Characterizing the Intelligence Analysis Process:
Informing Visual Analytics Design through a Longitudinal Field Study

Youn-ah Kang

Georgia Institute of Technology

John Stasko

Georgia Institute of Technology

ABSTRACT

While intelligence analysis has been a primary target domain for visual analytics system development, relatively little user and task analysis has been conducted within this area. Our research community's understanding of the work processes and practices of intelligence analysts is not deep enough to adequately address their needs. Without a better understanding of the analysts and their problems, we cannot build visual analytics systems that integrate well with their work processes and truly provide benefit to them. In order to close this knowledge gap, we conducted a longitudinal, observational field study of intelligence analysts in training within the intelligence program at Mercyhurst College. We observed three teams of analysts, each working on an intelligence problem for a ten-week period. Based upon study findings, we describe and characterize processes and methods of intelligence analysis that we observed, make clarifications regarding the processes and practices, and suggest design implications for visual analytics systems for intelligence analysis.

KEYWORDS: Intelligence analysis, qualitatvie user study.

1 INTRODUCTION

Visual analytics applies to many domains and problem areas, but one area of particular study since the beginnings of the field has been intelligence analysis. Intelligence analysis is a cognitively demanding process, one that seems ideal for the application of visual analytics tools. Accordingly, a growing number of systems have been built for it [2, 11, 25, 30].

Research in human-computer interaction also teaches us to deeply analyze and understand end-users and their problems in order to design appropriate computational solutions. We question whether visual analytics systems, including some of our own, have been based upon a deep enough understanding of the discipline. Relatively few studies of intelligence analysts, their tasks, and their work processes exist. Notable exceptions [4, 12, 19, 23] provide initial insights into the field, but we have frequently interacted with analysts who feel that their practices are misunderstood and that visual analytic systems often fail to address their most important problems.

To address these concerns and to learn more about the analysis process, we conducted a longitudinal, observational field study of intelligence analysis on real world problems. Unfortunately, getting access to working, professional analysts is challenging. Even if they are available, it is difficult or impossible to study them for an extended period of time while they work on real tasks without having some type of special access that simply was not available to us. As an alternative, we studied analysts-in-training who are soon to become working professionals. More specifically,

ykang3@gatech.edu
stasko@cc.gatech.edu

we studied groups of students from the Department of Intelligence Studies at Mercyhurst College as they conducted a term-long intelligence project.

We were given deep access to the students, the materials they examined, the tools they used, and their final intelligence products. We interviewed the teams multiple times and observed their group meetings. Additionally, we interviewed their instructor to learn his impressions of the process. Our goal was simply to better understand what these young analysts do, the challenges they face, and how we might be able to help them. Thus, the contributions of our research include a characterization of the processes and methods of intelligence analysis that we observed, clarification and reflection of several beliefs about intelligence analysis processes and practices, and resultant design implications for visual analytics systems for intelligence analysis.

Munzner has argued of the importance and the need for more domain characterization research like this [17]. She notes that such research is both difficult and time consuming to do properly, but the visualization community could benefit greatly from it.

2 BACKGROUND

One of the most widely used models in the visual analytics community is Pirolli and Card's sensemaking model [19] for intelligence analysis. While the model broadly characterizes processes used in analysis activities and has guided the design of visual analytics tools, the model does not provide rich details of how intelligence analysts work in the real world. More empirical, descriptive explanations of the intelligence analysis process are required to provide appropriate visual analytics system solutions.

Several studies have captured and characterized the work practices and analytical processes of individual or collaborative analysis through a qualitative approach. Chin et al. [4] conducted an observational case study with professional intelligence analysts in which participants worked on real-world scenarios. The researchers revealed various characteristics of the analytical processes of intelligence analysts. Gotz et al. [7] also recognized the lack of studies examining analyst behavior and conducted a user study to explore the ways in which analysts gather and process information. Another study by Robinson examined how analysts synthesize visual analytic results by studying domain experts conducting a simulated synthesis task using analytical artifacts [20]. Based on analysis of video coding results, he identified several characteristics in the process of collaborative synthesis. While these studies did not evaluate specific visual analytic tools or features per se, they provide valuable implications to inform design directions for future support tools.

Johnston [12], an anthropologist, conducted an ethnographic study of the CIA for a year and identified variables that affect intelligence analysis and requirements for techniques and procedures to reduce analytic error. While he made useful recommendations to improve analytic performance, his approach was primarily intended to understand organizational culture and describe current community practices, rather than identifying leverage points for designing support systems.

In our study, we aim to deeply understand the analysis process with an eye toward designers. We seek to provide meaningful implications for developing technological support for analysts.

3 METHODS

In order to investigate the intelligence analysis process in-depth, we conducted an observational study of teams of analysts conducting an in-class intelligence project. In the term-long (ten-week) project, each team addressed a real intelligence problem proposed by a client. We observed three teams, monitoring their status and process throughout the project. At the end of the project, each team had to produce final deliverables and present their findings and analysis to decision makers.

3.1 Participants

We recruited three groups of students, one team of four undergraduate students (Team A) and two teams of five graduate students (Teams B and C), from the Department of Intelligence Studies at Mercyhurst College [5]. Mercyhurst's Intelligence Program, started in 1992, provides education for students who want to pursue a career as an intelligence analyst. It is recognized as one of the top programs for intelligence studies in the United States, offering a broad range of classes and degrees for students seeking a career as an analyst in national security, law enforcement, or the private sector.

We recruited students who were taking the courses named "Strategic Intelligence" (undergraduate) and "Managing Strategic Intelligence" (graduate), in which teams are required to conduct an analysis project over a ten week term. The two courses are very similar with respect to the projects. The students all were close to graduation, with past internship experience, and most of whom had already received job offers.

While these student teams clearly are not practicing professional analysts, there was not a significant difference between the way the students worked and the way real analysts work, according to the instructor. The analysis process used in the class was modeled directly after the process employed by the US National Intelligence Council to produce its strategic reports, the National Intelligence Estimates [18]. The instructor also intentionally stayed relatively detached from the students, acting as a mentor and limiting his supervision so that the teams could autonomously work on the project. The teams were diverse in expertise on the subject matter, which is common for teams in the intelligence community. One key difference from real world practice was the relative absence of administrative and bureaucratic overhead affecting the student teams, as well as issues relating to security clearances. They operated in a much more "sanitary" environment than the real world.

3.2 Task

Different types of intelligence questions exist - we focused on one of the most common types, strategic intelligence. Strategic intelligence is "intelligence that is required for the formulation of strategy, policy, and military plans and operations at national and theater levels [8]." Strategic intelligence is exploratory and long-term in nature. The requirement for tasks within the class was that "the questions should be relevant and relatively important to the client's success or failure but outside their control." We served as a client/decisionmaker for team A in order to observe the process even closer, whereas Teams B and C worked with external organizations. The specific issues each team addressed were:

Team A
The strategic assessment of potentially influential factors to the evolution of computer-mediated undergraduate and graduate distance education: What aspects of computer-mediated distance education will likely influence R1 institutions during the next 5, 10, 20 years with specific, but not exclusive, emphasis on undergraduate education and computer science?

Team B
Who are the key people, technologies and organizations that likely currently have or will develop the potential to disrupt or replace traditional US national security Intelligence Community (IC) analytic work flows and products with commercially available products available over the next 24 months?: Criteria that will be used to identify these key players are:
- *Those that are not beholden to the IC or US Government as primary sources of funding.*
- *Those that are looking at future based events or actions that are outside the control of the forecaster/predictor.*

Team C
What are the most consistent and identifiable characteristics displayed by potential insider threats to (a defense department)?
- *An insider threat will be defined as an individual or collection of individuals employed directly or indirectly by the department who violate security or access control policies with the intent of causing significant damage to the department's personnel, operations, or information.*
- *Within the broad range of insider threats, special priority will be given to violent threats and improper diversion of information or physical assets.*

The teams updated the status and the process of the project on a wiki site. At the end of the semester, they needed to produce a final report that synthesizes analytical results, and strategies of the entire analysis process.

3.3 Study Protocol and Procedures

The analyst teams conducted the project for ten weeks - from the week of September 1 through the week of November 10, 2010 for Team A, and from the week of December 1 to the week of February 14, 2011 for teams B and C. Normally, strategic intelligence projects range from a couple months to years; ten weeks is short but within normal limits for strategic intelligence.

Before the project began, the external clients formulated a draft of their initial intelligence problem. In the first week of the project, the clients conducted a conference call with the analyst team to discuss the scope and requirements of the problem. During the next two weeks, the analysts refined the problem and wrote a formal statement of the intelligence question, which they call "Terms of Reference (TOR)". Upon approval from the decisionmakers, the teams began working on the problem, which took another seven weeks.

The wiki platform was used as a workspace for analysts to document their process and findings, and we were able to monitor the wiki's status throughout the project period. The final report of the projects also was placed on the wikis.

During the project period, we conducted two face-to-face meetings with each team – one in week 7 and the other in week 10. In the meetings, we interviewed each team as a group and the class instructor in order to learn more details about the project's status, process, difficulties, and future steps. Each interview took approximately an hour. While the interview was semi-structured, we followed an interview guide containing several key topics [3, 15], including:

- How do the analysts perceive their analysis process?
- What barriers and difficulties do they encounter?
- Tools and aids being used - where and why?

- Collaborative aspects in the analysis process
- Where in the process can technology help?

We also observed two team meetings firsthand, which took about 3 hours in total.

3.4 Data Collection and Analysis

Most of the process descriptions and produced artifacts were stored digitally. The teams reported methodologies, tools used, sources, as well as the findings on their own website (wiki). To further understand the process, we analyzed interview notes and audio recordings from the focus group interviews. We used the artifacts produced by the analysts, such as drawings, wiki pages, tables, and slides as further data. Additionally, we had access to history logs of wiki page changes.

We transcribed each interview's audio recording and then coded the transcripts based on the grounded theory approach [26]. We began by identifying major themes and categories from the text. One emergent theme focused on the analysis process, including methodology and challenges encountered. Another theme was collaboration, focusing on how and to what degree the analysts collaborated and what types of collaboration existed. Throughout the coding process we iteratively refined the categories. We then elaborated on supporting evidence from the data for each category through a deductive approach.

4 RESULTS & FINDINGS

4.1 Overall Analysis Process

Through the project, we found that four component processes were essential to the overall analysis: constructing a conceptual model, collection, analysis, and production. While the four processes did not occur linearly, this section describes the importance of each and how the analyst teams worked on each.

Phase 1: Constructing a Conceptual Model
Once the teams and clients/decisionmakers finalized the requirements of the intelligence question, the teams started to build a conceptual model, which is a map of issues and concepts that the team will be investigating to address the problem. The conceptual model illustrates the areas the analysts need to research by helping them to visualize the question at hand. The question is placed in the center, and then several high-level components of the question surround the question (Figure 1). Each component branches out and creates a bigger map, from which the team gains an idea of the areas with less/more information that they need to research. This allows the team to focus on collecting a set of data with an appropriate scope.

While the significance of the conceptual model differs depending on the question and the team, it plays a key role for the team to understand the domain area and determine the direction of research. We were told that analysts often construct this conceptual model implicitly, rather than externalizing it, which we found quite interesting. The instructor commented:

In most cases, it's implicit. People don't write it down. But it's the way they are actually doing it. There's a model in people's heads, and that's far more important than the data. There's research that says analysts' judgments are far more driven by the way they think about the problem than the data itself. So making the way you think about the problem explicit would allow analysts to identify whether they disagree about how to think about this model, and to merge their best thinking about this model. So the process happens, it's just the degree of to which it's made explicit, that is unusual.
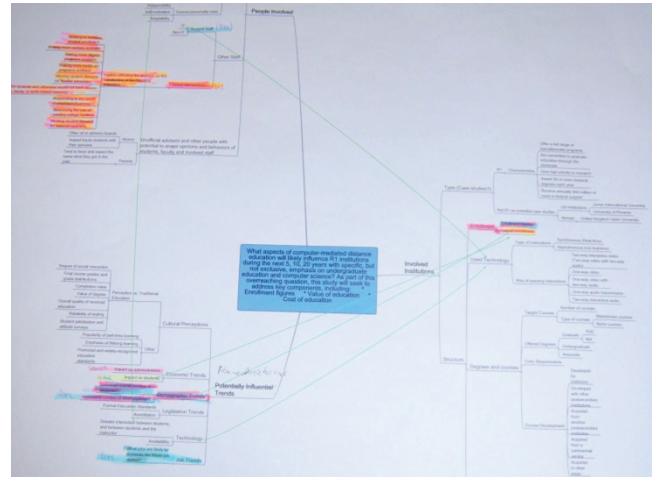

Figure 1. Conceptual Model. Printed from Mindmeister

Phase 2: Collection
While working on the conceptual model, the teams also assigned areas/concepts to each member. Next, they collected information from various sources including online and offline sources (e.g., interviews with experts), which they call "all-source intelligence". While each analyst was responsible for collecting data about their assigned topic(s), the team shared their sources using Zotero, a web browser plug-in for gathering and organizing source material. This allowed teammates to view the data like a common library – other team members might already have found information that they need. More specifically, the data collection process typically involved these steps:

- Once an analyst identifies needed information, they search the Internet using search engines and various keywords.
- The analyst also establishes RSS news feeds on websites of interest using Google Reader.
- Whenever they find useful sources, whether for their own topic or someone else's, the analysts place the link into a Zotero group library.

Phase 3: Analysis
The analysis phase exhibited various characteristics depending on the requirements and analytical methods used. In this phase, analysts processed data that they collected from many different sources in order to convert "information to knowledge." While team A directly began writing short format analytical reports on each topic, team B and C used a more structured format (e.g., spreadsheets) to quantify information and rank the significance of each topic or entity. No matter which method they used, the initial analysis of each topic/entity was undertaken and written by one person in accordance with the assigned topic. However, everyone on the team could review and comment on the others' work via the wiki pages. In all cases, the analysis phase was incorporated with the collection and the production phase.

Phase 4: Production
Once individual collection and analysis was almost finished, the teams met and tried to synthesize findings from each part, which led to the "key findings" – the major product of the analysis. Production was an intensive reading/writing process in which the team collaborated tightly with each other. This stage was more to prepare a presentation for the decisionmakers. Team members repeatedly checked their sources and findings to make sure that they were consistent and logical.

Reiterating, while we separated these four components for the sake of clarification, the process was not simple and it was not clear which phase the team was in throughout the project. Instead,

the characteristics of the question and the analytic method chosen most influenced the process. In our study, we observed two different styles of intelligence analysis process. The difference in approaches resulted from the type of the intelligence question.

4.1.1 Intuitive Analysis – Team A

Team A addressed potentially influential factors to online distance education in the near future. Because the requirements were rather broad and intuitive, the team decided to take a top-down approach, investigating meta-information sources such as research that forecasts future education trends.

Instead of using a specific analytic method, this team depended considerably on the conceptual model and used it as a guide throughout the entire project. They put significant time and effort into constructing it, revising the conceptual model until the seventh week of the project. Because of time constraints, they were not able to cover all the topics in the model. Through discussions, they chose a number of concepts they felt most worth exploring and divided up the concepts for each member.

After collecting and reading information for their designated topic, each analyst wrote a short format analytical report that synthesized the information. Most of the analysis simply involved reading. For a few topics that required careful weighing of alternative explanations, the team employed analysis of competing hypothesis (ACH) [9]. While documenting results, everyone was able to review and edit the others' drafts on the wiki page, and team members frequently discussed others' analysis (short write-ups) both online and face-to-face. Therefore, everyone was responsible for the reporting of each topic.

After working on the individual topics, the team met to write key findings together. This team invested considerable efforts in synthesizing their findings because their narrative was extremely important for their intuitive type of analysis.

4.1.2 Structured Analysis – Teams B and C

Teams B and C used structured analysis with quantified information because their research questions tended to be more specific and required rank-ordering of entities (e.g., top x indicators, key people/companies). Both teams built their conceptual model in the beginning as a base model. For these teams, however, the model was more of a collection plan rather than an actual conceptual model. Although they used the model to collect information and divide up the work, they did not refer to it for the remainder of the project. Instead, they started building a matrix in a spreadsheet to collect and analyze data from diverse sources. The matrix was rather a re-interpretation of the conceptual model, and each cell in the matrix indicated a collection requirement.

The purpose of the matrix was to evaluate each entity based on criteria chosen and identify the most influential ones, those of most interest to the decisionmaker. Team B, that was asked to identify key people, technologies, and companies that might affect intelligence community products, created a matrix and chose criteria while collecting information. They identified 180 entities and graded each based on the criteria, noting the ones with highest scores. Team C, that was asked to identify indicators displayed by potential insider threats to a defense department, analyzed data from the 117 case studies about crimes using a matrix (Figure 2). They used it to compare the relationship between crimes and motivations, as well as crimes and indicators.

In both teams, the matrix captured the conceptual model and how each team was thinking about the question. Filling in the cells was a time-consuming part as analysts needed to read and analyze each case/source to fill in one cell, addressing "the devil in the details."

However, this type of analysis required additional efforts in the production phase. Initially, the teams converted qualitative information from sources into quantitative information for rank-ordering. Once they had completed the matrix, the teams needed to transform its data into a story so that it could be made useful to decisionmakers.

Upon the completion of the projects, the instructor evaluated the teams' performances as being in the "top 10% of the projects over 8 years." He commented that all three teams performed the analysis well and in one case, the decisionmaker briefed the head of his organization with the team's results.
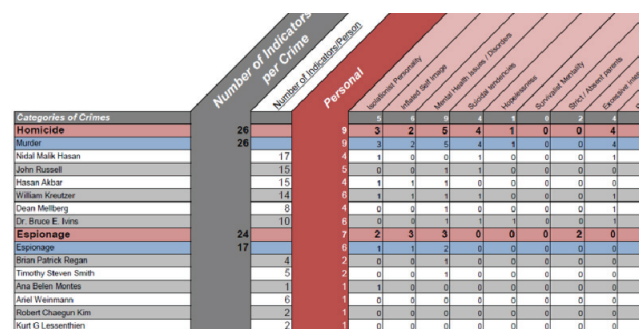

Figure 2. Case Study Matrix of Crimes

4.2 Tools and Methods Used

The teams used various software tools and analytical methods to develop hypotheses, arrive at analytic estimates, and create written reports and multimedia products.

Wikispaces/Google Sites: The teams used a wiki platform (Team A&B – Wikispaces, Team C- Google Sites) to exchange gathered information, aid administration, and share organizational details. The wiki sites became part of the final product, displaying the key findings, terms of reference, and all analytic reports.

Mindmeister (conceptual model): Mindmeister is an online mindmapping tool the teams used to build a conceptual model [16]. A conceptual model provides a revisable platform to view the requirements and their components. As research and facts begin to support or refute initial ideas, main ideas become more solidified and focused.

Zotero: The teams used Zotero as a source collection database [31]. Downloaded as an Add-on to Mozilla Firefox, Zotero allows the analyst to search websites and save the sites in a database that is accessible through the Zotero website. The teams used the Group Library feature to place their sources in a single database.

Website Evaluation Worksheet: To evaluate the credibility of the online sources, all the teams used the Dax Norman Trust Scale [19]. This matrix allows scores to be applied based upon criteria such as clear bias, corroboration of information, and the analyst's overall perception of the source. Based on the sum of scores, the source can score a High, Moderate, Low, or Not Credible rating.

Analytic Confidence: Each report includes an analytic confidence section that conveys to the decisionmaker the overall doubt connected with the estimative statement(s). While assessing the level of analytic confidence, the teams used Peterson's method [21]. Peterson identified seven factors that influence analytic confidence: the use of structured analytic methods, overall source reliability, source corroboration, level of expertise on subject, amount of collaboration, task complexity, and time pressure. In the analytic confidence section, the teams addressed these six factors as applicable to the particular estimate.

Social Network Analysis: Team C employed social network analysis using i2's Analyst's Notebook [11] to see relationships within industry. The team analyzed the social network analysis based on betweenness and eigenvector scores.

5 UNDERSTANDING THE INTELLIGENCE ANALYSIS PROCESS

Observing analyst teams helped us to better understand their goals and processes. In particular, the study highlighted a number of misconceptions we harbored about the intelligence process. Other visual analytics researchers may or may not share these preconceived beliefs, but we think that they have the potential for misunderstanding and are thus worth exploring.

Intelligence analysis is about finding an answer to a problem via a sequential process.
Some existing models of the intelligence analysis view it as an answer-finding process with a sequential flow, as noted in several models of the intelligence analysis process [12, 14, 27]. This perception presumes that the process is linear, sequential, and discrete by step. Pirolli and Card's sensemaking model [22] includes the notion of iterations and revisions between steps, but the fundamental assumption is that separate stages exist throughout the process and that analysts transition between stages.

However, this model was not the intelligence process we observed. Instead, the process appeared to be more parallel and organic, as one analyst described:

> *Intelligence analysis is not about getting from point A to point B along the route, but it is better associated with basic research where you don't necessarily know where you are going to go. You're cutting a path through the jungle that's never been explored. That's what you're doing in most intelligence analysis projects. It's not a mechanical process in a sense that an assembly line is. It's a very exploratory activity by nature.*

The key part of the intelligence process is the analysis of a specific set of data.
Visual analytics systems often manipulate pre-processed data for analysis. A primary misconception about intelligence analysis is that the data analysis process, in which investigators analyze a set of collected data, is the most difficult part and takes the most time. This belief assumes that analysis occurs after investigators collect all data required for the analysis.

This view, however, needs to be changed. Although analysis is important, we observed that the process of "constructing a frame," as described in the Data-Frame theory [13] is more important. In other words, intelligence is about determining how to answer a question, what to research, what to collect, and what criteria to use. This process becomes part of the analysis - analysis implicitly occurs during the process of the construction. Analysts also explore different sets of analytic techniques to address a problem. Deciding which method to use is important, but it often changes during analysis as the way that analysts think about the problem evolves, depending on information that constantly flows in.

Understanding that collection and analysis are integrated together in the process of building a frame is extremely important. Systems are not likely to be successful in supporting intelligence without acknowledging that fact. One analyst commented:

> *Intell analysis is not like that you have a set of data in hand and run a program. It's like a conundrum from the very beginning. You have to learn how to learn, how to frame the question, and how to answer it through collecting and evaluating sources.*

Analysts do not often collaborate.
One common perception of intelligence views analysts as isolated individuals who prefer to work alone, struggling with pieces of information, rather than as collaborative teams [4]. However, a faculty member at Mercyhurst countered this perception:

> *Collaboration is almost all intelligence analysts have done in the context of the team. In the CIA or DIA, working as a team is pretty normal. While working on a particular topic within an agency is typical, also typical is working on an interagency team that consists of analysts from different agencies such as state department teams, DIA teams, and NSA teams.*

> *Analysts are normally organized by function or geographical region. These typically operate as loose teams. Strategic projects almost always involve a team as do crisis projects (for example I am sure there are multiple Libya teams that did not exist a month ago). In short, teamwork is the norm although the teams differ in the degree of formality and to the degree that there is a designated leader.*

During our study, we also observed many collaborative elements of intelligence analysis. Collaboration is commonplace in intelligence analysis, and understanding how that occurs is important because it influences one's whole notion of the process. The intelligence community itself has recognized the importance of improved collaboration since 9-11 [6]. Although collaborative tools have been built and they are pushing users into tighter collaboration, it is still important to understand where tighter collaboration will be beneficial and where it may not help much.

We found that multiple layers of collaboration exist in intelligence analysis and that the degree of collaboration differs depending on the type of task and the group dynamics. We observed that analysts usually do not collaborate tightly on data and content – the actual collection and analysis. Although the teams had meetings frequently – twice or three times per week – the main purpose was to discuss their status, issues, and the next steps, as two analysts said:

> *We come up with an agenda before meeting, a list of what we're supposed to talk about - what we did, what we want to do, what the questions we need to solve as a group. We didn't really plan that way, but it just happened. It's the way it is.*

> *There was no detailed, content-based work during the meeting although sometimes we had discussions on controversial issues. We basically do work in our own time, get preliminary ideas of it, and get together and discuss what issues we had.*

The teams often worked together in the same lab, but it was rare that they worked on the same document or content. They worked on their own part, but they often talked to each other about issues and questions. Essentially, they took advantage of being in the same place at the same time. To better support analysts' work, we need to understand the collaborative aspects of the intelligence process and where technology can leverage collaboration. Some tasks are inherently done better by an individual.

We can help intelligence analysts by developing sophisticated analytic tools that assist their thinking process.
Visual analytics researchers often seek to help intelligence analysts by developing technologically advanced analytical tools, thereby assisting their cognitive processes. The tools support specific types of analysis, specific analytical methods, and specific stages of the process. Such tools certainly can be helpful, especially to assist analysts to handle a flood of information.

However, our study revealed that analysts want something more than that. Currently, more than 50 analytic methods exist in the intelligence community [10], and analysts try many different kinds of techniques depending on the problem. Consequently, their dependency on a specific analytical technique is relatively low. Instead, the ability to manage the intelligence process

effectively and employ various analytical methods and tools quickly is more important, as the instructor and an analyst said:

Everything is fragmented. I've got Mindmeister here, Mindmeister doesn't interface with my search technology and Google reader. I've got to manually go out and figure out what all those bullets are. No help from a computer…But there isn't any set that ties all these, the pieces are there, Mindmeister, Zotero, RSS Google reader, MS Word, the wiki, are here, but nothing links all that in one seamless thing so I can go from the requirement to a product in a single package, in a single way.

6 REFINING OUR THEORIES

6.1 Rethinking the Intelligence Analysis Process

6.1.1 Linear vs. Parallel

One might believe that the way intelligence analysts work is quite simple and straightforward. First they specify requirements, build a conceptual model of what to research, then collect information, analyze data using various techniques, and finally write a report. This belief is a common misconception about intelligence as mentioned in the previous section. The reality is quite different. Rather than working linearly, analysts work on *everything* during almost the entire project. That is, analysts do not hold writing until enough information is collected; they keep revising analysis and writing as new information flows in. Analysts do not decide what to research and move on to collecting information; they start searching for information even when they are not sure what to research. Analysts do not produce final products after they are done with analysis; they already have an idea or a structure of final products in the very beginning, although it may be rough.

This "parallelism" is portrayed well in Wheaton's model of the intelligence process (Figure 3). In each phase, one of the core processes is emphasized most but all other functions operate in parallel. Wheaton argues that "All four functions begin almost immediately, but through the course of the project, the amount of time spent focused on each function will change, with each function dominating the overall process at some point [29]."

Although several distinct elements exist in the analysis process, all are very closely coupled and the connection is very organic. One can easily observe an analyst working on collecting new information while analyzing and checking the credibility of previously collected sources at the same time. In our study, we observed that a team's conceptual model changed drastically in the middle of the process, that a new information source was added ten days before the deadline, and that a previous analysis report was discarded and new analysis began in a late stage. The matrices also kept changing as new information arrived. While the teams were working on the matrix, they were collecting information at the same time to make sure that they were familiar with the area. Several quotes better explain this:

But it isn't as rigidly isolated as it's on that (traditional) cycle because you can't build a good conceptual model without knowing what's out there. So there's little bit of collection as you're building the model and we refined it.

Our conceptual model is changing. It doesn't get set in phase 1 and we drive it, that's the difference between this process and an outline. An outline drives your production. But we are using it differently. As it changes, we're changing our analytic focus, we're making decisions about production, who's going to write something, who's going to do the analysis, based on how it's changing and that's being informed by new information that comes in.
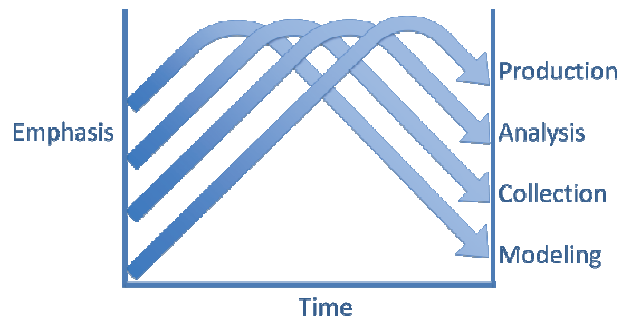

Figure 3. Wheaton's Multi-phasic Model of the Intelligence Process [29]

6.1.2 Pirolli and Card's sensemaking model

How does this new way of thinking about the intelligence process relate to Pirolli and Card's sensemaking model [19]? Because it is the most widely used model in the Visual Analytics domain, we were curious how well their model explains real-world intelligence analysis processes.

Pirolli and Card's model provides new insights about the intelligence process, suggests leverage points for analysis tools, and has guided the design of many visual analytics systems. However, we argue that the model still implies sequential, discrete stages of the intelligence process although it acknowledges that analysts can move either top-down or bottom-up or jump to different stages. For example, the model does not explain why analysts so frequently jump from one state to another state that is not adjacent. Many visual analytics tools thus support specific states only (e.g., shoebox and evidence file, evidence marshalling, foraging), and often they do not blend into the entire process of intelligence analysis.

More importantly, the model describes how information transforms and how data flows, rather than how analysts work and how they transition. It gives a very insightful illustration of how the form of information evolves from raw data to reportable results. However, it does not quite fit analysts' mental model of their work process because they do not work as information is transformed. Rather, information is transformed by how analysts proceed. Similarly, all different states of the model can exist at any point during the process. Analysts may have polished reports on certain sub-topics, drafts of analysis, structured matrices, and a collection of documents at a time.

The Pirolli-Card model identified various leverage points for visual analytics tools, but the linearity of the model could give researchers an inaccurate impression of the process. While models are inherently abstract and stage-based, it is important to understand the context and the purpose of the model. We would characterize their model as more of an information-processing process rather than intelligence analysis process. Pirolli-Card explicitly state that he model was suggested as a starting point to investigate the domain. While it has contributed to visual analytics researchers understanding of the domain, now we need to change our assumptions to build systems that better help intelligence analysts with their work.

6.2 Where and How Collaboration Occurred

6.2.1 Collaboration throughout the process

Throughout the project, the teams worked tightly together although the degree to which they collaborated differed depending on the phase of analysis. Once the project started, the team set up weekly meetings. The first thing they had to decide was to specify requirements of the problem, and then they collaboratively worked on building the conceptual model. Whether the team kept using this model or changed to a matrix, it played a role as a representation of their "group thinking," as an analyst described:

6

You want to say that this is the way I'm thinking about this problem. These are some of things I need to think about. And what we've done by building the conceptual model is to have that sort of group interaction, which is not necessarily harmonious action. There can be disagreements about how we should be thinking about this. And if there's shifting, moving it around, that represents an evolution of the way of our thinking.

Once the team had an idea of the areas to explore, they divided up the work and assigned concepts to each analyst. While each one worked on different concepts, they collaborated in collecting information by using a group library. Although this seems to be loose collaboration, the benefit the team gained was invaluable because it could significantly save time and effort in collection. An analyst explained how they worked in collection using Zotero:

Zotero is a good example of one way we collaborate. Each person creates a group library on the Zotero server. If I find a website that I think is useful, whether for my topic or someone else's topic, I add it to our group collection, and then other members can see it before they go searching the Internet for something. And if she doesn't find that in Zotero, then she might go out Google. So..try Zotero first, you might already have it.

While working on and analyzing their own topics, team members often met with each other to check status and discuss issues. When not all team members were available, they used typewith.me [28], a web-based collaboration tool for writing. When most of the areas they had planned to explore were covered and analyzed, they collectively wrote the key findings – the crux of the analysis project. Very tight collaboration occurred in this work. They met together and spent significant time to synthesize findings from all the topics and write the key findings.

6.2.2   Sharing vs. Content vs. Function

We found that three different types of collaboration exist when analysts discuss the topic: sharing, content, and function.

Sharing is a way to collaborate by sharing information. In our study, analysts shared sources to better assist their search process and understanding of the topics. At a higher level, however, this can be the sharing of analytical products as well as information sources. When people mention the importance of collaboration in the intelligence community, they primarily refer to the sharing of information sources and analytical results [1]. This type of collaboration can be significantly supported by technology.

Collaboration also occurs at the **content** level. This type of collaboration, in which analysts work together to create analytic products, can be seen more often in a small-size team. Examples in this project include constructing a conceptual model together, dividing concepts and assigning to each analyst, commenting on each other's analysis, working on ACH together, and writing the key findings together. However, in our study, once work was divided, then each part was done individually. The degree of tightness in this type of collaboration may directly affect the quality of analysis. The more closely the team works together, the more that output is coherent and logical. However, in reality, it is difficult to collaborate on content because of efficiency. This type of collaboration is also difficult to facilitate via technology because so many subtle issues – such as social dynamics, politics, teamwork, and motivations – are involved.

Functional collaboration is needed to execute practical tasks for completing the project, such as editing, creating a matrix structure, specialized analysis on a specific topic, and polishing deliverables. Whereas analysts work on the same thing and divide up the analytic product in the content level, functional collaboration naturally emerges at the later stage of the process as the team begins to think about allocating multiple functions. In this type of collaboration, analysts reinforce their strength. For example, if one is a good editor and has a detailed eye, then that person would do the editing, as one analyst explained:

There was a lot of collaboration. A spent a lot of time working on wiki stuff. B spent time doing the ACH stuff. C did a lot of technical stuff. So each of us spent extra amount of time doing something specific. Whichever parts of this you want to take on, those are the parts that get divided up. As different as I do this (analysis), I do this concept, I do this concept.

Olson et al. [20] similarly characterized collaborative activities of groups by analyzing design mettings from four software teams. Focusing on the time spent in meeting activities, they found similar patterns across design teams. In meetings, teams spent 40% of the time in direct discussions of design. 30% of meeting time was spent taking stock of their progress, and coordination activities consumed approximately 20% of the time. Clarification of ideas across these activities took one-third of the time, indicating that participants spent a large amount of time sharing and explaining expertise.

7   HOW VA CAN HELP: DESIGN IMPLICATIONS

How can visual analytics help intelligence analysis? Based on our study findings and reflections, we suggest several design implications for systems supporting intelligence analysis.

Externalize the thinking process - Help analysts continuously build a conceptual model
"Good analytic practices encourage continuous improvement upon the conceptual model throughout research, which continues through the end of the project."

The analysts in our study told us that the process of making sense of a problem and building a conceptual structure is one of the most important parts of intelligence analysis as it decides the direction of analysis. In most cases, analysts encounter a situation in which they need to learn about new subject matter, but it takes time and effort until they become familiar with the domain. Because they cannot build a good mental model of the problem without knowing what information is available, they struggle to know more about the domain until the later stage of analysis.

Using the power of representation, visual analytics systems can help analysts build a conceptual model or a structure of the problem and domain. For example, the system can take the main question the analyst has and suggest a number of possibly related concepts and keywords based on online encyclopedias, table of contents of books, tagging services, etc. The system should allow the analyst to refine the concepts so that it can repeat the search and suggest other relevant concepts. By connecting, grouping, and organizing concepts, analysts can continuously build up their conceptual model or structure of the area throughout the process. One analyst cited experience:

Ok, I got to model something, I've got to do a report on Ghana, I don't know anything all about Ghana, where's the tool that if I hit the button, it gives me a picture of what the relationship is, the model how to think about Ghana? It gives me 60-70% of the solution. But it gives me the ability to input and tweak and change those. Because I want to have a role in that, I can't allow the computers to do all my thinking, you know.

Support for this externalization should occur throughout the analysis process because as analysts learn more about the domain, they alter their way of thinking and refine their visual model.

Externalizing the thinking process also can assist analysts when they review their analysis after the project terminates. Supporting this activity would be especially useful because it will inform how the analysts could have done better and the areas that need to be examined if they did a similar project, as the instructor said:

The other thing this model helps you do is at the end of the project you can look back and go, "What did we not have time to do? And how does that impact our company, our estimates?" Because whatever reason we didn't get to it, this was important, we thought this define the space…We can sit back and go, ok, how confident are we on our estimates, knowing that our analysis is always at some level incomplete? And it's always incomplete, but how does it impact our confidence in our product? That's another way to use this representation.

Support source management - enable managing both pushed and pulled information and organizing sources meaningfully
One prominent characteristic of how analysts think about sources is that they have to be always vigilant of new sources. They often search for the same keywords again to see if any new materials have been added regarding the topic (pulled sources). They also receive news articles through RSS feeds everyday and check if they have received interesting information (pushed sources).

This process of searching sources takes more time than one may think, and systems should allow analysts to manage both pushed and pulled information associated with concepts they have identified. For example, a system could populate several concepts chosen by the analyst and store all the pulled sources in a database such as Zotero. Based on sources already found, the system also could recommend push resources such as blogs and news articles. For each source collected, the analyst could express if it is a useful source or not. Analysts commented on this functionality:

Sources are what we have to get, but where is the tool where I can integrate them? My RSS feeds dump into me every morning. But then I do searches as well. Where's the tool that allows me to integrate all data, the information that is useful for me?

If that kind of system exists, I have the ability to go back and find all my sources. Automatically, this (keywords, phrases) gets populated. And every point, I have the ability to say no or yes, no or yes to a source. But the actual extraction or the pulling, and the organization of that is automatic from that.

Then the list of sources can be organized in a meaningful way – for example, by keyword queries, by tags the analyst annotated, or by date the source was added. The system also could provide several ways of representing source results such as summary and tag clouds. Further support for analysis or visualization of collected sources as a group would be extremely beneficial.

All these technical capabilities currently exist in visual analytics systems. Now it is important that they be integrated together appropriately.

Support analysis with constantly changing information - integrate collection and analysis in a single system and help analysts use structured methods during collection
As described in the previous section, collection and analysis are not separate, but highly integrated processes. Analysts do not wait until all the data are gathered; rather, they start analysis even when they have only a few pieces of information. Through the repeated process of collection and analysis, they revise a frame and use the collected data as supporting evidence for the frame.

Currently, many systems provide analytical support assuming that processed data is available. If a system does not support a seamless transition between collection and analysis, it is likely to be less successful in assisting the analysis. Analysts collect during analysis and they analyze during collection. This differs from statistical analysis, in which a structure or a frame about how to analyze the data is clearly defined and analysis is done with clean dataset. An analyst mentioned:

If they had more reliable, structured data, I'd use statistical analysis. But intelligence data is unstructured and dirty. You don't know what the best way to analyze it is until the middle of the process, or even the end of the process.

Multiple visual analytics systems provide analytical capabilities. By supporting more flexible data manipulation so that analysts can easily import and remove data from the analysis pool, these systems will be more usable, with better integration into the analysis process.

If the processes of collection and analysis are integrated in a single system, this helps analysts apply structured analytic methods such as ACH, social network analysis, geospatial mapping, and decision matrix. In our interviews, two teams mentioned that if they had more time, they would have tried other analytic techniques. Analysts always want to push their findings and triage, aggressively reshuffling their analysis. One of the most effective ways to do this is to employ multiple analytic methods and compare and contrast findings from each. The ability to try various techniques with the data can help analysts find effective ways for addressing questions and strengthening their analysis.

We had this time crunch. We pretty much got rid of the process of re-evaluating our hypothesis, finding what's the most important to make it perfect, and hitting on that, and going back to the stuff that we didn't deem as important. If we had time, we would fill that in.

Help analysts create convincing production – support insight provenance and sanity checks of analytical products
Production is what differentiates intelligence analysis from general sensemaking which does not necessarily entail external representation. Even when analysts finish their analysis, they need to convert the results into a concise format so that decisionmakers can understand their findings. This can be tedious and time-consuming part of the intelligence process.

When asked about the most difficult part of their project, two teams mentioned production. Interestingly, this difficulty comes from sanity checking and insight provenance, not simply from formatting and writing issues. The sanity check, or qualitative double-check, takes time because data and findings are derived from many sources and analysts have meshed them through the process of collection and analysis. Analysts need to return to original sources and provide a rationale by which their statements are made. They also have to add references to their statements, for which they have to revisit original sources. The following quote from an analyst illustrates those difficulties:

Most difficult part…basically going back through all the sources we used to grade these technologies, people, and companies, then taking basic pieces from those and making a narrative out of it. So explaining why we thought they are the keys and then relating it to the rest of the other findings.

A system that promotes simple insight provenance during analysis could help analysts save their time in production.

Support asynchronous collaboration rather than synchronous collaboration for exploratory analysis

We discussed three different layers of collaboration in the intelligence process and that the degree to which technology can contribute varies. In particular, visual analytics systems seem to have the potential to help collaboration in "sharing" and "content."

From our study, we found that these types of collaboration tend to occur asynchronously, rather than synchronously. When meeting face-to-face, analysts did not work on actual tasks but spent time checking their status, coordinating next steps, and discussing issues. Even when they worked in the same lab for several hours, team members took their own computer and worked individually. Although they often talked to each other, it was for simple coordination issues or specific questions about the content. One analyst stated about his perception on collaboration:

We discussed how each of us interprets the data. We're very group-oriented when it comes to discussing to a consensus. Other than that, we prefer to work individually especially for the actual analysis. Of course we collaborate even when we work on our own parts, but there's no one who really knows about those concepts or entities like you do.

In a nutshell, analysts collaborate cognitively. Rather than trying to build a system that allows analysts to work at the same time in the same workspace, providing a system that promotes individual workspaces but also provides asynchronous collaborative features - such as the ability to share sources and data, view and comment on others' work, and merge individual work together - would appear to be more beneficial.

Note that our findings are based on strategic intelligence. In other types of intelligence such as tactical and operational intelligence, which form the basis for immediate action, real-time collaboration is also important because such intelligence must be shared and used quickly.

Unifying the pieces

Because their typical processes of requirements gathering, collection, analysis, and production are so intertwined, and it takes considerable time to coordinate between different software systems, it appeared to us that analysts want an all-in-one system that can streamline the analysis process and save their time. When asked about their 'dream' system, a few analysts answered:

If I had to go back to the beginning and start all the way over, I should be able to jump back and forth seamlessly between all of these processes. We need a tool that compensates for that.

It should be one program. We spend more time to make it work together. Nothing's compatible with others. We want a program that syncs all the documents. Help us do our visualization with the documents. A program that is compatible with Excel spreadsheet. Don't want to open 20 different programs.

Thus, a hypothetical tool that simplifies the intelligence analysis process would function as follows:

- The analyst enters requirements into the system.
- The system suggests various concepts associated with key terms, phrases, and ideas in the requirements.
- The system automatically draws connections between concepts, but it also allows the analyst to draw connections, group, and organize them.
- The system takes the concepts and starts populating them, collecting information sources using the concepts as keywords (pull sources).
- The system uses sources the analyst identified and suggests new articles relevant to the sources (push sources).

- All of these pulled and pushed sources are integrated into a source repository.
- For documents in the database, the analyst can highlight important facts and annotate his/her thoughts. On demand, the system extracts entities requested by the analyst.
- For intuitive analysis, the analyst can write reports in a preferred format, walking through each document.
- For structured analysis, the system helps the analyst try a variety of structured methods. It takes all the information identified by the analyst and integrates it directly into the methods.
- At the end of the process, when the analyst produces final output, the system automatically links each statement to relevant sources and the process by which the statement was derived.

Thus, analysts could flexibly move between conceptual model, collection, analysis, and production. The system accompanies the analyst from requirements to product in a single platform, speeding up the process, as expressed in one analyst's comment:

If I had something like that, I'd be blazingly fast. I mean I would be able to do this 10-week project in three weeks.

Interestingly, our suggestions reiterate the findings of other researchers who identified the importance of unifying disparate tools in a different domain. In an observational study of the scientific data analysis process, Springmeyer et al [24], concluded that "an effective data analysis environment should provide an integrated set of tools which supports not only visualization, but some of the additional functionality" such as capturing the context of analysis and linking materials from different stages of analysis.

8 CONCLUSION

In this paper, we described an empirical study to understand intelligence analysts and their processes. We observed three teams of student analysts working on typical intelligence problems. Our contributions include documentation of the processes and methods they followed, clarification of issues regarding the intelligence process, and design implications for visual analytics systems for intelligence analysis.

The study has several limitations. We followed only three teams (14 analysts). Also, the analysts were not working professional analysts, but were student analysts-in-training. The analytic questions studied were from strategic intelligence, one type of analysis. Possible future work includes the study of more cases, particularly with professional analysts working on similar or other types of intelligence problems. Of course, the design implications can serve as motivation for new visual analytics systems, ideally created through participatory design with analysts.

REFERENCES

[1] D. C. Andrus. The wiki and the blog: toward a complex adaptive intelligence community. In *Studies in Intelligence*, 49(3), September 2005.

[2] E. Bier, S. Card, and J. Bodnar, Entity-based collaboration tools for intelligence analysis. Proceedings of *IEEE VAST'08 (Columbus, Ohio, October 19-24, 2008)*, pages 99–106. IEEE Press, October 2008.

[3] K. Charmaz. Qualitative interviewing and grounded theory analysis. In *Handbook of Interview Research: Context and Method*, (J. F. Gubrium and J. A. Holstein, eds.), CA: Sage Publications, 2002.

[4] G. Chin, O. A. Kuchar, and K. E. Wolf. Exploring the analytical processes of intelligence analysts. Proceedings of *ACM CHI'09* (Boston, Massachusetts, April 4-9, 2009), pages 11–20. ACM Press, April 2009.

[5] Department of Intelligence Studies at Mercyhurst College. http://intel.mercyhurst.edu/ (accessed March 15, 2011).

[6] A. B. Eli and J. Hutchins. Intelligence after Intellipedia: improving the push pull balance with a social networking utility. Research Report in *Information Science and Technology Directoratet*. Defense Technical Information Center, February 2010.

[7] D. Gotz, M. X. Zhou, and Z. Wen. A study of information gathering and result processing in intelligence analysis. In *IUI 2006 Workshop on IUI for Intelligence Analysis*, 2006.

[8] J. G. Heidenrich. The State of strategic intelligence. In *Studies in Intelligence*, 51(2), 2008.

[9] R. Heuer. Psychology of intelligence analysis. Center for the Study of Intelligence, Central Intelligence Agency: Washington, DC, 1999.

[10] R. J. Heuer and R. H. Pherson. Structured Analytic Techniques for Intelligence Analysis, CQ Press College, 2010.

[11] i2 – Analyst's Notebook. http://www.i2group.com/us (accessed March 15, 2011).

[12] R. Johnston. Analytic culture in the United States intelligence community: an ethnographic study. Central Intelligence Agency: Washington, DC, 2005

[13] G. Klein, B. Moon, and R. R. Hoffman. Making sense of sensemaking 2: a macrocognitive model. In *IEEE Intelligent Systems*, 21(5), September/October 2006.

[14] L. Krizan. Intelligence essentials for everyone. Joint Military Intelligence College Occasional Paper Number 6, June 1999.

[15] D. R. Millen. Rapid ethnography: time deepening strategies for HCI field research. Proceedings of *ACM DIS'00 (New York, New York, August 17-19, 2000)*, pages 280-286. ACM Press, August 2000.

[16] Mindmeister. http://www.mindmeister.com/ (accessed March 15, 2011).

[17] T. Munzner. A nested model for visualization design and validation. In *IEEE Transactions on Visualization and Computer Graphics*, 15(6), pages 921-928. IEEE Press, 2009.

[18] National intelligence estimates. http://www.cfr.org/iraq/national-intelligence-estimates/p7758#p3 (accessed June 07, 2011)

[19] D. R. Norman. Websites you can trust. In *American Libraries*, 37(7), page 36. American Library Association, Aug 2006.

[20] G. M. Olson, J. S. Olson, M. Carter, and M. Storrøsten. Small group design meetings: An analysis of collaboration. In *Human-Computer Interaction*, 7, pges 347-374. 1992.

[21] J. J. Peterson. Appropriate factors to consider when assessing analytic confidence in intelligence analysis (Master's Thesis). Retrieved from http://scip.cms-plus.com/files/Resources/Peterson-Appropriate-Factors-to-Consider.pdf on March 15, 2011.

[22] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. Proceedings of *International Conference on Intelligence Analysis'05 (MacLean, Virginia, 2005),* May 2005.

[23] A. Robinson. Collaborative synthesis of visual analytic results. Proceedings of *IEEE VAST'08* (Columbus, Ohio, October 21-23, 2008), pages 67–74. ACM Press, October 2008.

[24] R. R. Springmeyer, M. M. Blattner, and N. L. Max. A characterization of the scientific data analysis process. Proceedings of IEEE VIS'92, pages 235-242. IEEE Press, 1992.

[25] J. Stasko, C. Gorg, and Z. Liu. Jigsaw: supporting investigative analysis through interactive visualization. In *Information Visualization*, 7(2), pages 118–132, 2008.

[26] A. Strauss and J. Corbin. Basics of qualitative research: Grounded theory procedures and techniques. Sage Publications: Newbury Park, Calif, 1990.

[27] G. F. Treverton. Reshaping national intelligence in an age of information. Cambridge: Cambridge University Press, 2001.

[28] Typewith.me http://typewith.me (accessed March 15, 2011).

[29] K. Wheaton. Wikis in intelligence. Unpublished manuscript, 2011.

[30] W. Wright, D. Schroh, P. Proulx, A. Skaburskis, and B. Cort. The sandbox for analysis: concepts and methods. Proceedings of *ACM CHI '06* (Montreal, Quebec, April 22-27, 2006), pages 801–810. ACM Press, April 2006.

[31] Zotero. http://zotero.org (accessed March 15, 2011).