be more or less sophisticated, but they can never be more than approximate for any non-trivial setting. This means that any machine problem solver is

potentially brittle; it can break down when circumstances go beyond the set explicitly considered by designers. In order to create systems which are robust under conditions of data overload, we must recognize and design to account for this limitation. This is not an argument against using machine processing, but it means we must deploy technology in ways such that the result is cooperative human-machine systems which are able to successfully adapt to unanticipated conditions (Billings, 1996).

TOWARDS CONTEXT-SENSITIVE METHODS

The preceding discussion leads to the conclusion that, because of the context-sensitivity problem, we must direct our efforts towards techniques which do not rely on knowing in advance what the relevant subset of data is. We have also argued that methods which rely centrally on machine processing are vulnerable to brittleness as a result of the context-sensitivity of data. This has led us to cast the problem as one of helping people to recognize or explore the portions of the data field that might be relevant for the current context so that they can focus attention on those areas.

Our approach involves two parallel strategies. The first is to use models of the domain semantics as the foundation for visualizations which provide a structured view of the data field for observers. The intent is to take advantage of the context-sensitive properties of human cognition by giving observers the perceptual leverage needed to focus in on relevant sub-portions of the data space. The second strategy is to use active machine intelligence in supplemental, cooperative roles to aid human observers in organizing, selecting, managing, and interpreting data.

Model-Based Visualizations

This tactic is similar in philosophy to other methods in the literature which use models of domain semantics as a way to structure displays of data (e.g., Vicente and Rasmussen, 1992). Taking advantage of the context-sensitive nature of human cognition presumes a structured data field on which our attentional processes operate. The idea therefore is to build a conceptual space for organizing the data based on a model of the fundamental relationships, objects, and events in the domain. In order to support skillful shifting of attention, these visualizations will have to include mechanisms allowing observers to perceive changes or potentially interesting conditions which are not necessarily in direct view, and to re-orient their attention to the new data. They must also emphasize anomalies and contrasts by showing how data departs from or conforms to expectations.

Because of the central importance of models in this method, it is important to develop a framework of models which can be used to understand different settings and to serve as the basis for choosing and instantiating appropriate classes of visualizations. Levels of the framework will vary in terms of how closely bound they are to a particular domain or scenario. Although there will be models that will generalize across all domains, such as models of events, a context-bound approach implies that some models will be instantiated that are specific to the domain and/or scenario, such as the data signature for a leak in a particular mechanical system on the Shuttle during take-off.

The characteristics of different application domains will impact the utility of the models that are employed in the design of visualizations to address the data overload problem. Monitored processes in supervisory control settings 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 1 illustrates this dimension by ordering three broad 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 or 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).

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 uncertainty and variability increases as we move from more to less "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.

For human/organizational processes, the models 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 of these models is what is responsible for the recognizable differences in performance between more and less experienced analysts.

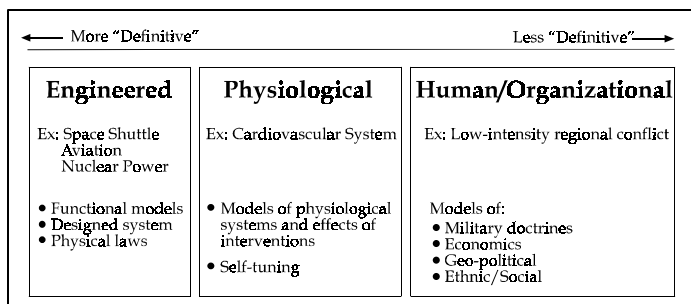


Figure 1. Nature of models in different domains

Cooperative Roles for Machine Intelligence

One prevalent view in the design of human-machine systems is to “allocate functions” between machine and human agents. With this approach, designers must choose which functions are better performed by machines or humans and tasks are assigned accordingly. This framing often leads either to over-commitment to an immature machine intelligence or over-reliance on human expertise. The philosophical underpinnings of Cognitive Systems Engineering have redefined this

debate of whether to trust the machine or the human as an issue of coordination between team players, where neither “does it all,” both are limited resource processors, and both are subject to brittleness in processing. This is not a rejection of technology, but rather a redefinition of how technology and people should interact as team players to make a system that is more robust than the individual elements.

There are several cooperative roles that machine intelligence could play in our approach that relax the need for machine intelligence to always be correct. One role is for active intelligence to allow people to interact with data at a high level of manipulation, thereby making it easier to interact with the computer, or “close the gulf of execution.” With this role, the user could initiate an interaction by marking data with a summary judgment such as “conflicting” which the machine intelligence could then use to organize the data field given an “understanding” of what the label implies. Another is for the intelligence to structure the data along dimensions that have been identified to be heuristically useful. For example, rather than having the machine intelligence sort by a summary, integrative judgment such as source quality, relevance, or similarity, the dimensions that make up those judgments could be used to organize the display of the data. With textual databases, for example, the computer could display documents along the dimensions of sources, dates, word counts and keywords so the machine can identify the tractable contributors to a high level judgment that the person can make, trading off the tensions and gaps among the contributors in a context-sensitive way. In addition, machine intelligence can play an active role as a critiquer to the human agent by suggesting ways to broaden the exploration of the data space and alternative hypotheses to explain the data.

CRITERIA FOR SUCCESS

Based on our previous research in this area (Patterson, 1999), we have identified a number of criteria that successful responses to the data overload problem need to satisfy. These criteria serve not only to guide design but are also useful in generating scenarios to test the effectiveness of proposed designs.

Broadening. First, solutions designed to deal with data overload should broaden the search for or recognition of pertinent information, break fixations on single hypotheses, and widen the hypothesis set that is considered to explain the available data. It is a general finding in inferential analysis tasks such as diagnosis that there is the potential for premature closure during the analysis process. This vulnerability is increased in data overload conditions because it is easy to miss

information that would challenge a leading hypothesis because it is hidden within a data field that is too large to be exhaustively browsed. When the process is prematurely closed, the potential is increased for erroneous or incomplete situation assessments.

Recognition of Unexpected Information. Second, solutions should bring practitioners' attention to highly informative or definitive data and relationships between data, even when the practitioners do not know to look for that data explicitly. Informative data includes data that deviates from expectations, data that eliminates potential hypotheses, and data that contradicts other data. A particularly difficult criteria to meet that should be designed into evaluation scenarios is to help practitioners recognize information updates that overturn or conflict with previous information.

Management of Uncertainty. Third, solutions should aid practitioners in managing data uncertainty. In particular, solutions should help practitioners identify, track, and revise judgments about data conflicts. A variety of breakdowns have been observed in previous research in identifying conflicts, tracking the conflicts, and revising judgments about the data in the face of new information.

These evaluation criteria are interesting, in part, because they are so difficult to address. They highlight the difficulties in helping people to find the significance of data when it is not known a priori what will be informative for the context. It is unlikely that simple, straightforward adjustments or feature additions to existing context-free approaches will address these criteria. In order to make progress on solving data overload, we must generate innovative design concepts that depict the informative relationships, events, and contrasts for a field of practice in a conceptual space in a way that is not dependent on brittle machine intelligence.

ACKNOWLEDGMENTS

This research was sponsored by the Air Force Research Laboratory Human Effectiveness Directorate, with exceptional support from Mr. Gil Kuperman, and by NASA under grants NAG-9-786 and NAG-9-1005. Additional funding was provided by a National Science Foundation Graduate Fellowship. Any opinions, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

- Billings, C. E. (1996). *Aviation Automation: The Search For A Human-Centered Approach*. Hillsdale, NJ: Lawrence Erlbaum.
- 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.
- 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.
- Patterson, E.S. (1999). A simulation study of computer-supported inferential analysis under data overload. *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting*.
- 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.
- Vicente, K. J. and Rasmussen, J. (1992). Ecological interface design: Theoretical foundations. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(4), 589-606.
- Woods, D.D., Patterson, E.S., and Roth, E.M. (1998). Aiding the intelligence analyst in situations of data overload: a diagnosis of data overload. *Institute for Ergonomics/Cognitive Systems Engineering Laboratory Report, ERGO-CSEL 98-TR-03*, The Ohio State University, Columbus OH.
- 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. (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. and Watts, J. C. (1997). How not to have to navigate through too many displays. In Helander, M. G., Landauer, T. K. and Prabhu, P. (Eds.) *Handbook of Human-Computer Interaction*, 2nd edition. Amsterdam: Elsevier Science.
- Zhang, J. & Norman, D.A. (1994). Representations in distributed cognitive tasks. *Cognitive Science*, 18, 87-122.