Cognition, Technology & Work (2001) 3:224–237
Ownership and Copyright
Springer-Verlag London Limited © 2001

Cognition
Technology &
Work

Predicting Vulnerabilities in Computer-Supported Inferential Analysis under Data Overload

E. S. Patterson[1], E. M. Roth[2] and D. D. Woods[1]

[1]Cognitive Systems Engineering Laboratory, Institute for Ergonomics, Ohio State University, Columbus, Ohio;
[2]Roth Cognitive Engineering, Inc., Brookline, Massachusetts, USA

Abstract: Data overload is a condition where a practitioner, supported by artefacts and other practitioners, finds it extremely challenging to focus in on, assemble and synthesise the significant subset of data for the problem context into a coherent situation assessment, where the subset is a small portion of a vast data field. In order to predict vulnerabilities in intelligence analysis that might arise when traditional strategies for coping with data overload are undermined, we conducted an observational study in a simulated setting. Ten professional intelligence analysts analysed the causes and impacts of the Ariane 501 accident. When study participants performed a time-pressured analysis outside their base of expertise based on sampling reports from a large set, some made inaccurate statements in verbal briefings. Participants that made no inaccurate statements spent more time during the analysis, read more documents, and relied on higher-quality documents than participants who made inaccurate statements. All participants missed potentially available relevant information and had difficulty detecting and resolving data conflicts. Sources of inaccurate statements were: (1) relying upon default assumptions, (2) incorporating inaccurate information and (3) incorporating information that was considered accurate at one point in time. These findings have design implications and point to evaluation criteria for systems designed to address the data overload problem in intelligence analysis.

Keywords: Cognitive task analysis; Data overload; Information retrieval; Intelligence analysis; Observation; Simulation

1. INFERENTIAL ANALYSIS UNDER DATA OVERLOAD

This research is driven by a formidable, ubiquitous problem: assessing the status of a dynamic, evolving situation by focusing in on pertinent data from a vast data field. Data overload is a function of work demands, practitioner strategies and supporting artefacts. We define data overload to be a condition where a domain practitioner, supported by artefacts and other human agents, finds it extremely challenging to focus in on, assemble and synthesise the significant subset of data for the problem context into a coherent situation assessment, where the subset of data is a small portion of a vast data field.

Intelligence analysis is an outstanding natural laboratory for studying inference-making, or inferential analysis, under data overload. The demands of intelligence analysis have always included the need to cope with data overload, but the scale of the problem has greatly increased with the explosion of accessible electronic data and widespread reductions in staff. In addition, 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 required to step outside their areas of expertise to respond quickly to targeted questions.

The goal of this research was to predict what vulnerabilities in inferential analysis might arise when traditional strategies for coping with data overload are undermined by technological and organisational change. Specifically, this research was designed to answer the question: What are potential vulnerabilities in computer-supported inferential analysis under data overload for professional analysts working on a short deadline outside their immediate base of expertise?

In order to answer this question, 10 professional intelligence analysts were observed while analysing the causes and impacts of the failure of the maiden flight of the Ariane 501 rocket launcher. Most participants had some related expertise, but none were able to fully answer the question prior to searching for information. A customised set of approximately 2000 reports from sources such as *Aviation Week and Space Technology* could be searched. A baseline support system, similar to tools used by intelligence analysts, was provided.

Four protocols for each study participant were used to identify vulnerabilities in inferential analysis under data overload in an iterative, exploratory fashion. These findings have design implications and point to evaluation criteria for systems designed to address the data overload problem.

2. INTELLIGENCE ANALYSIS

Intelligence analysis has many similarities to traditional supervisory control, although there are also important distinctions because the intelligence analyst monitors a mix of technological and human/organisational systems rather than an engineered system. Intelligence analysis is a complex task in which it takes approximately seven years to be considered expert, it is driven by events, conducted under time pressure, and involves monitoring the potential for surprise by adversarial nations. Intelligence analysis is an instantiation of inferential analysis, which we define as determining the best explanation for uncertain, contradictory and incomplete data. In addition, there are potentially high consequences for failure. For example, on 7 August 1998, the United States was surprised by Embassy bombings in Tanzania and Nigeria. Because the bombs were not detected, 224 people died, including 12 US citizens.

Intelligence analysis, like other domains, is undergoing changes that stress the ability to meet task demands. First, there is an explosion in the amount of data available to intelligence analysts. On an average day, an analyst will receive hundreds of text messages through electronic mail that are selected by keyword 'profiles' from sources such as the National Security Agency. These messages update analysts on topics related to particular technologies and countries, and they are often sorted into personal databases. In addition to these resources, there are massive organisational databases that are accessed when a question is asked about something that has not been actively tracked. For example, an analyst described that he was asked to 'Tell me everything you know about the command and control structures in Country X in the next 24 hours.' Since no analyst had ever monitored that country, he performed keyword searches in an on-site database that generated 42,000 documents. Theoretically, he could also have searched other databases, such as Lexus Nexus^{TM}, and classified and unclassified sites on the World Wide Web. He estimated that he could scan 15,000 messages in a day, making it impossible to read all the documents in the allotted time.

Second, while the number of monitored countries and technologies has greatly increased in the post-Cold War environment, there are also widespread reductions in staff and loss of expertise. For example, at the National Air Intelligence Center (NAIC), 30% of the analysts are eligible for retirement in five years. Active duty military personnel with no prior analysis experience will replace many of these analysts. Military personnel are assigned for three to four years, of which the first year is normally spent obtaining clearance to work with classified information. The net result is that less experienced intelligence analysts are increasingly asked to analyse situations that are outside their bases of expertise on short time horizons.

3. SIMULATED TASK: ANALYSING THE ARIANE 501 ACCIDENT

The Ariane 501 accident was analysed by study participants under the conditions of data overload and a short deadline of several hours. The maiden launch on 4 June 1996 of the Ariane 5 vehicle ended in a complete loss of the rocket booster and the scientific payload, four Cluster satellites, when it exploded 30 seconds after lift-off. The Ariane 5 rocket was a new European rocket design by Arianespace that was intended to eventually replace the successful Ariane 4 rocket. The Ariane 5 rocket was designed to be larger, more powerful and to carry multiple payloads. The Ariane 501 accident was significant in how it departed from typical launch failures. First, the explosion was due to a software design problem rather than the more classic mechanical failure – there was numerical overflow in an unprotected horizontal velocity variable in the embedded software that was reused from the Ariane 4, a slower rocket. Additionally, it was the first launch of a new rocket design, which raised concern about the design's viability. Overall, however, launch failures were relatively common in the industry, and first launches in particular were prone to fail, so the reputation of the Ariane programme was not greatly damaged.

During a prior interview, an experienced intelligence analyst stated that the Ariane 501 scenario captured critical aspects necessary for high face validity. First, the scenario was challenging to analyse in a short time, with opportunities for the study participants to make inaccurate statements based on misleading and inaccurate information in the provided database. Second, the analysis required technical knowledge about aerospace vehicles, which is prototypical of tasks performed by US air force intelligence analysts. Third, although all study participants had some pertinent experience that helped them to perform the analysis, none had been directly monitoring the particular country or technologies. Fourth, open sources such as *Aviation Week and Space Technology* closely paralleled classified sources in reporting style, analytic depth and technical detail.

The Ariane 501 scenario also involved an accident investigation, which allowed the investigators to leverage models of responses to accidents in designing the simulated task. For example, the dates for documents in the electronic database ranged from 1994 until 1999 (Fig. 1). These dates were selected so that the database included distractors prior to the accident, the Ariane 501 accident, the Inquiry Board Report detailing the findings of the accident investigation, and the next landmark event after the accident, the Ariane 502 launch. The naturally emerging structure of reports in the database mirrored structures from accident investigations in other high-consequence settings. The initial flurry of reports about the accident tended to be sensationalistic, included quotes from eyewitnesses about the causes and immediate reactions to the accident from affected parties, and contained details not available in later reports, some of which later turned out to be inaccurate. These early reports emphasised contributors to the accident that were closest in time and space (e.g., decision by ground operator to blow up the rocket). The second main flurry of reports summarised the findings of the Inquiry Board about the causes of the accident. Intermittently after the inquiry board report was released, comprehensive, in-depth analyses and long-term impacts could be found. Reports at this time tended to have less diversity in the descriptions about the causes and impacts of the accident and contained fewer details. These later reports included contributors that were farther in space and time from the accident – limitations with the design and testing of the rocket and the organisational context for the rocket design. Finally, another small flurry of reports was seen immediately following the second attempted launch of the Ariane 5 rocket, which was the next landmark event after the accident. These reports briefly summarised the accident and provided updates on sub-themes.

In addition, discrepancies in the data set followed patterns seen in other accident investigations (boxed items in Fig. 2 had inaccurate information in the database about that item). There were inaccuracies in early reports about the causes of the accident because all data were not yet available. In addition, there were inaccuracies about in-depth, technical topics such as the numerical overflow.

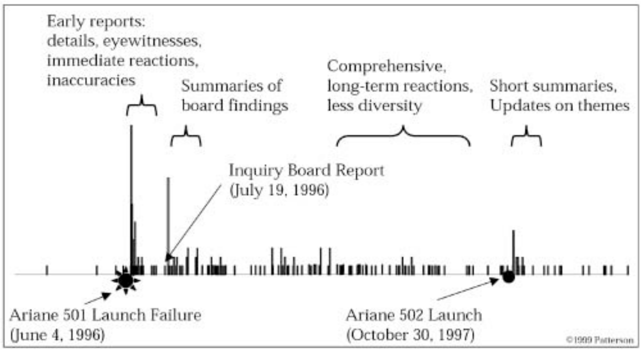
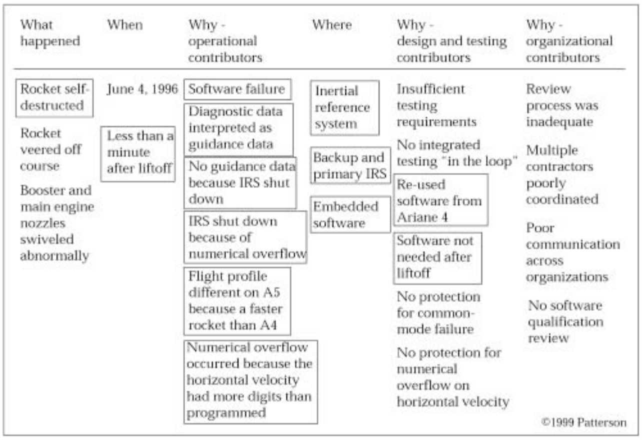


Fig. 1. Report database reflected response to accident.



What happened	When	Why - operational contributors	Where	Why - design and testing contributors	Why - organizational contributors
Rocket self-destructed	June 4, 1996	Software failure	Inertial reference system	Insufficient testing requirements	Review process was inadequate
Rocket veered off course	Less than a minute after liftoff	Diagnostic data interpreted as guidance data	Backup and primary IRS	No integrated testing "in the loop"	Multiple contractors poorly coordinated
Booster and main engine nozzles swiveled abnormally		No guidance data because IRS shut down	Embedded software	Re-used software from Ariane 4	Poor communication across organizations
		IRS shut down because of numerical overflow		Software not needed after liftoff	No software qualification review
		Flight profile different on A5 because a faster rocket than A4		No protection for common-mode failure	
		Numerical overflow occurred because the horizontal velocity had more digits than programmed		No protection for numerical overflow on horizontal velocity	

©1999 Patterson

Fig. 2. Discrepancies in the causes of the Ariane 501 accident.

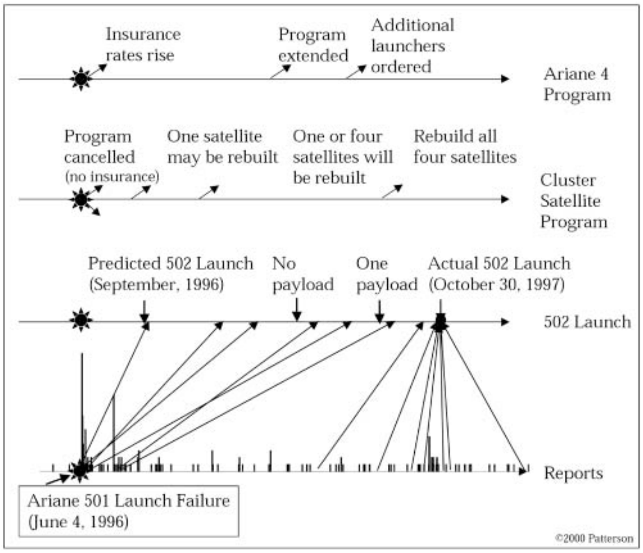


Fig. 3. Updates about impacts of the failure.

Finally, information about impacts on the Ariane 4 rocket programme, Cluster scientific programme and the next launch in the Ariane 5 rocket programme, 502, came in over time, causing information from different points in time to conflict (Fig. 3). For example, the original predictions of the second launch of the Ariane 5 vehicle (502) of September 1996 were overly optimistic, and predictions gradually became nearer to the actual date of 30 October 1997. As is expected following costly accidents, impact predictions radically changed over time. For example, immediately after the 501 accident, it was reported that the Cluster scientific programme would be shut down because of the uninsured loss of the $500 million scientific satellites. A month later, it was reported that one of the four satellites might be rebuilt. Two months later, it was reported that either one or four satellites would be rebuilt. Seven months later, it was reported that all four satellites would be rebuilt.

The database contained enough information to support a comprehensive analysis of the causes and impacts of the Ariane 501 accident. There were approximately 2000 documents selected from open source literature. The majority (~60%) of the documents were 'on target' in that they contained information about the causes and impacts of the accident. Some documents (~35%) contained information that helped to provide context, such as information about other rocket launch failures, but were not directly related. As would be expected by intelligence analysts from searches of their organisational databases, in contrast to keyword searches on the World Wide Web, only a small portion contained completely irrelevant information (~5%), such as articles about women named Ariane. Nine documents in the database were identified as particularly high quality, classified as 'high profit' by the investigators. The high profit categorisation was based on both high topicality and utility, which are often used in relevance definitions in the information retrieval literature (see Mizzaro 1997 for a review of factors in relevance definitions; cf. Blair and Maron 1985 for their distinctions between vital, relevant, partially relevant and not relevant documents in legal analysis). An example of a high profit document was 'Board Faults Ariane 5 Software', published on 29 July by *Aviation Week and Space Technology*.

4. METHODOLOGY

The target situation was inferential analysis conducted by experienced analysts under data overload, on tight deadlines, and outside their immediate bases of expertise. The study was designed to simulate this target situation:

- 10 professional intelligence analysts, ranging from 7 to 30 years of analytic experience, representing diverse areas of expertise that were related to portions of the simulated task;
- analysing a face valid task that they had not previously analysed and was not in the immediate base of expertise: the cause and impacts of the 4 June 1996 Ariane 501 rocket launch failure on the Ariane 5 rocket's maiden flight;
- given 2000 text documents in a mostly 'on topic' database generated by representative searches in Lexus Nexus and DIALOG by the investigators and a professional search intermediary from the intelligence agency;
- in 3–4-hour sessions; and
- 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 participants were asked to think aloud during the process and provide a verbal briefing in response to the written question: 'In 1996, the European Space Agency lost a satellite during the first qualification launch of a new rocket design. Give a short briefing about the basic facts of the accident: when it was, why it occurred, and what the immediate impacts were.' Two investigators directly observed this process for all study participants, which was also audio and video taped. The investigators noted during the session the unique database number for the opened

Table 1. An excerpt of study participant 5's article trace protocol

Article no.	Query	Name and source info	Why selected	Important?	Notes
1380	1	ARIANE 5 EXPLOSION CAUSED BY FAULTY SOFTWARE; SATELLITE NEWS	Wants to work backwards so wants a late article		Faulty software
1274	1	NEW CLUES TO ARIANE-5 FAILURE; DEFENSE DAILY	Title and looking for date of event		4 June 1996(limits query results to after June 1 since event is June 4)
253	1A	STRIDE: FIRING TESTS OF NEW H IIA ROCKET ENGINE COMPLETED	Time of article close to event		Of no interest – recognises the HIIA rocket engine is from Japan
1855	1A	European space rocket explodes: Work continues with 14 similar models; Ottawa Citizen		Cuts and pastes	5 km from launch site 40 seconds 14 rockets on production line – if fault is not generic, the programme won't suffer too much (software would classify as not generic according to him)
1223	1A	False computer command blamed in Ariane V failure; Aerospace Daily	6-6-96 date, also title	Cuts and pastes, marks, says good article	Computer command Aerospace Daily as a good source Says article is 'remarkably good' and takes a while reading it 6 June knew false signal and looking closer at it Says what causes were eliminated

documents. The investigators also captured the queries, the documents that were returned by the queries, the 'marked' documents, the workspace configuration of the screen, and snapshots of electronic notes.

Four protocols which emphasised different aspects were generated: the search strategies, the selection and interpretations of documents, the strategies for resolving conflicts in the data and the construction of the verbal briefing. An excerpt of the protocol focusing on document selection and interpretation is provided in Table 1. This protocol was generated by a single investigator and then verified and expanded by a second investigator. Differing interpretations were resolved through additional searches for evidence and debate.

The data analysis was an iterative, discovery-oriented process. As the protocols were generated, areas for more detailed investigation were noted. In addition, the verbal briefings were transcribed and items were coded as not mentioned, accurate, vague and inaccurate. The process that an individual participant followed that arrived at the inaccurate statement was analysed and emerging patterns used to identify the cognitive challenges that led to inaccurate statements across participants.

Overall, the analysis process involved bottom-up searching for patterns combined with top-down conceptually driven investigations (see Woods 1993 for a description of the process tracing methodology used in the data analysis). The base protocols served as a detailed account of the process from the perspective of different conceptual frameworks, including strategies to cope with data overload in supervisory control, information retrieval strategies, and resolving data interactions in abductive inference. The protocols were used to identify patterns on particular themes. These patterns were then represented across participants in ways that highlighted similarities and differences along relevant dimensions.

5. FINDINGS

The analysis process generally followed the pattern shown in Fig. 4. Reports were selected from the database through the refinement of keyword queries and by browsing the returned reports by title or date. A small number of their sampled reports were heavily relied upon. These documents made up the skeleton of the analysis product. Excerpts from supporting documents were then used to corroborate some information and fill in details. Conflicts in the data were flagged and judgements about which data to include in the developing story were revisited as new information was discovered. When the study participants felt ready, they organised their notes and generated a coherent story.

During the analysis, several patterns in process vulnerabilities were identified for the challenging simulated conditions: study participants asked to analyse something outside their immediate base of expertise, tasked with a tight deadline, and under data overload. These patterns included information sampling strategies that reduced the amount of data to manage at the cost of missing critical information and conflict identification and resolution strategies that led to sources of inaccurate statements in the verbal briefings. Note that two of the study participants' data were not included in the analysis. One participant attempted to analyse a different satellite failure (SPOT-3), which was not well supported by the database. Another participant would not complete the task because a printer was not available and he relied on viewing printed documents in parallel and noting discrepancies directly on the documents.

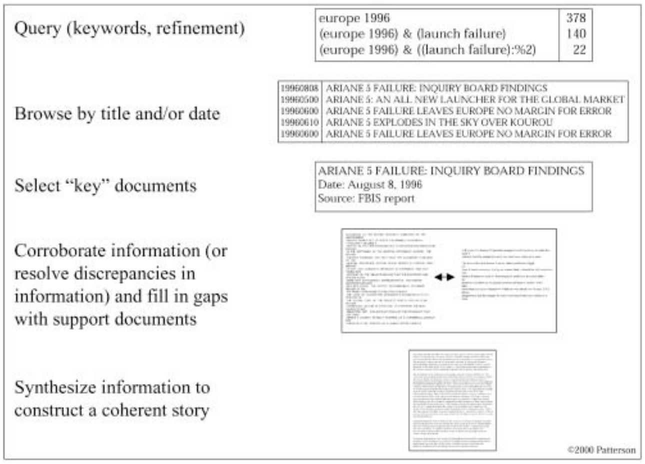

Query (keywords, refinement)
europe 1996 | 378
(europe 1996) & (launch failure) | 140
(europe 1996) & ((launch failure):%2) | 22
Browse by title and/or date
19960808 ARIANE 5 FAILURE: INQUIRY BOARD FINDINGS
19960500 ARIANE 5: AN ALL NEW LAUNCHER FOR THE GLOBAL MARKET
19960600 ARIANE 5 FAILURE LEAVES EUROPE NO MARGIN FOR ERROR
19960610 ARIANE 5 EXPLODES IN THE SKY OVER KOUROU
19960600 ARIANE 5 FAILURE LEAVES EUROPE NO MARGIN FOR ERROR
Select "key" documents
ARIANE 5 FAILURE: INQUIRY BOARD FINDINGS
Date: August 8, 1996
Source: FBIS report
Corroborate information (or resolve discrepancies in information) and fill in gaps with support documents
Synthesize information to construct a coherent story
©2000 Patterson

Fig. 4. Typical analysis process.

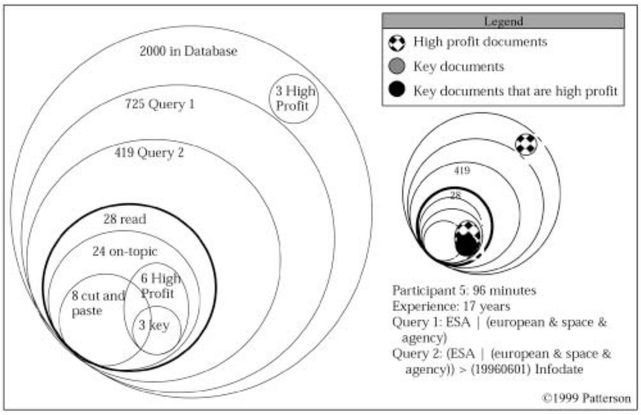
5.1. Sampling by Narrowing In

In inferential analysis under data overload in baseline electronic environments with textual databases, information is effectively sampled, generally through querying and browsing. In our study, participants were observed to begin the analysis process by making queries with standard inputs such as keywords and date limits. If a returned set of documents was judged to be too large, the search was narrowed rather than starting with a new set of search terms. Typical narrowing strategies included adding a keyword, limiting to a date range or enforcing a proximity requirement on a set of keywords. The search was then further narrowed through browsing by summary information about a document, typically dates and titles. Documents were then opened by double-clicking a report title.

A subset of the opened documents was judged to be relevant to the analysis. Of this set of documents, a small number was heavily relied upon during the analysis, which we refer to as 'key' documents. Key documents were identified by a combination of behavioural and verbal data.

To illustrate, consider the information sampling process employed by study participant 5 (Fig. 5). The participant started with a Boolean keyword search (esa OR (european AND space AND agency)). This search returned 725 hits, so he narrowed the search to documents published after 1 June 1996 after determining that the date of the accident was 4 June 1996 from scanning three articles. 419 documents remained after this narrowing criteria, which became his 'home query' in that he did no more keyword searches. Twenty-eight documents were opened, 24 of which were on-topic, or relevant to the analysis. Six opened documents were 'high-profit' in that they were judged to be highly informative by the investigators. The other three high-profit documents were available in the database but were not returned by either query. The participant cut and pasted portions of eight documents along with references into a word-processing file and used a marking function to highlight two documents: one because he stated that it was a remarkably good article and one in case he needed to refer back to it later. Three articles were identified as his 'key' documents: (1) document 1223 because he highlighted it with a marking function, said that it was 'remarkably good' and spent a long time reading it; (2) document 1301 because he spent a long time reading it and said immediately after reading it that he now had a good idea of what had happened; and (3) document 1882 because he said that it was 'a definite keeper', that it was like briefings by professional analysts in its quality, spent a long time reading it, and electronically selected text from it. Note that all three key documents were high-profit documents.

The information sampling strategy for study participant 5 could be characterised as progressive narrowing. An initial query was refined to reach a document set that was judged manageable based on the number of hits. A small subset of these documents was heavily relied upon in generating the analysis product. All study participants employed a similar search process (Fig. 6). All participants narrowed their queries to a manageable number (22–419 documents), from which they opened documents based on dates and titles (4–29 documents). They relied heavily on a small number of documents (1–4 key documents) to generate the verbal briefing.



Participant 5: 96 minutes
Experience: 17 years
Query 1: ESA | (european & space & agency)
Query 2: (ESA | (european & space & agency)) > (19960601) Infodate

©1999 Patterson

Fig. 5. Information sampling by study participant 5.

Under data overload, narrowing in on a small number of reports through keyword additions to an initial query is a common strategy (Bates 1979; Blair 1980; Olsen et al 1998). This is also similar to the filtering strategy to cope with information overload observed by Miller (1960) in a laboratory task.

The finding that the study participants used relatively primitive search strategies is similar to other domains, where domain experts conduct their own searches but do not learn sophisticated search tactics (e.g., legal analysts, Blair and Maron 1985). One implication is that there is a need to train intelligence analysts in effective search techniques or to involve professional search intermediaries in the search process. If intermediaries perform the searches, it would be important to include analysts with domain knowledge in selecting search terms, as that is an important factor in the ability to locate relevant information (Saracevic et al 1988), and search and domain expertise is only partially decomposable (Shute and Smith 1992).

In addition, the interface could be redesigned to encourage better search tactics. For example, faceted search is a recommended search strategy where synonyms are combined with 'OR' commands within a facet and crossed orthogonally with conceptually distinct facets using the 'AND' command. Study participant 6 used a query that crossed (destr* OR explo*) with fail* in order to narrow the number of documents to browse. Instead, these three terms should be included as synonyms within a facet and crossed with a conceptually distinct facet such as the payload. The interface could afford faceted search by displaying synonyms together that are automatically combined with 'OR' and displaying different synonym groups separately that are automatically crossed against other groups with 'AND'. Extensions to this design direction might include options to use synonym dictionaries, interface elements that encourage narrowing by document attributes rather than keywords, or having the machine critique a user's query formulation.

It is not surprising, given crude search tactics and the provided interface, that all participants missed high-profit documents without being aware of it. This is similar to the findings of Blair and Maron (1985), where legal analysts were poorly calibrated to the amount of relevant information that they missed after searching an electronic database. Samples returned by keyword search were opaque about how they related to what was available, such as what high-profit documents were left out of the query results. Documents were then sampled based on dates and titles,

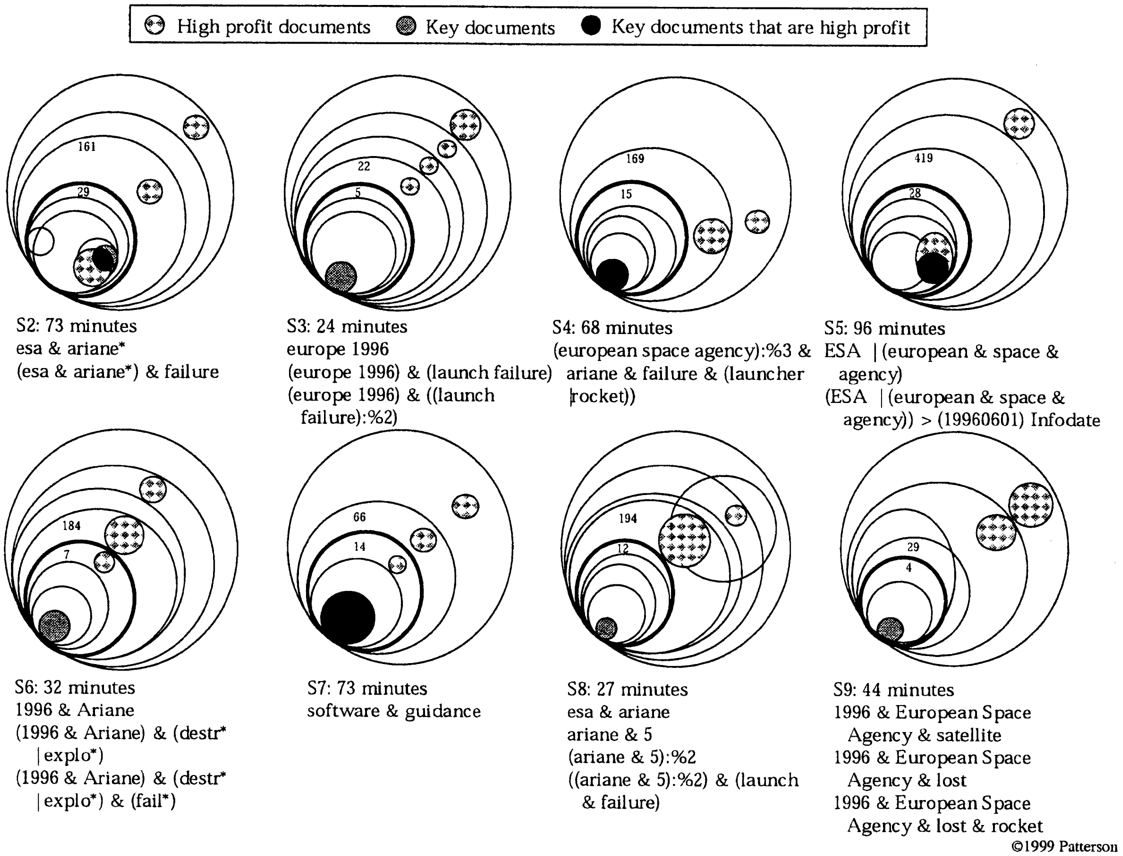

High profit documents
Key documents
Key documents that are high profit
161
29
S2: 73 minutes
esa & ariane*
(esa & ariane*) & failure
22
5
S3: 24 minutes
europe 1996
(europe 1996) & (launch failure)
(europe 1996) & ((launch failure):%2)
169
15
S4: 68 minutes
(european space agency):%3 & ariane & failure & (launcher |rocket))
419
28
S5: 96 minutes
ESA |(european & space & agency)
(ESA |(european & space & agency)) > (19960601) Infodate
184
7
S6: 32 minutes
1996 & Ariane
(1996 & Ariane) & (destr* |explo*)
(1996 & Ariane) & (destr* |explo*) & (fail*)
66
14
S7: 73 minutes
software & guidance
194
12
S8: 27 minutes
esa & ariane
ariane & 5
(ariane & 5):%2
((ariane & 5):%2) & (launch & failure)
29
4
S9: 44 minutes
1996 & European Space Agency & satellite
1996 & European Space Agency & lost
1996 & European Space Agency & lost & rocket
©1999 Patterson

Fig. 6. Searching processes for all study participants.

which were weak quality indicators (Table 2). The first 'low-profit' article was a translated description of an article originally published in Italy that contained inaccuracies about the cause of the failure. The second article was a short abstract that contained little information. The third article contained inaccuracies because it was published soon after the event occurred.

Note that during the simulated task some study participants verbalised that they should conduct new searches for specific information. In addition, comments made by some study participants indicated that they did not know what was available in the database and how their queries related to what was available. In spite of these statements, the study participants did not generate additional queries, either to search for information or to characterise the database. With each new query, information such as what documents had been opened and marked, as well as workspace tailoring such as resized windows and sorting reports by date, was lost. If the navigation and workspace tailoring costs were reduced, it might be more likely that analysts would sample more areas of the database.

Table 2. Dates and titles of low- and high-profit articles

'Low-profit' articles	'High-profit' articles
Europe: Causes of Ariane 5 failure (5 July 1996)	Software design flaw destroyed Ariane V; next flight in 1997 (24 July 1996)
Ariane 5 Failure: Inquiry Board findings (25 July 1996)	Board faults Ariane 5 software (29 July 1996)
False computer command blamed in Ariane V failure (6 June 1996)	Ariane 5 loss avoidable with complete testing (16 September 1996)

5.2. Basing Analyses on High-Profit Documents

In Fig. 6, the black circles represent when the key documents were also high-profit documents, i.e., when the documents that were heavily relied upon were the best available documents. Comparing the four participants that used some high-profit documents as key documents versus the four that did not, there are some interesting differences between the two groups (Tables 3 and 4). The participants that used high-profit documents as key documents spent more time during the analysis, read more documents and read more of the high-profit documents.

We believe that the best explanation for the differences

Table 3. Comparison of participants that read high-profit documents versus those that did not

Participant	Experience (years)	Time (minutes)	Final query (no. hits)	Documents (no. read)	High-profit docs (no. read)
Participants whose key documents were not high-profit documents					
3	7	24	22	5	0
6	8	32	184	7	2
8	11	27	194	12	0
9	18	44	29	4	0
Average:	11	32*	107	7*	0.5*
Participants whose key documents were high profit documents					
2	8	73	161	29	3
4	8	68	169	15	2
5	17	96	419	28	2
7	9	73	66	14	5
Average:	10.5	78*	204	22*	3*

*Significant difference using Wilcoxon–Mann–Whitney non-parametric test.

Table 4. Comparison of querying and browsing breadth

	Final 'home' query	No. of hits in query	No. of high-profit hits in query	Per cent of query docs that are high profit	No. of documents read	No. of high-profit documents opened	Per cent of 'key' docs that are high profit
Participants whose key documents were not high-profit documents							
3	(europe 1996) & ((launch failure):%2)	22	1	5%	5	0/9	0% (0/1)
6	(1996 & Ariane) & (destr* | explo*) & (fail*)	184	7	4%	7	2/9	0% (0/3)
8	((ariane & 5):%2) & (launch & failure)	194	8	4%	12	0/9	0% (0/1)
9	1996 & European Space Agency & satellite & lost & rocket	29	0	0%	4	0/9	0% (0/1)
	Average:	107	4	3%	7*	0.5/9*	0%
Participants whose key documents were high-profit documents							
2	(esa & ariane*) & (failure)	161	6	4%	29	3/9	50% (1/2)
4	(european space agency):%3 & ariane & failure & (launcher | rocket))	169	7	4%	15	2/9	100% (2/2)
5	(ESA | (european & space & agency))> (19960601) Infodate	419	7	2%	28	2/9	33% (1/3)
7	Software & guidance	66	7	11%	14	5/9	100% (4/4)
	Average:	204	7	5%	22*	3/9*	71%

*Significant difference using Wilcoxon–Mann–Whitney non-parametric test.

between these two groups is that participants who found the high-profit documents were more 'persistent' in that they took longer and read more documents. We investigated nine alternative explanations and found little evidence to support them (Patterson et al 1999), including effectiveness of search strategies, ability to recognise high-quality documents, and domain and scenario-specific knowledge.

If the explanation is persistence, then this indicates that one of the ways, given a baseline electronic toolset of keyword querying and browsing by dates and titles, to find the high-profit documents in the database would be to cast a wider net by sampling more, either by performing more queries or by opening up more documents. Training or machine 'reminders' to broaden information search might be helpful. Nevertheless, given increasing pressure to do analyses on shorter deadlines, these interventions might be ineffective.

A potentially useful intervention might be to use machine processing to help an analyst quickly locate high-profit documents. For example, machine processing can use combinations of document attributes such as source, length, language, abstract and how many times it has been opened to identify likely candidates. Similarly, a user could mark a set of documents as 'good' and the computer could then search for documents with similar attributes and content.

Because machine processing is unlikely to reliably and exhaustively identify high-profit documents, we feel that it would be important to use a 'model' of a high-profit document that does not rely heavily on the machine processing being correct and is easily inspectable and redirectable. Which 'weak commitment' architecture to employ (e.g., reminder, critiquer, visualisation) would depend 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 and expertise-specific characteristics.

Table 5. Summary of types of statements in verbal briefings

Participant	Accurate	Vague	Inaccurate	Nothing
Participants whose key documents were not high-profit documents				
3	5	2	2	11
6	11	1	3	5
8	9	0	0	11
9	5	3	1	11
Average:	7.5	1.5	1.5*	9.5
Participants whose key documents were high-profit documents				
2	5	2	0	13
4	11	2	0	7
5	12	3	0	5
7	8	1	0	11
Average:	11	2	0*	6.75

*Significant difference using Wilcoxon–Mann–Whitney non-parametric

5.3. Impact of Basing Analyses on High-Profit Documents

An important question is whether the study participants who used the high-profit documents as key documents in their analyses performed better than those that did not. The study participants' verbal briefings were coded on 20 topic items from the Ariane 501 case as accurate, vague, inaccurate or no information (Table 5, intercoder reliability $\kappa = 0.84$). As expected, the participants who did not use high-profit documents had more inaccurate statements in their verbal briefings than participants who did.

5.4. Sources of Inaccurate Statements

Two conceptual frameworks guided the data analysis. The first framework was information sampling strategies, generally referred to as search tactics in the information retrieval literature. The second framework was evidence interactions in abductive inference (Schum 1994; Josephson and Josephson 1994), defined as inference to the best explanation. Diagnosis is an example of a well-known abductive inference process, where a diagnostic reasoner selects an explanatory hypothesis to explain observed symptoms. The abductive process involves observing deviations from a nominal state, proposing explanatory hypotheses to account for the deviations, and selecting the 'best' or most warranted explanation from the set of candidate hypotheses.

Determining the cause of the Ariane 501 accident could be characterised as an abductive inference task. Anomalous data could be explained by several hypotheses (Fig. 7). For example, the observation that the rocket swivelled abnormally could have been due to inaccurate guidance data, a mechanical failure or a software failure. The main observation that pointed to a software failure hypothesis was that both the primary and backup inertial reference systems (IRS) shut down simultaneously. Although this finding made the software failure the most plausible explanation, there was an additional finding not covered by this hypothesis: unexpected roll torque during ascent. The full set of observations was explained by a combination of two hypotheses: a software failure and an unrelated mechanical problem.

We were surprised that there was little evidence from the think-aloud protocols and decisions regarding data conflicts for a traditional abductive inference process. Rather than gathering a collection of data, determining what hypotheses would explain the data, and comparing the plausibility for different combinations of hypotheses in order to come up with a best explanation, the study participants appeared to be following a different process. The main difference between the theoretical pattern of abductive inference and the empirical evidence was that the study participants were not dealing with elemental observations and hypotheses. They were dealing with a 'second-order' data set where interpretive frames already existed in which the report writers presented data embedded within their hypotheses. The main task of the study participant, therefore, was to improve the veracity of the analytic product by corroborating multiple reports of others who had already performed the task of mapping explanatory hypotheses to a dynamically changing data set.

Analysis revealed that the 'hypothesis space' was better represented by Fig. 8 than Fig. 7. Rather than the 'elemental' data and hypotheses for the Ariane 501

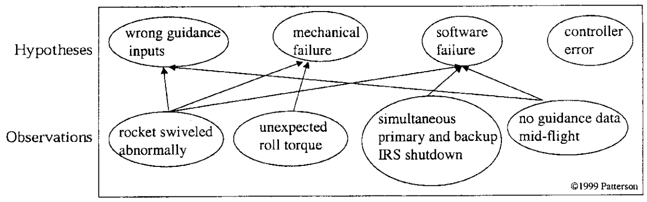


Fig. 7. Hypothesis space in Ariane 501 scenario.

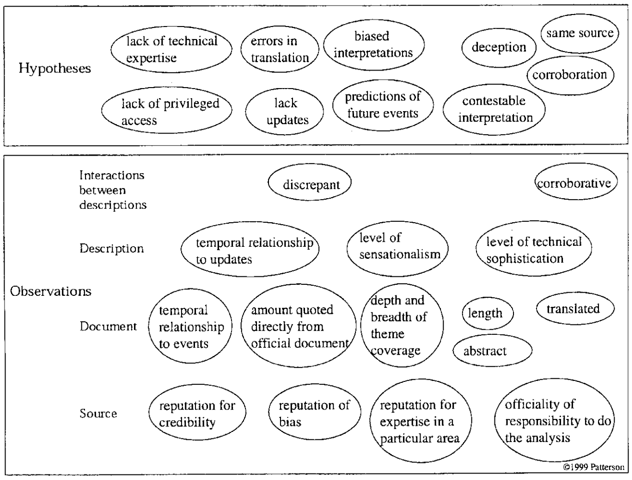

Hypotheses
lack of technical expertise
errors in translation
biased interpretations
deception
same source
corroboration
lack of privileged access
lack updates
predictions of future events
contestable interpretation
Observations
Interactions between descriptions
discrepant
corroborative
Description
temporal relationship to updates
level of sensationalism
level of technical sophistication
Document
temporal relationship to events
amount quoted directly from official document
depth and breadth of theme coverage
length
translated
abstract
Source
reputation for credibility
reputation of bias
reputation for expertise in a particular area
officiality of responsibility to do the analysis
©1999 Patterson

Fig. 8. 'Second-order' hypothesis space.

scenario, the think-aloud protocols gave evidence for the study participants dealing at the 'second-order' level of using cues from the text, document and source to evaluate how to resolve data conflicts. The study participants displayed expertise in recognising the cues that were used in evaluating the information and in relating those cues to possible hypotheses. Note that this expertise is scenario-independent, but is not expected to be available to novice intelligence analysts or undergraduate students.

In addition to seeing how the study participants mapped observations to hypotheses, we investigated why inaccurate statements in the verbal briefings were made. Three sources of inaccurate statements were identified: (1) relying upon default assumptions; (2) incorporating inaccurate information; and (3) incorporating information that was considered accurate at one point in time.

5.4.1. Relying on Assumptions that did not Apply

Several inaccurate statements in the verbal briefings did not come from any of the opened documents. In these cases, the participants appeared to be relying on default assumptions that did not apply in this case. For example, during the verbal briefing, one participant stated that the monetary loss of the Cluster satellite payload would be recovered by insurance. Although payloads are often insured, the Cluster satellites, as a unique design for a solar wind experiment to be launched on a maiden voyage of a new rocket design, could not be.

It is unclear how this vulnerability could be addressed other than by making it easier to corroborate information. Relying on assumptions is a heuristic that is normally useful in filling in missing information. For example, participant 2 assumed the Ariane 5 rocket would eventually replace the Ariane 4 as the standard launch vehicle in his estimation of the impacts. In addition, default assumptions can be valuable in knowing what information to seek. For example, participant 4 stated that he assumed that there was a payload on the flight and then explicitly looked to see what it was.

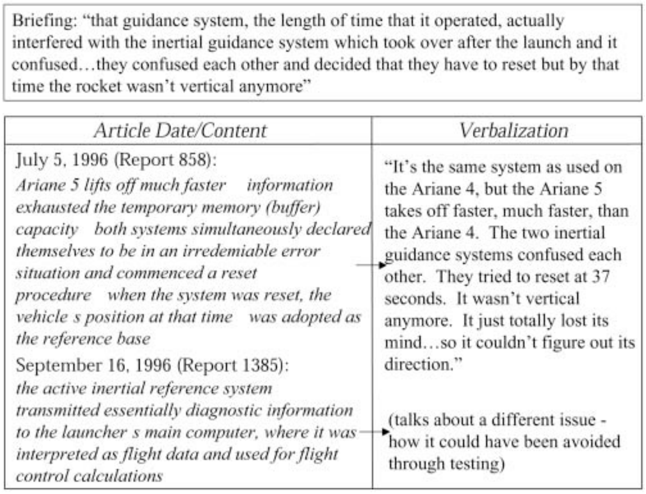
5.4.2. Incorporating Information that was Inaccurate

Some inaccurate statements were repeats of inaccurate descriptions in opened documents. Intelligence analysts clearly view the elimination of inaccuracies by finding converging evidence across independent sources as a major component of the value of an analytic product. The participants described and employed a variety of strategies for tracking and resolving discrepant descriptions in order to reduce their vulnerability to incorporating inaccurate information. Partly because this cognitively difficult process of corroborating information and resolving conflicting information was largely unsupported by the provided tools, nearly every participant experienced some breakdowns in this process. Breakdowns included failing to corroborate information, missing conflicts in opened documents, forgetting how many corroborating and conflicting descriptions had been read from independent sources, forgetting from which source information originated and treating descriptions that stemmed from the same source or document as corroborating.

To illustrate the difficulties in eliminating inaccuracies, consider the example of determining the cause for why the rocket swivelled abnormally. Interestingly, participants 6 and 7 both read the same two documents that contained discrepant descriptions but ended up with different assessments in their verbal briefings (Figs 9 and 10).

Participant 6 based his analysis of why the rocket swivelled mainly on report 858, which described the cause as a reset of the inertial reference frame following a numeric overflow (Fig. 9). As he read 858, he verbalised why the rocket swivelled based on what he was reading. Later, he read 1385, which had a contradictory description of why the rocket swivelled. At that point, however, it was the last document that he looked at, and he was focused on a different issue – why testing did not reveal the software error. He gave no evidence that he recognised the conflict. In addition, when asked how he decided to stop searching for information, he explained: 'It doesn't look like anybody will have any different opinions. From looking at the other titles, it looks like I won't come up with anything new.'

Therefore, not only did this participant not explicitly conduct the step of corroborating the information through an independent source; he also did not recognise a conflict in what he read. This indicates that recognising conflicts is a non-trivial task. Direct attention must be given to interpreting the information, remembering what had been read in other articles and recognising the conflict. In the electronic environment, this task is particularly challenging because only one report can be viewed at a time on the computer screen. Furthermore, in addition to the difficulty



Briefing: "that guidance system, the length of time that it operated, actually interfered with the inertial guidance system which took over after the launch and it confused…they confused each other and decided that they have to reset but by that time the rocket wasn't vertical anymore"

Article Date/Content	*Verbalization*
July 5, 1996 (Report 858): *Ariane 5 lifts off much faster information exhausted the temporary memory (buffer) capacity both systems simultaneously declared themselves to be in an irredemiable error situation and commenced a reset procedure when the system was reset, the vehicle s position at that time was adopted as the reference base*	"It's the same system as used on the Ariane 4, but the Ariane 5 takes off faster, much faster, than the Ariane 4. The two inertial guidance systems confused each other. They tried to reset at 37 seconds. It wasn't vertical anymore. It just totally lost its mind…so it couldn't figure out its direction."
September 16, 1996 (Report 1385): *the active inertial reference system transmitted essentially diagnostic information to the launcher s main computer, where it was interpreted as flight data and used for flight control calculations*	(talks about a different issue - how it could have been avoided through testing)

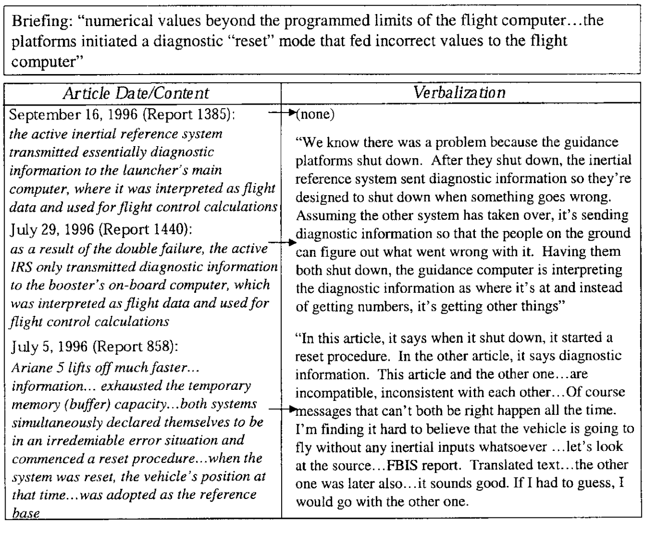
Fig. 9. Participant 6's process trace on why the rocket swivelled.

of determining conflicts in data that is read, it is even more difficult to detect data conflicts in unopened documents.

In contrast, participant 7 described the cause of the abnormal rocket swivel as diagnostic information interpreted as command data (Fig. 10). This explanation was incompatible because participant 7's description said that there was no command data at all because the guidance platforms had shut down whereas participant 6's description said that there was command data, just that it was inaccurate because the guidance platforms had been reset mid-flight.

Participant 7 recognised the conflict in the descriptions in documents 858 and 1440 and resolved it based on a judgement of source quality. He decided to base his analysis on the description in 1440 because it was later and therefore more likely to have more accurate information, not translated and from a more authoritative source. Note, however, that even though this was the accurate judgement to make, he did not notice that a previously opened article corroborated the hypothesis that he selected, which would have made the judgement easier. This would have been particularly helpful in this case because, as he pointed out: '[The inaccurate description] sounds good.' The inaccurate description was written in a way that sounded as if the reporter had sufficient technical expertise to understand the cause in detail. If he had only read article 858 and not detected the conflict, he likely would have believed the inaccurate description.

In blood banking, Guerlain et al (1999) found that experts in antibody identification routinely collect independent, converging evidence to both confirm the presence of hypothesised antibodies and to rule out other potential antibodies. When asked, the study participants described and demonstrated similar strategies to protect against the vulnerability of incorporating inaccurate information in their analytic products. During the simulated task, however, the study participants did not use or used greatly reduced versions of these strategies, and similarly described that under high workload conditions they tended to drop this in the workplace as well. One likely explanation is that the strategies were highly resource-intensive, such as printing out and iteratively using highlighter pens on specific themes to check that information was corroborated from multiple, independent sources. In addition, these strategies were generally not easy to perform within the electronic environment. These observations point to design concepts that would allow the easy manipulation, viewing and tagging of small text bundles, as well as aids for identifying, tracking and revising judgements about relationships between data.



Briefing: "numerical values beyond the programmed limits of the flight computer…the platforms initiated a diagnostic "reset" mode that fed incorrect values to the flight computer"

Article Date/Content	*Verbalization*
September 16, 1996 (Report 1385): *the active inertial reference system transmitted essentially diagnostic information to the launcher's main computer, where it was interpreted as flight data and used for flight control calculations*	(none)
July 29, 1996 (Report 1440): *as a result of the double failure, the active IRS only transmitted diagnostic information to the booster's on-board computer, which was interpreted as flight data and used for flight control calculations*	"We know there was a problem because the guidance platforms shut down. After they shut down, the inertial reference system sent diagnostic information so they're designed to shut down when something goes wrong. Assuming the other system has taken over, it's sending diagnostic information so that the people on the ground can figure out what went wrong with it. Having them both shut down, the guidance computer is interpreting the diagnostic information as where it's at and instead of getting numbers, it's getting other things"
July 5, 1996 (Report 858): *Ariane 5 lifts off much faster… information… exhausted the temporary memory (buffer) capacity…both systems simultaneously declared themselves to be in an irredemiable error situation and commenced a reset procedure…when the system was reset, the vehicle's position at that time…was adopted as the reference base*	"In this article, it says when it shut down, it started a reset procedure. In the other article, it says diagnostic information. This article and the other one…are incompatible, inconsistent with each other…Of course messages that can't both be right happen all the time. I'm finding it hard to believe that the vehicle is going to fly without any inertial inputs whatsoever …let's look at the source…FBIS report. Translated text…the other one was later also…it sounds good. If I had to guess, I would go with the other one.

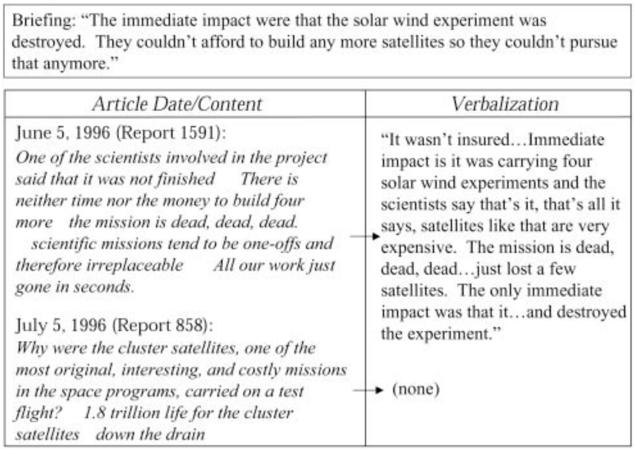
Fig. 10. Participant 7's process trace on why the rocket swivelled.

5.4.3. Relying on Outdated Information

The third source of inaccurate statements was outdated information that at one time was considered correct but then later was overturned when new information became available. This type of inaccurate information was the most difficult to detect and resolve. Because the 'findings' or data set on which to base an analysis came in over time, there was always the possibility of missing information that would overturn or render previous information 'stale'. This occurred both for descriptions of past events where the information about the event came in over time as well as for predictions about future events that changed as new information became available on which to base the predictions. When these updates occurred on themes that were not central enough to be included in report titles or newsworthy enough to generate a flurry of reports, it was very difficult to know if updates had occurred or where to look for them.

To illustrate how easy it is to rely on outdated information, consider study participant 6's conclusion that the Cluster satellite programme would be discontinued as a result of the Ariane 501 accident (Fig. 11). This is an inaccurate statement because the Cluster programme was fully reinstated, although the prediction was accurate at the time.

Essentially, participant 6 did not open any documents that contained updates on the impact to the Cluster satellite programme. The participant opened seven documents. Only two of the documents contained descriptions that predicted what the impact to the Cluster satellite programme as a result of the Ariane 501 failure would be. In the first description, a scientist working on the project directly stated that the project would be discontinued. While reading this report, the participant verbalised that the scientific mission was dead and that the experiment was destroyed. The second description was more vague about the impact and does not directly make any predictions, but could be viewed as weakly converging evidence that the



Briefing: "The immediate impact were that the solar wind experiment was destroyed. They couldn't afford to build any more satellites so they couldn't pursue that anymore."

Article Date/Content	Verbalization
June 5, 1996 (Report 1591): *One of the scientists involved in the project said that it was not finished There is neither time nor the money to build four more the mission is dead, dead, dead. scientific missions tend to be one-offs and therefore irreplaceable All our work just gone in seconds.*	→ "It wasn't insured…Immediate impact is it was carrying four solar wind experiments and the scientists say that's it, that's all it says, satellites like that are very expensive. The mission is dead, dead, dead…just lost a few satellites. The only immediate impact was that it…and destroyed the experiment."
July 5, 1996 (Report 858): *Why were the cluster satellites, one of the most original, interesting, and costly missions in the space programs, carried on a test flight? 1.8 trillion life for the cluster satellites down the drain*	→ (none)

Fig. 11. Participant 6's process trace on the impact to the satellite programme.

Cluster satellite programme would be discontinued. It is no surprise that the participant included in the verbal briefing a description similar to the one from the 5 June 1996 article that the experiment was destroyed and that the programme would no longer be pursued.

As a result of basing an analysis on 'stale' information that had been turned over by later updates, study participants made several inaccurate statements at varying levels of importance. The vulnerability to missing critical information is particularly troubling because it is so difficult for practitioners to determine when they have missed critical information. It is the *absence* of information, either from not sampling the information or having attention directed on a different theme while reading a document, that creates the vulnerability. A strategy of corroborating information from two independent, authoritative sources would likely eliminate the first two sources of inaccuracies but would not eliminate the third source.

Few study participants specifically looked for updates or described strategies to do so. It is possible that this observation has training implications, although many study participants verbalised that reports immediately following the event lacked information that came out later. Perhaps few strategies have been developed to deal with this vulnerability because the problem is intractable with current support tools. Updates could be reported hours, days, weeks, months or years after an event. Updates on minor themes did not generate a flurry of reports and were not reflected in the date/title view of the reports. It is possible that 'agents' that suggest targeted query formulations and/or 'seed' representations with updates on a theme would be useful. This is particularly true if advances in natural language processing would enable a machine to reliably recognise updates on a theme and instances where a previous interpretation is overturned.

6. DISCUSSION

By observing experienced intelligence analysts perform a relatively complex, face valid task using a baseline set of querying and browsing tools, we were able to greatly increase our understanding of the challenges of intelligence analysis under data overload. When study participants performed a time-pressured analysis outside their base of expertise based on sampling reports from a large set, some made inaccurate statements in verbal briefings. Participants that made no inaccurate statements spent more time during the analysis, read more documents and relied on higher-quality documents than participants who made inaccurate statements. All participants missed potentially available relevant information and had difficulty detecting and resolving data conflicts. Sources of inaccurate statements

Table 6. Summary of observed behaviour and design implications

Observed behaviour	Design implications
Some wanted to know what was in the database (but did not characterise it)	Visualisations for interactive, real-time exploration of information attributes of sets
Analysts used primitive search strategies	Training; Use search intermediaries; Design interface with affordances for recommended search tactics (e.g., narrow by document attributes, faceted search)
All opened documents (4–29 documents) from a single query (22–419 documents)	Improve query formulation usability and workspace design; Visualisations to browse larger document sets
Participants who made inaccurate statements did not read high-profit documents; All participants missed some high-profit documents	Suggest 'high-profit' candidates; Algorithms to find 'same as' documents
Some missed critical events	Event recognition algorithms based on flurries of reports in short time; Overview visualisation where holes in story are visible
Some forgot sources for selected text; Time-intensive strategies to track source information	Link text to source document; Identify duplicates from the same source
Some missed data conflicts	Support identifying and tracking conflicts; 'Bookmark' function; Suggestions for new queries to find conflicting or corroborating data
Some missed updates	Suggest 'update' candidates; Support tracking updates on a theme; Suggestions for new queries to find updates on a theme

were: (1) relying upon default assumptions; (2) incorporating inaccurate information; and (3) incorporating information that was considered accurate at one point in time. These findings and associated design implications are summarised in Table 6.

These observations indicate that baseline tools do not adequately support analysts in meeting the challenges of performing inferential analysis under data overload, leaving them open to making inaccurate statements and missing critical events. The understanding that was gained from this moderately high-fidelity simulated task is much richer and more detailed than was gained from prior interviews. In addition, we were able to gain insights that we might have missed if we had controlled more variables in order to establish causal relationships. For example, we might have missed the importance of 'key' documents during analysis, the correlation between participants that used the high-profit documents in the database as their key documents and the amount of time that they spent and the number of documents that they opened, process breakdowns while corroborating information, and the vulnerability of missing updates even when using high-quality documents. This understanding provides valuable insight for what might be useful support tools, as well as variables to pursue in a more targeted way in follow-up design intervention studies.

In addition, these findings point to a set of challenging design criteria that human-centred solutions to data overload in intelligence analysis should meet in order to be useful. These criteria can serve, not only to guide the next cycle in design, but also to generate scenarios to test the effectiveness of any proposed organisational, design or training intervention:

1. Bring analysts' 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 'high-profit' documents, data that indicates an escalation of activities or a disrupting event and data that deviates from expectations.
2. Aid analysts in managing data uncertainty. In particular, solutions should help analysts identify, track and revise judgements about data conflicts and aid in the search for updates on thematic elements.
3. 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.

Although these criteria might appear obvious, note that they are different from criteria that are implicitly assumed in many proposed solutions. Alternative criteria include (1) to have an analyst be able to read it all, (2) to find the relevant information needed to perform an analysis, (3) to visualise the landscape of the information space, (4) to have the machine tell an analyst when an important message is received, (5) to see an overview of events in an area that have not been monitored for some time, and (6) to have the machine summarise the important points in each message. Our evaluation criteria are interesting, in part, because they are more difficult to meet than these alternatives. Meeting these criteria will likely require innovative design concepts rather than simple, straightforward adjustments or feature additions to current tools.

Acknowledgements

We thank the study participants for donating their valuable time and expertise, and in particular Jay Finley for his dedication to this effort. We thank Janet Miller for sharing her findings from a cognitive task analysis (CTA) of air force intelligence analysts and David Tinapple for creating an animated visualisation of design concepts. Funding was provided by the Human Effectiveness Directorate of the Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio, with exceptional support from Gil Kuperman. Additional funding was provided by a National Science Foundation Graduate Fellowship. Any opinions, conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of the National Science Foundation.

References

Bates MJ (1979). Information search tactics. Journal of the American Society for Information Science 30:205–214.

Blair DC (1980). Searching biases in large interactive document retrieval systems. Journal of the American Society for Information Science 31:271–277.

Blair DC, Maron ME (1985). An evaluation of retrieval effectiveness for a full-text document retrieval system Communications of the ACM 28(3):289–299.

Guerlain S, Smith PJ, Obradovich J et al (1999). Interactive critiquing as a form of decision support: an empirical evaluation. Human Factors 41(1):72–89.

Josephson J, Josephson S (1994). Abductive Inference. Cambridge University Press, New York.

Miller J (1960). Information input overload and psychopathology. American Journal of Psychiatry 116:695–704.

Mizzaro S (1997). Relevance: the whole history. Journal of the American Society for Information Science 48(9):810–832.

Olsen KA, Sochats KM, Williams JG (1998). Full text searching and information overload. International Information and Library Review 30(2):105–122.

Patterson ES, Roth EM, Woods DD (1999). Aiding the intelligence analyst in situations of data overload: a simulation study of computer-supported inferential analysis under data overload. Institute for Ergonomics/Cognitive Systems Engineering Laboratory Report, ERGO-CSEL 99-TR-02, Ohio State University, Columbus, OH.

Saracevic T, Kantor P, Chamis AY, Trivison D (1988). A study of information seeking and retrieving (3 parts). Journal of the American Society for Information Science 39:161–216.

Schum DA (1994). The evidential foundations of probabilistic reasoning. Wiley, New York.

Shute SJ, Smith PJ (1992). Knowledge-based search tactics. Information Processing and Management 29(1):29–45.

Woods DD (1993). Process tracing methods for the study of cognition outside of the experimental psychology laboratory. In Klein G, Orasanu J, Calderwood R (eds). Decision making in action: models and methods. Ablex, Norwood, NJ, pp 228–251.

Correspondence and offprint requests to: E. S. Patterson, Cognitive Systems Engineering Laboratory, Institute for Ergonomics, Ohio State University, 210 Baker Systems, 1971 Neil Ave., Columbus, OH 43210, USA. Email: patterson.150@osu.edu