Analysis tools in the study of distributed decision-making: a meta-study of command and control research

Ola Leifler and Henrik Eriksson

Ola Leifler and Henrik Eriksson, Analysis tools in the study of distributed decision-making: a meta-study of command and control research, 2012, Cognition, Technology & Work, (14), 2, 157-168.
http://dx.doi.org/10.1007/s10111-011-0177-4

Cognition, Technology, & Work manuscript No.
(will be inserted by the editor)

Analysis tools in the study of Distributed Decision-Making

A Meta-Study of Command and Control Research

Ola Leifler · Henrik Eriksson

Abstract Our understanding of distributed decision making in professional teams and their performance comes in part from studies in which researchers gather and process information about the communications and actions of teams. In many cases, the data sets available for analysis are large, unwieldy and require methods for exploratory and dynamic management of data. In this paper, we report the results of interviewing eight researchers on their work process when conducting such analyses and their use of support tools in this process. Our aim with the study was to gain an understanding of their workflow when studying distributed decision making in teams, and specifically how automated pattern extraction tools could be of use in their work. Based on an analysis of the interviews, we elicited three issues of concern related to the use of support tools in analysis: *focusing* on a subset of data to study, *drawing conclusions* from data and *understanding tool limitations*. Together, these three issues point to two observations regarding tool use that are of specific relevance to the design of intelligent support tools based on pattern extraction: *open-endedness* and *transparency*.

Keywords command and control, text analysis, interview study, exploratory sequential data analysis

1 Introduction

Understanding and assessing the performance of professional decision makers who operate in teams is central to training and research purposes. For training purposes, it is critical to understand how command teams work in order to help them improve their performance in crisis management. For researchers in decision making and team cognition, the context of command and control (C^2) offers fertile ground for studying the fundamental processes involved. In both cases, however, there are great technical challenges.

The technical challenges when studying distributed decision making in C^2 or similar settings concern setting up the proper instrumentation and subsequently collecting sufficient data to enable the study of the selected aspect of the command team (Andriole, 1989). Due to the amounts of data generated in these settings, researchers need flexible methods for studying data sets that can accomodate different approaches to visual representation and reasoning. Researchers C^2 study decision processes with different goals, perspectives, and methods. Common to most studies of distributed decision making, however, is the issue of how to efficiently extract useful information from large assemblages of research data. The objective of this study is to provide a better understanding of the tasks involved in studying distributed decision making for the purpose of eliciting design criteria for support systems based on automated pattern extraction systems. Our study has been based on the hypothesis that such pattern extraction systems can provide significant contributions to the process of analyzing decision-making processes in teams.

1.1 Outline

The remainder of this paper is organized as follows: Section 2 provides a background to C^2, research methods in data analysis and tools for generating and verifying hypotheses about C^2 team performance. In Section 3, we present the method used during the interviews and the main themes discussed. Section 4 presents the main results from the interview study, organized around three critical issues that

O. Leifler · H. Eriksson
Dept. of Computer and Information Science
Linköping University
SE-581 83 Linköping, Sweden
Tel. +46 13 281000
E-mail: {olale,her}@ida.liu.se

emerged in our analysis of the interview results. Finally, Section 6 concludes this paper with a description of the critical issues and design criteria for tool support that are of relevance to the design of automated pattern extraction systems.

2 Background

In this paper, we study how researchers conduct analyses of decision making in command teams. Here, C^2 is used as a term for describing what people commanding others do: directing the work of subordinate units and coordinating their efforts toward a common goal (command), making sure that orders are carried out and monitoring the outcome of all actions (control).

Research in C^2 during the last 15 years has been stimulated by visions of the affordances of new technology (Alberts et al., 2000) and new results on what it means when people command others (Klein, 1998). In contrast to the traditional view of C^2 as a highly structured process of identifying crisp problems, generating a set of options for solving them and finally determining a course of action according to statically defined notions of utility and risks, newer model of decision making from the naturalistic decision making paradigm take into account the dynamic interplay between people, technology, organizations and tasks when managing problems in command Klein et al. (1993); Stanton et al. (2008); Jensen (2009) .

According to results from naturalistic studies of decision making, decision makers rarely set up and evaluate multiple plan options explicitly and may even be better of for it, as they are more involved with building an understanding of a complete situation (not just a problem), and communicating this understanding to others to form an common intent and monitoring their environment for relevant changes (see e.g., (Jensen, 2009)). Several models or frameworks for reasoning about commanders' activities have been proposed with varying foci and backgrounds. Shattuck and Woods (2000) describe command as the process of establishing common intent and monitoring whether subordinate units act according to this intent, Brehmer (2005) expands on the cybernetic control loop of Boyd to accomodate more aspects of interplay and feedback in the command loop, and Stanton et al. (2008) provide a model, divided in reactive and proactive activities in C^2 settings, that is grounded in empirical observations of command teams from several domains and attempts to bridge the conceptual differences of several other models. These new appreciations of what characterizes C^2 has resulted in an increased interest in methods for studying what it means to successfully make sense of the environment and communicate intent.

2.1 C^2 Research

To measure how well commanders perform their key functions, and consequently measure the effect of C^2 functions, researchers have adopted methods from cognitive psychology for training staff and reasoning about staff work, such as role-playing simulations and the use of micro-world simulations.

In role-playing simulations, the roles and responsibilities of those studied are similar to naturalistic settings which facilitate the study of group dynamics between staff (Rubel, 2001). As part of a role-playing simulation, a group of staff members assemble to form a team assigned with a task that they are likely to encounter in their profession. When operating in a role-playing simulation environment, participants play a scenario with much the same tools they would be expected to use in real situations which means there are many data sources from the simulated environment available for analysis afterwards. However, important parts of a team's work may be performed as part of discussions between members of staff, and thus, aspects of their work may not be well articulated in terms of interactions with the computer systems at all. To capture these important aspects of staff work, researchers typically complement simulation logs with human observers that periodically give accounts of what the staff are doing and evaluate their performance with respect to predefined categories (Thorstensson et al., 2001; Jenvald and Eriksson, 2009). Different human observers tend to tag events differently, however, reducing the reliability of human observations.

Micro-world simulations often form the basis of role-playing scenarios as they combine the controlled environment of laboratory experiments with the domain fidelity of natural settings for decision making (Wærn and Cañas, 2003). They have been used both for research (Johansson et al., 2003) and training (Kylesten and Nählinder, 2010) and in both settings, they have demonstrated that they can provide key insights on decision making.

When researchers set up and analyze simulation-based scenarios, they typically follow a set of procedures for determining appropriate instrumentation and analysis tools (Morin, 2002). As the data sets are usually rather large and diverse, containing both observer logs, simulation logs, communication data, screen recordings and possibly even video recordings of the participants, researchers must use exploratory, data-driven methods in their studies. Understanding more exactly *how* they focus their search for patterns in such data sets was the justification for the interview study presented in Section 4.

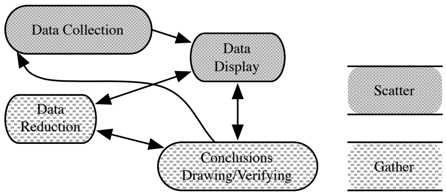

Fig. 1 Four stages in qualitative data analysis, adapted from Miles and Huberman (1994), compared to the scatter/gather paradigm of data mining.

2.2 Data analysis methods

The process of analyzing data after an exercise can be characterized in several ways. One way is to describe it is by using the model of Miles and Huberman (1994) with four stages of qualitative data analysis (Figure 1). The description presents four stages with the interplay between them. How data is selected affects the display of data, and the conclusions drawn guide further explorations (reduction/collection/display). In research on distributed decision making in command and control, the iterative nature of Miles' and Huberman's description resonates well with the naturalistic decision research paradigm.

Another common characterization of research processes such as those involved in the analysis of simulation-based scenarios is exploratory sequential data analysis (ESDA) (Sanderson and Fisher, 1994). In ESDA analyses, researchers devise indicators of team performance given the outcome of the scenario at hand, which are used the next time there is a similar exercise to focus the goals of the exercise and direct the data collection. This process leads to successively improved understanding of team performance, methods for analysis, and tool use. Sanderson and Fisher (1994) describes data collection and analysis in ESDA as a process in which the researcher iteratively selects parts of the available data sets (typically logs and recordings), conducts analyses of the transformed products that are derived from the data sets (typically transcribed speech and annotated events) and uses the results to guide the selection of another set of data to study closer until conclusions can be drawn. Figure 2 shows these two processes, guided by a set of formal concepts or questions.

As an example, one researcher described how she was initially tasked to study role assignments within an emergency management team when only some of the team members had the competencies required by the mission. The focus on improvisation lead to he selection of the communication to and from the two team members who were assigned tasks that they were not initially trained for. Later, their communications were analyzed with respect to annotation schemas, that were in turn developed as part of the transcription process. As episodes of special interest were found, the researcher went back to video recordings to look

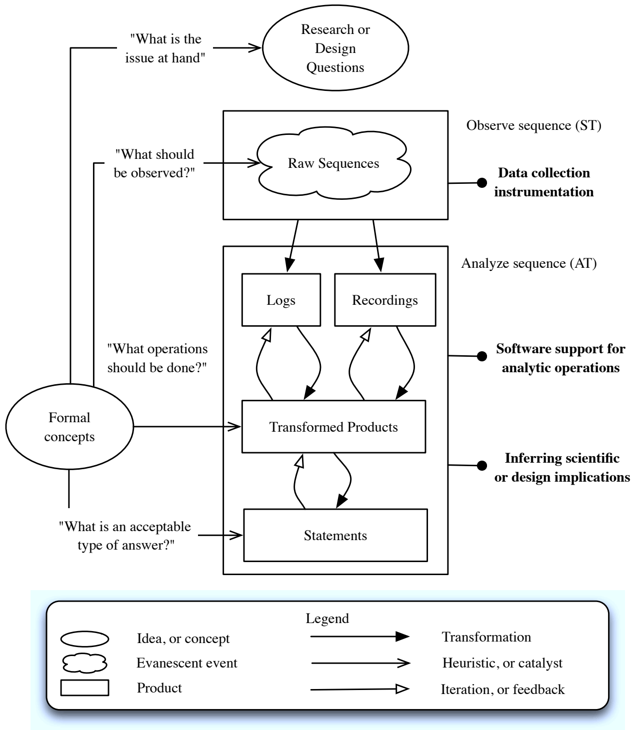

Fig. 2 Exploratory sequential data analysis, adapted from Sanderson and Fisher (1994).

for interactions that could reveal how the two team members got and interpreted information. With both transcribed speech and video logs, the researcher could make inferences about how the new assignments had affected the team members' abilities to assess information and do their job.

The ESDA process of data-driven hypothesis generation as well as Miles' description of qualitative data analysis closely resemble one from the data mining community, where Cutting has characterized the process as iterating two main activities: scattering and gathering (Cutting et al., 1992). In general, *scattering* is defined as the act of creating a set of distinct objects of study by using some metric for comparison of objects, and *gathering* as the act of treating some objects as similar according to some criteria. Scattering and gathering may be iterated so that objects considered similar may again be scattered according to some new metric and so on.

The scatter/gather paradigm could be viewed as an alternative formulation of the iterated work process in ESDA: when analyzing logs from a team scenario, researchers may use transcribed communications as an entry point to further analysis and annotate the transcribed text according to a certain annotation schema. The annotation schema creates a set of distinct objects of study in the form of a set of episodes, which in turn direct further analysis by making video logs or observer reports at specific points in time relevant to study.

Due to the similarities between these two paradigms, we had previously evaluated the technical soundness of using data mining methods in text analysis (Leifler and Eriksson, 2010) to support the exploration of patterns. We argue that, as the work process described by the ESDA model coincides well with the scatter/gather paradigm of how data mining tools are to be used, tools for data mining could probably fit the tasks in ESDA well. A pattern exploration tool based on machine learning techniques could possibly help researchers formulate hypotheses about connections between textual data and team performance, and we had conjectured that command and control research projects would have a better chance at capturing more of the factors that determine the outcome of command and control given better tools for reducing and analyzing large text-based data sets.

In our technical evaluations, we had found that, of all the meta-data found in communications from a set of scenarios, the message *texts* were the most significant factors when attempting to emulate human classifications by machine classification. Also, automatic text classification emulated human classifications well enough that we believed it to be justified to incorporate it in a support tool for scenario analysis. It was not clear from our technical evaluations exactly *how* such support systems should be devised, however, which made it imperative to conduct the interview study. We did not consider entirely automatic classification to be viable as a basis for a support tool, but rather hypothesized that data mining methods based on texts could provide guidance for searching for possible patterns in large data sets.

During the interviews, we probed specifically into the participants' use of data analysis tools for analyzing simulation scenarios, with the intent of understanding how such tools shape their work and how a data-mining-based support tool might be used to facilitate the search for significant patterns in team behavior or communication.

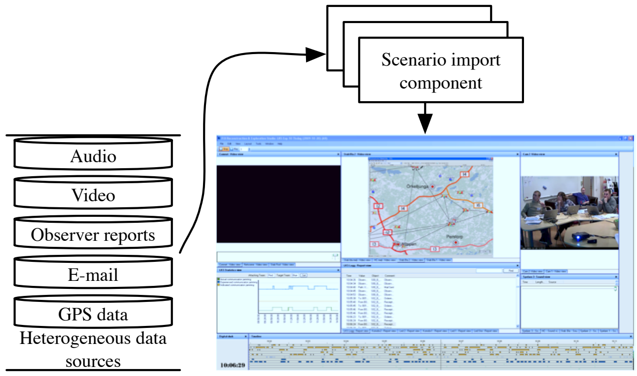

Fig. 3 An ESDA tool used for analysis of C^2 scenarios by the participants in the study.

2.3 Data analysis tools

In qualitative content analysis, researchers seek to categorize data from interviews and other sources in a common framework. The framework can either be taken from previous literature from the research field the study concerns (a priori coding), or be developed as part of the analysis (emergent coding) (Lazar et al., 2010). Especially when an emergent coding is needed for analyzing data, it can be very labor-intensive. To support the coding and analysis of communication data from command teams, ESDA tools can be used for understanding patterns in the sequence of messages exchanged between the members of a group.

In the scenarios we have studied, at least some of the researchers use ESDA tools capable of merging and viewing many different data sources. Figure 3 displays some of the capabilities of one such tool, F-REX, a re-implementation of the MIND (Thorstensson et al., 2001) system. In F-REX, MIND, and similar ESDA support tools, several heterogeneous data sources are imported after a scenario and made available as a series of events along a common scenario timeline. For every scenario, particular configurations can be made to emphasize a particular data source of importance to analysis by providing a specific layout of the graphical components. In the middle of Figure 3, a screenshot displays how screen-captured video, radio communications, text messages and other data sources are available through a graphical interface with a timeline at the bottom. A central aspect of the exploration phase during the analysis of a scenario is the analysis of team communications. With multiple actors involved in a scenario and several parallel courses of events unfolding, it is important that the sequence of communication events can be managed efficiently. For this purpose, it has proved to be especially useful to have tools such as MIND for annotating, searching and visualizing the flow of information (Albinsson et al., 2004; Morin and Albinsson, 2005).

The use of reconstruction and exploration tools has opened up new possibilities for researchers in formulating hypotheses about team performance since the amounts of data that can be treated has increased greatly. However, this has accentuated the problem of data reduction or, by using the terminology of Sanderson, the interplay between sequences of data and the transformed products one can create from them. F-REX/MIND typically make data available for direct, visual inspection and provide direct navigation facilities along a timeline together with annotation of events. Other tools such as MacSHAPA (Sanderson et al., 1994) are more directed to support the manual categorization of communications but are typically less capable with respect to managing large sets of heterogenous data sources. To facilitate the process of reducing data sources to manageable and comprehensible chunks, researchers have devised tools for visual exploration of patterns (Albinsson and Morin, 2002) to find critical incidents by using explicitly available attributes of communications to elicit patterns.

Apart from dedicated ESDA support tools, many researchers including those in our study, use other support tools as well. Some use specialized factor-analysis software tools for LISREL analyses (Jöreskog, 1973) or multidimensional scaling (Kruskal, 1964), but it was also very common to use plain spreadsheets for creating and manipulating communications categories. Using the workflow description of Miles, tools for factor analysis support verification and conclusions drawing stage of the process but not the other stages. ESDA tools are typically stronger on data collection and data display, but weaker on data reduction and conclusions drawing, whereas spreadsheets can be said to support data display and data reduction but not data collection, as the collection and adaption of data needs to be done in advance to suit the cell-based data representations in spreadsheets.

2.4 Pattern extraction

In our previous study on classification of messages and observer reports from decision-making scenarios, we found text clustering in particular to be useful as a pattern extraction technique (Leifler and Eriksson, 2010). Text clustering can be used to relate texts to one another based on *distance metrics*. Such distance metrics, when suitably used in a framework for clustering texts based on them, can be used to guide a manual search for patterns between texts and terms in texts (Rosell and Velupillai, 2008). Although the existence of statistically valid patterns in texts may not be enough to draw conclusions, it may be of great help in finding related sets of texts that are difficult to find otherwise (Rosell and Velupillai, 2008).

3 Research method

In this study, we conducted a series of semi-structured interviews with command and control researchers who study distributed decision making to understand their work process and the tools they use for support. We also investigated data from three different exercises they had been involved in and the current tools used to process these data. The researchers were from different backgrounds but were all involved in C^2 studies.

In all, we performed a series of 8 interviews with people involved in C^2 analysis. The participants had experience from conducting research on C^2 and training staff and they were interviewed to establish how they perform data analysis with a particular focus on how they find patterns in text-based data sources. All interviews were semi-structured and conducted using critical-incident interview techniques (Flanagan, 1954) where we probed for situations in which the participants had engaged in critical and specific activities typical to the analysis of large data sets.

Their work was studied through interviews because the work they carry out is infrequent and distributed over longer periods of time which, on a practical level, makes it difficult to investigate the context of their work over a given period in a more situated manner. Semi-structured interviews were therefore considered a valuable tool for extracting information about their methods when analyzing data from decision making studies and their relationships to tools used for such analyses. Each interview lasted approximately 60 minutes and followed a script in which three main themes were discussed:

- *What is the purpose of conducting analysis of team communications and behavior?* The participants were given a chance to elaborate on the purpose and nature of exercises or experiments they had been involved in and how those purposes directed the analysis of scenario data.
- *How exactly is analysis performed?* The participants were asked to answer this question by relating to their own personal experience from one or more scenarios in which they had conducted analysis. Depending on the role of each participant, they had either been involved in the analysis of scenario outcomes or the construction of support tools for such analyses.
- *What are the most time-consuming, challenging or precarious stages of analysis?* In describing their work with analyzing data, all participants were asked to elaborate specifically on challenges with communication analysis, with respect to tool usage as well as other factors, as such analyses are commonly used to establish patterns in team behavior.

Although there was an initial, clear focus on communications from us, other types of data, including reports from observers, questionnaire data and simulation logs were brought up by the participants and discussed during the interviews.

4 Interview study

The participants were chosen based on their expertise in the analysis of communications and development of technical systems for the support of such analyses. The people interviewed are listed in Table 1, where their names have been anonymized.

Jane, 28, (M.Sc. in cognitive science) had experience from one study concerning team improvisation. She had used a tool similar to F-REX/MIND during her analysis along with spreadsheets for categorizing communications.

Charlotte, 63, (M.Sc. in behavioral science) had more than fifteen years' experience in doing systems evaluations

Table 1 The roles of people interviewed regarding communication analysis in C^2.

Person	Role
Jane	Junior C^2 researcher
Charlotte	Senior C^2 researcher
John	Research project leader
James	Junior C^2 researcher
Charlie	Senior C^2 researcher, C^2 training expert
Sebastian	Junior C^2 researcher
H-G	Senior technical researcher
Freddy	Junior technical researcher

and studying methods for team training. The latest project concerned a usability evaluation of the decision support system. She has used special tools for LISREL analyses and other factor analyses as well as spreadsheets.

John, 36, (Ph.D in cognitive science) had five years' experience of doing team communication analyses with a particular focus on the shared situation awareness of a team. He had experience from using ESDA tools such as F-REX as well as spreadsheets in his work.

James, 33, (Ph.D in informatics) had planned and managed team exercises for studying C^2 in crisis management and specifically inter-organizational aspects of collaboration. He had participated in several exercises with command staffs. Mostly, his work was conducted with spreadsheets and stand-alone tools for transcribing and annotating speech.

Charlie, 43, (Ph.D. in computer science) had ten years of experience with planning, conducting and analyzing team training, during which both spoken and written communication was logged and analyzed. His work mostly concerned training teams. He had used ESDA tools extensively to process and present information from exercises he had managed.

Sebastian, 34, (M.Sc. in cognitive science) had conducted one communication analysis of a tactical scenario in which fighter pilots collaborated in dog fight scenarios against an opponent team. The analysis was conducted using transcription tools and spreadsheets.

H-G, 34, (M.Sc. in computer science) was a computer science researcher working with ESDA support tools for C^2 analysis and had participated in constructing the tools used by researchers in their analyses of C^2 scenarios.

Freddy, 29, (M.Sc. in computer science) was a computer scientist who, like H-G, worked with support tools for communication analysis and had a special interest in exploring patterns in communication data.

4.1 Interview analysis

All interviews were recorded and transcribed. The interviews were in Swedish and the quotes that follow have therefore been translated to English. They were divided in dialogues for clarity and annotated using the main categories of *purpose*, *method*, *tools* and *challenges*, which related to the themes of the questions during the interviews. The annotated materials were later tagged with sub-categories for each category. Later, the sub-categories with special relevance for the interplay between tools and methodological issues were selected for further analysis. We selected the critical incidents mentioned by the participants in which these sub-categories appeared and grouped these together as three *issues of special concern* related to the use of technical systems in research.

The first issue concerned the activity of *focusing* the research question, restricting the overall research question to a specific question or selecting a subset of the available research material for further study. The second issue concerned how researchers *draw conclusions* from data and how different representations of the material help them in this work, and how the representations and tools they use affect conclusions they draw. The third issue concerned their *understanding of limitations* in tools and data used.

4.2 Focusing

The participants stated that the goals in their studies were concerned with either understanding how people work when they solve a particular task or evaluate their performance. One of the participants, Jane, described a research project on the behavior of a command team in the absence of specific competencies. Her task was primarily to study the general behavior of the team under the specific conditions of the study. However, *the amount of data* recorded soon restricted what could be studied.

> The reason why I chose not to [study other aspects] was that there was such a huge amount of material. [. . .] it was an issue of time as well. *(Jane, line 565–567)*

The type of data to be recorded during the exercise studied by Jane was known in advance, but it was only during the initial analysis that they saw how much data had been collected and decided to restrict what was to be studied. They focused the study by removing all communication trails that did not have to do with two specific staff members who were assigned duties they were not trained for, with no regard to what those communication trails could include. Sebastian described a similar situation in which he was tasked with analyzing pilot communications from a fighter pilot training session. As in Jane's case, the amount of data made it necessary to reduce the scope of the analysis to narrow time frames prior to when weapons were launched:

> sometimes [the course of events] go quite fast in these situations and sometimes things are rather slow [...] so then we chose a minute before the shot as a good compromise. *(Sebastian, line 371–375)*

The one minute time frame was chosen by Sebastian and his supervisors based on experience, though he could not remember any more details regarding the rationale, why one minute before would be better than two minutes, or one minute before and one after. The reason why a portion of the communication logs would be relevant for analysis was because of a hypothesized relation between the communication, the situation awareness in the team and the joint performance of the team, which was assessed by both performance metrics in the simulation as well as subjective evaluations by a senior officer.

> [The participants] have to be synchronized and you have to coordinate every little step like this and be sure that others have understood. I mean, before you move on with the next part of the procedure. *(Sebastian, line 408–413)*

All efforts at focusing a study by selecting subsets of the communication data to analyze were conducted with the intention that the data studied should be an optimal subset for correlating team performance with communications. In Charlie's interview, he explained that when distilling a set of reports into a few key observations to highlight during team debriefing after training, he narrowed down the set of available observations by including only those that had enough reflective remarks in them and selected those reflections that matched his own subjective evaluations. So, the agreement between his own impressions and the data collected by others was vital for his conclusions.

Focusing the communications analysis effort towards a particular question was described in the interviews as a process guided either by experts who had their own hypotheses regarding the parts of team communications that were relevant, or as a result of manual work transcribing communications. In focusing their work, the participants used mostly spreadsheets for collecting communication trails and categorizing communications according to a number of categories. They mentioned the use of ESDA tools mostly in the context of clarifying what someone was talking about in cases where a spoken conversation related to objects visible on a computer screen or on a table. When asked if there were specific theories guiding the hypothesis generation, the participants answered by relating to doctrine (Jane), personal evaluations and experience (Charlie) or subject expert evaluations (Sebastian). The tool support available enabled them to create tables with communications and classify those communications according to categories, but there was no automatic support for extracting statistical patterns from communication features such as message length, direction of messages or co-related terms.

4.3 Drawing conclusions

The second issue of special concern during the interviews was how they drew conclusions from the data they had selected for analysis. Jane explained specifically how and when she could draw conclusions regarding the effect of competence loss on staff performance. She had manually annotated episodes (conversations, threads) in the communication flow when the staff members talked about a specific topic which they had no prior experience in dealing with and then began to search for how they had managed that lack of competence.

> [My assessment of their performance] probably started with these somewhat obvious errors [...] when he informed [his colleagues] incorrectly given the directions from the scenario management team. Then I could see more things that had gone wrong. [How I reached my conclusions] is difficult to say, it was a process and difficult to remember exactly. *(Jane, line 732–737)*

Another participant noted that the process of understanding the communication structure of a team became obvious given annotated communication and the amount of communication sent in the different annotation categories. No further help was needed to understand the structure of a team's communication given figures extracted from the amount of messages in each of the categories used.

The participants described the process of drawing conclusions as one in which researchers narratively describe what has happened in a scenario which has transpired: who communicated with whom, what actions were taken and so forth. A generalization of such a narrative may generate a model of people's responsibilities, key communicators, pivotal events and typical responses to those events, maybe in the form of a team workflow. Relating a model of team communications to team performance was described as rather difficult. A descriptive model can be written in many different ways, depending on the aspects the researcher is interested in describing. A prescriptive model, that is, one directly related to team performance, needs to model the aspects of teamwork that most concretely influence some type of performance. Establishing a metric for performance is maybe the most challenging task in creating a model for prescribing team performance.

However, all participants also noted that assessing performance in a joint team is very challenging due to the difficulty in describing the nature of what a team does, and that performance assessments therefore tend to be concerned with metrics that are constructed to be simple to measure, and that those measurements can be used together with others to triangulate some understanding of the concept of performance. In one case, John described how they studied the effects of command styles on communication and performance:

> We have looked at the frequency of direct orders [in this simulated environment] versus communication of intent [and] by looking at that you get an idea of what kind of style the commanders in this exercise have. If you could do statistical analyses and establish a connection between a particular style and performance that would be very exciting. We have not been able to do that. *(John, line 69–78)*

Correlating that which can be analyzed to the more elusive concept of team performance is challenging. Written communication is an accessible form for analysis, in contrast to video and other media. It is therefore natural that the researchers look for patterns in such data to guide their analyses. However, because performance measures are difficult to specify clearly enough to be measured unambiguously, researchers do one of two things. Either, they measure success by a proxy variable (communication style), with an hypothesized but unverified connection between the proxy variable and the outcome of the scenario. Alternatively, performance is defined subjectively by experts, which can lead to difficulties if the reasoning conducted by the expert is not well understood. In a study Sebastian participated in, experts helped him construct a classification scheme which would identify problems believed to be associated with low team performance. The subject experts were also involved in establishing the performance gradings for the teams.

> There is a positive correlation between the number of [communication] problems per minute and grading, that is, the more problems the higher grade. There is another [communication] category which is higher for the best team and that is "unclear information". *(Sebastian, lines 1179–1181, 1188–1189)*

The relationship between communication issues and scenario outcome could not be determined as the expert had suggested. During the interview, Sebastian discussed the possibility that increased communication of problems could indicate a willingness to discuss issues instead of avoiding them, a willingness which might be positive to the joint situation awareness of the team.

The descriptions all interview participants gave of the process of drawing conclusions from data centered on the representations used. John noted that the tabular representations of *messages and categorizations of messages* directly led to conclusions regarding communication style. Jane could not pinpoint exactly when she could draw conclusions regarding how the team had managed performances, but she reasoned in terms of how the communication already selected for analysis was *color-coded in episodes* (dialogues, communication threads) and how those episodes formed a direct basis for drawing conclusions. All representations cited in the interviews were the direct result of using spreadsheets and some paper calculations. In the process of drawing conclusions from data, existing ESDA tools were only cited as useful in training, where Charlie explained how they used their ESDA tool similar to F-REX for after-action debriefing of staffs they trained in emergency management. The visual, integrated presentation was the central aspect of using an ESDA tool in their case and a source for reasoning about joint staff behavior.

4.4 Understanding tools and data

Several researchers had used F-REX or similar ESDA tools for data exploration that allowed multiple data sources to be shown simultaneously. ESDA tools can offer significant advantages for analysts when they search for key events in a scenario with many data sources, but several of those interviewed described that the tools are difficult to use for communicating results.

> Even if you as an analyst [understand the tool], it is difficult to demonstrate [your results] in a good way to someone else without access to the tool itself. *(John, line 181–183)*

Having data available as tables in a spreadsheet was considered much easier when communicating results to others. Especially for categorizing messages, sorting, selecting them and presenting simple summations of results, spreadsheets fill many roles in research.

> You work a lot with software that makes it easier to sort and mark things, so it has become a lot of Excel and then I can have a column next to [the messages] where I enter their category codes. [...] When it comes to visualization you often use Excel because then I can create my tables right there and show them. *(John, line 494–496)*

The visualization and direct representation of synthesized results inspired one of the researchers involved in creating support tools to develop an annotation component to be used when reasoning about key events:

> You are looking at some kind of data source in a window and wonder how it is related to, ah that, and then you have a map there [...] Then you want to like save this, just as it is right so that the next person doesn't have to do that all over again. *(H-G, line 730–747)*

The most important use of the annotation component was to communicate important events to other researchers. Several participants described the process of communicating results through the tools they used to manage data sources. The difficulty when using tools like F-REX to communicate results could be caused by the fact that using audio and video sources are difficult in themselves and could be the real reason why people resort to formats that are easier to analyze such as text. John described how they used mostly text because of the amounts of audio and video generated during multi-day scenarios. Those amounts could simply not be managed within the timeframes commonly available for analysis. Jane described how she had only selected a small subset of the scenario episodes related to lack of competence in a team out of all the telephone logs for transcription. She decided on a certain subset of episodes to study before she transcribed any audio simply because of the time required to analyze all data.

Understanding the limitations of the data available and the tools was considered critical by all participants, and especially two areas of concern were highlighted: the reliability of human observers and the transparency of statistical modeling tools. Regarding inter-observer reliability, that is, the degree to which independent observers describe the same course of events using the same categories, John went as far as suggesting that a computer system for annotating events that was at least consistent between multiple scenarios would be preferable to human observers. It would not have to annotate using the same categories as a human, but if it could at least behave in a consistent manner, that would make analyses possible, as opposed to when annotations could not be used due to the differences in how people evaluate the same situation in the same scenario.

When using statistical tools, Charlotte described how she used LISREL (Jöreskog, 1973) modeling to capture patterns in questionnaire data as well as behaviors and communication. LISREL modeling can reveal several statistically valid equation models with the variables measured. Only some of the models constructed would actually be contextually reasonable though, which made the work of interpreting them difficult without knowledge of the work context, the mathematical properties of the underlying variables and the distribution of possible solutions. However, the ability to explore several possible relations in data was considered very valuable in her research.

5 Discussion

The interviewed researchers noted that the act of identifying interesting objects of study from the data sets they had collected was not straightforward, despite the use of structured tools and methods. In fact, they described several cases in which the decision to restrict a study to a certain subset of data was based on tacit knowledge. They rarely used advanced pattern exploration techniques as part of the workflow for selecting possible hypotheses regarding patterns in data but they did use specialized systems for automatically generating factor models of team communications and behavior.

In qualitative data analysis, scientists use tools for data analysis as part of both the collection and display of data (the first two stages in Figure 1). The interview participants described a process in which they collect information using tools such as F-REX or MIND, but select subsets of data for closer study only based on their own judgments, not based on emergent properties revealed by tools for statistical analysis. The *software support for analytic operations* as it is called in Figure 2 did not play a major role in their work with going between transformed products (e.g., transcribed speech or annotated video) and logs or recordings.

We called this issue *focusing*, in that they made decisions to select subsets of the initial data collections based primarily on their own or others' personal judgments, and not guided by statistical analysis techniques. Neither was the activity of *drawing conclusions* from data supported by pattern extraction techniques or statistical analysis tools well enough. When considering how support tools fit the work process, we made the observation that the four stage analysis process, as described by Miles and Huberman (1994) (Figure 1), seemed to capture both the focusing and conclusions drawing stages of the ESDA process. In a sense, both the iteration between logs, reports and refined products (the first stage of the analysis sequence of the analysis sequence of ESDA), and the iteration between statements (such as hypotheses) and refined products (the second stage of the ESDA analysis sequence) show great resemblance to Figure 1. Also, the paradigm of analysis presented by data mining researchers seems to, in turn, coincide well with the description of Miles and Huberman (1994), which might suggest that support tools that are based on the use of pattern extraction and data mining techniques could be suitable as foundations of support tools in ESDA analysis.

However, in both the activities of focusing and drawing conclusions, the participants found that advanced support tools were lacking in transparency and open-endedness with respect to how they supported data reduction and drawing conclusions or verifying results. Specialized, statistical analysis techniques for extracting patterns from data require an intimate understanding of the requirements for using them, the possible outcomes and how to interpret the models constructed (Fornell and Bookstein, 1982; Steiger and Schönemann, 1978). When confronted with the data sources most often available from a role-playing simulation, the process of validating data sources and making sure that they are valid as a basis for statistical analyses can be a serious impediment. Some data sources may be textual notes from observations made by human observers that have categorized their observations. Such manual categorizations may differ among observers and thus have poor reliability as a basis for statistical analyses. Other condensed metrics such as communication density are unreliable and incomprehensible predictors of team performance in real situations (Gorman et al., 2003).

Data exploration and analysis tools such as F-REX do not make use of automated techniques for pattern generation in the data sets they manage but primarily support users with a unified interface to several data sources. Although some attempts have been made to augment these tools with metadata and annotation capabilities on raw data sequences, these capabilities have not been extended to include automated reasoning about data.

Our goal with this study was to improve our understanding of the interplay between tools and methods in communication analysis. Trough the interview study, we identified three critical issues in communication analysis (focusing the study, drawing conclusions and understanding the limitations of tools and data) that were consistent with the observation that the research activity that enjoyed least tool support was the selection and reduction of data to manageable units of analysis (see Figure 1).

6 Conclusions

Taken together, the interviews illustrated two observations of tool use that we considered relevant for the construction of support systems for communication analysis:

Open-endedness Several research projects described during the interviews started with open questions on how to characterize teamwork, irrespective of performance metrics or the relation between team performance and manifested behavior. Therefore, tools to support analysis of communications must not make or require any specific assumptions about team communications. The interview participants mentioned that an automated approach for annotating communications would possibly be useful even if it did not use the same tags for annotation as a human observer. Also, they noted that the exploration of possible patterns in data, both when using spreadsheets and statistical modeling tools, was very useful for their understanding. The utility of the tools they used was described not so much in the level of automation provided but the freedom to choose how to operate on data.

Transparency Many of the participants described that they used tools for analysis that allowed a direct and visible relationship between synthesis and data. H-G had constructed a component for one of the ESDA tools used to make the association between reasoning and data transparent by adding annotations directly to the timeline of events, and one of the main arguments for the use of simpler tools such as Excel was that the connection between data and statistics was much easier to make. The use of specialized tools was described as dependent on the ability to use the tool for communicating results, and the primary risk with specialized tools was considered to be the risk of not being able to show the insights gained through them to other researchers or clients.

These two observations relate to our earlier observations regarding criteria for intelligent decision support systems, where we identified *graceful regulation* (allowing different uses of a tool in open-ended scenarios) and *transparency* as central conditions for success (Leifler, 2008) for tools that assist in military planning. Some participants described that using any special-purpose system for analysis was problematic. In their descriptions of why, they attributed the difficulties to the task performed by explorative multimedia management tools (managing large, heterogeneous data sources), a discrepancy between researchers' work and the specific requirements of the tool, and the fact that any special-purpose program requires too much dedicated work with that particular program to be used frequently enough.

All three main areas of concern elicited in the interviews revealed that the data sets were not reduced or classified using automated methods for pattern generation to any great extent. Although the interview participants indicated that great care must be taken when implementing special-purpose systems that implement advanced analysis techniques, they also recognized that systems for revealing patterns, such as LISREL modeling system, were of great use in their work. It is a great challenge to construct support tools that fit with the work process of researchers and provide adequate support for the tasks they typically engage in as well as ones they refrain from due to the current lack of tools. Support tools based on automatic pattern extraction are built on advanced mathematical models but must still be transparent enough that their requirements and underpinnings are made comprehensible to users of different professional background.

Nevertheless, due to the importance of understanding how people who manage high-risk enterprises in crises actually make their decisions, and understanding decision-making in distributed settings in general, we argue that it is a challenge well worth undertaking.

7 Acknowledgments

This work was supported by the Swedish National Defense College. We would like to thank the participants at the Swedish Defense Research Agency and VSL Systems AB for participating in this study.

References

Alberts, D. S., Gartska, J. J., and Stein, F. P. (2000). *Network Centric Warefare: Developing and Leveraging Information Superiority*. National Defense University Press, Washington, DC.

Albinsson, P.-A. and Morin, M. (2002). Visual exploration of communication in command and control. In *Proceedings of the Sixth International Conference on Information Visualisation*, London, UK.

Albinsson, P.-A., Morin, M., and Thorstensson, M. (2004). Managing metadata in collaborative command and control analysis. In *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society*.

Andriole, S. J. (1989). *Handbook of Decision Support Systems*. TAB Books Inc.

Brehmer, B. (2005). The Dynamic OODA Loop: Amalgamating Boyd's OODA Loop and the Cybernetic Approach to Command and Control. In *Proceedings of the 2005 Command and Control Research and Technology Symposium*.

Cutting, D. R., Karger, D. R., Pedersen, J. O., and Tukey, J. W. (1992). Scatter/gather: A cluster-based approach to browsing large document collections. In *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51(4).

Fornell, C. and Bookstein, F. L. (1982). Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *Journal of Marketing Research*, 19(4):440–452.

Gorman, J. C., Foltz, P. W., Kiekel, P. A., and Martin, M. J. (2003). Evaluation of latent semantic analysis-based measures of team communication content. In *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting*.

Jensen, E. (2009). Sensemaking in military planning: a methodological study of command teams. *Cognition, Technology & Work*, 11:103–118.

Jenvald, J. and Eriksson, M. (2009). Structured reflective observation in continuing training. In *Proceedings of the 8th WANO Human Performance Meeting*.

Johansson, B., Persson, M., Granlund, R., and Mattsson, P. (2003). C3fire in command and control research. *Cognition, Technology & Work*, 5(3):191–196.

Jöreskog, K. G. (1973). A general method for estimating linear structural equation systems. In Goldberger, A. S. and Duncan, O. D., editors, *Structural Equation Models in the Social Sciences*. Seminar Press.

Klein, G. (1998). *Sources of Power. How People Make Decisions*. MIT Press, Cambridge, Massachusetts.

Klein, G. A., Orasanu, J., Calderwood, R., and Zsambok, C. E., editors (1993). *Decision Making in Action: Models and Methods*. Ablex Publishing corporation.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1).

Kylesten, B. and Nählinder, S. (2010). The effect of decision-making training: results from a command-and-control training facility. *Cognition, Technology & Work*, pages 1–9. 10.1007/s10111-010-0157-0.

Lazar, J., Feng, J. H., and Hochheiser, H. (2010). *Research Methods in Human-Computer Interaction*. Wiley.

Leifler, O. (2008). Combining Technical and Human-Centered Strategies for Decision Support in Command and Control — The ComPlan Approach. In *Proceedings of the 5th International Conference on Information Systems for Crisis Response and Management*.

Leifler, O. and Eriksson, H. (2010). Message classification as a basis for studying command and control communications - an evaluation of machine learning approaches. Submitted for publication.

Miles, M. B. and Huberman, A. M. (1994). *Qualitative data analysis: an expanded sourcebook*. SAGE.

Morin, M. (2002). *Multimedia Representations of Distributed Tactical Operations*. PhD thesis, Institute of Technology, Linköpings universitet.

Morin, M. and Albinsson, P.-A. (2005). *Creating High-Tech Teams: Practical Guidance on Work Performance and Technology*, chapter Exploration and context in communication analysis, pages 89–112. APA Press.

Rosell, M. and Velupillai, S. (2008). Revealing relations between open and closed answers in questionnaires through text clustering evaluation. In *Proceedings of LREC 2008*, Marrakesh, Morocco.

Rubel, R. C. (2001). War-gaming network-centric warfare. *Naval War College Review*, 54(2):61–74.

Sanderson, P. and Fisher, C. (1994). Exploratory sequential data analysis: Foundations. *Human-Computer Interaction*, 9:251–317.

Sanderson, P., Scott, J., Johnston, T., Mainzer, J., Watanabe, L., and James, J. (1994). MacSHAPA and the enterprise of exploratory sequential data analysis (ESDA). *International Journal of Human-Computer Studies*, 41(5):633–681.

Shattuck, L. G. and Woods, D. D. (2000). Communication of intent in military command and control systems. In McCann, C. and Pigeau, R., editors, *The Human in Command: Exploring the Modern Military Experience*, pages 279–292. Kluwer Academic/Plenum Publishers, 241 Borough High Street, London.

Stanton, N., Baber, C., Walker, G., Houghton, R., McMaster, R., Stewart, R., Harris, D., Jenkins, D., Young, M., and Salmon, P. (2008). Development of a generic activities model of command and control. *Cognition, Technology & Work*, 10:209–220. 10.1007/s10111-007-0097-5.

Steiger, J. H. and Schönemann, P. H. (1978). *Theory construction and data analysis in the behavioral sciences*, chapter A History of Factor Indeterminacy. Jossey-Bass Inc.

Thorstensson, M., Axelsson, M., Morin, M., and Jenvald, J. (2001). Monitoring and analysis of command post communication in rescue operations. *Safety Science*, 39:51–60.

Wærn, Y. and Cañas, J. J. (2003). Microworld task environments for conducting research on command and control. *Cognition, Technology & Work*, 5(3):181–182. 10.1007/s10111-003-0126-y.