# Helping Users Recall Their Reasoning Process

Heather Richter Lipford[*], Felesia Stukes[*], Wenwen Dou[*], Matthew E. Hawkins[*], and Remco Chang[*†]

[*]University of North Carolina at Charlotte          [†]Tufts University

**ABSTRACT**

The final product of an analyst's investigation using a visualization is often a report of the discovered knowledge, as well as the methods employed and reasoning behind the discovery. We believe that analysts may have difficulty keeping track of their knowledge discovery process and will require tools to assist in accurately recovering their reasoning. We first report on a study examining analysts' recall of their strategies and methods, demonstrating their lack of memory of the path of knowledge discovery. We then explore whether a tool visualizing the steps of the visual analysis can aid users in recalling their reasoning process. The results of our second study indicate that visualizations of interaction logs can serve as an effective memory aid, allowing analysts to recall additional details of their strategies and decisions.

**KEYWORDS:** Visual analytics, visualization, reasoning process.

**INDEX TERMS:** H.5.2 [Information Interfaces And Presentation (e.g., HCI)]: User Interfaces—Evaluation/methodology; H.5.2 [Information Interfaces And Presentation (e.g., HCI)]: User Interfaces—Graphical user interfaces (GUI)

## 1  INTRODUCTION

The use of visual analytical tools is becoming increasingly important in a variety of domains, aiding analysts in understanding and drawing conclusions from complex data. Yet, utilizing a visualization tool involves more than merely viewing a visual representation. A user's analysis process is comprised of forming and testing hypotheses in order to make decisions based on the knowledge gained from the visualization. This process is often iterative and can account for hours of exploration. Intermediate findings may lead to new strategies and decisions, resulting in a cyclic progression of knowledge building and understanding.

In most cases, the analyst creates a report at the end of the process, documenting both the knowledge discovered and the paths of the investigation [19]. The actual reasoning steps employed are important to record alongside the knowledge in order for the analysis to be explainable, reproducible and trustworthy [3]. In particular, for analyses such as those for national security [3] or financial fraud detection [2], these reasoning steps are often used as a roadmap for potential criminal investigations; the accuracy and degree of detail of these reasoning steps are therefore of critical importance. Yet, in analysis of such complex problems, the amount of knowledge

---

9201 University City Blvd. Charlotte, NC 28223
{Heather.Lipford, f.stukes, wdou1, mhawki30, rchang} @ uncc.edu

gained and applied during the investigation can be tremendous. We hypothesize that without additional support, analysts are unlikely to be able to recreate their exploration and discovery paths, resulting in unintentional errors and gaps in their reporting, potentially limiting the use of that report.

We are exploring several fundamental questions to inform the design of systems that help users document their analysis process: how do people conceptualize their strategies and interactions with a visualization? How much can they remember, and what do they forget? What kinds of tools will aid in the recovery of their process? In this paper, we present two studies exploring these questions. First, we performed a formative evaluation using a financial analysis visualization to examine the memories that analysts have of their strategies and methods when performing an analysis. The results indicate that the analysts' recall of their findings, strategies, and methods often deviate significantly from the actual analyses. Based on these results, we then performed a second study examining how a tool that visualizes the user's interactions with the visualization can aid in that recall. Our results reveal that recall of their reasoning, particularly of their rationale, does improve with the use of visualizations of their interaction logs. The combined results of our two studies have implications for the design of tools to support the recording, documenting, and communicating of the results of visual analysis.

## 2  BACKGROUND

There are a number of research efforts in visualization that focus on the externalization of reasoning processes of an analyst's investigation. Some of these efforts focus on capturing a user's interactions as a means to record the analysis process. The GlassBox project by Cowley et al. aims to extract the analyst's reasoning processes through the examination of low-level user interaction [4, 11]. Recently, Gotz and Zhou incorporated automatic tracking of semantic-level user interactions in the HARVEST system [10]. From a theoretical perspective, Jankun-Kelly et al. propose a comprehensive model for capturing user interactions in visualization tools [13]. Heer et al. presented methods for logging user interactions as well as mechanisms to review, edit and annotate the interactions [12]. History models that have been employed by researchers are comprehensively presented in their work.

Other research efforts consider artifacts of a user's reasoning process. Shrinivasan and van Wijk created the Aruvi framework that allows the user to annotate his/her process during an investigation [18]. Pike et al. facilitate collaboration between multiple analysts in their Scalable Reasoning System by allowing each analyst to record their reasoning processes using a node-link diagram [15]. Lastly, a user's annotations retain semantic connections to their corresponding events in GeoTime [7].

The emphasis of the aforementioned work has been on the creation of effective tools and methods for identifying and encoding reasoning processes and the visual communication of the results. In contrast, our work is examining the motivations behind such research.

Despite significant work in the psychology and cognitive science communities on memory, there has been considerably less

investigation of human's memories of their own reasoning processes, particularly in the complex, exploratory, insight discovery, and knowledge-building tasks [16] that are part of visual analytics. Clinical decision researchers have demonstrated that doctors have little recall of their diagnosing policies (decision processes) [14], and recommend using video-cued interviews in order to uncover their reasoning about a patient's case [20]. However, researchers have demonstrated that retrospective verbalization, having users verbalize their thoughts about a task after the task is completed, is prone to forgetting and fabrication as people inadvertently access similar memory structures [9]. Prompted recall, such as by showing people where they were looking on a webpage during a task using eye-tracking, may reduce such memory failures [8].Thus, we expect that users will not remember their entire reasoning processes, and aim to understand specifically how they conceptualize their own investigation, what they tend to remember and forget and how those memories are affected by the prompted recall provided by interaction logs [5]. To the best of our knowledge, there has been no research in the visualization community that investigates analysts' recall and reconstruction of their reasoning process.

## 3 STUDY 1: RECALL OF ANALYSIS PROCESS

We first examined how well users can recall their reasoning of a visual analysis. While it is not surprising that people would not be able to recall every detail of their process, we wanted to investigate what aspects they would be able to recall, where and why they had memory errors, and how users thought about their investigative process. To that end, we performed an observational study of a visualization of financial transactions. We chose this particular visualization because it was functional enough for detailed use, yet learnable with a brief tutorial. Additionally, we have access to experts in finance and fraud, and we had a realistic but limited data set that was complex enough for a study investigation, but not too overwhelming. We first describe the visualization before discussing the details of the study.

### 3.1 The WireVis Visualization

WireVis [2], see Figure 1, is a hierarchical, interactive visualization with multiple linked views. Financial institutions must implement procedures for detecting and reporting suspicious transactions to U.S. governmental agencies. WireVis was developed to replace a text-based process of wire transaction analysis, helping bank analysts visualize activities that may be indicative of fraud and associated illicit financial activity. WireVis depicts the relationships among accounts, time, and keywords within wire transactions. The heatmap view depicts the relationship between keywords and accounts, the keyword network view shows how keywords relate to each other, the string and beads view shows users the relationship between transaction amounts over time, and the search-by-example tool helps user discover similar accounts.

Due to security and privacy concerns, we created a transactional dataset with no personally identifying information for use in our study. Although none of the transactions in the dataset were real, we captured the characteristics and statistics from real financial transactions. The dataset was designed so that the domain experts in our study could easily identify with the information presented and effectively explore this dataset within the provided time frame. It was also complex enough that interesting and complicated patterns could be discovered. There were a total of
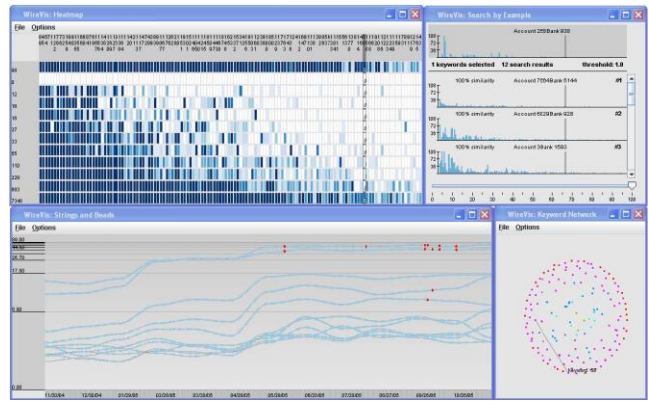


Figure 1. The WireVis system, showing (in clock-wise order from upper left) the Heatmap, Search by Example, Keyword Network, and Strings and Beads views.

300 wire transactions, and 29 keywords in our sample dataset. The keywords represented names of countries, goods, or services. Sample keywords include Mexico, Minerals & Gems, and Pharmaceuticals. We constructed four threat scenarios to inject into the dataset: transactions in which keywords should not appear together, accounts with dual roles, keywords with unusually high transaction amounts, and accounts with suspicious transactional patterns appearing over time. For details of the dataset and the data generation process, please refer to [5]. However, these were not the only transactions that users might consider suspicious and participants had no difficulty identifying suspicious transactions in our studies.

### 3.2 Methodology

We used a mixed qualitative and quantitative approach for our data collection and evaluation. In this first study, we recruited 10 participants from our local business community, with an average of 9.9 years of financial industry experience. Each participant was either currently working as a financial analyst or had financial analyst expertise in a large financial firm. 8 of the 10 participants were trained specifically in fraud detection analysis. Their job titles include, but are not limited to, Anti Money Laundering Analyst, Financial Crime Analyst, Transaction Data Analyst, and Information Securities Analyst.

Evaluation sessions were performed by one of two researchers, following a detailed script to ensure consistency. Initially, we spent approximately 12 minutes training the participants on the concept of wire fraud and the use of WireVis. We introduced each of the four views and features, and then walked each participant through one scenario of finding a suspicious transaction.

Next, participants were instructed to spend 20 uninterrupted minutes to use the visualization to find transactions that they felt were suspicious. The users were asked to think aloud during their exploration. We specifically encouraged the users to describe the steps they were taking, as well as the data or information used to locate what they felt was a suspicious transaction in an effort to reveal their strategies and goals as they went. They were encouraged to ask questions, should they need assistance with a WireVis command. They were also told that there were no right or wrong answers, so they would not second guess their individual strategies and freely explore.

| | | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Findings** | Total | 7 | 8 | 5 | 10 | 9 | 5 | 3 | 7 | 6 | 4 | 6.6 |
| | Repeated | 1 | 4 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0.8 |
| **Strategy** | Correct | 2 | 4 | 3 | 3 | 2 | 3 | 1 | 1 | 1 | 3 | 2.3 |
| | Incomplete | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Guess | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0.2 |
| | No Recall | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0.3 |
| **Method** | Correct | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.1 |
| | Incomplete | 1 | 3 | 1 | 1 | 2 | 2 | 2 | 2 | 5 | 2 | 2.1 |
| | Guess | 1 | 0 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0.5 |
| | No Recall | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Table 1. Recall of the findings, strategies, and methods of the analyses. For findings, "total" represents the number of recalled findings, and "repeated" indicates the number of duplicated findings. For strategies and methods, "correct" denotes the number of accurate recalls, "incomplete" indicates recalls that are partially accurate, "guess" represents when participants were completely inaccurate in their recall, and "no recall" means that the participant explicitly stated he/she could not recall the steps at all.

Once users drilled down to a specific transaction, they were asked to write it down on a Transaction Discovery Sheet because we considered this a key decision point in the user's analysis process. Users were also asked to write down their determinations as to whether the transaction was suspicious, not suspicious, or inconclusive. Once participants documented a specific transaction, they were encouraged to continue looking for others until the time limit was reached.

Following the 20 minutes of discovery, the participants were given a 15 minute semi-structured interview with consistent questions designed to find out what they thought they did during their analysis. They had no visual prompts except the Transaction Discovery Sheet. They were asked to describe their strategies and the steps involved in them, with questions including: "*What was your strategy in searching for suspicious activity?*", "*Describe the 1st step to finding suspicious activity using Strategy X,*" and "*Did Strategy X change during your analysis?*"

Next, we interviewed the participants about each of the suspicious findings they wrote down, and asked them to give the step-by-step descriptions of how they reached that specific suspicious transaction. The participants were also asked questions regarding documenting their overall process, including: "*Which parts of your analysis would you want to record for the purpose of sharing your discovery methods with others?*" and "*What other areas in the recording would be useful to others?*"

### 3.2.1 Data Capture and Analysis

We captured a recording of the screen, mouse clicks, keystrokes, as well as audio of the session with professional screen capturing software. We also logged all WireVis interactions. Using the think-aloud comments, we created a detailed textual timeline for all participants describing their strategies and goals of the analysis tasks. This transcript included descriptions of the user's mouse clicks, the sequence of WireVis views used during the analysis, the users' goals based on the think-aloud comments, and their strategies. We also transcribed the interviews of each participant. With these transcriptions we created a mapping worksheet that provided a side by side comparison of the textual content of what they actually performed during the analysis against the transcription of what each participant said they did in order to identify the differences.

Open coding was first performed independently by two researchers, without any initial categories. Based on this, they discussed and agreed on a categorization to characterize the different aspects and accuracy of a user's process, explained next. The transcripts were then coded a second time using this categorization. This coding was performed by one researcher and

checked by a second; disagreements were verbally discussed and resolved between the two coders.

## 3.3 Study 1 Results

In evaluating our participants' memories, we analyzed the results along three different aspects of the analysis process: findings, strategies, and methods. A finding is the knowledge gained – the transactions recorded on the discovery sheet. A strategy is the analyst's plan to discover a suspicious transaction, while a method is the steps taken in carrying out the current strategy. To help quantitatively analyze our data, we used the mapping worksheet to identify how many of the participant's strategies and methods were accurately remembered, how many were remembered incompletely, and how many were guessed or not remembered at all.

### 3.3.1 Findings

We asked participants to write down their findings, which they did an average of 6.4 times during their session, 3 of which were deemed suspicious. While we did not ask users to recall findings after the session as they were written down, we did notice memory problems during their analysis tasks. As Table 1 summarizes, several participants duplicated findings, writing down the same transaction more than once. Not all duplicates are necessarily negative. 2 participants found the same transaction through two different strategies, providing additional evidence for that particular finding. However, 3 participants repeated the finding using the same strategy, which did not add additional knowledge and was inefficient and redundant. Thus, users may have had difficulty remembering their knowledge and process even during the analysis task itself, leading to wasted time.

### 3.3.2 Strategies

The participants developed strategies during the training as to how to begin their undirected searches. We grouped their resulting strategies into 3 main categories. The most popular kind of strategy was keyword-based, locating an account with two or more keywords that they determined should not appear together. For example the keyword 'Pharmaceuticals' represented a known business front for illicit activities to several analysts. When paired with the geographical keyword, 'Mexico', 'Pharmaceuticals' was more likely to lead to suspicious activity through further analysis.

A second kind of strategy was to group cells on the heatmap based on their shade. The darker the cell shading, the more frequent that keyword occurs in that account or group of accounts. Many of our participants concluded that if they looked at the account groupings with the most activity indicated by the darker

shading, then they would be more likely to discover a suspicious finding. Alternatively, some participants chose to look at the lightly shaded cells to investigate transactions that occurred less frequently, which often indicates an anomaly within accounts. The final kind of strategy was to view the String and Beads view to find transaction amounts that appeared unusually high over time. Most users described this strategy in terms of looking for 'spikes.'

Participants tended to describe their strategies in terms of the visualization and interface, instead of the semantic meaning of their actions. So to describe their strategies participants described clicking on keywords (to highlight them), looking for dark boxes (indicating multiple transactions), and looking for spikes in the line (indicating a change in transaction value). For example: "*My eyes are like a counter clockwise... Then start out here and go around. (gesture spiral) ... What are the financial anomalies?*"

As Table 1 shows, users were in general aware of and accurate in identifying their high level strategies. They were able to relate their strategy to real world examples, and were often quite passionate and colorful in explaining them. However, they could not recall if those strategies changed or evolved during their analysis. One user responded by stating, "*I didn't intentionally change it…. But there was variation to my approach I'm sure.*"

### 3.3.3 Methods

We believe that effectively communicating how users carry out their strategies to reach their findings is a very important source of expert knowledge, particularly for analyses involving criminal investigations. Our participants were asked to recall their methods and exact paths they took in carrying out each strategy. As Table 1 shows, users were almost completely unable to accurately recall their analysis paths. And not surprisingly, the more complex the path, the more inaccurate the memory was. While the users attempted to provide a sequential account of what they thought they did, the ordering of those steps was inaccurate. The recall was more free form, identifying some of the steps that occurred, but not in the correct sequential order. Participants often interjected and re-ordered steps from their entire analysis session, in addition to the steps involved in investigating the specific transaction in question.

For example, P4 recalled her method regarding a suspicious finding: "*I went on the map (Heatmap) and something on there did not match up. I don't remember exactly what it was. From there I saw in the String and Beads view at the bottom there was an abnormal spike, then I pulled it out.*"

P4's recollection was partially true; however she did not remember that she saw an unusual relationship between three keywords (Raw Materials, Food, Arts & Crafts) which ultimately led to the finding. This was common – participants often remembered their initial or intended strategy and keywords, but could not recall the subsequent keywords they used to lead to a particular finding. So they often remembered the portion of the method that used their initial strategies, but then could not recall additional steps. When participants struggled with their memories, they returned to talking about their initial strategy, describing "*What I was trying to do is…*", rather than what they were actually doing at the time and what actually led to the finding.

## 3.4    Study 1: Summary

To summarize the results from Study 1: Participants were able to accurately recall their starting strategies, and tended to conceptualize these strategies in terms of the information representation in the visualization. They were unable to accurately recall the methods they took within those strategies, and how their strategies and methods evolved and changed through the analysis.

We also found that our participants tended to overlook steps involving unfruitful explorations and dead end paths. In addition, they sometimes repeated findings, indicating memory problems during analysis itself.

Thus, if analysts strictly rely on their memories to report their results, those reports may be inaccurate and not contain sufficient details to effectively communicate the knowledge and decisions gained from the analysis session. These results led us to investigate what might help with this recall, and specifically the impact on memories of the reasoning process when analysts are provided with a visualization of their interaction logs.

## 4    STUDY 2: AIDED RECALL

In this study, we hypothesize that if a visualization tool can display each step of the analyst's interaction with the WireVis tool, the analyst should be able to retrace her analysis process and thereby improve the recall of her findings, strategies, and methods. To validate our hypothesis, we utilized a previously developed interaction-log visualization tool called the Operation Analysis Tool [5] in the second study. We have previously examined whether this tool can help other users reconstruct some of the reasoning process of an analysis [6]. Our goal now is to determine how analysts themselves recall their reasoning when using the tool. To our knowledge, this is the first exploration of recall when using such a tool.

### 4.1    The Operation Analysis Tool

To view analysis activities, we utilized the Operation Analysis Tool shown in Figure 2, showing interactions with WireVis over time [2]. The x-axis of the main view represents time, with a striped background indicating the length of a fixed time duration (defaulted to 60 seconds per strip). The y-axis is divided into 5 sections, with each section supporting one aspect of the participant's investigation process [5]. Sections A, B and C are made up of rows that correspond to the different views in WireVis: the top row shows the points at which the user entered a new transaction into the Transaction Sheet. The next row is the heatmap view showing which keywords were hovered over or clicked on. Below is the strings-and-beads view showing the transactions being viewed. In addition, the depth of the analysis is shown via the number of transactions that are visible, as well as the areas the user is exploring. Hovering over an event shows the transaction details on the right-hand panel.

This tool was not meant to represent an ideal visualization for an end-user analyst. Instead, before we begin that work, we wanted to evaluate whether such a tool would be useful and provide additional reasoning information to analysts.

### 4.2    Methodology

In the second user study, approximately one year after the first, we again recruited 10 participants from our local business community. We first attempted to recruit participants from the first study to enable cross-study comparisons, and were able to recruit two. Enough time had passed that they did not remember their previous activities and had to be retrained on WireVis. Our participants had an average of 7.95 years of financial industry experience, working in either an analyst capacity or in a risk management role. Their job titles include, but are not limited to Compliance & Operational Risk Developer, Fixed Income Risk Manager, and Fraud Detection Business Analyst.

Most of this study was identical to the first study, with only minor adjustments. Participants were first trained on WireVis, but we were able to cut this training to 7 minutes by omitting the
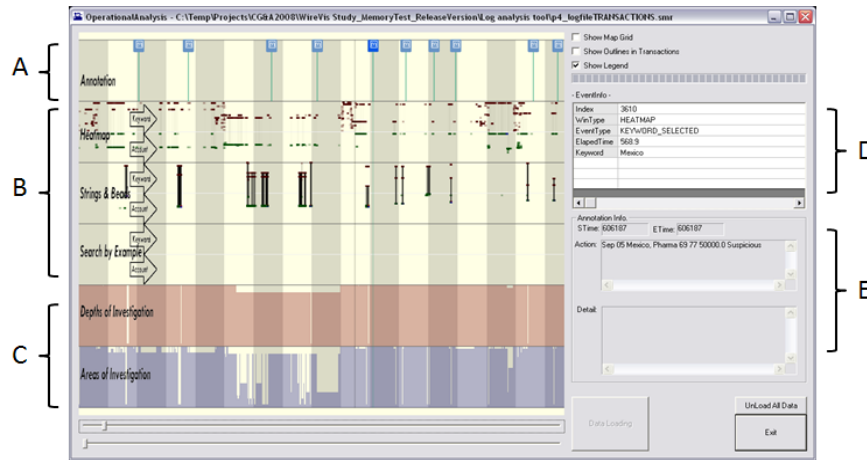
Figure 2.The Operation Analysis Tool. (A) shows the recorded transactions. (B) shows the participant's interactions with the views in WireVis. (C) shows the depth and areas of investigation. (D) shows the event details while (E) shows the transaction details.

details of the WireVis functionality that were not useful for participants. For example, we left out the Search By Example view as no one used it in the previous study. Next, participants were instructed to spend 20 uninterrupted minutes to use the visualization to find transactions that they felt were suspicious, as before. The only difference this time was when the user recorded their transactions, we provided an electronic form instead of the paper transaction sheet so that the findings could be logged and included in the operation analysis visualization.

Following the 20 minutes of discovery, the study participant was again interviewed concerning the methods they used to reach the findings listed on their electronic transaction recorder. The user's recall at this time was unaided and focused on their suspicious transactions. As with the first study, we asked participants to recall what they did to arrive at each suspicious finding. However, this part of the interview was shorter than the first study in order to accommodate the aided recall portion below.

The final part of the second study was entirely new. We closed WireVis and opened the Operation Analysis Tool. First we loaded the users' interaction log file and gave each participant a brief overview of how to zoom and pan around the Operation Analysis Tool window. In this overview, we also highlighted how the tool visualized their WireVis interaction data through the use of color, annotation and time to represent what they did. Next, we asked each participant to start at the beginning of this visual history and describe what they did during their analysis based on what they saw on the visualization.

### 4.3 Study 2 Results

Similar to the first study, our analysis involved coding each participant's strategies, methods, and findings and mapping their recalled methods to the transcript of what they actually performed. This was again performed by one researcher, and checked by a second with verbal resolution of disagreements. In addition, we looked at the types of information participants shared while using the operation analysis visualization and how this information compared with their unaided recall. The coding for the rationale was performed as open coding by one researcher.

#### 4.3.1 Findings, Strategies, and Methods

Table 2 reports on the same statistics as Study 1 (although since the interview was shorter in Study 2, the numbers are not directly

comparable between the two studies). Participants had 8.4 findings on average, of which 4.2 were deemed suspicious. When asked to recall how they found a suspicious transaction, once again the participants were very accurate in recalling their overall strategies but were not accurate in completely recalling their methods. In this interview we did not press the participants to report each and every step they took, thus they provided fewer guesses of their methods in this study and only reported what they thought they remembered.

#### 4.3.2 Aided Recall

In Study 1, it was highly unlikely for analysts to correctly describe how they reached a decision point during their analysis process. We found that using the Operation Analysis Tool to view the WireVis interaction logs reduced the user's memory load for recalling their strategies and methods. This allowed them to instead report additional details of the strategies and methods they had not previously reported or remembered.

For example, one user initially did not recall considering a particular keyword, but while using the tool he discovered that he had: "*I didn't add anything, obviously, from Mexico and Pharmaceuticals, although wait a minute I did, I did add one from pharmaceutical.*"

The operation analysis visualization aided participants' recall of new strategies that were not previously recovered, as shown in Table 3. For example, while using the tool, several participants pointed out times when they were exploring WireVis without a direct strategy in mind. These periods of exploration were not revealed during the participant's unaided recall portion of their interview. The use of the Operation Analysis tool prompted participants to make comments such as: "[The] *first part is me exploring the interface, had to learn how the system worked*", "*Here I go through all the different keyword categories, just scanning..*,"and "*Well when I started I don't think I had a very clear plan of what to look at.*"

Additionally, participants in Study 2 discussed more of their false paths that did not lead to a decision point when viewing the interaction logs. Most participants just started talking through their process using the operation analysis tool from beginning to end. And when they ran across a strategy that did not result in a finding, they were able to identify it as such: "*then I guess I took a look at some different transactions and didn't result in anything.*"

| | | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Findings** | Total | 5 | 7 | 4 | 11 | 18 | 8 | 14 | 3 | 9 | 5 | 8.4 |
| | Repeated | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 3 | 0 | 0.4 |
| **Strategy** | Correct | 1 | 7 | 2 | 3 | 9 | 4 | 1 | 3 | 5 | 4 | 3.9 |
| | Incomplete | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0.4 |
| | Guess | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | No Recall | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Method** | Correct | 1 | 4 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0.9 |
| | Incomplete | 0 | 3 | 3 | 3 | 9 | 3 | 1 | 2 | 6 | 3 | 3.3 |
| | Guess | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.1 |
| | No Recall | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Table 2. Results of findings and unaided recall of strategies and methods in Study 2.

| | | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Strategy** | New Recall | 2 | 0 | 0 | 0 | 1 | 0 | 1 | 2 | 1 | 1 | 0.8 |
| | New Details | 2 | 2 | 3 | 4 | 6 | 2 | 5 | 4 | 3 | 2 | 3.3 |
| **Method** | New Recall | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | New Details | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.1 |

Table 3. Number of comments providing new details while using the Operation Analysis Tool. New Recall is the reporting of additional strategies and methods, whereas new details are additional comments about strategies and methods that were previously recalled while not using the Operation Analysis Tool.

| Rationale | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P8 | P10 | Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Strategy-based** | 4 | 1 | 1 | 0 | 3 | 0 | 2 | 3 | 1 | 0 | 1.5 |
| **Keyword-based** | 4 | 1 | 5 | 6 | 3 | 0 | 0 | 1 | 0 | 1 | 2.1 |
| **Method-based** | 0 | 0 | 0 | 0 | 0 | 3 | 2 | 1 | 0 | 0 | 0.6 |
| **View-based** | 0 | 1 | 0 | 1 | 2 | 0 | 2 | 1 | 0 | 0 | 0.7 |
| **Total # Rationale** | 8 | 3 | 6 | 7 | 8 | 3 | 6 | 6 | 2 | 1 | 5 |

Table 4. Number of comments of reasoning rationale provided while using the Operation Analysis Tool. Strategy-based and Method-based comments explained why a particular strategy or method was utilized. Keyword-based were comments explaining the use of particular keywords. View-based comments explained why a WireVis view was utilized.

### 4.3.3 Rationale

During the interviews, we did not ask participants to identify why they were taking the steps they took to reach a decision point. Our questions just asked them to discuss what they were doing while looking at their interaction logs using the operational analysis tool. Yet, in doing so, every participant mentioned their rationale behind different strategies and decisions, on average five comments per participant. The participants' rationale-related comments were compared with their unaided recall to verify that the details had not been previously mentioned. As shown in Table 4, we categorized the types of rationale recovered into strategy-based rationale, method-based rationale, keyword-based rationale, and WireVis view-based rationale. We also attempted to verify whether this rationale was remembered correctly by comparing against the think aloud comments from the analysis. None of the rationale appeared incorrect, but much of it was not reported in the think aloud. Thus, we were able to verify approximately 50% of these comments.

Strategy-based rationale were the reasons why the participant used a particular strategy to reach a decision point. In the first study, participants tended to talk about their intended strategy – what they were trying to look for. With their steps in front of them, this time participants were able to talk more about the actual strategy as it evolved and why. We were also able to learn much more about what aspects of those methods or strategies really mattered to them, and what it was that actually led them to the next step or a particular transaction. One user demonstrates this by saying: "*I think my reaction was, or the reason I kept going to that was because when we have attacks on the financial scene [it] is very concentrated by regions*."

Participants also frequently identified keyword-based rationale as the reasons why they used a particular keyword or combination of keywords to reach a decision point. We found that our participants made comments to support this type of rationale most frequently, such as: "*Then I jumped into Transportation, Car Parts. I was thinking maybe some sort of chop shop,*" and "*Transportation, Raw Materials, and Electronics, I was thinking there with the transportation and electronics sort of a black-market operation.*"

Method-based rationale were reasons why the participant chose a particular sequence of steps to reach a decision point. We also labeled view-based rationale as reasons why the participant used a particular view in WireVis. One user asserted, "*here I was looking at different transactions on the string and beads and that is what was very helpful because you can look at each string [to] look at a number of transactions.*"

We revisited the unaided recall of the first and second studies looking for similar rationale-related comments and found almost no instances of such comments.

### 4.3.4 Confidence

An interesting observation we did not anticipate was the impact that the operation analysis tool had on the participants' confidence in the strategies and methods they used to reach a decision point. To examine this, we looked at the qualitative comments that

revealed information related to the participants' self-assuredness in their judgment and ability.

In the first study, and in the unaided recall portion of the 2nd study, participants often expressed feelings of unease or inadequacy when being interviewed. Sample comments that revealed a lack of confidence are: "*I'm not entirely sure*", "*I probably got off track*" and "*I don't know that I'm grouping things correctly.*" Some of this lack of confidence may be due to having to talk to an interviewer about their process, feeling like a memory test. Yet, this lack of confidence with unaided recall may have important implications for reporting.

Conversely, when reviewing their own logs in the 2nd study, the participants used noticeably different language with less concern about their accuracy and more emphasis on recalling details to support their strategies and methods. The participants did not hesitate to report when they were exploring something that did not contribute to a finding and did not seem to associate failure with these false paths.

Confidence is of particular importance to analysts because their role is often linked to making vital recommendations to others based on their findings. The visual representation of their interaction logs seemed to act as a digital witness, validating their strategies and methods beyond the user's personal, and potentially flawed, unaided recall of how they reached each decision point. Moreover, if an analyst is not highly confident in how he reached a decision point, he may choose not to make the recommendation due to the fear of being wrong.

## 4.4     Study 2: Summary

To summarize the results of this study: Participants doing unaided recall had similar results as the first study. When these participants then used the Operation Analysis Tool, they were able to remember more details and in particular provided rationale of their analysis strategies and methods, all with greater confidence. The result was that the analysts were able to more accurately report not just their intended and initial strategies for discovering a particular transaction, but what aspects of the knowledge evolved those strategies and actually led to their findings.

## 5     DISCUSSION

Our results demonstrated that our participants were able to easily recall their initial analysis strategies. However, they were very inaccurate in identifying their methods for achieving those strategies and the changes in strategy that occurred during their analysis. Thus, users remembered their intent, and some of the story constructed around the suspicious transaction. But they did not recall how they actually got there.

So why were the participants good at remembering high level strategies? "People never reason from scratch [17]", they utilize the help of past experience to formulate plans of action and decide between them. As many participants mentioned in both studies, their strategies were formed based on their experience as financial analysts. Thus, the cues of those strategies were already stored in their long-term memories. The few initial strategies that were not remembered were strategies based on the visual representations, such as "looking for keywords far away from each other in the Keyword Network View". These strategies were formed based on interacting with the visualization, and recalling them without any visual cues was more difficult.

One possible reason for users' lack of method recall was that they were focused intently on the task at hand. They appeared immersed in the visualization, deeply concentrating. Thus, the users' cognitive abilities were dedicated to the task, and not on their process and committing those details to memory. This lack

of memory was also evident in the repeated examination of the same transaction by multiple users.

In the first study, we also asked participants what they would have liked to record or document to understand what they considered important to remember. Many felt that the steps that resulted in the location of a suspicious transaction would be the most useful and worthy of reporting to others. Thus, while users were concerned with finding and constructing explanations for suspicious transactions, they were not paying as much attention to the dead ends, false paths, and changes to their plans that occurred along the way. Yet, sometimes these false paths resulted in a new insight that did lead to a finding. Thus, they often were not accurate about all of the knowledge that led them to determine that a transaction was fraudulent. They also dismissed certain views as not useful, even though they actually used them in their analysis process.

There are several interesting implications. First, users conceptualize their analysis process as a set of intended strategies that lead to successful findings, in this case, finding suspicious transactions. They are able to accurately recall their initial strategies, and use them to explain their reasoning at a high level. Yet, they were unaware of how their processes evolved and changed. This would make it difficult to modify or learn better strategies over time. It may make it more likely that they or others will repeat unsuccessful strategies and methods if they do not track or communicate them effectively. And it may make it more difficult to recreate the process by those who need to further investigate those findings.

An additional implication is that if users were charged with documenting their reasoning along the way, they may still struggle to create an accurate record. Having to keep track of their steps and rationale during analysis would break the user's flow, potentially interfering with the analysis itself. Our participants struggled to even think aloud at times. And, users would likely choose to mostly document the high level strategies that led to suspicious findings, and not all the rest. While accurate, this incomplete record may not provide the reasoning needed to fully communicate and learn from the analysis. It may also make it more difficult in this particular domain of financial fraud for those pursuing criminal investigations to sufficiently replicate the reasoning steps.

Our second study demonstrated that with cued recall, when the users were freed from the burden of having to remember what actions they took, they were able to think about and remember more of their rationale and decision points. The story of the analysis was no longer about the intended strategy, but instead the evolving factors that actually led the analyst to make decisions about transactions.

We believe our results illustrate that interaction logs are useful for the analysts themselves, and can provide valuable cues to spark recall about thoughts, decisions, and rationale of an analysis. And with greater confidence than unaided recall. Although the operation analysis visualization sufficiently communicated the actions the analyst had taken for the purpose of our study, it was not designed for the needs of the end user. We believe our results suggest several design guidelines for iterating on our Operation Analysis Tool.

First, our users mostly concentrated on sets of actions, looking for where their strategy or methods changed, such as moving from exploring a number of keywords, to focusing on several specific keywords, to looking at accounts in the string and beads view. Thus, the tool could simplify the interface by more directly visualizing such changes. The tool could also allow users to annotate the log visualization with their rationale to further

document their reasoning, and enable better communication and comparison across groups of analysts to enhance decision making.

The most difficult strategies and methods for our participants to remember had to do with visual cues in WireVis that were no longer present on the screen during recall. Thus, the tool could also be improved by linking the actions to snapshots or even video so that users could view exactly what they saw at the time, possibly sparking additional reasoning recovery. However, we also believe that a tool should not solely consist of video, the actual action logs are more easily interpreted. In a pilot study, we had originally asked users to view a video of their analysis session as part of the interview. We found that users had a hard time following the video as actions were hard to discern. None of the views change drastically with any particular action, so following along with just video was challenging.

## 6 LIMITATIONS

There are several limitations to our study methodology. Our participants were learning new tools, and may have behaved differently or had improved memory with more experience. Their analysis was short, and their recall occurred immediately after their analysis. However, longer sessions with more time in between would likely lead to even less recall. We need additional studies of longer term, real world analyses in order to extend our results. Yet, such field studies are also extremely difficult to perform given the sensitive nature of many analysts' work and the heavy demands on their time. We would also like to compare the usage of interaction histories with traditional, manual methods for annotating and tracking reasoning processes, such as pen and paper and screen captures. Despite the limitations, we believe our results do advance our understanding of the analysis process and suggest the importance for further study of the use of interaction histories for recovering and reporting analytic reasoning.

## 7 CONCLUSION AND FUTURE WORK

A number of visual analytics tools have been suggested to help analysts document their reasoning process both during and after analysis [3]. We believe we are one of the first to examine the participants' recall of their analysis processes to more deeply inform these efforts. Our studies demonstrate that even in a short time span, participants are unable to accurately recall the steps in their analysis process from memory alone. The visualization of their interactions not only relieved this memory burden, but allowed users to also recall and discuss their rationale and decision points more confidently and clearly. This rationale is likely important to document and communicate, showing how the insights were derived. Thus, being able to view the interaction logs is valuable even if the exact steps of every method are themselves not important to report for a particular analysis.

These results need further investigation to inform real world use of analysis tools and long term memory and documentation of results. As one of our participants commented about the Operation Analysis Tool: "*if you brought me back next week, I'd have no idea what I did until I looked at this. Especially if I came back every day this week and explored different data sets and then you said 'What did you do Monday?', it would be impossible until I went back to this.*" Analysts may view a variety of different data sets for a much longer time period over many days, which may require additional features to help with recall and documentation. However, we believe interaction logs are promising for aiding in the reporting of larger scale analyses of visualizations. We continue to examine the automated capture and annotation of users' interactions with complex visualizations.

## REFERENCES

[1] D.A. Bowman, E.T. Davis, L.F. Hodges and A.N. Badre. Maintaining Spatial Orientation during Travel in an Immersive Virtual Environment. Presence: Teleoperators and Virtual Environment. P618-631, 1999.

[2] R. Chang, M. Ghoniem, R. Kosara, W. Ribarsky, J. Yang, E. Suma, C. Ziemkiewicz, D. Kern, and A. Sudjianto. Wirevis: Visualization of categorical, time-varying data from financial transactions. In Proc. *IEEE Visual Analytics Science and Technology (VAST)*, pages 155–162, 2007.

[3] N. Chinchor and W.A. Pike. The science of analytic reporting. *Information Visualization*, 8:286-293, 2009.

[4] P. Cowley, L. Nowell, and J. Scholtz. Glass box: An instrumented infrastructure for supporting human interaction with information. In *Proceedings of HICSS '05*, pages 296c–296c, January 2005.

[5] J. H. Dong, W. Dou, F. Stukes, W. Ribarsky, H. R. Lipford, and R. Chang. "Evaluating the Relationship between User Interaction and Financial Visual Analysis." In *Proc. IEEE Visual Analytics Science and Technology (VAST)*, Nov. 2008.

[6] W. Dou, D.H. Jeong, F. Stukes, W. Ribarsky, H.R.Lipford, and R. Chang. Recovering reasoning processes from user interactions, *IEEE Computer Graphics and Applications*, 29(3):52–61, 2009.

[7] R. Eccles, T. Kapler, R. Harper, W. Wright, Stories in GeoTime, *IEEE Visual Analytics Science and Technology (VAST),* 2007.

[8] N. Eger, L.J. Ball, R. Stephens, and J. Dodd. Cueing Retrospective Verbal Reports in Usability Testing Through Eye-Movement Replay. In *Proceedings of BCS HCI,* 2007.

[9] A.K. Ericsson, and H.A. Simon. *Protocol Analysis: Verbal Reports as Data.* MIT Press, Cambridge, MA 1993.

[10] D. Gotz and M.X. Zhou. Characterizing Users' Visual Analytic Activity for Insight Provenance. *IEEE Visual Analytics Science and Technology (VAST)*, Columbus, Ohio, 2008.

[11] F. Greitzer. Methodology, metrics and measures for testing and evaluation of intelligence analysis tools. PNWD-3550, Battelle-Pacific Northwest Division, Richland, WA, 2005.

[12] J. Heer, J. D. Mackinlay, C. Stolte, and M. Agrawala. Graphical histories for visualization: Supporting analysis, communication, and evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1189-1196, 2008.

[13] T. Jankun-Kelly, K.-L. Ma, and M. Gertz. A model and framework for visualization exploration. *IEEE Transactions on Visualization and Computer Graphics*, 13(2):357–369, March/April 2007.

[14] J.R. Kirwan, D.M. Chaput De Saintonge, C.R.B. Joyce, J. Holmes, and H.L.F. Currey. Inability of rheumatologists to describe their policies for assessing rheumatoid arthritis. Annals of the Rheumatic Diseases, 45, 156-161, 1986.

[15] W. Pike, R. May, and A. Turner. Supporting knowledge transfer through decomposable reasoning artifacts. In *Proceedings of HICSS 2007*, pages 204c–204c, January 2007.

[16] W. Ribarsky, B. Fisher, and W.M. Pottenger. Science of analytical reasoning. *Information Visualization*, 8: 254–262, 2009.

[17] R. Shank and R. Farrell: Memory. Understanding Cognitive Science. 120-133, 1988.

[18] Y. B. Shrinivasan and J. J. va Wijk. Supporting the analytical reasoning process in information visualization. In Proceedings of the *ACM Conference on Human Factors in Computing Systems (CHI 08)*, 1237-1246, 2008.

[19] J. Thomas and K. Cook. Illuminating the Path: The Research and Development Agenda for Visual Analytics. National Visualization and Analytics Center, 2005.

[20] C.A. Unsworth. Using a Head-Mounted Video Camera to Explore Current Conceptualizations of Clinical Reasoning in Occupational Therapy. *The American Journal of Occupational Therapy*. 59(1), 31-40, 2005.