USING COGNITIVE TASK ANALYSIS (CTA) TO SEED DESIGN CONCEPTS FOR INTELLIGENCE ANALYSTS UNDER DATA OVERLOAD

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

Emilie M. Roth
Roth Cognitive Engineering

This paper describes how a Cognitive Task Analysis (CTA) was used to jumpstart the exploration of useful design aids to combat data overload in intelligence analysis. During a simulated analysis task, we observed how professional intelligence analysts were vulnerable to making inaccurate statements when they were under time pressure and working in a topic area outside their area of expertise. From these observations, we generated design recommendations and criteria for evaluating the usefulness of any effort aimed at reducing data overload. Then, we used CTA insights to trigger the development of modular design concepts, or "design seeds," that leverage advances in machine processing to address vulnerabilities. Nine design seeds were integrated into a "Visual Narratives" workspace visualization concept. Feedback about the usefulness of the design seeds was obtained during an elicitation session following an animated fly-through, or "Ani-mock," demonstration.

BACKGROUND

This report explains how we used findings from a Cognitive Task Analysis (CTA) to jumpstart the exploration of useful design aids to combat data overload in intelligence analysis. Data overload is a fundamental, ubiquitous problem. In intelligence analysis, as in many other domains, the data overload problem is expected to become even more difficult due to the explosion of accessible electronic data and substantial reductions in staff. As intelligence agencies transfer from Cold War paradigms of monitoring a small number of adversarial countries to tracking large numbers of developing "hot spots" and supporting peacekeeping operations, analysts are increasingly stepping outside their areas of expertise to respond quickly to targeted questions.

FRAMING THE PROBLEM

The challenge presented to the team was to develop information visualizations that would help intelligence analysts deal with the avalanche of electronic, textual data that is available for analysis. Specifically, there was interest in helping analysts find appropriate information to formulate answers to questions outside their immediate expertise under a tight deadline such as 24 hours.

In order to provide information visualizations that would be useful as well as usable, an important first step was to identify the abstract properties of the intelligence analysis domain that would allow us to leverage previous research:

- Similar to traditional human factors domains, intelligence analysis is conducted in a complex, dynamic, high consequence, socio-technical setting, requires significant expertise, and is supported by specialized, computerized artifacts.
- Intelligence analysis is an instantiation of inferential analysis, which is defined as determining the best explanation for uncertain, often contradictory and incomplete data. Inferential analysis uses abductive reasoning (Peirce, 1903) in the sense that a conclusion is inherently contestable due to uncertainty, but that experts agree that certain conclusions have higher warrant than others.
- An intelligence analyst is analogous to a supervisory controller, although there are several important distinctions. Rather than monitoring an engineered system, the analyst monitors a mix of technological and human/organizational processes. As a result, the monitored system adapts based on knowing it is observed, and diagnostic interventions can be difficult to conduct. In addition, the process data is mostly freeform text rather than parameter data, which makes it difficult to alarm setpoint crossings.
- We defined data overload to be a condition where a domain practitioner, supported by artifacts and other human agents, finds it extremely challenging to focus in on, assemble, and synthesize the significant subset of data for the problem context into a coherent assessment of a situation, where the subset of data is a small portion of a vast data field. The starting point for this definition was recognizing that large amounts of potentially available data stressed one kind of cognitive activity: focusing in on the relevant or

interesting subset of data for the current problem context.

CONSTRAINTS ON EFFECTIVE SOLUTIONS

Based on previous research in supervisory control, we "diagnosed" the data overload problem and developed domain-independent constraints on effective solutions (Woods *et al.*, 1999).

1. All approaches to data overload involve some sense of selectivity.
2. Organization precedes selectivity.
3. All techniques to cope with data overload must deal with context sensitivity.
4. Observability is more than mere data availability.
5. To cope with data overload, ultimately, will require the design of conceptual spaces.

CTA IDENTIFIES LEVERAGE POINTS

Although we feel that the constraints on effective solutions to data overload are relevant to intelligence analysis, we conducted a Cognitive Task Analysis (CTA) to target design efforts at leverage points. Leverage points were selected based on a calibrated understanding of the most prominent aspects of the data overload problem, the format and representation of data, the unique cognitive challenges and strategies that are employed to address those challenges in a work setting, the nature of tools available to practitioners, the risks of tradeoff options when goals conflict, and the consequences of particular types of failures.

The CTA was developed from a study with the target of: inferential analysis conducted by experienced analysts under data overload, on tight deadlines, and outside their immediate bases of expertise. The test that was designed to map to this target was:

1. ten professional intelligence analysts, ranging from 7 to 30 years of analytic experience,
2. analyzing a face valid task in 3-4 hours that they had not previously analyzed and was not in the immediate base of expertise: the cause and impacts of the June 4, 1996, Ariane 501 rocket launch failure on the Ariane 5 rocket's maiden flight,
3. given 2000 text documents in a mostly "on topic" database generated by representative searches in Lexus Nexus™ and DIALOG™,
4. using a "baseline" toolset that supported keyword queries, browsing articles by dates and titles sorted by relevance or date, and cutting and pasting selected portions of documents to a text editor.

The summary of observed behavior across the study participants and preliminary design recommendations is provided in Table 1.

Table 1. Summary of observed behavior and recommendations

Observed Behavior	Preliminary Recommendations
None characterized the database	Interactive exploration of database characteristics
All narrowed in on a small document set	Reminders to search other areas; "Longshot" visualizations to browse larger document sets
All read initial documents more carefully	Recommend initial documents based on quality; Identify data conflicts and updates; Identify new information in a document
All did not conduct new searches	Recommend queries; Improve query usability; Visualization to compare search results
Participants who made inaccurate statements did not read high profit documents; All missed some high profit documents	Identify high profit documents; Critiquer to narrow by document attributes not keywords
Some missed important events	Event recognition; Event-based displays
Some missed data conflicts	Identify and track data conflicts
Some could not remember sources for data; Time-intensive strategies to track sources	Track sources; Identify duplicate information from the same source
Some missed updates	Identify and track updates
Wide variation in confidence estimates about accuracy of verbal briefings	Identify events, conflicts, and updates that were missed; Reminders of data conflicts; "Longshot" visualizations that show "holes" in analysis

EVALUATION CRITERIA

In addition to generating preliminary design recommendations, the CTA also generated prescriptive evaluation criteria that effective solutions to the data overload problem in intelligence analysis should satisfy.

1. *Recognize Unexpected Information.* Bring analysts' attention to highly informative or definitive data and relationships between data, even when they do not know to look for it.
2. *Manage Uncertainty.* Aid analysts in managing data uncertainty. In particular, solutions should help analysts identify, track, and revise judgments about data conflicts and aid in the search for updates on thematic elements.
3. *Broaden.* Help analysts to avoid prematurely closing the analysis process. Solutions should broaden the search for or recognition of pertinent information, break fixations on single hypotheses, and/or widen the hypothesis set that is considered to explain the available data.

DESIGN SEEDS

Because the data overload problem is unlikely to be solved by design recommendations deriving from a single cognitive task analysis (CTA), we developed "design seeds" with the explicit purpose of supporting further exploration of what might prove useful to intelligence analysts. A design seed instantiates a modular strategy for aiding users around a domain-specific leverage point that is expected to usefully support performance. Rather than develop and implement an integrated support system only to discover when it is fielded that analysts do not find it useful, or that only portions of the system are useful, a design seed can be individually evaluated for its usefulness, as well as iterated based on feedback from participatory design sessions, in parallel with implementation in multiple systems.

Based on insights from the CTA, we created 9 design seeds. The documentation of each design seed included 1) the design concept, 2) the vulnerability the seed addresses or the strategy the seed supports, 3) criteria for how much to rely on accurate machine inferences, 4) an instantiation of the seed in the Ariane 501 scenario, and 5) a discussion of how the seed might generalize to other settings.

SEED 1: MANIPULATING HIGH PROFIT DOCUMENTS

An example of a design seed is identifying high quality documents based on document attributes. A central insight from the CTA study was the performance impact of locating a small number of documents judged to be of great importance by study participants, which we refer to as "high profit" documents. Therefore, the design seed:

- helps analysts quickly locate high profit documents
- provides useful support even though high profit documents would not be reliably and exhaustively identified by machine processing of document attributes
- uses visualizations to characterize documents

Because machine processing is unlikely to reliably and exhaustively identify high profit documents, we feel that it is important to use a "model" of a high profit document that does not rely heavily on the machine processing being correct. The requirement to have "weak" commitment to machine identification of a high profit document comes from the research base on human-machine interaction in complex, high-consequence systems. Which "weak commitment" architecture to employ (e.g., reminder, critiquer, visualization) depends on the capabilities of the algorithms, the preferences of the user, the preferences of the design team, the amount of time that users have in performing an analysis, and other domain-specific or expertise-specific characteristics. Some alternative instantiations of the design seed include:

- the user marks a document as "good" and the computer then displays, categorizes, and finds similar documents (Figure 1),
- the computer "seeds" high profit candidates for the user to browse based on a designer-defined model of a high profit document,
- the computer reminds the user to search for high profit documents, and
- the computer critiques the documents opened by the user for profitability.

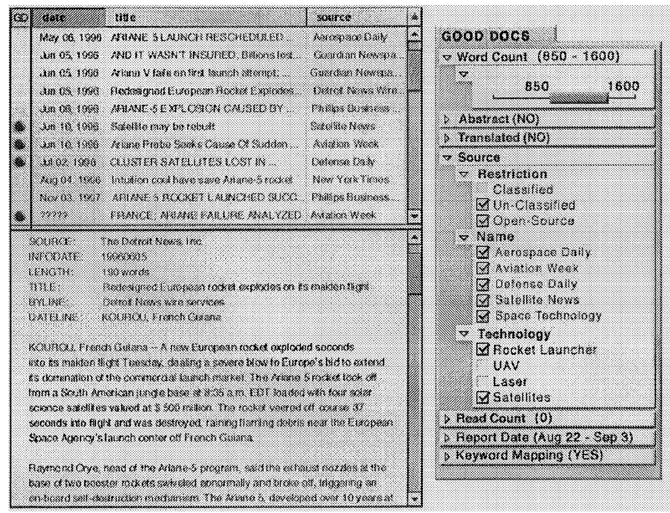

Figure 1. "Good" Documents Identified by the User

SEED 2: FINDING UPDATES ON A THEME

Another example of a design seed is finding updates on a theme. A surprising observation was that the study participant who had the most prior knowledge of the Ariane 501 scenario, the most knowledge about rocket launcher technology, and spent the most time, made an inaccurate statement in a briefing. This inaccurate statement provides insight about a particularly challenging vulnerability in analysis: missing updates that overturn previous interpretations. Because the data set is dynamic, there is always the possibility of missing information that renders previous information "stale" (see Figure 2 for examples in the Ariane 501 scenario). When these updates occurred on themes that were not central enough to be included in report titles or disrupting enough to generate a flurry of reports, it was extremely difficult to know if updates had occurred or where to look for them. Updates could be hours, days, weeks, months, years, or decades later. In addition, even "high profit" documents contained information that was believed accurate at one time, but then was later overturned.

This design seed:

- Helps analysts to locate updates that overturn or substantially change an analytic conclusion
- Helps analysts to calibrate judgments of analytic accuracy to the likelihood that updates were missed

It is interesting to note that most study participants never specifically looked for updates or described strategies to do so. It is possible that training might be useful, although the reaction of one novice analyst was that it would be very hard to find updates with the tools available to him.

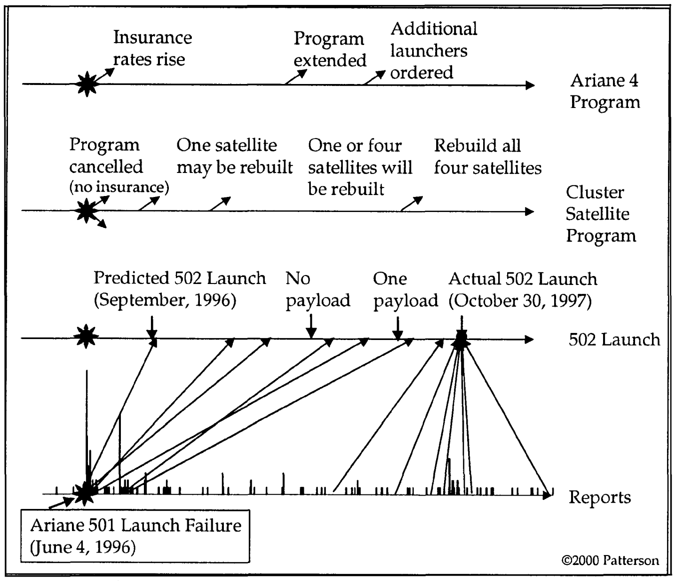

Figure 2. Updates on themes in the Ariane 501 Scenario

Therefore, "agents" that suggest targeted query formulations and/or "seed" representations with updates on a theme would be useful. Advances in natural language processing that allow recognition of updates on a theme or words that indicate that a previous interpretation has been overturned would greatly enhance this design seed.

Finding updates over time is a difficult challenge for human intelligence, and so it is also likely to be challenging for machine intelligence, raising "brittleness" and interaction issues that require further investigation. How much can the vulnerability to missing updates be reduced simply by having the machine intelligence remind the human partner to look for updates? Are there cues to informative areas where updates might be found, such as a flurry of setpoint crossings in a short amount of time on interrelated systems? Should the machine intelligence suggest possible candidate updates, either by "seeding" a visualization or by requiring the human to explicitly consider recommended items? What advancements in machine intelligence are required to make more accurate seeding recommendations? It is through exploration of these and similar issues that we could stimulate research that would increase the usefulness of the design seed, such as advances in natural language understanding, as well as add to the "design base" about effective ways of supporting human performance in complex domains.

VISUAL NARRATIVES

Although design seeds are modular so that they can be individually evaluated for usefulness and reused across projects, there are interactions between seeds. In order to investigate how the design seeds interact when coordinated in a workspace, we examined their integration under an overarching "Visual Narratives" concept (Figure 3).

Visual Narratives are built upon sequences of events in time organized by themes. An event is defined as a meaningful pattern of change for an observer in some environment (Christoffersen, 1999). Event structures are hierarchical with at least three levels of inspectability: episodes, events, and elements. Event categories are distinguished implicitly through their functional roles in the visualization. Two landmark events define an epoch. Disrupting events have repercussions to ongoing plans. Predictions of future events are less certain than pointers to past events.

In intelligence analysis, the Visual Narrative is constructed dynamically from information "packed" into text reports. The workspace includes both a "report space" and "theme space," and interactions between the two against a temporal frame of reference. The report space displays reports by date, which are selected by a mechanism such as keyword queries. Flurries of reports in time naturally emerge from the histogram display as indicators of disrupting events. The event space is a continuously constructed and modified narrative composed of interwoven, partially decomposable threads in time. Sequences of events occur at multiple levels in a hierarchical theme space. These events are displayed against a backdrop of ongoing plans and expectations. The configuration of the event space, including the level and focus of the events, is modifiable.

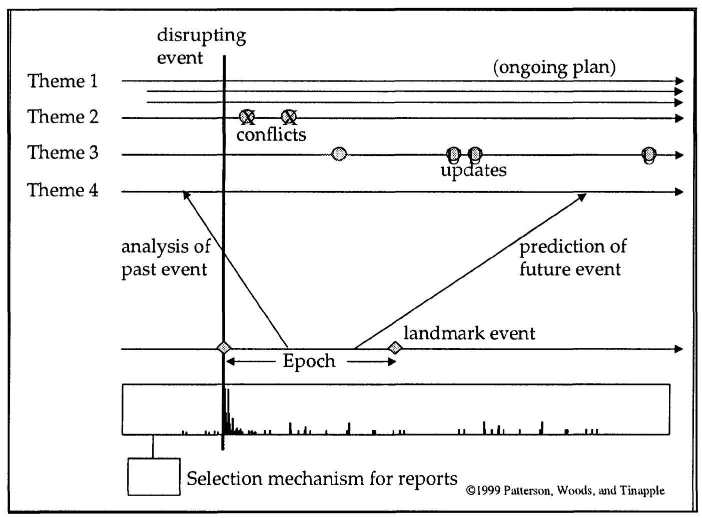

Figure 3. Visual Narratives in Intelligence Analysis

ANI-MOCKS

Each design seed was instantiated as an animated fly-through, or ***Ani-mock***, based on the Ariane 501 scenario (Figure 4). An ***Ani-mock*** is a scenario-based, dynamic visualization of a concept that grounds discussion in a particular area of a design space, yet leaves many object-level details unspecified. With ***Ani-mocks***, domain practitioners can get a tangible sense of how a design seed would support their activities in a realistic scenario. The goal of a participatory design session with an ***Ani-mock*** is to obtain user feedback about the usefulness of the concept (including the possibility that it is not), increase the understanding of the vulnerabilities and strategies that need to be supported, and elicit scenario modifications and extensions. The goal of an ***Ani-mock*** session with a design team is to generate a visible "usefulness" target when

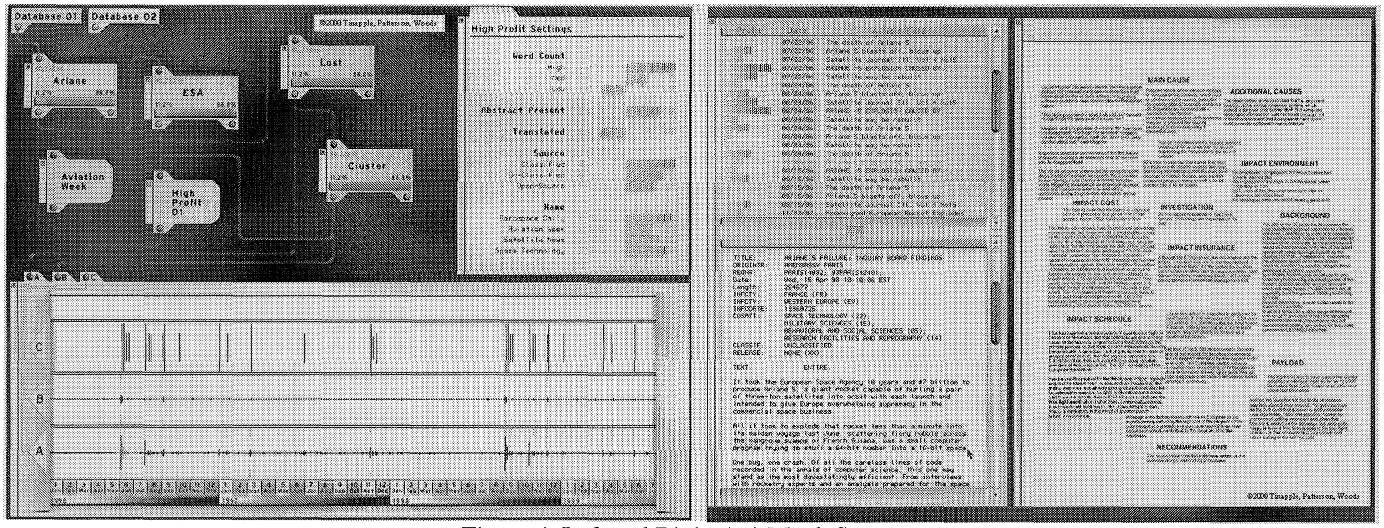

Figure 4. Left and Right Ani-Mock Screens

negotiating object-level details and the type and reliability of "backend" machine processing capability.

Although the creation of ***Ani-mocks*** is currently somewhat labor-intensive (although certainly less labor-intensive than creating an interactive prototype), we believe it is a promising direction for better integrating research and development efforts. In order to conduct research that is relevant to end users, as well as to prevent engaging in individual, one-off design endeavors without learning how to improve systems over time, we can develop, evaluate, and iterate design seeds that could be incorporated into multiple design projects. An area for further research is how to better support developing and reusing design seeds and instantiating them as ***Ani-mocks*** for particular scenarios.

ANI-MOCK PARTICIPATORY DESIGN SESSION

During a two-hour participatory design session, the ***Ani-mock*** was played several times to a small number of professional intelligence analysts to elicit feedback to improve the usefulness of the design seeds. Some representative examples of this feedback include:

- The need for radically different timescales for different scenarios (i.e., other than Ariane 501)
- The need to support regrouping, renaming, and associating themes after they have been created
- The need to support judging what has been missed in documents that have not been looked at (i.e., more support than reading titles of messages)
- The need to reuse themes from previous analyses and share themes with other analysts
- The need to have an overview visualization to see "holes" in a developing analysis
- Opportunities for using "theme" information to perform new queries

CONCLUSION

This paper describes how CTA results were used to jumpstart the exploration of "design seeds" that could prove useful in combatting "data overload" in intelligence analysis. In order to "bridge the gap" from cognitive task analysis to design, we generated preliminary design recommendations, identified prescriptive evaluation criteria that any solution should need to meet, innovated nine modular design concepts, or "design seeds," that could be incorporated into both ongoing and future design efforts, integrated the design seeds under an overarching workspace concept called Visual Narratives, and instantiated them as an animated fly-through, or ***Ani-mock*** demonstration, based in a face valid scenario.

ACKNOWLEDGMENTS

Funding was provided by the Human Effectiveness Directorate of the Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio, with truly exceptional support and guidance from Gil Kuperman. Many NAIC employees generously dedicated their time and expertise throughout the project. In particular Jay Finley, James Homer, and Mickey Wilkinson provided excellent guidance and feedback.

REFERENCES

Christoffersen, K. (1999). Event representations: Communicating change information in supervisory control. *Proceedings of the Human Factors and Ergonomics Society 43rd annual meeting.* Houston, TX

Peirce, C. S. (1903). Abduction and Induction. In J. Buchler (Ed.), *Philosophy of Pierce: Selected Writings.* (pp. 150-156). New York: Dover (1955).

Woods, D.D., Patterson, E.S., Roth, E.M., and Christoffersen, K. (1999). Can we ever escape from data overload? A cognitive systems diagnosis. *Proceedings of the Human Factors and Ergonomics Society 43rd annual meeting.* Houston, TX.