Data visualization for fraud detection: Practice implications and a call for future research

William N. Dilla a,⁎, Robyn L. Raschke b

a Department of Accounting, Iowa State University, USA
b Department of Accounting, University of Nevada, Las Vegas, USA

A R T I C L E  I N F O

Article history:
Received 8 June 2012
Received in revised form 7 January 2015
Accepted 11 January 2015
Available online 27 January 2015

A B S T R A C T

Analysis of data to detect transaction anomalies is an important fraud detection procedure. Interactive data visualization tools that allow the investigator to change the representation of data from text to graphics and filter out subsets of transactions for further investigation have substantial potential for making the detection of fraudulent transactions more efficient and effective. However, little research to date has directly examined the efficacy of data visualization techniques for fraud detection. In this paper, we develop a theoretical framework to predict when and how investigators might use data visualization techniques to detect fraudulent transactions. We use this framework to develop testable propositions and research questions related to this topic. The paper concludes by discussing how academic research might proceed in investigating the efficacy of interactive data visualization tools for fraud detection.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

Fraud investigators have recently recognized the importance of data visualization for fraud detection, and are starting to implement this technique in practice (Deloitte, 2011; Clopton et al., 2014). Data visualization is especially important in the early stages of fraud investigation, where the investigator is attempting to perform an efficient and effective data analysis and desires to better understand the relationships that may be present in a complex data set. The fraud investigator may employ data visualization as a proactive detection approach, using it to search for data patterns that suggest fraudulent activity (Albrecht et al., 2012). Alternately, the investigator might be evaluating a predication of fraud, using data visualization to develop a fraud theory.
that is subsequently used to select additional investigative procedures (Wells, 2003). In either case, the investigator is following a hypothesis testing approach, developing preliminary hypotheses about fraud and analyzing relevant data to see if they appear to be true. If the data analysis phase of a fraud investigation does not support the hypothesis that fraud has occurred, then the investigation ends. On the other hand, if data analysis produces results consistent with the hypothesis that fraud has occurred, the investigator will proceed to conduct other investigative procedures, such as examination of documents and interviewing possible witnesses (Wells, 2003).

Since fraudulent actions are deliberate and non-random, traditional audit methods involving the use of statistical sampling are often ineffective for discovering fraud. Hence, fraud examination professionals recommend the use of data mining procedures for detecting fraudulent transactions (Kranacher et al., 2011; Albrecht et al., 2012). In performing data mining procedures, investigators may brainstorm about possible irregularities that could occur in the business processes or transactions they are examining. Alternately, they may have been given a predication that fraud has occurred. Regardless of whether they identify possible fraudulent activity through brainstorming or receive a predication of fraud, investigators then outline ways that the schemes might show up in data patterns. For each indicator thus identified, investigators design a data mining query or procedure intended to identify whether there are individual transactions that need to be examined more closely.

Audit software packages such as ACL and IDEA facilitate data mining for fraudulent transactions within organizations (Lanza, 2004). Additionally, custom-designed software may be used to identify items for further investigation in complex, high-risk transaction environments (e.g., Chang et al., 2008; Pryke, 2010). However, interpreting the output from these tools may require considerable skill, as anomalies in data may not be readily apparent, except to the expert investigator. Graphical analysis may facilitate identifying suspicious patterns of transactions in data (Lanza, 2005a). While spreadsheet programs can facilitate graphical analysis, such analysis can be cumbersome— if the user wants to change the variables being graphed or focus on a subset of the data, it is usually necessary to generate a new graph. Interactive data visualization programs that allow the user to more easily change the data being graphed or its format have now become readily available (e.g., Centrifuge Systems, Inc., 2015; SAP, 2010; Tableau Software, 2010; TIBCO, 2010). Given the potential for interactive data visualization to assist investigators in seeing and understanding data patterns that are consistent with fraudulent activity, forensic accounting practitioners have recently recommended the use of this technology as an investigative tool (Deloitte, 2011; Clopton et al., 2014).

Proponents of data visualization software contend that it facilitates better decisions by supporting visual thinking. For example, data visualization consultant and author Stephen Few states that

“Visual analysis software allows us to not only represent data graphically, but to also interact with those visual representations to change the nature of the display, filter out what's not relevant, drill into lower levels of detail, and highlight subsets of data across multiple graphs simultaneously. This makes good use of our eyes and assists our brains, resulting in insights that cannot be matched by traditional approaches (Few, 2007).”

Further, the developers of Tableau Software claim that

“Genuine data visualization supports visual thinking. The human brain can process a picture much faster than a table of numbers. The right presentation, using the best practices of information visualization, makes organizing and understanding information simple. Features, trends and outliers show up the way they never do in rows and columns (Tableau Software, 2010).”

While these statements make a plausible argument for why data visualization software might facilitate detecting fraudulent transactions, research evidence suggests that the efficacy of interactive graphical displays for decision making depends on task and user characteristics (Lurie and Mason, 2007; Baker et al., 2009; Dilla et al., 2010; Yigitbasioglu and Velcu, 2012). To date, little research has examined whether and how data visualization software might allow investigators to see patterns in data that are indicative of fraudulent activity. Conducting research on this topic is important, given that there are substantial training costs associated with adopting data visualization software, and it is important that practitioners know under what circumstances such software might facilitate more efficient and effective detection of fraudulent transactions. Therefore, the
objectives of this paper are to describe data visualization techniques, discuss how they might be useful in detecting fraudulent transactions, and build a framework that identifies opportunities for future research.

The remainder of this paper is organized as follows. First, we describe data visualization techniques and provide an example of how they might be applied to detect fraudulent transactions. Second, we identify and describe areas where data visualization techniques might be applied to facilitate detecting fraudulent transactions. Third, we present a framework for investigating factors that affect the efficacy of data visualization techniques for detecting fraudulent transactions and use this framework to develop testable propositions and research questions. The paper closes with a summary and discussion of implications.

2. Data visualization techniques

2.1. Overview

We use the term data visualization to refer to the “use of computer-supported, interactive, visual representations of data to amplify cognition, or the acquisition and use of knowledge” (Card et al., 1999, 6). Throughout this paper, we assume the use of interactive data visualization techniques. In contrast to static data visualization, interactive data visualization enables decision makers to specify the format used to display information (i.e., interactive visual representation), select the information they view as most relevant for decision making (i.e., interactive selection), or both (Dilla et al., 2010). Thus, interactive data visualization is an “on demand” visualization process that allows decision makers to navigate to selected data and display it at various levels of detail or in various formats.

We adapt Yi et al.’s (2007) taxonomy of data interaction techniques to further describe the various tools used to select data views and change the manner in which data are represented. As shown in Table 1, interactive data representation tools allow decision makers to change the encoding of data (i.e., from tables to graphs or vice versa), reconfigure displays of graphical data, and connect data items in large or complex displays. Data selection tools perform functions such as selecting and marking data items of interest for further examination, changing the level of elaboration (i.e., altering the data view from an overview down to details of individual cases) or abstraction (i.e., shifting the view from individual cases back to an overview), filtering data through query tools, and exploring a large set of data through hyperlinks or visual panning techniques.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representation tools</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Encode</td>
<td>Show different representations of data.</td>
<td>Convert tabular representations to graphs or vice versa. Change graph type (e.g., from pie chart to histogram).</td>
</tr>
<tr>
<td>Reconfigure</td>
<td>Show different arrangements of data.</td>
<td>Adjust baselines or axis scales, reverse attributes displayed on x- and y-axes.</td>
</tr>
<tr>
<td>Connect</td>
<td>Show related data items.</td>
<td>View leveled set of data flow diagrams or entity relationship diagrams. Highlight patterns in complex transaction data.</td>
</tr>
<tr>
<td>Data selection tools</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Select</td>
<td>Mark data items of interest.</td>
<td>Highlight selected items in large spreadsheets or graphical displays. Item remains highlighted, so it can be located even after rearranging the display.</td>
</tr>
<tr>
<td>Elaborate/abstract</td>
<td>Show more or less detail.</td>
<td>Move cursor over screen to view more or less detailed information (e.g., show data underlying a segment of a graphical display).</td>
</tr>
<tr>
<td>Filter</td>
<td>Show data based on specific condition(s).</td>
<td>Query tools embedded in database and spreadsheet products, enterprise computer programs (e.g., Oracle, SAP, PeopleSoft), or specialized audit programs (e.g., ACL, IDEA).</td>
</tr>
<tr>
<td>Explore</td>
<td>Show other data.</td>
<td>Panning or movement of cursor across a graphical display to view different segments of a display. Clicking on hyperlinks to navigate within large, complex textual documents.</td>
</tr>
</tbody>
</table>
2.2. Fraud detection examples

We use two cases to illustrate how data visualization techniques might be used in data mining to identify fraudulent transactions. The first case is adapted from Albrecht et al. (2012, 196) and involves an investigation of purchasing practices at a janitorial services company. The second case involves detection of money laundering activity at a large commercial bank (Chang et al., 2008). The cases represent two different levels of complexity. The purchasing investigation case is typical of a relatively simple transaction analysis task that might be addressed by downloading data into a commercially available interactive viewing program, then analyzing it. The money laundering case involves analyzing complex transaction patterns in a very large data set. Therefore, it requires the use of a custom-developed software package designed specifically for this problem.

The investigator in the purchasing investigation case (Albrecht et al., 2012, 196) has two major concerns: (1) an employee or employees may have developed inappropriate relationships with vendors and are purchasing supplies at above market price and (2) an employee or employees may be generating payments to a fictitious entity by using a name very similar to that of a real vendor. In the first step of this analysis, the investigator imports transaction data into an interactive data viewing program and generates a textual summary of purchases tabulated by vendor and purchaser (See Fig. 1—Panel A). The textual display immediately suggests that one vendor (José) may be generating payments to a fictitious entity by using a name very similar to that of a real vendor. By changing the representation of data from a table to a graph (See Fig. 1—Panel B), one can also see that two purchasers have rather large proportions of purchases from individual vendors (José from Master Cleaning Supply and Sally from Cleaners R Us). To further explore relationships between vendors and purchasers, the investigator may decide to change the graphical representation of a data from a bar graph to a scatterplot (See Fig. 1—Panel C). This allows one to see that purchases made by Jose from Master Cleaning Inc. and Master Cleaning Supply and by Sally from Cleaners R Us have a substantially higher total dollar amount per transaction than the other observations, which tend to cluster around the trend line for number of records processed and total dollar amount purchased. Returning to the bar graph representation, the investigator may further disaggregate the purchase data, displaying it by purchaser, vendor, and product to determine if purchases of a specific item are causing the observed anomalies. The resulting display suggests that José has made an unusually large amount of industrial push broom purchases from Master Cleaning Supply and Master Cleaning Inc. and Sally has made an unusually large amount of 30 count trash bag purchases from Cleaners R Us. The data visualization software also allows one to see the details of total purchases underlying any rectangle on the graph through a simple abstract/elaborate tool. By right-clicking on a graph rectangle, one can filter and view transactions related to the rectangle and sort these according to a variable of interest, as shown with Sally's purchases of 30 count trash bags from Cleaners R Us, sorted by unit price (Fig. 1—Panel E).

In summary, this relatively simple review of purchasing data for fraudulent transactions involves at least five of the interactive data visualization techniques described in Yi et al. (2007). The encode tool allows one to change textual to graphical representations and vice versa. It also allows one to shift among different graphical representations, for example, from a bar graph to a scatterplot. By moving the cursor over items of interest on a graph, the user may select the item for further examination and elaborate, or show more detail underlying the item. At the same time, one can filter the data to show a textual display of data related to an item or category of interest. While not shown in Fig. 1, the reconfiguration tool would also be helpful in this case, for example, if the investigator wanted to change the graph in Fig. 1—Panel B to show columns for each vendor with different colored blocks for each purchaser. The advantage of interactive data visualization software for identifying data patterns suggestive of fraud is that it allows one to change representations and explore data on a single platform. Otherwise, the investigator would need to use multiple platforms, i.e., extracting data using an audit software program, then exporting it to a spreadsheet program for graphical analysis.

The money laundering detection application described in Chang et al. (2008) illustrates additional visualization techniques (See Fig. 2—Panel A.). The data used in this application contain a variety of numeric and textual cues such as sender and receiver identities, transaction frequency and amount, and keywords used by sender and receiver. The investigator uses a custom-designed visualization page

---

1 One of the authors adapted the data downloaded from the book's web site at http://www.cengage.com for this example. Tableau Software (2011) Version 6.0 was used to develop the example analysis.
to analyze these data, searching for patterns that are suggestive of illicit activity. For example, the connect tool is used to identify transaction descriptions which contain keywords that should not be related in the context of a money transfer (See Fig. 2—Panel B.). The most frequent keywords appear in the middle of the view, while the less frequent ones appear on the outskirts of the circle. When a user highlights a specific keyword, lines are drawn from that keyword to all relating keywords, thus facilitating the identification of suspicious transactions. Finally, the explore tool allows the investigator to navigate through the application’s data displays. For example, the user might pan across the heatmap of keyword combinations in the upper left corner of Fig. 2—Panel A or use an exploration tool to change the dates displayed in the “strings and beads” time-series graph of keyword occurrences in the lower left corner of Fig. 2—Panel A.

3. Data visualization applications for fraud investigation and detection

3.1. Investigating suspected fraud

The purchasing transaction analysis described in Section 2.2 provides an example of how data visualization might be applied to investigate suspected fraud. A fraud investigation starts with a predication of fraud, or “circumstances, taken as a whole, that would lead a reasonable, prudent professional to believe a fraud has occurred (Albrecht et al., 2012, 80).” Once a fraud investigator is presented with a predication, he or she must develop a theory of how the fraud occurred and who may have committed the fraud (Wells, 2003; ACFE, 2010). In this case, the investigator has already hypothesized that fraud is being committed in a specific department (purchasing), through purchases made above market prices, payments to a fictitious vendor, or both. Data visualization enables the investigator to further refine the fraud theory by first finding analytical anomalies suggestive of fraud, then identifying specific suspicious transactions and the persons responsible for them.

3.2. Detecting fraudulent transactions

Practitioners recommend using proactive fraud detection methods as part of an effective fraud risk management program (Lanza, 2005b; Albrecht et al., 2012). One important fraud detection method is to identify areas where there is a high risk of fraudulent transactions (e.g., purchases of supplies, travel and entertainment

Fig. 1. Example of interactive data visualization analysis for fraud detection.
Fig. 1 (continued).
expense, or purchasing card use), then employ computer-aided data analysis methods to search for anomalies. These procedures might be performed periodically by internal auditors, or on a continuous basis by personnel whose primary responsibility is fraud detection. For example, the purchasing transaction analysis described in Section 2.2 might be performed on a periodic basis as a means of proactive fraud detection. Thus, the data analysis process for investigating suspected fraud and detecting fraudulent transactions is similar, in that the investigator evaluates a set of preliminary hypotheses regarding fraud (Albrecht et al., 2012). The primary difference between the two scenarios is that when investigating suspected fraud, the investigator starts with hypotheses based on a predication, whereas when performing proactive fraud detection procedures, the investigator develops hypotheses based on a risk analysis of where and when fraudulent transactions are likely to occur.

The bank transfer visualization application described in Section 2.2 is an example of a fraud detection method performed on a continuous basis to detect illegal transactions soon after they occur. There are, however, important differences between the problem of detecting money laundering or other anomalies in very large data populations, as opposed to detecting suspicious transactions in smaller data sets. First, the money laundering detection task requires specific visualization functions to effectively support the analysis of complex relationships. For example, Chang et al. (2008) describe functions that are specifically designed to track transaction description keywords that are indicative of money laundering activity. Second, due to data storage limitations, the visualization tool may require a specialized database structure that allows it to efficiently access only the data that is required for effective visualization and analysis (Chang et al., 2008). Therefore, it is often more effective to build a specialized visualization tool for analysis of large, complex data sets than to rely on a generalized data visualization package. Indeed, specialized transaction visualization...
tools have been developed for detecting insurance (Insurance Services Office, 2010), telecommunications (Cox et al., 1997), mortgage (Visual Analytics, 2004), wire transfer (Chang et al., 2008; Jeong et al., 2008), and securities Pryke (2010) fraud.

3.3. Substantive testing in financial statement audits

Auditing standards state that auditors should respond to identified fraud risks by increasing the extent of audit procedures to be performed (AICPA, 2013, AU-C 240; IAASB, 2012, ISA 240; PCAOB, 2013, AS 13). One technique developed in response to these recommendations is software-aided journal entry testing (Lanza, 2006; Lanza and Gilbert, 2007; CAQ, 2008). Debreceny and Gray’s (2014) analysis of data mining applications for fraud detection in financial statement audits not only identifies journal entries as a key source of information, but also suggests that auditors consider mining information such as XBRL disclosures, social media postings, and supporting documents for revenue transactions. Mining journal entries and other financial statement-related information produces complex textual and numeric output. Therefore, interactive data visualization has the potential to improve the effectiveness of data mining techniques by making it easier for auditors to understand and analyze the data that these techniques produce.

Mining journal entries involves using generalized audit software, database queries, or spreadsheet analysis to identify items which suggest attempts to fraudulently misstate financial statements (Lanza, 2006; Lanza and Gilbert, 2007; CAQ, 2008). These entries may have characteristics such as being (AICPA, 2013, AU-C 240.A49):

(a) made to unrelated, unusual, or seldom-used accounts, (b) made by individuals who typically do not make journal entries, or (c) recorded at the end of the period or as closing entries with little or no explanation.
Auditors may apply a scoring rule to transactions thus identified to further distinguish between false positives and journal entries that indicate a high likelihood of fraud. This approach, however, produces a unidimensional score and does not clearly display unusual relationships that might suggest fraud (e.g., an individual posting journal entries to an account or accounts that they normally do not handle). To better see unusual relationships among journal entry data, Lanza (2006) suggests graphing the number of transactions for each general ledger account (debits and credits separately). Similarly, one could graph amounts versus accounts or the number of entries approved by each individual. Graphs that display multiple attributes at the same time (as in Fig. 1—Panels B and D) would also be useful for this type of analysis. Since there is no one “best” sequence of graphical displays for analyzing journal entry data, the ease with which data visualization software allows one to transition between different graphical displays makes data visualization an especially useful tool for analyzing journal entry data.

4. A framework for fraud data visualization research

Even though data visualization techniques have the potential to be a powerful fraud detection tool (Deloitte, 2011), there is little research that directly examines the efficacy of these techniques in fraud detection contexts. Behavioral research in accounting focuses on the ability of decision makers to use textual and numeric cues for fraud risk analysis and detection, and this research focuses almost entirely on fraudulent financial reporting (e.g., Hogan et al., 2008). Research that directly examines the efficacy of interactive visualization tools typically consists of small-sample studies focused on describing the operation of a specific application (e.g., Chang et al., 2008; Jeong et al., 2008). Given the lack of judgment research on data visualization techniques for fraud detection, we develop a set of propositions for future research in this area.

Fig. 1 (continued).
We rely on Vessey’s (1991) theory of cognitive fit to develop our research propositions. Cognitive fit theory is a special case of cognitive cost–benefit theory, which specifies that decision makers will choose strategies that trade off the effort required to make a decision versus outcome accuracy (Beach and Mitchell, 1978; Payne, 1982). As originally developed, cognitive fit theory specifies that more efficient and effective decision
making results when the problem representation matches the task to be accomplished (Vessey, 1991; Vessey and Galletta, 1991). Specifically, graphical representations are more efficient and effective for spatial tasks (i.e., assessing relationships among data), while textual representations are more efficient and effective for symbolic tasks (i.e., extracting individual data values). Cognitive fit theory was subsequently extended to consider both the external problem representation and the decision maker’s internal representation of the problem domain, as well as the interaction between them (Shaft and Vessey, 2006; Vessey, 2006). Extended cognitive fit theory therefore suggests that the fit between external representation and task may be contingent on individual characteristics such as expertise in a given domain.

We adapt the frameworks developed by Dilla et al. (2010) in their review of interactive data visualization techniques in accounting and by Yigitbasioglu and Velcu (2012) in their review of dashboards in performance management to organize our discussion. As shown in Fig. 3, the framework used in this paper considers relationships between (1) task and fraud investigator characteristics, (2) data visualization characteristics, and (3) decision outcomes.

Before proceeding, it is important to define what is meant by efficiency and accuracy in the context of detecting fraudulent transactions with interactive data visualization tools. Vessey (2006) suggests that total time spent on a judgment task may not be a reliable measure of efficiency when comparing graphical versus tabular representations of data, since there are conflicting results regarding whether problem solving occurs more quickly with graphs or tables (Lohse, 1993; Dennis and Carte, 1998). Vessey’s (2006) conclusion, however, is based on research with static information representations, where participants typically examine one, or a limited set of data views. With respect to interactive data visualization tools, it appears reasonable to consider total time spent on the task as a measure of efficiency. Since interactive visualization tools allow easy transitions between different representations, arrangements, and levels of detail of data, they should result in faster decisions than conventional visualization tools, assuming that the user examines the same number of data views with both technologies. If, however, a decision maker examines a greater number of data views using interactive visualization, the decision process may become less efficient in terms of total time spent. As Fig. 3 suggests, task, fraud investigator, and data visualization characteristics may influence the efficiency of a user’s search processes. Therefore, we use overall time spent to reach a conclusion about the presence or absence of fraudulent transactions in a given data set as a measure of efficiency in the research propositions stated below.

Accuracy is making a correct conclusion about whether significant fraudulent transactions are present or absent in a given data set. An investigator following up on a suspicion of fraud engages in an iterative process, in which he or she develops one or more fraud theories (Wells, 2003; ACFE, 2010). If preliminary data validate a fraud theory, then the investigation continues with additional evidence gathering and eventually, with interviews of witnesses and perhaps with the perpetrator himself. Otherwise, if none of the fraud theories appears viable, the investigation terminates. Thus, there are two possible types of errors in a fraud investigation: (1) the investigator stops gathering evidence, even though fraudulent transactions are present or (2) the investigator continues gathering evidence when fraudulent transactions are not present. The costs of these two errors are typically asymmetric, as failure to find significant fraud is usually more costly than conducting unnecessary investigative procedures.

The decision-making process with respect to fraud in a financial audit is somewhat different. The financial auditor is not working with a predication of fraud; rather, he or she is required to reach a conclusion about whether the financial statements contain material misstatements due to fraud (AICPA, 2013, AU-C 240.A56). Thus, the two possible types of errors with respect to fraud detection in a financial audit are: (1) the auditor does not recognize that fraudulent transactions may be present, does not make appropriate modifications in audit procedures, and fails to detect material fraud and (2) the auditor incorrectly concludes that fraudulent transactions are present, and performs unnecessary extended audit procedures. Similar to fraud investigation, the costs of these two types of errors are usually asymmetric.

Tradeoffs between the two types of fraud detection errors are complex and will vary across contexts. For example, a company’s management might decide that the potential financial penalties and damage to a company’s reputation from United States Foreign Corrupt Practices Act violations are so great, that they are willing to expend resources to investigate even weak evidence of such activity. On the other hand, while financial auditors in a competitive environment are cognizant of the potential litigation costs associated with the failure to find material fraudulent misstatements, they also must avoid the costs associated with
over-auditing. Since detailed analysis of cost tradeoffs between the two types of fraud detection errors is beyond the scope of this paper, the development of research propositions will focus on a single accuracy construct. In addition, since “correct” decisions in the fraud investigation context are binary, the term “accuracy” in the research propositions refers to the proportion of cases in which the investigator correctly concludes that fraudulent transactions are or are not present in a given data set.

We organize the following development of propositions and research questions into three subsections. The first addresses the linkages between task characteristics and visualization interactivity, and between interactivity and decision outcomes. The second addresses the linkages between fraud investigator

---

**Panel A**

<table>
<thead>
<tr>
<th>Task and Investigator Characteristics</th>
<th>Data Visualization Characteristics</th>
<th>Decision Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Task Characteristics</td>
<td>Interactivity</td>
<td>Efficiency</td>
</tr>
<tr>
<td>- spatial</td>
<td>- transitioning between representations</td>
<td>- time spent on decision</td>
</tr>
<tr>
<td>- symbolic</td>
<td>- information navigation/selection</td>
<td>- number of information acquisitions</td>
</tr>
<tr>
<td>Task Complexity</td>
<td>- transitioning among display configurations</td>
<td></td>
</tr>
<tr>
<td>- component</td>
<td>Structure</td>
<td>Accuracy</td>
</tr>
<tr>
<td>- coordinative</td>
<td>- pre-selected representations</td>
<td>- decision to continue or stop investigation is correct</td>
</tr>
<tr>
<td>- dynamic</td>
<td>- limits on information navigation/selection</td>
<td></td>
</tr>
<tr>
<td>Investigator Characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- fraud detection expertise</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i.e., knowledge and experience)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- cognitive style and abilities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- personality</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Panel B**

<table>
<thead>
<tr>
<th>Propositions</th>
<th>Brief Description</th>
<th>Panel A Diagram Paths</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a, 1b</td>
<td>Less time required / more accurate with interactive, as opposed to static data visualization</td>
<td>1, 6</td>
</tr>
<tr>
<td>2a, 2b</td>
<td>Reduction in time / increase in accuracy with interactive data visualization greater in complex than in simple fraud detection scenarios</td>
<td>2, 6</td>
</tr>
<tr>
<td>3a, 3b</td>
<td>Reduction in time / increase in accuracy with interactive data visualization in simple fraud detection scenarios greater for non-expert than for expert investigators.</td>
<td>2, 4, 6</td>
</tr>
<tr>
<td>4a, 4b</td>
<td>Reduction in time / increase in accuracy with interactive data visualization in complex fraud detection scenarios greater for expert than for non-expert investigators.</td>
<td>2, 4, 6</td>
</tr>
<tr>
<td>5a, 5b</td>
<td>Non-expert investigators require less time / are more accurate with structured than with unstructured data visualization tools.</td>
<td>5, 7</td>
</tr>
<tr>
<td>6a, 6b</td>
<td>Expert investigators require less time / are more accurate with structured than with unstructured data visualization tools in complex fraud detection scenarios.</td>
<td>3, 5, 7</td>
</tr>
</tbody>
</table>

**Fig. 3.** Relationships between task, investigator, and data visualization tool characteristics with implications for decision outcomes.
characteristics and visualization interactivity, and between interactivity and decision outcomes. Finally, the third subsection discusses how structured data visualization tools might affect decision outcomes, depending on task and investigator characteristics.

4.1. Task characteristics

4.1.1. General characteristics of the fraud detection task

The process of discovering anomalies suggestive of fraud involves detecting unusual patterns in data (a spatial task supported by graphical representations), then drilling down into the data and selecting individual items for further examination (a symbolic task supported by tabular representations). As path 1 in Fig. 3 indicates, interactive data visualization tools are a good fit for fraud investigation tasks, given that they allow the investigator to easily transition between graphical and tabular representations of data. As path 6 in Fig. 3 indicates, these transitions are likely to occur more quickly than with conventional visualization tools, resulting in less time being spent on the fraudulent transaction detection task. Further, interactive visualization allows the investigator to more easily choose representations that have cognitive fit with the different steps of this task, resulting in greater accuracy in detecting fraudulent transactions in a given data set.

In addition, the search for data patterns suggestive of fraud involves viewing multiple spatial configurations of data as the investigator disaggregates or explores different data relationships. Researchers have not directly investigated the efficacy of functional features such as graphical display interactivity for viewing and exploring different data relationships (Yigitbasioglu and Velcu, 2012). However, the process of exploring data for fraudulent transactions should require less time with interactive than with static data visualization, given that interactive visualization supports quick transitions among different data configurations. Also, by allowing the investigator to choose data configurations that highlight anomalies within data, interactive visualization should support greater accuracy in detecting fraudulent transactions.
This discussion therefore suggests the following propositions:

**Proposition 1a.** Investigators will require less time to reach a conclusion about the presence or absence of fraudulent transactions in a given data set when they use interactive, as opposed to static data visualization tools.

**Proposition 1b.** Investigators will make accurate conclusions about the presence or absence of fraudulent transactions in a given data set more often when they use interactive, as opposed to static data visualization tools.

### 4.1.2. Task complexity

Judgment researchers typically identify three major components of task complexity: (1) component complexity, or the number of information cues that must be processed, (2) coordinative complexity, or the number of distinct processes that must be executed, and (3) dynamic complexity, or the interdependence and changing relationships over time between cues and processes (Wood, 1986; Bonner, 1994; Speier, 2006). Extensions of cognitive fit theory suggest that as task complexity increases, decision makers will rely more on holistic (i.e., considering relationships among data) than on analytical (i.e., extracting data values and performing computations) strategies. This in turn indicates that graphical representations will be more effective than tables for complex judgment tasks, regardless of whether they involve spatial or symbolic processes (Speier and Morris, 2003; Wheeler and Jones, 2003; Speier, 2006; Vessey, 2006).

The purchases fraud detection task described in Section 2.2 has relatively low component complexity—while the investigator must consider various disaggregations of data (i.e., by vendor, purchaser, and product type), there are only two measures to be considered (i.e., the number and dollar volume of transactions processed). The task has moderate coordinative complexity, since the investigator must execute a series of data viewing operations before reaching a conclusion. It has low dynamic complexity, as it is unlikely that the relationships between information cues will change quickly over time. Thus, the task overall is of low complexity.

In contrast, the money laundering detection task described in Section 2.2 (Chang et al., 2008) is substantially more complex. It has high component complexity, as the investigator must consider a set of time-series data containing a variety of numeric and textual cues such as sender and receiver identities, transaction frequency and amount, and keywords used by sender and receiver that are suggestive of illicit activity. The task has high coordinative complexity, as a large number of viewing operations may be necessary to identify transactions for further investigation. These include using heatmaps and query tools to identify transactions with keywords suggestive of money laundering (upper windows in Fig. 2—Panel A), a strings and beads tool to display transaction trends across time (lower left-hand window in Fig. 2—Panel A), and a connect tool to display suspicious relationships between senders and receivers (lower right-hand window in Fig. 2—Panel A and Fig. 2—Panel B). Finally, the task has high dynamic complexity, as the investigator is attempting to find transactions suggestive of money laundering activity in real time and relationships among data change as sophisticated perpetrators vary their transaction patterns in order to avoid detection. Thus, the task is of high complexity overall.

In a low complexity fraud detection task such as the purchases example, the investigator may review unusual patterns in data (a spatial task) and drill down into the data to select individual items for further examination (a symbolic task). In more complex detection tasks, he or she may rely more on pattern detection through the use of graphical representations than on analysis of individual transactions, in order to minimize cognitive effort (Vessey, 2006). Thus, path 2 in Fig. 3 indicates that the efficacy of interactive data visualization for fraud detection may be contingent on task complexity. Specifically,
data visualization tools that facilitate the display of data in a variety of appropriate graphical formats should improve decision speed and accuracy in complex fraud decision tasks. Further, interactive data visualization tools may be more effective in complex than in simple fraud detection tasks. This leads to the following propositions:

**Proposition 2a.** The reduction in time required to reach a conclusion about the presence or absence of fraudulent transactions in a given data set achieved by interactive, as opposed to static visualization tools will be greater in complex than in simple fraud detection scenarios.

**Proposition 2b.** The increase in the proportion of accurate conclusions regarding the presence or absence of fraudulent transactions in a given data set achieved by interactive, as opposed to static visualization tools will be greater in complex than in simple fraud detection scenarios.

4.2. Decision maker characteristics

4.2.1. Expertise

Extended cognitive fit theory (Shaft and Vessey, 2006; Vessey, 2006) maintains that the decision maker’s internal representation of the problem domain and the external problem representation, as well as the interaction between them, contribute to the mental representation developed to solve the problem. The internal representations of individuals who have greater expertise in fraud detection (i.e., have greater domain-specific knowledge and experience in this area) are likely to differ from those of individuals with less expertise. Consequently, the effects of interactive data visualization tools on fraud detection information search strategies and judgments will likely be contingent not only on task difficulty, as discussed above, but also on investigators’ experience and knowledge levels. This suggests that the effects of paths 2 and 4 in Fig. 3 need to be considered jointly. Therefore, we do not propose main effects for expertise; rather, our propositions reflect a potential interaction between expertise and task complexity.

Research suggests that graphs are more effective in improving the performance of less knowledgeable or experienced decision makers in accounting-related tasks. Anderson and Mueller (2005) show that improvements in ability to assess relationships between financial variables in an analytical procedures task are more pronounced for students using graphical displays than for experienced auditors. Cardinaels (2008) finds that less knowledgeable participants performing a judgment task using managerial accounting data are more accurate when using graphs, while more knowledgeable participants are more accurate using tables. More knowledgeable participants spend more time in information search with tables, but information display format does not affect less knowledgeable participants’ information search time.

Dilla et al. (2013) find evidence that the influence of graphs on investor judgments depends on both expertise and task difficulty. Graphical displays of pro forma earnings measures influence the current year earnings evaluation judgments of nonprofessional investors, but not those of professional investors. For the more complex tasks of evaluating earnings potential and making an investment decision, graphical displays influence the judgments of both types of investors. These results indicate that nonprofessional investors rely on graphical displays to support decision strategies that involve reviewing data patterns, regardless of task complexity. Professional investors tend to focus on strategies that involve evaluating individual data items when performing relatively simple tasks and rely on graphical displays to support reviewing data patterns only for more complex tasks.

Thus, the efficacy of interactive data visualization tools in detecting fraudulent transactions is likely to be contingent on both task difficulty and investigators’ fraud detection expertise. In a less complex fraud detection task, expert investigators may focus on analytically extracting and calculating differences among individual values. The opposite may be true for non-expert investigators, who may find even a simple fraud detection task difficult and tend to reduce cognitive effort by holistically evaluating data patterns. Therefore, to the degree that they support the holistic review of data patterns, interactive data visualization tools are more likely to improve information search efficiency and judgment accuracy.
in simple fraud detection tasks for non-expert, as opposed to expert fraud investigators. This leads to the following propositions:

**Proposition 3a.** In simple fraud detection scenarios, the reduction in time required to reach a conclusion about the presence or absence of fraudulent transactions in a given data set achieved by interactive, compared to static visualization tools will be greater for non-expert than for expert investigators.

**Proposition 3b.** In simple fraud detection scenarios, the increase in the proportion of accurate conclusions regarding the presence or absence of fraudulent transactions in a given data set achieved by interactive, compared to static visualization tools will be greater for non-expert than for expert investigators.

The opposite may occur for more complex fraud detection tasks. Expert investigators may find it necessary to rely on the holistic visualization of data trends as task complexity increases. Interactive data visualization tools will support this strategy to a greater extent than static representations. On the other hand, a ceiling effect may occur for non-expert investigators faced with a complex fraud detection task. Decision support in the form of interactive data visualization may therefore not improve search efficiency or accuracy for such individuals. This leads to the following propositions:

**Proposition 4a.** In complex fraud detection scenarios, the reduction in time required to reach a conclusion about the presence or absence of fraudulent transactions in a given data set achieved by interactive, compared to static visualization tools will be greater for expert than for non-expert investigators.

**Proposition 4b.** In complex fraud detection scenarios, the increase in the proportion of accurate conclusions regarding the presence or absence of fraudulent transactions in a given data set achieved by interactive, compared to static visualization tools will be greater for expert than for non-expert investigators.

### 4.2.2. Cognitive styles and abilities

Unlike the literature on the interactive effects of expertise and information representations in accounting settings, research on cognitive styles and abilities and their relationship to information representation efficacy is less well-developed. Consequently, this research literature does not suggest testable propositions with respect to fraud data anomaly detection, only a research question.

Two streams of research have examined the interaction between cognitive styles, interactive information display characteristics, and decision-making performance. The first stream uses Myers–Briggs Type Indicator (MBTI) personality traits as an indicator of cognitive style. Wheeler (2001) argues that the MBTI provides a useful framework for accounting research, however, only a few studies have examined the relationship between MBTI traits, information display characteristics, and decision-making in accounting contexts. Results from these studies are mixed. Chenhall and Morris (1991) find that MBTI traits influence whether decision makers consider opportunity costs in a resource allocation task, and Kelliher and Mahoney (2007) find that MBTI traits influence the variability and efficiency of information search processes in an attribute by alternative search task. In contrast to these studies, Bryant et al. (2009) find that MBTI traits do not influence identification of either within-cycle or out-of-cycle internal control cues.

On average, accountants tend to have dominant sensing perception types and judgment personality styles (Vaassen et al., 1993; Wheeler, 2001). Given that these traits are associated with individuals working in management and administration who possess general personality characteristics such as practical, sensible, decisive, logical, detached, systematic, and objectively critical (Myers et al., 1998, as quoted in Wheeler, 2001), it is likely that fraud detection specialists possess similar MBTI traits. Even so, it is premature to suggest designing interactive data visualization tools to fit the supposed personality characteristics of individuals who might be engaged in fraud detection tasks, because of the lack of strong evidence for an association between MBTI personality characteristics and interactive information display characteristics (Yigitbasioglu and Velcu, 2012).

The other stream of research that examines the interactive effects of cognitive styles and information display characteristics focuses on field dependency, or “individual differences in ease or difficulty in separating an item from an organized field or overcoming an embedding context” (Witkin and Goodenough, 1981).
Field–independent individuals are better able to differentiate parts of a field as discrete from an organized background. A field–independent cognitive style is often assumed to imply higher analytical reasoning, whereas a field–dependent style implies lower analytical or heuristic reasoning (Benbasat and Dexter, 1979). Mahoney et al. (2003) find that when task type (i.e., spatial or symbolic) and representation (i.e., graphs or tables) do not match, field–independent accountants are more accurate, but take more time to make judgments than field–dependent accountants. Field dependency does not affect accuracy or decision time when individuals are given decisional guidance to select a representation that matches task type. Turetken and Schuff (2007) find that context-aware data flow diagrams (DFDs) with fisheye views improve information search performance for field–dependent individuals, relative to conventional, context–free DFDs. DFD format does not influence information search performance for field–independent individuals.

Unfortunately, additional research which examines the interactive effects of cognitive styles and information display characteristics is limited, due in large part to Huber’s (1983, 571) argument that “we do not know if DSS designs should (1) conform to the user’s cognitive style or (2) complement the user’s cognitive style” (Davern et al., 2012). At the time Huber (1983) made this statement, however, DSSs that allow the user to select and modify features to fit task characteristics and personal preferences were only beginning to emerge. Current interactive data visualization tools now incorporate these features. Thus, users may choose interactive display features, which may or may not be consistent with their own cognitive style (Mahoney et al., 2003).

A more recent stream of research in human–computer interaction (Conati and Maclaren, 2008; Toker et al., 2012) examines the influence of various cognitive abilities such as perceptual speed, visual working memory, and verbal working memory on individuals’ decision efficacy using bar graphs versus radar graphs (See Fig. 4). Conati and Maclaren (2008) find that perceptual speed and graph format have an interactive effect on accuracy in a complex multivariate spatial task: individuals with high perceptual speed are more accurate with bar graphs, while those with low perceptual speed are more accurate with radar graphs. Further, Toker et al. (2012) find that while task completion time is always greater with radar graphs, the difference in time performance between bar and radar graphs is greater for low perceptual speed individuals. Toker et al. (2012) also report that users with higher visual working memory have higher preference ratings for radar graphs, while users with lower verbal working memory rate bar graphs as easier to use.

One possible reason for the renewed interest in research on the interactive effects of individual cognitive differences and interactive information display characteristics on decision–making performance is that technologies are being developed which can sense user characteristics such as perceptual speed, and adapt by providing the user with visualizations that are best suited to his or her individual characteristics (Steichen et al., 2013). Research on the interactive effects of individual cognitive differences and interactive information display characteristics is in its early stages, and results so far are inconclusive. Thus, it remains to be seen whether efforts to design data visualization technologies which adapt to individual user characteristics in business–related contexts such as fraud detection are justified.

The lines of research that examine the interaction between individual cognitive styles and abilities and the efficacy of information visualization tools are not well–developed enough to make specific predictions regarding the effects of a given combination of decision maker characteristics and information representation. Even so, it would still be interesting to examine whether cognitive styles and abilities influence: (1) an investigator’s choice of interactive visualization tools and (2) the efficacy of such tools for the specific task of identifying anomalies suggestive of fraudulent activity. Therefore, we propose the following research question.

---

4 Perceptual speed is the speed with which one carries out simple visual perception tasks, such as comparing figures or symbols. Visual and verbal working memory are the ability to remember the configurations, location, and orientation of figural and textual materials, respectively.

5 For example, Conati and Maclaren (2008) assessed several other cognitive abilities in addition to processing speed (e.g., visual memory, special visualization ability, need for cognition, and learning styles), but did not find that any of these influenced the efficacy of bar versus radar graphs.
Research Question: The choice and efficacy of a given interactive data visualization tool for fraud data anomaly detection will depend on individual decision maker characteristics such as personality traits, field dependency, and cognitive abilities.

4.3. Data visualization tool characteristics

This analysis so far has developed propositions that predict under what circumstances decision outcomes may differ with and without interactive data visualization tools. It is also likely that imposing structure on data visualization processes in fraud detection tasks, such as limiting the set of available interactive data visualization tools (Chang et al., 2008, also see Fig. 2), or restricting the order in which the tools might be used may affect decision outcomes under certain circumstances. This is depicted by path 7 in Fig. 3.

Further, the efficacy of structured data visualization tools may be contingent on investigator expertise, as indicated by path 5 in Fig. 3. For example, there is evidence that more experienced financial analysts follow a more directive (as opposed to sequential) search strategy, and that directive search strategies are correlated
with earnings forecast accuracy (Hunton and McEwen, 1997). In addition, research indicates that more experienced accountants are able to choose an appropriate internal problem representation for a cash flow analysis task, regardless of data format (Vera-Munoz et al., 2001). These results suggest that expert investigators who use an interactive data visualization tool for fraud investigation will follow a directive search strategy and will be able to choose appropriate information representations for each investigation subtask. On the other hand, non-expert investigators may not be able to follow a directive search strategy or choose appropriate information representations. This suggests that non-expert investigators may benefit from using structured interactive data visualization tools, leading to the following propositions:

**Proposition 5a.** Non-expert investigators will require less time to reach a conclusion about the presence or absence of fraudulent transactions in a given data set with structured than with unstructured interactive data visualization tools.

**Proposition 5b.** Non-expert investigators will make more accurate conclusions about the presence or absence of fraudulent transactions in a given data set more often when they use structured, as opposed to unstructured interactive data visualization tools.

In contrast, the degree to which an expert investigator might need to rely on structured interactive data visualizations to support an efficient and effective search strategy is likely to be contingent on task complexity, as depicted by path 3 in Fig. 3. Expert investigators are likely to employ a directed search strategy when performing low-complexity analytical fraud detection tasks, hence, any attempts to structure the data visualization processes they follow are likely to be of limited benefit. On the other hand, as fraud detection task complexity increases, even expert investigators may not have a well-developed strategy for selecting from and evaluating a larger set of information cues. Since it may be more difficult for expert investigators to follow a directive strategy in such cases, they are likely to benefit from imposing structure on data visualization processes. Thus, we predict that structured data visualization techniques will influence expert fraud investigators’ decision efficiency and accuracy only for complex fraud detection tasks.

![Fig. 6. Relative frequency of fraud keyword hits by employee (Torpey et al., 2009b).](image)
Proposition 6a. In complex fraud detection scenarios, expert investigators will require less time to reach a conclusion about the presence or absence of fraudulent transactions in a given data set with structured than with unstructured interactive data visualization tools.

Proposition 6b. In complex fraud detection scenarios, expert investigators will make more accurate conclusions about the presence or absence of fraudulent transactions in a given data set more often when they use structured, as opposed to unstructured interactive data visualization tools.

5. Summary and conclusions

Successful fraud detection depends on the investigator's ability to detect patterns in data that are suggestive of fraudulent transactions. Interactive data visualization tools have substantial potential for making the fraudulent transaction detection process more efficient and effective. The analysis presented in this paper indicates that these tools should support the different cognitive processes required at various stages of the detection process, as they allow the user to navigate large data sets, change the representation of data, and filter subsets of data for further examination. We use cognitive fit theory (Vessey, 1991; Vessey, 2006) to describe a theoretical framework and develop a set of propositions and research questions for future research regarding the circumstances under which interactive data visualization might increase investigators' efficiency and effectiveness in detecting fraudulent transactions. Conducting future research to investigate these circumstances is important, as it should provide practitioners with evidence regarding when it might be appropriate to use interactive data visualization to detect fraudulent transactions.

While this analysis has focused on detecting fraudulent transactions, interactive visualization may also be useful at other stages of the fraud investigation process. Once a suspected perpetrator or perpetrators have been identified, a social network analysis graph (Heer and Boyd, 2005) might be used to identify collusion among the suspected perpetrators and other individuals (See Fig. 5). Further, data mining of suspected perpetrators' e-mails might be used to detect keywords indicative of the fraud triangle components of pressure, opportunity, and rationalization (Torpey et al., 2009a,b; Debreceny and Gray, 2011). Fig. 6 displays a multi-dimensional graph of these measures. It clearly shows that three individuals' e-mail communications have high proportions of keywords related to all three fraud triangle components. While this graph is a static representation, adding an interactive element that allows investigators to drill down from the summary graph to view additional details about these individuals' communications would likely increase the efficiency of this investigative process.

Research into the efficacy of interactive data visualizations for fraud detection is extremely limited to date. This is in large part due to the limited number of available professionals for judgment-based fraud detection studies. There are, however, alternatives to using expert investigators for research in this area, if the objective of the research is to test basic propositions about the effects of domain knowledge on the use of interactive data visualization tools in low- and moderate-complexity fraud detection tasks. A number of the propositions and research questions stated in this paper might be tested using participants with appropriate levels of accounting knowledge, even if their specific expertise with detecting and investigating fraud is limited. Indeed, a number of studies which examine cognitive fit theory and related information visualization phenomena employ students with basic knowledge in a given domain as surrogates for professionals (e.g., Speier et al., 2003; Speier, 2006; Fang and Holsapple, 2007).

Further, researchers in accounting and information systems have tended to discount the value of small sample studies. This is in contrast to earlier accounting research (e.g., Biggs and Mock, 1983; Biggs et al., 1988; Johnson et al., 1993), and to the approach often taken in the human–computer interaction literature (e.g., Chang et al., 2008; Jeong et al., 2008), which has studied in-depth the decision processes of small samples of experts. These studies suggest that there is much to be learned by detailed examination of fraud detection experts' interactions with data visualization tools, without recruiting large numbers of subjects.

See Kang et al. (2010) for an illustration of how data visualization software might have been used to detect patterns of collusion among Enron executives and Debreceny and Gray (2011) for a detailed analysis of how social network analysis diagrams might have been used to uncover revenue recognition fraud in Enron's electrical business segment.
References


Cenfriuge Systems, Inc., 2015


Debreceny RS, Gray GL. A taxonomy to guide research on the application of data mining to fraud detection in financial statement audits. Working paper. Northridge: University of California and Hawaii State University; 2014.


Wheelbar PR, Jones DR. The effects of exclusive user choice of decision aid features on decision making. J Inf Syst 2003;17(1):63–84.


