Perceptions of Auditor Negligence: The Effects of Big Data Visualisations on Jurors' Decisions

Travis L. Christensen

Supervised by Professor Ian Eggleton
and Professor Jake Rose

School of Accounting and Commercial Law
Victoria University of Wellington

2020

Contents

Declaration of Originality

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material that have been accepted for the award of any other degree or diploma at this or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at this university or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception in style, presentation and linguistic expression is acknowledged.

Acknowledgements

I would not have been able to finish such a momentous project without a village for support. First, I would like to thank my supervisors, Professors Ian Eggleton and Jake Rose, for their support. Second, I will be forever indebted to my parents, who were always willing to read my work and give me suggestions on how I could improve the quality of the work. I must further acknowledge Qiao Rui, who walked me home and encouraged me after many late nights working at Rutherford House.

Overview

This study analyses the effects of Big Data visualisations on jurors' decisions in audit litigation cases. Specifically, the study investigates the effects of different types of Big Data visualisations (word clouds or bar graphs) and different sources of Big Data (emails or social media posts) on jurors' perceptions of auditors' work and the size of the negligence awards that jurors recommend. The study theorises that the emotions elicited and the reliability of the data used to create visualisations such as word clouds will have dramatic effects on jury verdicts in audit negligence trials. There is considerable literature to support this assertion. However, after data collection, it was discovered that jurors are not influenced by the emotions elicited by visualisations. Rather, participants were very sceptical of more novel types of visualisations, such as word clouds, but could be persuaded by the inherent emotions elicited and the reliability of the data if they found the visualisation useful.

List of Abbreviations

AMT	Amazon Mechanical Turk
ANCOVA	analysis of covariance
ANOVA	analysis of variance
MANCOVA	multivariate analysis of covariance
MANOVA	multivariate analysis of variance
US	United States
LLC	Limited Liability Company

List of Figures

List of Tables

Chapter 1
Introduction

Chapter Preview

This chapter gives a general overview of the issue of Big Data visualisations included in audit papers and their possible effects on United States (US) juries during audit negligence trials. It explains why the issue is important and that there is a gap in the literature on the subject.

Audit litigation is of grave concern for public auditors. Donelson (2013) estimates that from 1999 to 2007 (excluding insurance), the six largest public accounting firms incurred 10 billion US dollars in litigation costs. Within that period, litigation costs steadily increased, and accounting firms claimed that they were subject to unreasonable litigation risk (Donelson 2013). As audit processes advance to incorporate Big Data, much of which is from new and emerging sources such as social media, litigation risks and costs may continue to rise. Prior to this study, jurors' responses to evidence presented in the form of Big Data visualisations were unknown.

Jurors generally have very little accounting experience and are often incapable of comprehending the complexities of an auditor negligence lawsuit (Palmrose 2006). Further, juries are willing to award large settlements to defendants with little consideration for the quality of the audit (Reffett, Brewster and Ballou 2012). For auditors to protect themselves from litigation, it is important that they understand the factors that influence juror decisions. Several recent studies have examined juror decisions in audit litigation settings (see Appendix A for a summary of the auditor negligence literature). However, a critical issue that has not been studied and could potentially leave an auditor vulnerable to litigation is the use of Big Data visualisations by audit firms. Big Data visualisations help auditors to recognise important patterns in data and are expected to rapidly grow in usage (Rose et al. 2017). However, the use of Big Data visualisations as part of the audit process may create significant threats to audit firms because of their potential effects on juror decisions. In fact, firms are beginning to become aware of the potential threat, as there appears to be reluctance among partners of public company audit firms to fully integrate analytics of emerging data sources into the auditing process or to include visualisations of Big Data in audit workpapers (Gepp et al. 2018). This hesitance derives from concerns that include potential litigation risk (Franzel et al. 2018).

Earlier studies that predate the current widespread availability of Big Data from digital sources have shown that data presentation formats affect human judgment, decision accuracy and information processing (Stock and Watson 1984). Yet, an auditor's vulnerability to litigation risks resulting from Big Data visualisations has not been studied, and there are several reasons to believe that visualisations included in audit work papers have significant capacity to alter juror decisions.

This is an important issue because the amount of data that firms collect is growing exponentially, and visualisations of these data will permeate auditors' work. In 2000, 25 per cent of all stored information was digital—this increased to more than 98 per cent by 2013 (Cukier and Mayer-Schonberger 2013). In the past two years, organisations collected more data than was collected during the previous 2000 years (Syed, Gillela and Venugopal 2013). Ninety per cent of these data are unstructured and consist of data sources such as social media posts, email messages, phone calls, website traffic and video streams (Syed et al. 2013). While collecting these data is now easy and inexpensive, processing it to extract meaningful information is difficult due to the sheer volume involved. The processing challenge is especially relevant to financial statement audits (Alles 2015). These data are often beyond the capabilities of currently used query tools that examine both financial and non-financial data sources (Brown-Liburd and Vasarhelyi 2015). Therefore, auditors must rely on significant professional judgment when dealing with Big Data (Brown-Liburd, Issa and Lombardi 2015). To help auditors identify patterns in, analyse and interpret Big Data, audit firms are using visualisations that allow large datasets to be quickly viewed and interpreted.

A critical issue that has not been considered is the influence that these Big Data visualisations could have on jurors' judgments in audit negligence lawsuits. Visualisations often evoke emotional responses that could potentially overwhelm jurors' decision-making processes. The existing literature shows that emotions are significantly linked to verdicts of auditor negligence. Kadous (2000, 2001), Reffett (2010) and Braseletmal (2016) all found that jurors who experience negative emotional reactions towards auditor defendants are more likely to find them negligent. Some visualisations, like word clouds (a visual depiction of text in which the most frequently used terms are displayed the most prominently), have the potential to create strong negative emotional responses because of their depiction of emotionally laden words; these strong emotional responses could ultimately turn juror decisions against auditors.

The literature shows that auditors, who are highly skilled professionals, often struggle to find patterns in non-financial and financial data (O'Donnell and Perkins 2011), which suggests that jurors, who may have no knowledge of accounting, may have even greater difficulties understanding complex visualisations. Brown-Liburd et al. (2015) argue that the large volume of outputs from Big Data analyses could overwhelm cognitive processes, making it difficult for auditors to recognise patterns. They further argue that using Big Data visualisations could cause auditors to pursue inefficient or irrelevant investigations.

Many Big Data visualisations, such as word clouds, may evoke strong emotional responses. If highly trained auditors struggle to understand Big Data visualisations, unskilled decision-makers, such as those who would serve on a US jury, might struggle even more to comprehend Big Data visualisations and may, therefore, be significantly influenced by the emotions they evoke. Therefore, the emotional effect that Big Data visualisations have on juries could leave auditors vulnerable to negligence verdicts. Further, because auditors often miss patterns when using large datasets, audit firms may feel compelled to use visualisations that evoke strong emotional responses to help auditors recognise patterns more clearly (Rose et al. 2019).

A second critical issue that has not been considered is how the data source used to create visualisations may influence jurors evaluating auditor negligence. Where jurors evaluate many different types of evidence, the source of the evidence may play an important role in determining their findings regarding the auditor's level of negligence. The literature shows that the persuasiveness of information depends on the credibility of its source (Birnbaum and Stegner 1979; McGinnies and Ward 1980; Eagly and Chaiken 1993; Chaiken and Maheswaran 1994). For example, Kaufman, Stasson and Hart (1999) found that articles from the *Washington Post* are rated as more factual, believable and accurate than articles from the *National Enquirer*. Therefore, the source of data used to construct a Big Data visualisation could play an important role in persuading a jury that an auditor is guilty—or not—of negligence. The purpose of this study is to empirically evaluate the effects of emotional responses to visualisations and the source of the data used to create the visualisations on jurors' decisions in auditor negligence lawsuits.

Research Question

What is the impact on jury verdicts in audit negligence cases of (i) the emotions evoked by a visualisation from Big Data and (ii) the perceived reliability of the underlying data?

Chapter Review

This chapter introduced the problem that including Big Data visualisations in audit workpapers may affect US jury decisions. It explained two possible reasons, which will be studied in this experiment: the emotional response that some visualisations evoke and the credibility of the data source.

Chapter 2
Literature Review and Hypotheses Development

Chapter Preview

The following is a thorough review of the Big Data, audit, psychology and legal literature. Two theoretical constructs: neural affective decision theory and source credibility theory are used to form four hypotheses. Finally, an interaction hypothesis combining the two theories is presented.

Big Data in the Audit

Although there are many different definitions of Big Data, the term generally refers to datasets of enormous size (Moffitt and Vasarhelyi 2013). Big Data consists of very large, unstructured datasets that are beyond the processing capabilities of traditional query tools and that include data from both financial and non-financial sources (Brown-Liburd, Issa and Lombardi 2015). However, Big Data are also defined by the velocity, variety and veracity of the data. Velocity means real-time over-batch processing and variety refers to different types of data, such as audio files, social media posts, global positioning measures and video streams. Veracity involves the reliability of data, which requires reducing noise and gaining truthful information from the data (Yoon, Hoogduin and Zhang, 2015). This study operationalizes big data by focusing not on the volume of data but rather the veracity of the data.

Big Data are potentially problematic for auditors because they require auditors to document their rationale for each judgment made during the audit process. This rationale must show that they were professionally sceptical and applied critical thinking to each judgment (Public Company Accounting Oversight Board Auditing Standard 3, 2004). Further, auditors are also required to document in the work papers evidence discovered during the audit that disconfirms management assertions. New requirements by the Public Company Accounting Oversight Board (PCAOB Auditing Standard 3) created in 2004 require auditors to examine disconfirming evidence for all management estimates. PCAOB AS 1215.8 states: 'In addition to the documentation necessary to support the auditor's final conclusions, audit documentation must include information the auditor has identified relating to significant findings or issues that is inconsistent with or contradicts the auditor's final conclusions.' Therefore, if an auditor finds disconfirming evidence related to an audit issue in

the vast amount of Big Data related to a client, the auditor is required to document these findings and such documentation will be available to jurors if there is future auditor litigation.

Much of the research concerning auditor negligence lawsuits has found that the consideration of accounting alternatives can potentially make auditors vulnerable to litigation. Reffett (2010) presents empirical evidence that documenting and investigating accounting alternatives can increase an auditor's litigation liabilities. The documentation of disconfirming evidence provides jurors with evidence that there were reasons to question the auditors' conclusions, which makes them vulnerable to litigation threats. As a result, requirements to document disconfirming evidence have the capacity to create new threats to auditors and Big Data visualisations are a likely source of significant disconfirming evidence. Further, in the near future, the results of analyses of Big Data will often be recorded in the work papers in the form of visualisations and graphics. I propose that such visualisations will strongly influence juror decisions. Rose et al. (2017) found that Big Data visualisations represent a significant opportunity for auditors to detect disconfirming patterns of evidence. They indicate that Big Data visualisation groups are among the fastest growing practice areas at 'Big Four' firms and that visualisations are being introduced at different points in the audit engagement, both before and after traditional audit evidence is examined. While these visualisations represent new opportunities to identify patterns (Rose et al. 2017), the emotional responses they trigger may lead to an increase in juries ruling against auditors. When jurors see visualisations that disconfirm an auditor's conclusions, they have new reasons to believe that auditor's judgments were inappropriate—these visualisations may create powerful emotional reactions in jurors even when the data sources are unreliable.

The accounting literature related to Big Data suggests potential applications for Big Data in company audits, where it can generally be used to complement audit evidence when other types of evidence are lacking (Yoon et al. 2015). For example, if an auditor needs missing sales forecasts, they can use Big Data to perform text analysis on product discussion blogs, news articles and social media posts to better understand sales trends. Big Data can also be used to assist in fraud detection. Obtaining fraud evidence is a challenge because components involving morality, rationalisation and conduct are not always observable. Evaluating social media posts or emails can help an auditor to better understand an individual's motivations, rationalisations and feelings, such as animosity towards the firm (Yoon, et al. 2015). Big Data analytics can be used to identify patterns in the data to help

detect fraud risk. Indeed, 'Detecting fraud risk and Qualifications of Risks' is considered one of the top five benefits of analysing Big Data (Russom 2011).

Big Data can also help to enhance the reliability of audit evidence. Rather than examining traditional shipping forms to verify shipments, for example, an auditor can review truck GPS data to verify that the shipments were delivered. Big Data can further be used to improve audit efficiency by reducing the likelihood of false positives and identifying anomalies requiring further investigation (Issa and Kogan 2014; Cao, Chychyla and Stewart 2015). While it is clear Big Data will be increasingly important to audits, I posit the increased use of Big Data and visualisations will generate new auditor liability risks.

The Importance of Big Data Visualisations

Because human beings are not cognitively capable of sifting through enormous volumes of data, auditors are using advanced audit tools to analyse data from non-traditional sources. Yoon, et al. (2015, 432) state: 'Sophisticated data-mining techniques such as visualisations, predictive modelling, association, and clustering, are required to analyse Big Data'. Further, studies have found that visualisation strategies can help an auditor to recognise patterns. In their diagram design system-thinking tool approach, O'Donnell and Perkins (2011), found that visualisations help auditors to better identify fluctuation patterns and determine misstatement risk. These studies show that visualisations are important to the audit. Rose et al. (2017) interviewed Big Four accounting firm partners and found that visualisations are already being used at many points throughout the audit. Visualisations are becoming more and more important to the audit and their use is growing rapidly.

Visualisations of Data and Auditor Negligence Lawsuits

As Big Data visualisations are more commonly used by auditors, it is important to be aware of the significant effects these visualisations can have on juror emotions and decisions. Multiple studies have shown that jurors consider legally irrelevant factors when making decisions. For example, Hastie (1993) and Kadous (2000, 2001) found that when the consequences of audit failures are severe, juries are likely to find auditors negligent, regardless of the quality of the audit. Jurors are considered lay evaluators because typical jurors in the US have little accounting experience. In fact, a juror with accounting experience is likely to be dismissed by the plaintiff or defence attorney during *voir dire*. Given that jurors lack accounting knowledge, they often find it difficult to combine relevant case facts

into a coherent story (Reffett, Brewster and Ballou, 2012). Thus, their decisions are driven by easily understood emotional cues triggered by factors such as plaintiff losses (Kadous 2001). It is even questionable whether jurors can differentiate between levels of audit quality (Arel et al. 2012). There is a wealth of psychology literature stating that people are 'cognitive misers' who prefer to rely on time-effective strategies when making decisions (Fiske and Taylor 1984). The emotional response elicited by a visualisation may provide a shortcut for jurors in making decisions about auditor negligence, and the audit literature already demonstrates that jurors rely on their emotional responses to evidence. One topic that permeates the psychology literature on persuasion is the concept of vividness. The literature defines vividness as information 'that is as likely to attract and hold our attention and to excite the imagination, to the extent that is: (a) emotionally interesting, (b) concrete and imagery provoking, and (c) proximate in a sensory, temporal or spatial way' (Nisbett and Ross 1980, 45).

Many psychological studies have been conducted on vividness. For example, it has been found that more vividly presented information can enhance the ability to recall information (Shedler and Mantis 1986). Vividness has also been found to affect judgment. Lichtenstein at al. (1978) found that when examining frequencies of death, more vivid causes of death, such as botulism, flood and homicide, were overestimated and less vivid forms of death, such as asthma, stroke and heart disease, were underestimated. Lichtenstein et al. (1978) found that this phenomenon is explained by secondary bias, which suggests that the ease with which an event can be imagined or recalled depends on its vividness. People believe that they are more likely to be killed in a dramatic way and overestimate the possibility of that event occurring than they do for less vivid deaths. The results of this study were more recently replicated (Hertwig, Pachur and Kurzenhauser 2005). Further, the literature shows that individuals are better able to cognitively process visualisations that are more emotionally stimulating and vivid, which allows them to better use the information in the decision-making process (Keller and Block 1997). The literature clearly shows that vividness is an important factor in recall ability and can clearly affect judgment. Therefore, more vivid visualisations that induce greater emotions such as word clouds that emphasize words like "stressful" and "unethical" would be more likely to persuade a juror that a defendant is guilty of negligence.

Framing

Another important topic regarding emotions and visualisations is that of framing. The way in which visualisations, such as graphs or word clouds, display Big Data can dramatically affect the decisions of a jury. Neural affective decision theory is a psychological theory that consists of four principles: affect, brain, valuation and framing (Litt, Eliasmith and Thagard 2008). The first of these principles, 'affect', considers decision-making as closely tied to emotions. The second and third principles, 'brain' and 'valuation', posit that neural processes in the brain and valuing outcomes have positive or negative influences on decisions. The fourth principle is 'framing', which states that judgments and decisions vary depending on the context and manner in which information is presented (Litt et al. 2008). Framing a decision as a loss or a gain can dramatically affect the choices that people make (Tversky and Kahneman 1981, 1986). Therefore, if a visualisation such as a word cloud dramatically emphasises terms laden with negative emotional associations (e.g., 'unethical' or 'frustrating'), then the emotional reactions and subsequent decisions made by juries can be profound—and correspondingly negative.

Reffett et al. (2015) compared lay evaluators to auditors in their assessment of auditor negligence. Their experimental results show that among lay evaluators, the best predictor of auditor negligence verdicts was the extent to which plaintiff losses were considered, whereas among auditor evaluators, the best predictor was their emotional reaction to defendant auditors. These findings show that negligence lawsuits are not determined by the quality of the audit alone, but also by the jury's perceptions of the size of the loss and their emotional connection to the plaintiff or defendant. The emotional response triggered by a plaintiff who lost a significant amount of money is a greater factor in determining auditor guilt than the quality of the audit. Similarly, consistent with social identity theory, auditors are likely to empathise with other auditors who are put on trial (Reffett et al. 2015). Thus, positive or negative emotional responses play a role in determining auditor negligence.

Taken together, the audit and psychology literature show that emotions are a significant factor in determining auditor guilt in a trial situation. Big Data visualisations, such as word clouds, have the capacity to trigger strong emotional responses in jury members. Given that emotional responses have the capacity to increase auditors' vulnerability to negligence lawsuits, I hypothesise:

H1a: When visualisations of Big Data contradict auditor conclusions, visualisations that produce stronger emotional responses will cause jurors to be more likely to find auditors negligent than will visualisations that produce weaker emotional responses.

H1b: When visualisations of Big Data contradict auditor conclusions, visualisations that produce stronger emotional responses will cause jurors to award larger damage awards than will visualisations that produce weaker emotional responses.

Source Credibility

In addition to the emotional effects of different types of visualisations, the data sources used by auditors to create visualisations are also likely to have substantial effects on jurors' negligence decisions. Big Data are unlike other forms of traditional client-provided data. These data are often messy and come from sources that may lack credibility or perceived credibility relative to traditional client-provided data. One study found that 90 per cent of Big Data are unstructured, consisting of information such as social media posts, email messages, phone calls, website traffic and video streams (Syed et al. 2013). As auditors incorporate Big Data from digital sources such as social media into their audit processes, it may influence jurors' perceptions of the quality of both the data and the audit. If the audit evidence comes from a source that is perceived to be unreliable (e.g., social media posts), then jurors may be less convinced that the auditor was negligent when visualisations contradict management assertions.

Source credibility theory is a psychological theory that has been used extensively in marketing, advertising and political research to discover whether high credibility sources are more effective than low credibility sources in changings the beliefs, attitudes or behaviours of an audience (Pornpitakpan 2004). Several accounting studies used source credibility theory in relation to the audit process (DeZoort, Hermanson and Houston 2003; Brown and Popova 2016) and many accounting studies rely on source credibility theory, as developed by Chaiken and Maheswaran (1994). In these credibility studies, subjects read either an ambiguous or unambiguous message from a high or low credibility source. The studies found that source credibility affects persuasion when evidence is ambiguous. Chaiken and Maheswaran (1994, 460) stated that 'numerous experiments have shown that the attitude judgments of low-motivation or low-capacity subjects are influenced very little by the calibre of a message's persuasive argument but are influenced quite substantially by heuristic cues such as source credibility.' I propose that jurors are typically low-capacity and low-motivation subjects, who are likely to be significantly influenced by the source of evidence.

Source credibility theory consists of two parts: source expertise and source bias. Source expertise involves the perceived competence of the source providing information and source bias refers to perceived impartiality of the source (Birnbaum and Stegner 1979). For example, in the US, a Republican might be considered a biased source of information about a Democrat running for office (Birnbaum and Stegner 1979). Social media posts are likely to be perceived as more biased and lacking in expertise relative to audit evidence collected directly from the client. Information shared on social media tends to reflect individuals' personal, political, religious and other values and ideologies. Thus, real biases exist in social media posts. Further, the sources of information in social media are often unknown and therefore, the expertise of the sources is also unknown. There is also inherent informality in social media. Twitter, for example, limits postings to 140 characters or, at most, 280 characters, forcing many users to use creative spelling and emojis to express themselves. In contrast, data from audit clients have known sources, which make assessments of source expertise more reliable. Client data also exhibit greater levels of formality, suggesting higher levels of expertise. Professionalism further requires individuals to keep their personal opinions private; thus, they appear less biased than may be the case. Kruikemeier and Lecheler (2016) investigated how consumers evaluated news from different digital sources. They found that consumers believed that information gathered from email messages was significantly more credible than information gathered from Facebook and Twitter. Thus, it is likely that data in corporate emails will be perceived as more reliable than data obtained from social media sources such as Facebook and Twitter and therefore, juror decisions can be expected to be guided by the reliability of the data source used to create the Big Data visualisations.

Reliability has been used in the definition of source credibility theory. McCroskey (1966) specifically conceptualises reliability in his definition of authoritativeness when using source credibility. Frasca and Edwards (2017) drew on the McCroskey definition in applying source credibility when examining different social media graduate recruitment campaigns. Thus, it is justifiable to use the expanded definition of source credibility that includes reliability in hypothesis formation. Thus, I hypothesise:

> *H2a: When visualisations of Big Data contradict auditor conclusions, visualisations produced from more reliable data sources will cause jurors to be more likely to find auditors to be negligent than will visualisations produced from less reliable data sources.*

H2b: When visualisations of Big Data contradict auditor conclusions, visualisations produced from more reliable data sources will cause jurors to award larger damage awards than will visualisations produced from less reliable data sources.

Interaction Effects

An important interaction to consider in this study involves the potential for the effects of emotions in response to visualisations to overwhelm those of source credibility. What is most worrisome is the idea that jurors may choose to ignore the source credibility and reliability of data if visualisations create strong emotional responses. If this is the case in practice, then auditors will be at significant risk any time a visualisation that creates a strong emotional response is used. It is important to examine the interaction between data sources and the emotional effects of visualisation induced by the data because of the risk that emotional responses to visualisations could cause jurors to ignore the credibility and reliability of the underlying data. Thus, I hypothesise:

H3: When visualisations create a stronger emotional response, jurors will be influenced less by the reliability of the data used to generate the visualisation than when visualisations create a weaker emotional response.

Chapter Review

Big Data have become an important part of the audit process and auditors now use visualisations to recognise patterns. These visualisations are often vivid and create memorable effects. How a visualisation is framed can lead to different judgments. Big Data are often biased and unstructured. People tend to consider information they perceive to be less biased as more reliable and credible. I predict that more vivid and emotional visualisations would ultimately force one to ignore the credibility of the source.

Chapter 3
Research Method

Chapter Preview

This chapter explains the experimental design, the case used in the experiment, the demographics of the participants and the variables used.

Overview

This research examines how jurors in an auditor negligence case respond to two types of Big Data visualisations (bar graph v. word cloud) drawn from two data sources (social media v. email). This is a 2 x 2, full-factorial between-participant experiment, where the participants act as jurors. Participants read a case on their computer or electronic devices and then responded to a series of questions related to their emotional response to the visualisation, their perceptions of the credibility of the data source and their judgment on the degree of guilt of the auditor. Every participant read the same case and answered the same questions related to the content of the case. The only difference, as will be explained later, was the presentation of the evidence by the plaintiff attorney. The data were collected with Qualtrics, which is subscription-based software for collecting and analysing data for market and academic research.

Participant Selection

Participants were randomly selected from a population of qualified American jurors using Amazon Mechanical Turk (AMT). AMT is an online marketplace where individuals complete tasks for compensation. Recent studies in accounting have shown that AMT is reliable and provides diverse representation of the US population (Buhrmester, Kwang and Gosling 2011). AMT has also been used in accounting studies as a proxy for jury participants (see e.g., Grenier et al. 2015). Further, AMT is a better alternative to other popular methods of sampling such as using current university students and commercial focus groups. The problem with using current university students is that the external validity is low (Mason and Suri 2012). Conversely, commercial panels are expensive and usable responses are not guaranteed (Daly and Nataraajan 2015). Thus, AMT is an effective and affordable means of collecting data. AMT allows researchers to require potential study participants to undergo a screening process to eliminate candidates who would have been eliminated during *voir dire*. Therefore, extensive screening was used to ensure that only participants who could

understand the basics of the decision case were included in the study. Subjects who failed the screening and training questions were eliminated from the study.

Qualification Questions

As part of the screening process, participants at the beginning of the experiment were asked if they were eligible to serve on juries and were told that to be eligible they had to be: a US citizen, at least 18 years old and not convicted of a felony. Participants who were not eligible to serve on a jury were eliminated from the study. Participants were then asked four qualification questions. Three of the questions were simple but required some intellectual thought. The purpose of these questions was to eliminate subjects who were not serious about understanding the case and were simply attempting to click through the case without giving it the intellectual rigor required. The qualification questions include: 'Library is to book as book is to: a) binding, b) copy c) page d) cover'; 'Please enter the next number in the following sequence 11, 9, 7, 5' and 'Leaf is to tree as tree is to: a) plant, b) pine cone c) forest d) rock'. The final qualification question asked for the number of accounting classes taken by the participant. It is unlikely that a person with significant accounting experience would serve on the jury because one side would eliminate the subject during *voir dire*. Therefore, subjects who had taken six or more accounting classes were disqualified from the study.

Attention Check

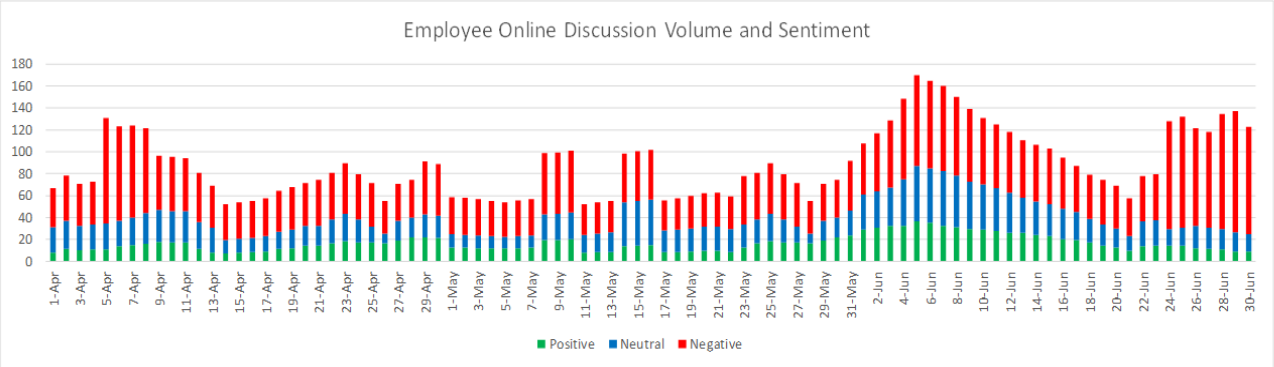
To ensure that subjects thoroughly read the case, they were told that they would receive a $1 bonus if they answered certain questions designed to check whether they were paying attention to the case. In the middle of the case text, the subjects were instructed to answer 'green' to the question 'what colour is the sky?'. If participants answered 'green', then they were awarded the bonus. If they answered any other colour, they did not receive the bonus. The purpose of the bonus was to give participants an incentive to thoroughly read the case and answer the questions thoughtfully and to remove those who did not adequately perform this task. It could be argued that including attention checks distracts participants or draws their focus to the attention check questions rather than those relating to the case. However, recent psychology studies show that attention checks do not affect scale validity and that using attention check questions is well justified (Kung, Kwok and Brown 2018).

To further ensure that participants observed the independent variables, participants were asked whether they viewed a bar graph or a word cloud, and further, whether they

observed Smith and Larson LLC's social media feeds or internal emails. These questions served as attention checks for the manipulations. The attention check questions were:

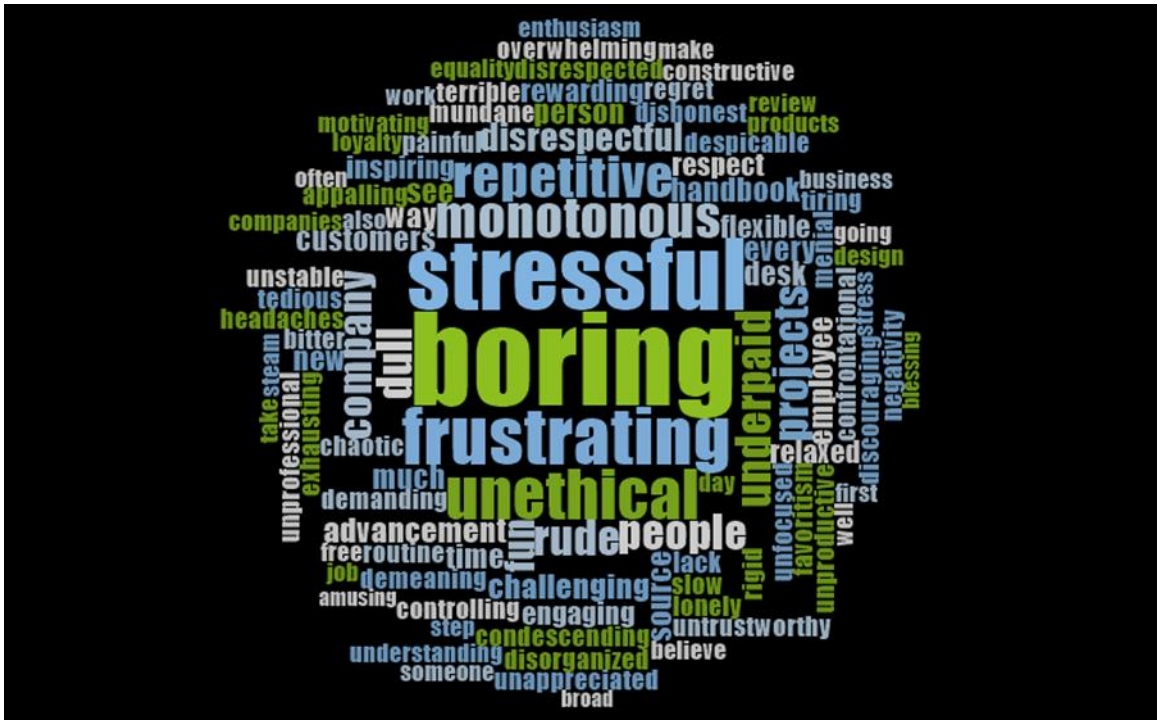
In the case, evidence about Omega employees' feelings was presented using a (Click on one image to indicate your response)

a. *Bar Graph*



Source	Mentions	Positive	Positive%	Neutral	Neutral%	Negative	Negative%	Net Sentiment
Total	8117	1496	18%	2130	26%	4492	55%	-37%

b. *Word Cloud*



In the case, Smith and Larson LLC analysed:

(a) The social media postings that are believed to have been made by Omega employees

(b) The internal emails sent by Omega employees to the Omega Human Resources department and upper management

Participant Demographics

Participants who met the requirements to serve on a US jury and passed the qualification questions were allowed to complete the experiment. A potential 529 participants were paid $2 for completing the entire case plus the $1 bonus for answering the 'What colour is the sky?' attention check question correctly. Anyone who failed any of the three attention check questions mentioned above was eliminated from the sample. Therefore, of the original 529 potential participants, only 410 were retained in the sample; 119 failed the attention checks and were eliminated (see Table 1 for a frequency table of the participants included in each treatment).

Table 1 Number of Participants per Treatment

Treatment	Number of Participants	Percentage
1 (Bar Chart/Social Media)	101	24.6
2 (Bar Chart/Internal Emails)	100	24.4
3 (Word Cloud/Social Media)	111	27.1
4 (Word Cloud/Internal Emails)	98	23.9
Total	**410**	**100**

Responses to each treatment were made by participants with diverse gender, education and professional backgrounds (see Tables 2–4). There were no statistically significant effects of gender, age or education on the dependent variables of interest, and gender, age and education levels did not vary significantly by treatment condition. Interestingly, the majority of participants were highly educated, with the majority being college graduates. Also, a disproportionate number of participants were young (see Table 5). As Table 6 suggests, participants from diverse backgrounds completed the experiment.

Table 2 Gender by Treatment

Treatment	Female	Male	Prefer Not to Say	Total
1 (Bar Chart/Social Media)	54	47	0	101
2 (Bar Chart/Internal Emails)	53	46	1	100
3 (Word Cloud/Social Media)	53	57	1	111
4 (Word Cloud/Internal Emails)	50	48	0	98
Total	**210**	**198**	**2**	**410**

Table 3 Education by Treatment

Treatment	Some High School	High School Graduate	Some College	College Graduate	Postgraduate Degree	Total
1 (Bar Chart/Social Media)	1	16	29	44	11	101
2 (Bar Chart/Internal Emails)	0	8	25	50	17	100
3 (Word Cloud/Social Media)	0	8	25	59	19	111
4 (Word Cloud/Internal Emails)	0	11	23	52	12	98
Total	**1**	**43**	**102**	**205**	**59**	**410**

Table 4 Education by Treatment Expressed as a Percentage

Treatment	Some High School	High School Graduate	Some College	College Graduate	Postgraduate Degree	Total
1 (Bar Chart/Social Media)	0.99	15.84	28.71	43.56	10.89	**24.63**
2 (Bar Chart/Internal Emails)	0.00	8.00	25.00	50.00	17.00	**24.39**
3 (Word Cloud/Social Media)	0.00	7.21	22.52	53.15	17.12	**27.07**
4 (Word Cloud/Internal Emails)	0.00	11.22	23.47	53.06	12.24	**23.90**
Total	**0.24**	**10.49**	**24.88**	**50.00**	**14.39**	**100**

Table 5 Age Ranges by Treatment

Treatment	Age Range (years)	Frequency	Percentage	Cumulative Percentage
1 (Bar Chart/Social Media)	18–35	49	48.5	48.5
	36–50	36	35.6	84.2
	Over 50	16	15.8	100.0
	Total	101	100.0	
2 (Bar Chart/Internal Emails)	18–35	59	59.0	59.0
	36–50	30	30.0	89.0
	Over 50	11	11.0	100.0
	Total	100	100.0	
3 (Word Cloud/Social Media)	18–35	55	49.5	49.5
	36–50	29	26.1	75.7
	Over 50	27	24.3	100.0
	Total	111	100.0	
4 (Word Cloud/Internal Emails)	18–35	42	42.9	42.9
	36–50	43	43.9	86.7
	Over 50	13	13.3	100.0
	Total	98	100.0	

Table 6 Number of Participants with each Professional Background by Treatment

Treatment	**1** (Bar Chart/ Social Media)	**2** (Bar Chart/ Internal Emails)	**3** (Word Cloud/ Social Media)	**4** (Word Cloud/ Internal Emails)	Total
Accounting/finance	7	6	4	3	20
Administration/office	6	11	13	6	36
Architecture/ engineering	2	6	1	1	10
Biotech/science	2	1	1	3	7
Business/management	9	10	12	9	40
Customer service	3	3	2	2	10
Food/beverage/hospitality	7	3	3	2	15
Education	6	9	6	8	29
General labour	14	3	4	10	31
Homemaker	5	4	2	1	12
Human resources	2	2	3		7
Legal / paralegal	1		2	2	5
Manufacturing	1		1		2
Marketing/public relations/advertising	1	2	2	2	7
Medical/health	9	6	8	10	33
Non-profit sector					0
Real estate			2		2
Retail/wholesale	1	2	2	3	8
Retired	2	2	2	3	9

Sales/business development	2	4	8	3	17
Salon/Spa/fitness		2		3	5
Security					0
Self employed	5	5	4	5	19
Skilled trade/craft	1	3	6		10
Student	1	2	4	3	10
Software/system network	3	8	8	3	22
Technical support	3			3	6
Transportation	2	1	1	3	7
TV/film/video/artist	3		2	1	6
Web/information design	1	1	2		4
Writing/editing		2	5	5	12
Unemployed	2	2	1	4	9
Total	103	100	111	98	

Unfortunately, there is very little data on the demographics of a typical jury. The reason for the lack of data is that there are thousands of state and local courts, and their methods and practices of selecting jury members differ widely. In addition, the people who are selected to serve on a jury are largely contingent on the subject matter of the case (Martin, 2015). Given that the typical demographics of US juries are unknown, it is not possible to determine how closely the participants in the experiment match a typical jury. However, participants in this study were drawn from a similar population used to produce actual juries, and the processes employed to select participants mirror those implemented in published research on jurors in accounting cases.

Decision Case

The negligence case is a modified version of the cases used by Backof (2015) and Rose et al. (2017). The case involves Nelson Inc., who relied on Absolute Company's audited financial statements in deciding to give Absolute a loan. It was later determined that the financial statements were misstated, which resulted in a $10 million loss for Nelson Inc. Nelson Inc. is suing Absolute's auditor, Smith and Larson LLC, for audit negligence. The case begins with the plaintiff's attorney making an opening statement. The opening statement explains Smith and Larson's duty as an auditor and states that Smith and Larson was negligent in its audit.

PLAINTIFF'S ATTORNEY OPENING STATEMENT

This case is about auditor negligence. The defendant, Smith and Larson, audited Absolute's 2016 financial statements. My client, Nelson, Inc., received and relied on Absolute's audited 2016 financial statements, but later found out that those audited financial statements were misstated. Smith and Larson's negligence in the conduct of its audit of Absolute cost Nelson, Inc. $10,000,000 in lost principal and interest on its loan to Absolute.

Financial statements are summaries of a company's financial information that are given to investors and creditors to help them make informed decisions. Auditors investigate the financial records of a company to determine whether the financial statements are a valid summary of the economic events and transactions that affected the company during the year. The result of auditors' work is a report that states whether or not the financial statements of a company are accurate, or, put another way, that the financial statements are not materially misstated. 'Material' means important, and is often measured in dollars. Although auditors are hired and paid by the companies whose financial statements they examine, an auditor's primary duty is to the general public, investors, and creditors to whom it matters that the financial statements are not materially misstated.

It is my job to prove to you, on behalf of Nelson, Inc., that Smith and Larson was negligent in its performance of the audit of Absolute's 2016 financial statements. Smith and Larson reported that the 2016 financial statements of Absolute were not materially misstated. In other words, Smith and Larson gave Absolute a 'clean' report. However, after the audited financial statements were released, it came to light that Absolute's financial statements listed revenues that were $5,000,000 too high. Smith and Larson failed to find this huge inaccuracy because the auditors did not perform an audit of sufficient quality; that is, they did not exercise the same degree of care in their conduct of the audit that other auditors in their position would have used. The $5,000,000 overstatement of revenue hid Absolute's financial problems from Nelson, Inc., and from others as well. Nelson, Inc. relied on the misstated financial statements when it decided to loan Absolute $10,000,000.

When Absolute's financial problems came to light, Absolute declared bankruptcy. The company closed and all 100 employees lost their jobs. Because of the terrible job market, many of them are still unemployed now, and others had to accept far less attractive jobs. The stockholders of Absolute suffered large losses when the company declared bankruptcy, and my client has received nothing in return for its loan to Absolute. Nelson, Inc. feels that the auditor, who negligently allowed the overstatement of revenue to occur, should reimburse it for its loss.

After I present my case, the defense will present its case. The defense will claim that Smith and Larson satisfied professional auditing standards with the work that it did and that Smith and Larson's judgments were reasonable given the facts available at the time of the audit. Consider carefully whether you believe this to be true and whether the auditors performed their duties in an appropriate manner in this particular case. I am confident that after weighing the evidence you will find for the plaintiff as Smith and Larson was negligent in performing its audit of the 2016 financial statements of Absolute and my client suffered as a result of Smith and Larson's negligence.

Next, the defense attorney makes an opening statement. The defense attorney states that Smith and Larson performed a quality audit and used good professional judgment when evaluating the facts available at the time of the audit. He also states that the plaintiff must prove that Smith and Larson was negligent and that the plaintiff will not be able to prove negligence because the audit evidence proves that Smith and Larson performed a quality audit.

DEFENSE'S ATTORNEY OPENING STATEMENT

The plaintiff has alleged that my client, Smith and Larson, was negligent in its audit of Absolute's 2016 financial statements. The plaintiff makes a point of mentioning the loss to his client, Nelson, Inc. That loss is not relevant in determining whether Smith and Larson was negligent in performing its audit of Absolute's 2016 financial statements. Only the actions and decisions made by Smith and Larson, as compared with those that would have been made by other competent Certified Public Accountants (CPAs) in similar circumstances, are relevant. Further, the losses to parties other than Nelson, Inc. are not relevant to this trial.

Negligence can be established only when an auditor fails to exercise the usual judgment, care, skill, and diligence employed by other CPAs in the community. CPAs use the guidance provided in the professional auditing standards to plan and perform their audit work, but the professional auditing standards also require auditors to use their professional judgment throughout an audit. According to the professional auditing standards, auditors must plan and conduct an audit so that they can provide reasonable assurance that the financial statements are free of material misstatements. In other words, an audit conducted in accordance with the professional auditing standards reduces, but cannot completely eliminate, the chance that people receive misstated financial statements. It is the defense's position that if an auditor complies with professional auditing standards and makes reasonable professional judgments given the facts available at the time of the audit, he or she has not been negligent. It is my job to prove to you that Smith and Larson did just that. I will present evidence that proves that Smith and Larson conducted a quality audit in accordance with the auditing standards and used good professional judgment when evaluating the facts available at the time of the audit.

The plaintiff must prove its allegations that Smith and Larson was negligent by a preponderance of the evidence. This means that it must show that the charges are more probably true than not true. The plaintiff cannot do so. The audit evidence proves that Smith and Larson conducted a quality audit, made a reasonable professional judgment, and in no way violated professional auditing standards. Smith and Larson is a competent, esteemed accounting firm, and I am confident that you will find in its favor.

In the section of the decision case that follows, the plaintiff, Nelson, presents evidence by explaining how the Absolute Company expected to be compensated for their services based on a five-year contract with Omega Company. If Absolute Company received the

expected revenue from Omega Company, it is unlikely that Absolute would have gone bankrupt. Omega compensates Absolute based on its own employee turnover. For example: if employee turnover is less than 5%, Absolute receives a bonus of four million dollars. Absolute estimates their chances of receiving a four million dollar bonus to be 90 per cent. The plaintiff's attorney then presents evidence showing that Absolute had very little chance of achieving this goal. The primary evidence is a visualisation prepared by the auditor that contradicts the conclusion that the bonus should have been received. The experimental manipulations are related to this visualisation.

Independent Variables

The experiment involves two manipulations. The first manipulation is the type of visualisation (bar graph v. word cloud) presented to the jury, and the effects of this manipulation are tested in hypotheses 1a (whether the auditor was negligent) and 1b (the size of the damage award). The manipulations are presented in Figures 1 and 2. Both graphics are produced from the same underlying data (social media posts/internal emails), but their presentation differs. The theoretical construct underlying this manipulation involves the emotional effect of the visualisation. The word cloud was expected to have a stronger emotional effect because of the sentiment elicited upon viewing emotion-laden words emphasised. Prior research demonstrated that this word cloud produces a stronger negative emotional response than this bar graph (Rose et al. 2019). Seeing words like 'boring,' 'stressful' and 'frustrated' evokes a strong emotional response. The emotional response to a bar graph, which only displays positive and negative sentiment, evokes a less emotional response than the word cloud. Importantly, the emotional response of the instrument was measured to provide evidence that the manipulation is effective in the context of the experiment and to demonstrate that the measured effects were caused by differences in emotional responses. When the plaintiff's attorney presented evidence in the decision case, half of the participants viewed the word cloud, while the other half viewed the bar graph.

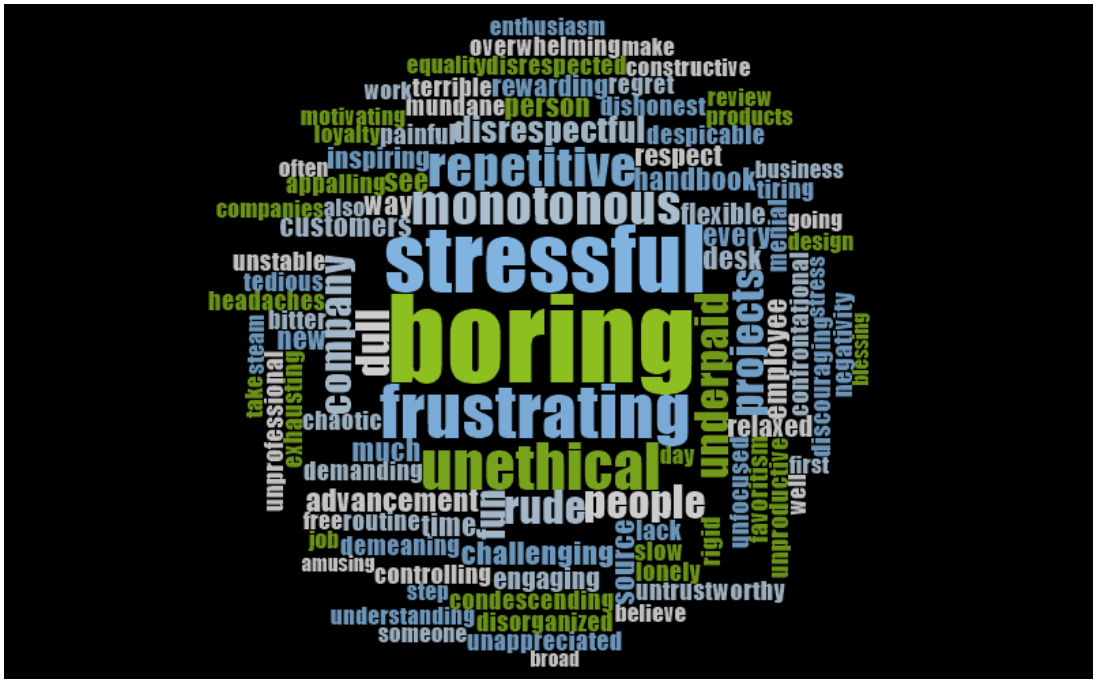
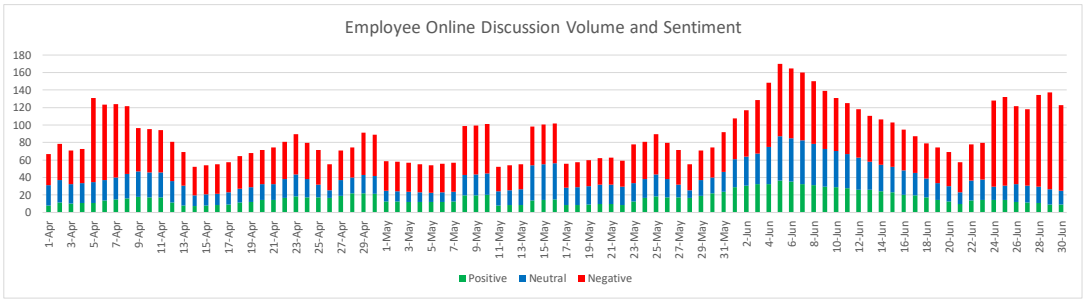


Figure 1 More Emotional Word Cloud



Source	Mentions	Positive	Positive%	Neutral	Neutral%	Negative	Negative%	Net Sentiment
Total	8117	1496	18%	2130	26%	4492	55%	-37%

Figure 2 Less Emotional Bar Graph

Manipulation two was the data source for the visualisation (Omega employees' social media v. internal emails) presented to the jury. The effects of this manipulation were evaluated in tests of hypotheses 2a (whether the auditor was negligent) and 2b (the size of the damage award). The theoretical construct underlying this manipulation involves the reliability of the data source. The internal emails are considered more reliable because the sources are known and the expertise associated with them can be evaluated. Importantly, the reliability of the data source used in the experiment was measured to provide evidence that the manipulation was effective and to demonstrate that the measured effects are caused by the credibility of the data source. Half of the participants received information that the source of the Big Data visualisation was social media and the other half were told that that the source of the Big Data visualisation was employee email messages when the plaintiff presented

evidence. To protect the quality of the data, participants were not allowed to change answers once they were entered into the experiment. However, as the decision case was relatively complex, participants were allowed to reread the case to build a better understanding of it.

Dependent Variables

The following are descriptions of the dependent variables used to test the hypotheses. The first dependent variable measures the participant's perception of auditor negligence:

> ***How likely it is that Smith and Larson LLC was negligent in performing the audit? (circle a number on the scale to indicate your response)***
>
> *0 10 20 30 40 50 60 70 80 90 100*
>
> *Not Likely At All* ... *Extremely Likely*

Consistent with the prior literature that examines juror decisions (Reffett 2010; Kaudus and Mercer 2016), I employed a binary response variable to capture a determination of negligence or no negligence:

> ***You will now cast a vote for the verdict. If the jury took a poll before deliberations, how would you vote? (check one response to indicate your response)***
>
> _____ *Smith and* Larson LLC *was* ***negligent***
>
> _____ *Smith and* Larson LLC *was* ***not negligent***

After measuring participants' determinations of negligence, the instrument asked participants to determine a damage award. The damage award question is also consistent with prior studies of juror decisions (Reffett 2010; Reffett et al. 2012). Only participants who found the auditor negligent in the binary response question were asked to provide a damage award amount:

> ***If you voted negligent, indicate the amount of damages (from $0 to $10 million) that you would require the auditors to pay to Nelson, Inc., the plaintiff. Please write an amount in the range of $0 to $10,000,000 to indicate your response. If you voted not negligent, please leave this line blank.***
>
> *$*_____

Participants are then asked one of the two questions below, depending on the manipulation received. These two questions capture the jurors' emotional responses to the auditors' performance after they viewed the visualisations.

For treatments 1 and 3:

In thinking about the auditor's performance in light of what you learned about the feelings of Omega employees from the word cloud were you?

0 *50* *100*
/– – – – /– – – – /– – – –/– – – –/– – – –/ – – – –/ – – – – /– – – –/ – – – –/ – – – –/
Not At All Angry *Indifferent* *Very Angry*

For treatments 2 and 4:

In thinking about the auditor's performance in light of what you learned about the feelings of Omega employees from the bar graph were you?

0 *50* *100*
/– – – – /– – – – /– – – –/– – – –/– – – –/ – – – –/ – – – – /– – – –/ – – – –/ – – – –/
Not At All Angry *Indifferent* *Very Angry*

Manipulation Checks

To demonstrate the effectiveness of the manipulations in the pilot and main experiment, it is important to show that they had the intended effects on the underlying theoretical constructs of interest. Participants answered the following questions to ensure that their perceptions of the manipulations were consistent with the theories (neural affective decision theory and source credibility theory) that are proposed to drive the responses. The first five questions were designed to measure the emotional response to visualisations and were derived from Rose et al (2017).

Do you believe that the Omega employees were happy?

0 *50* *100*
/– – – – /– – – – /– – – –/– – – –/– – – –/ – – – –/ – – – – /– – – –/ – – – –/ – – – –/
Not At All *Completely*

Do you believe that the Omega employees were discouraged?

0 *50* *100*
Not At All *Completely*

Do you believe that the Omega employees were angry?

0 *50* *100*
Not At All *Completely*

Do you believe that the Omega employees were frustrated?

0 *50* *100*
Not At All *Completely*

Do you believe that the Omega employees were depressed?

0 *50* *100*
Not At All *Completely*

The next four questions captured participants' beliefs about the reliability of email and social media as sources of audit evidence.

For Treatments 2 and 4:

Do you believe that emails sent by Omega employees to the Human Resources department and management of Absolute are reliable sources of audit evidence?

0 *50* *100*
Definitely Not Reliable *Definitely Are Reliable*

(Treatment 1 and 3)

Do you believe that social media postings about Absolute are reliable sources of audit evidence?

0 *50* *100*
Definitely Not Reliable *Definitely Are Reliable*

Do you believe that emails sent by Omega employees to the Human Resources department and management of Absolute are credible sources of audit evidence?

0 *50* *100*
Definitely Not Credible *Definitely Are Credible*

Do you believe that social media postings about Absolute are credible sources of audit evidence?

0 *50* *100*
Definitely Not Credible *Definitely Are Credible*

Debriefing Questions

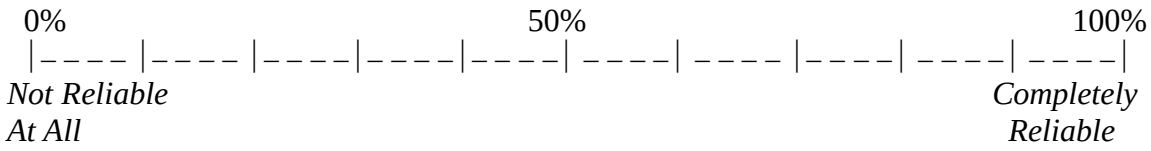
Following the experiment, the debriefing questions were used to build a better understanding of the beliefs and demographics of participants so that these measures could be used as statistical control variables to determine if they had significant influences on the dependent variables. First, participants were questioned about their perception of the visualisation that they examined during the experiment:

In your opinion, how useful was the visualisation of employee sentiment?

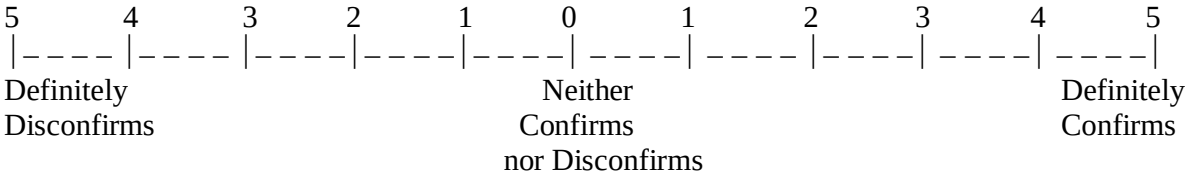
0% *50%* *100%*
Not Useful *Very Useful*

Next, participants were asked about their beliefs regarding the underlying reliability of the data (an essential aspect of credibility) and their perceptions regarding whether the visualisation disconfirms or confirms Absolute's estimate of receiving the bonus for low employee turnover.

In your opinion, how reliable were the data used to create the visualisation of employee sentiment?



In your opinion, do you believe that this visualisation confirms or disconfirms Absolute's estimate that it is 90% likely to achieve the bonus for low employee turnover?



Finally, participants in all treatments provided demographic information:

1) *What is your gender? _____Male _____Female _____ Other*
2) *Are you currently a student? _____ No _____Yes*
3) *What is your age _____*
4) *What is your primary occupation? _____*

Chapter Review

A 2 x 2 full-factorial between-subjects experiment was conducted. The participants were selected from diverse backgrounds using the online market place AMT. Participants were randomly assigned into four treatments and were shown either an unemotional stacked bar graph or an emotional word cloud. They were either told that data to create the word cloud or bar graph came from the company's (more reliable) internal emails or from (less reliable social media posts. To ensure that the participants gave the task the intellectual capital required, attention and manipulation checks were used.

Chapter 4
Results

Chapter Preview

The following chapter examines the statistical tests used in this experiment. These include outlier, normality, randomisation and hypotheses tests. Only one outlier was found. The sample followed a largely normal distribution and was randomised. Respondents' understanding was found to be a significant covariate. The results of the initial hypotheses testing found the opposite of what was expected for hypotheses 1 and 2. No significant results were found for the remainder of the hypotheses. After a brief discussion of debriefing questions and theory, mediation testing found that usefulness was an important mediation variable.

Outlier Check

As described in Chapter 3, 410 of 529 participants successfully answered all three attention check questions (77%) correctly and their answers were included in the data analysis. Both dependent variables (jurors' assessment of negligence and determination of the damage award) were checked for extreme outliers (see Appendix B). The tests were performed in SPSS using a boxplot. A boxplot or box-whisker plot (see Figures 3 and 4) displays the data such that the middle line represents the median of the dataset and the surrounding top and bottom blue boxes indicate where 50 per cent of all observations fall. The whiskers or top and bottom lines represent the least and most extreme scores. If the data are too extreme, it will not be marked by the whiskers but by a circle (see Figure 4, treatment two). Thus, boxplots support determination of whether there are any extreme outliers in the dataset.

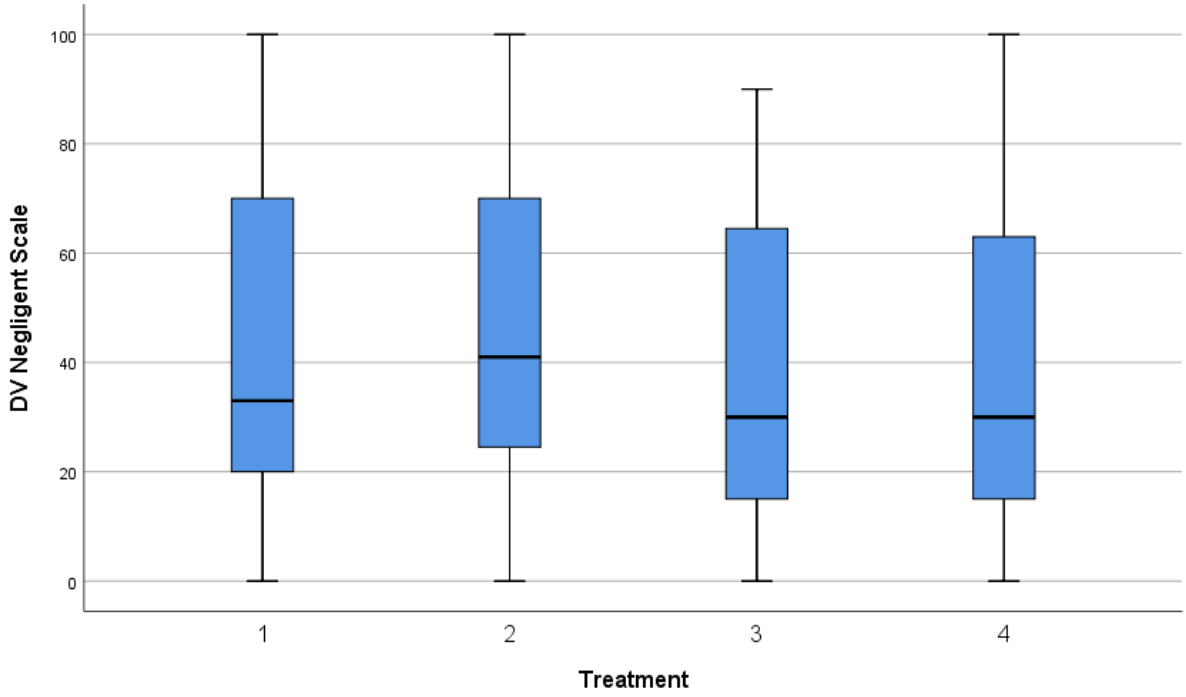


Figure 3 Boxplot Median Perception of Auditor by Treatment
The DV Negligent Scale shows the likelihood that Smith and Larson was negligent (0 = Not Likely At All; 100 = Extremely Likely). Participants were divided into four treatment groups (Treatment 1 = Bar Chart/Social Media, 2= Bar Chart/Internal Emails, 3= Word Cloud/Social Media, 4 = Word Cloud/Internal Emails).

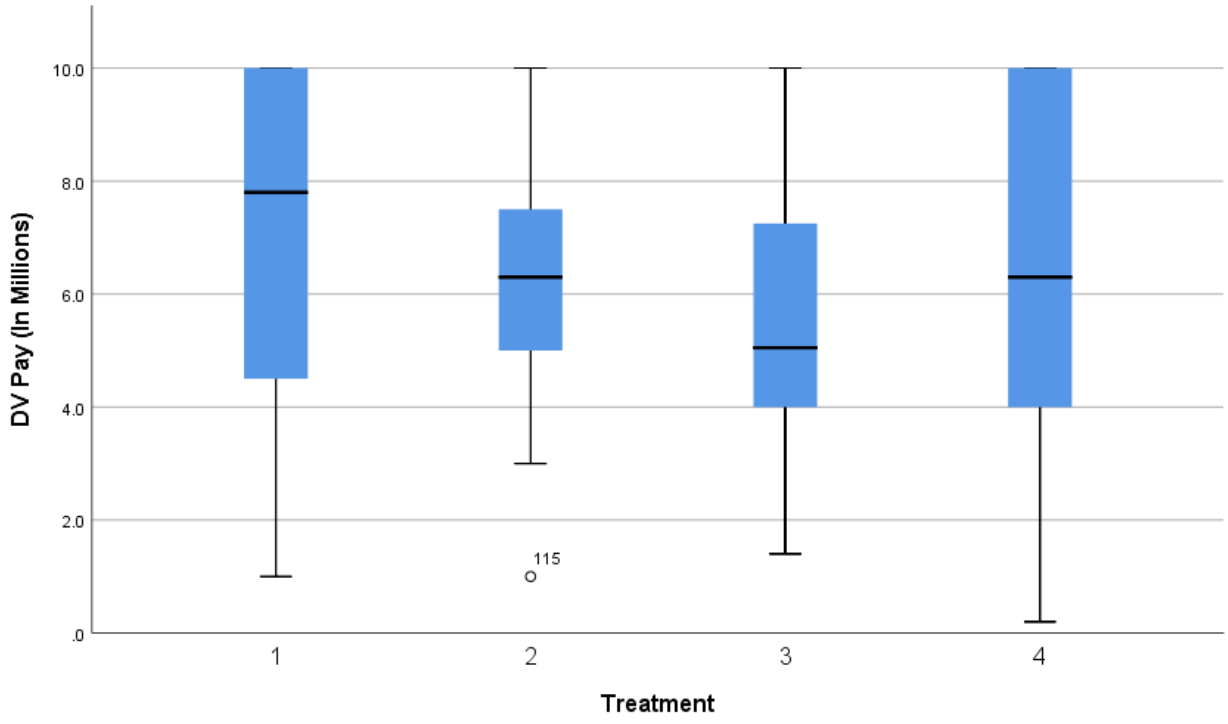


Figure 4 Boxplot Median amount defendant should be awarded by treatment
The DV Pay axis shows the damage amount participants stated that auditors should pay Nelson, if auditors were negligent ($0 to $10 million). Participants were divided into four treatment

groups (Treatment 1 = Bar Chart/Social Media, 2= Bar Chart/Internal Emails, 3= Word Cloud/Social Media, 4 = Word Cloud/Internal Emails).

There were no extreme outliers with respect to the likelihood of any of the jurors finding the auditor negligent in any of the four treatments. However, there was one extreme outlier in treatment 2 (see Figure 4), where one juror indicated the award should be only $178 out of a possible $10 million. This participant was removed from the sample, given their nonsensical response.

Normality Check

Next, the basic assumptions of the analysis of variance (ANOVA) analyses were checked for the continuous dependent variables (jurors' assessment of auditor negligence and determination of the size of the damage award, if negligent) by examining the normality of the distribution of the errors from each ANOVA. To test normality, the Shapiro-Wilk test was used (see Tables 7 and 8).

Table 7 The Shapiro-Wilk Test for DV/Negligent Scale

	Statistic	Degrees of Freedom	Sig
Negligent Scale	.937	410	.000

DV Negligent Scale = Likelihood Smith and Larson was negligent (0 = Not Likely At All; 100 = Extremely Likely)

Table 8 The Shapiro-Wilk Test for DV/Damage Award

	Statistic	Degrees of Freedom	Sig
Damage Award	.965	156	.001

DV Damage Award = Amount of damages auditors should pay Nelson, if auditors were negligent ($0 to $10,000,000).

If the results of the Shapiro-Wilk test are not statistically significant, the alternative hypothesis that the data are not from a normal distribution can be rejected. It can be assumed that the data came from a normal distribution and that the assumptions for using ANOVA and analysis of covariance (ANCOVA) to analyse the data are met. The p-values (significance values of the test) (see Tables 7 and 8) indicate that the results are significant, suggesting that the errors are not normally distributed. However, the Q-Q (quantile-quantile) plots (See Figures 5–8) show only a slight deviance from a normal distribution of errors (if the data are normally distributed, the dots on the chart should fall on the straight line). Therefore, further analysis of the error distribution was not required, given that ANOVA and ANCOVA are robust for departures from normality when samples are reasonably large and of similar sizes.

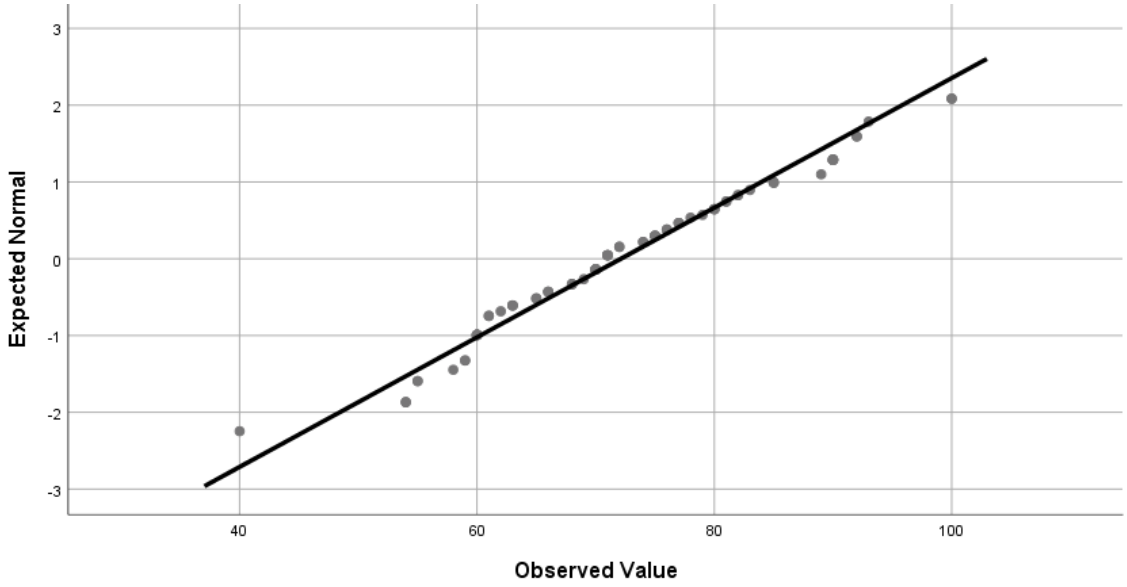


Figure 5 Normal Q-Q Plot of DV Negligent Scale for Internal Email Source

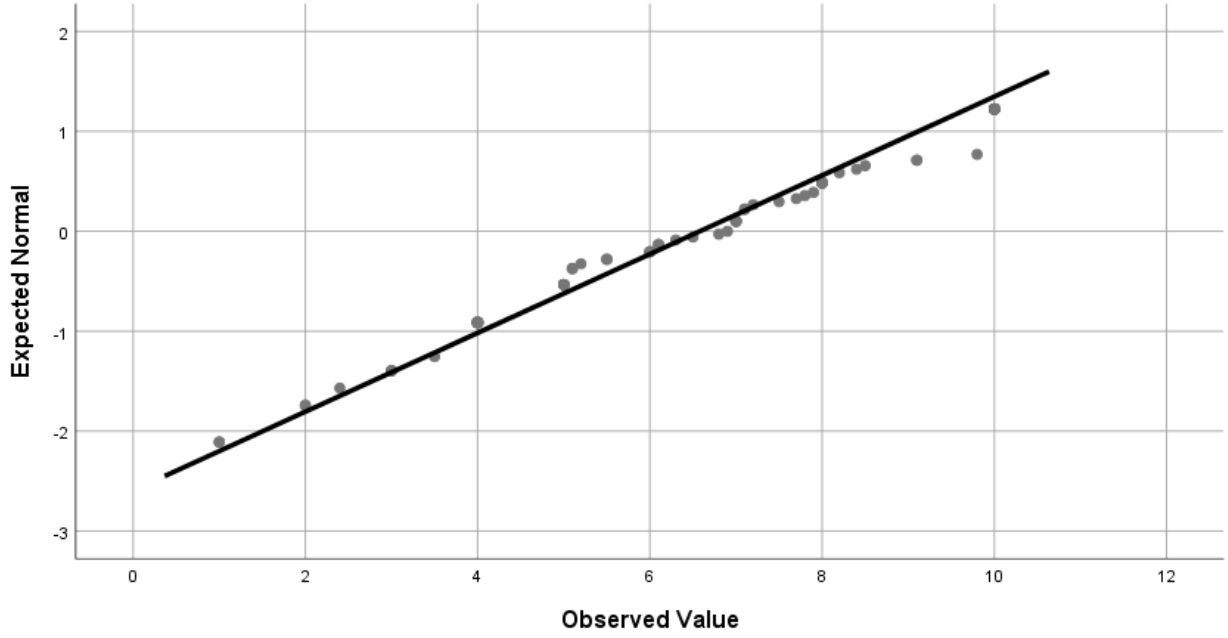


Figure 6 Normal Q-Q Plot of DV Damage Award for the Bar Graph Visualisation

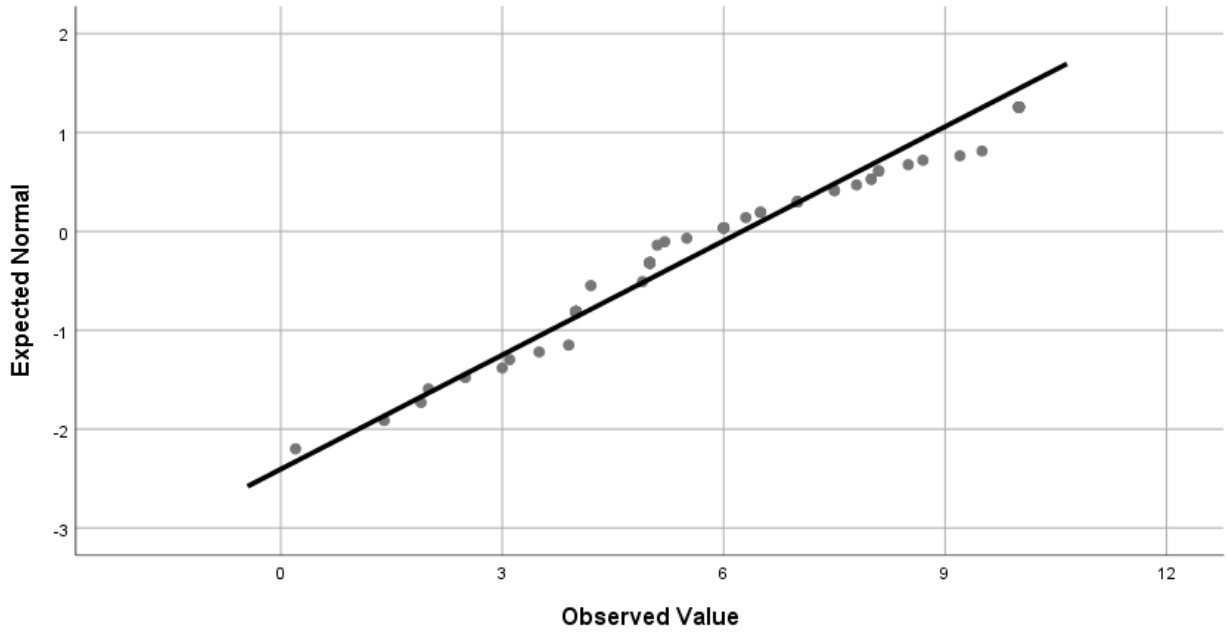


Figure 7 Normal Q-Q Plot of DV Damage Award for the Word Cloud Visualisation

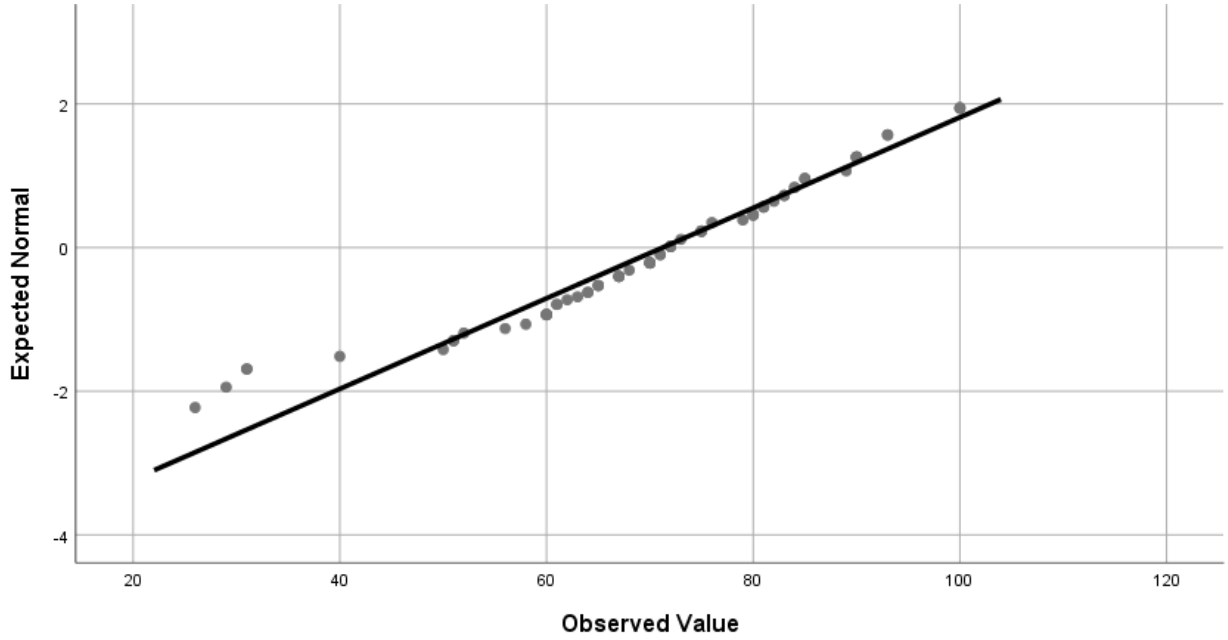


Figure 8 Normal Q-Q Plot of DV Negligent Scale for the Social Media Post Source

Randomisation Test

Randomisation is crucial in experiments because it eliminates most sources of systemic variance and thus, supports the ultimate objective of determining whether or not the independent variables are associated with the dependent variable. Generally, more precise measures are achieved when systematic variation is minimised. Scientists consider

randomisation an effective means of reducing systemic variance between treatment conditions and thus, researchers may conclude that changes to the dependent variable are most likely the result of the independent variable (Field 2009).

The randomisation process was checked to determine whether the key demographic variables were reasonably randomised. Therefore, a multivariate analysis of variance (MANOVA) model with the independent variables of the two treatments, plus their interaction, and the dependent variables of age, student and gender was evaluated (see Table 9).

Table 9 MANOVA of Randomising Key Demographic Variables

Source	Dependent Variable	Type III Sum of Squares	Degrees of Freedom	Mean Square	F-value	Sig.
Visualisation Type	Gender	.155	1	.155	.615	.433
	Student	.027	1	.027	.263	.608
	Age	418.947	1	418.947	2.970	.086
Reliability of Data Source	Gender	.022	1	.022	.086	.770
	Student	.144	1	.144	1.381	.241
	Age	356.421	1	356.421	2.5626	.113
Visualisation Type x reliability of Data Source (interaction)	Gender	.020	1	.020	.077	.781
	Student	.024	1	.024	.232	.630
	Age	58.61	1	58.610	.415	.520
Error	Gender	101.708	404	.252		
	Student	42.159	404	.104		
	Age	56995.325	404	141.078		
Total	Gender	198.000	408			
	Student	48.000	408			
	Age	664286.000	408			
Corrected Total	Gender	101.912	407			
	Student	42.353	407			
	Age	57832.814	407			

Variable Definitions: Manipulation of Visualisation Type = Word Cloud v. Stacked Bar Graph; Manipulation of Reliability of Data Source = Internal company email v. Social Media; Gender = Participant response to the question ‘What is your gender? (Male/Female/Prefer not to say); Student = Participant response to the question ‘Are you currently a student?‘ (Yes or No); Age = Participant response to the question ‘What is your age?’.

None of the demographic variables were significant, although age as it related to the graph manipulation tended towards significance with a p-value of 0.086. This shows that most demographic factors were effectively randomised across treatments. Further, this provides evidence that it is not necessary to statistically control for these variables in the hypothesis testing models. To determine whether it is appropriate to statistically control for age, a multivariate analysis of covariance (MANCOVA) was run with age as a covariate and negligence scale and damage award as dependent variables (see Table 10).

Table 10 MANCOVA with Age as a Covariate

Source	Dependent Variable	Type III Sum of Squares	Degrees of Freedom	Mean Square	F-value	Sig.
Age	Negligent Scale	2.169	1	2.169	0.011	0.917
	Damages	0.259	1	0.259	0.040	0.842
Visualisation Type	Negligent scale	0.316	1	0.316	0.002	0.968
	Damages	4.640	1	4.640	0.711	0.400
Reliability of Data Source	Negligent Scale	38.626	1	38.626	0.195	0.659
	Damages	0.298	1	0.298	0.046	0.831
Visualisation Type x Reliability of Data Source (interaction)	Negligent Scale	112.743	1	112.743	0.570	0.451
	Damages	27.445	1	27.445	4.208	0.042
Error	Negligent scale	29863.778	151	197.773		
	Damages	984.738	151	6.521		
Total	Negligent scale	831673.000	156			
	Damages	7465.400	156			
Corrected Total	Negligent Scale	30009.609	155			
Vividness	Damages	1016.624	155			

Variable Definitions: Manipulation of Visualisation Type = Word Cloud visualisation (more emotional response) v. Stacked Bar Graph visualisation (less emotional response; Manipulation of Reliability of Data Source = reliability of source underlying Big Data visualisation. Company email (more reliable) v. Social Media (less reliable); Age = participant response to the question, 'What is your age?; DV Negligent Scale = Likelihood Smith and Larson was negligent (0 = Not Likely At All; 100 = Extremely Likely); DV Damage Award = Amount of damages auditors should pay Nelson, if auditors were negligent ($0 to $10,000,000).

It is clear that age is not statistically significant and thus, does not need to be included as covariate in the statistical tests of the models.

Covariate Evaluation

The next test evaluated the need for covariates to be included in the hypothesis testing models. Testing for covariates investigates whether there are any variables that are not determined by the independent variable that act as a covariate. Thus, two ANCOVA models, one for each dependent variable, were run (see Tables 11 and 12). The potential covariates evaluated were: 'did you understand the case?' and 'have you seen word clouds or graphs before?'. These covariates were chosen because it was suspected that they may influence responses to the dependent variables in addition to the main manipulations (type of visualisation and source credibility) of the experiment.

Table 11 ANCOVA Model for the Covariate Understanding on the DV Negligence Scale

Source	Type III Sum of Squares	Degrees of Freedom	Mean Square	F-value	Sig.
Corrected Model	20,759.656	5	4151.931	5.568	0.000
Intercept	54,919.111	1	54919.111	73.654	0.000
Understand Case	15,525.404	1	15525.404	20.822	0.000
Previously Seen Visualisations	106.935	1	106.935	0.143	0.705
Visualisation Type	3,460.627	1	3460.627	4.641	0.032
Reliability of Data Source	379.535	1	379.535	0.509	0.476
Visualisation Type x Reliability of Data Source (interaction)	124.377	1	124.377	0.167	0.683
Error	301,235.456	404	745.632		
Total	1,024,222.000	410			
Corrected Total	321,995.112	409			

Variable Definition: Manipulation of Visualisation Type = Word Cloud visualisation (more emotional response) v. Stacked Bar Graph visualisation (less emotional response); Manipulation of Reliability of Data Source = reliability of source underlying Big Data visualisation—Company Email (more reliable) versus Social Media (less reliable); DV Negligent Scale; Understand Questionnaire =answers to the question, 'Did you understand the case and questionnaire?' (Responses are anchored on an 11-point scale ranging from 0 (Not at All) to 100 (Completely)); Previously seen visualisations = Answers to the question, 'Have you

previously seen visualisations like the presentation of words of varying sizes (these are often called word clouds)?‘ (Yes/No).

Table 12 ANCOVA Model for the Covariate Understanding on the DV Damage Award Scale

Source	Type III Sum of Squares	Degrees of Freedom	Mean Square	F-value	Sig.
Corrected Model	92.721[a]	5	18.544	3.011	0.013
Intercept	42.737	1	42.737	6.939	0.009
Understand Case	56.504	1	56.504	9.174	0.003
Previously Seen Visualisations	6.858	1	6.858	1.113	0.293
Visualisation Type	.039	1	0.039	0.006	0.937
Credibility of Data Source	2.863	1	2.863	0.465	0.496
Visualisation Type x Reliability of Data Source	38.200	1	38.200	6.202	0.014
Error	923.903	150	6.159		
Total	7465.400	156			
Corrected Total	1016.624	155			

Variable Definition: Manipulation of Visualisation Type = Word Cloud visualisation (more emotional response) v. Stacked Bar Graph visualisation (less emotional response); Manipulation of Reliability of Data Source = Reliability of source underlying Big Data visualisation (Company email (more reliable) versus Social media (less reliable)); DV Damage Award = Amount of damages auditors should pay Nelson, if auditors were negligent ($0 to $10,000,000); Understanding Covariate = Responses to the question, ‘Did you understand the case and questionnaire?’ on a scale from 0 (Not at all) to 100 (Completely); Previously seen visualisations; Covariate =Answers to the question, ‘Have you previously seen visualisations like the presentation of words of varying sizes (these are often called word clouds)?‘(Yes/No).

As Tables 11 and 12 show, participants understanding of the case and questionnaire is a significant covariate for both the negligence ($p < 0.001$) and damage award ($p < 0.003$) scales and thus, was included in models for hypothesis testing related to the continuous dependent variables. Having previously viewed graphs or word clouds was not statistically significant and was, therefore, not included for the negligence scale ($p < 0.705$) or the damage award ($p < 0.293$).

Hypotheses Testing

The first two hypotheses (H1a and H1b) involve the effects of emotional responses to visualisations on jurors' decisions.

Test of Hypothesis 1

H1a: When visualisations of Big Data contradict auditor conclusions, visualisations that produce stronger emotional responses will cause jurors to be more likely to find auditors to be negligent than will visualisations that produce weaker emotional responses.

H1b: When visualisations of Big Data contradict auditor conclusions, visualisations that produce stronger emotional responses will cause jurors to award larger damage awards than will visualisations that produce weaker emotional responses.

The hypotheses posit that the negative emotions elicited by a more emotional visualisation will influence jurors, leading them to make conclusions about the case based on these emotions. H1a and H1b suggest that jurors, when faced with more emotional visualisations that do not support the auditors' conclusions, will be more likely to find the auditors negligent and will require the payment of higher damages. Both hypotheses were tested using ANCOVA models, where the independent variables were the emotional effect of the visualisation and the source of the data. The word clouds are associated with stronger emotional responses and the stacked bar graphs are associated with weaker emotional responses. The source of the data was either the company's internal emails (more reliable) or social media posts (less reliable). The covariate is the juror's understanding of the case. The dependent variable is the assessment of the auditor's negligence (100-point scale) for H1a and the amount of damages awarded for H1b. There is a statistically significant effect of visualisation type (word cloud/stacked bar graph) on negligence assessment ($p = 0.021$), which supports H1a (see Table 13).

Table 13 ANCOVA Results for Juror Assessment of Negligence

Factor	Degrees of Freedom	F-value	p-value
Visualisation Type	1	5.339	0.021
Reliability of Data Source	1	0.582	0.446
Visualisation Type x Reliability of Data Source	1	0.186	0.666
Understanding	1	21.093	< 0.001
Error	405		

Variable definitions: DV Negligent Scale = Likelihood Smith and Larson was negligent (0 = Not Likely At All to 100 = Extremely Likely); Manipulation of Visualisation Type = Word Cloud visualisation (more emotional response) v. Stacked Bar Graph visualisation (less emotional response); Manipulation of Reliability of Data Source: Reliability of source underlying Big Data visualisation (Company email (more reliable) versus Social media (less reliable)); Understanding Covariate = answer to the question, 'Did you understand the case and questionnaire?' on a scale from 0 (Not at all) to 100 (Completely).

When considering descriptive statistics, the overall results were the opposite of what was hypothesised. As Table 14 indicates, the mean for the more emotional visualisation (word cloud) was equal to 38.40 and the mean for the less emotional visualisation (graph) was equal to 44.49. Thus, participants were more persuaded by the less emotional, less vivid bar chart than by the more vivid and emotional word cloud.

Table 14 Descriptive Statistics for Juror Assessment of Negligence Means

	More Emotional Visualisation	**Less Emotional Visualisation**	**Total**
More Reliable Evidence	39.60 (28.61) [98]	46.60 (26.12) [100]	43.14 (27.54) [198]
Less Reliable Evidence	37.33 (27.98) [111]	42.41 (28.98) [101]	39.75 (28.51) [212]
Total	38.40 (28.23) [209]	44.49 (27.60) [201]	41.39 (28.06) [410]

(Standard Deviation) [Number of Participants]

A logistic regression using an alternative dependent measure of negligence (Negligent/Not Negligent) was further used to test hypothesis H1a. The alternative test examined the dichotomous answer to the question: 'was Smith and Larson negligent or not negligent?' The model included the visualisation eliciting emotional effect, data source reliability and the control variable for understanding. The results were consistent with the ANCOVA results using a continuous measure of negligence. The negative significant coefficient for the visualisation type ($b = –0.381$, $p = 0.069$) approaches significance and indicates that jurors found the auditor to be less likely to be negligent when the visualisation was more emotionally affective (see Table 15).

Table 15: Binary Logistic Results for Juror Negligence Decision

Variable	B	S.E.	Wald	Degrees of Freedom	Sig.
Visualisation Type	–0.381	0.21	3.298	1	0.069
Reliability of Data Source	0.104	0.21	0.243	1	0.0622
Understanding	–0.038	0.009	17.519	1	<0.001
Constant	3.164	0.872	13.160	1	<0.001

Dependent Variable = Participant response to the following question, 'You will now cast a vote for the verdict. If the jury took a poll before deliberations, how would you vote?' Participants could respond 'negligent' or 'not negligent'.; Manipulation of Visualisation Type = Word Cloud visualisation (more emotional response) v. Stacked Bar Graph visualisation (less emotional response); Manipulation of Reliability of Data Source: Reliability of source underlying Big Data visualisation (Company email (more reliable) versus Social media (less reliable)); Understanding Covariate = answer to the question, 'Did you understand the case and questionnaire?' on a scale from 0 (Not at all) to 100 (Completely).

The same ANCOVA test was used to test H1b, except the dependent variable for this test was the value of the jury's damage award. There were fewer subjects in this analysis because only those who found the defendant negligent were asked to determine damages owed to the plaintiff. Therefore, the sample size decreased to 156 subjects. No significant effect was found for the visualisations' emotional effect on the damage awards ($p = 0.686$). However, there was a significant interaction between visualisation type and data reliability ($p = 0.018$), which will be discussed later. The next two hypotheses examined the effects of data reliability on juror decisions.

Test of Hypothesis 2

> *H2a: When visualisations of Big Data contradict auditor conclusions, visualisations produced from more reliable data sources will cause jurors to be more likely to find auditors to be negligent than will visualisations produced from less reliable data sources.*
>
> *H2b: When visualisations of Big Data contradict auditor conclusions, visualisations produced from more reliable data sources will cause jurors to award larger damage awards than will visualisations produced from less reliable data sources.*

H2a and H2b predict that when jurors are presented with Big Data visualisations that were created from less reliable data sources and that contradict auditors' conclusions, they will find auditors less negligent and will award lower damages. H2a is tested in the model in Table 13, which indicates no significant effects of data source reliability on negligence determinations ($p = .446$). The logistic regression model in Table 15 produces a slightly

different conclusion because there is a marginally significant effect of reliability on the binary decision (p = .0622).

Table 16 ANCOVA Results for Juror Assessment of the Damage Award

Factor	Degrees of Freedom	F-value	P-value
Visualisation Type	1	0.164	0.686
Reliability of Data Source	1	0.622	0.431
Visualisation Type x Reliability of Data Source	1	3.837	0.018
Understanding	1	8.799	0.004
Error	151		

DV Damage Award = Amount of damages auditors should pay Nelson, if auditors were negligent ($0 to $10,000,000). Manipulation of Visualisation Type: Word Cloud visualisation (more emotional response) v. Stacked Bar Graph visualisation (less emotional response); Manipulation of Reliability of Data Source: Reliability of source underlying Big Data visualisation (Company email (more reliable) versus Social media (less reliable)); Understanding Covariate = answer to the question, 'Did you understand the case and questionnaire?' on a scale from 0 (Not at all) to 100 (Completely).

The ANCOVA model presented in Table 16 that examined damage award determinations produced similar conclusions. There is no main effect of reliability on damage award decisions (p = .431, see Table 16). However, there is a significant disordinal interaction between reliability and the emotional effect. A disordinal interaction is when two or more group means cross. Thus, the results of main effects tests cannot be reliably interpreted without evaluation of the interaction. The third hypothesis in this study concerns this interaction and suggested that highly emotional visualisations would influence jurors, leading them to place less weight on the reliability of the data when making negligence decisions.

Test of Hypothesis 3

> *H3: When visualisations create a stronger emotional response, jurors will be influenced less by the reliability of the data used to generate the visualisation than when visualisations create a weaker emotional response*

No significant interaction was found between the emotional response and reliability on evaluations of negligence. It appears that only the emotional response influences negligence decisions. However, the results for the damage award are significant (p = .018) and will be discussed later (see Figure 9).

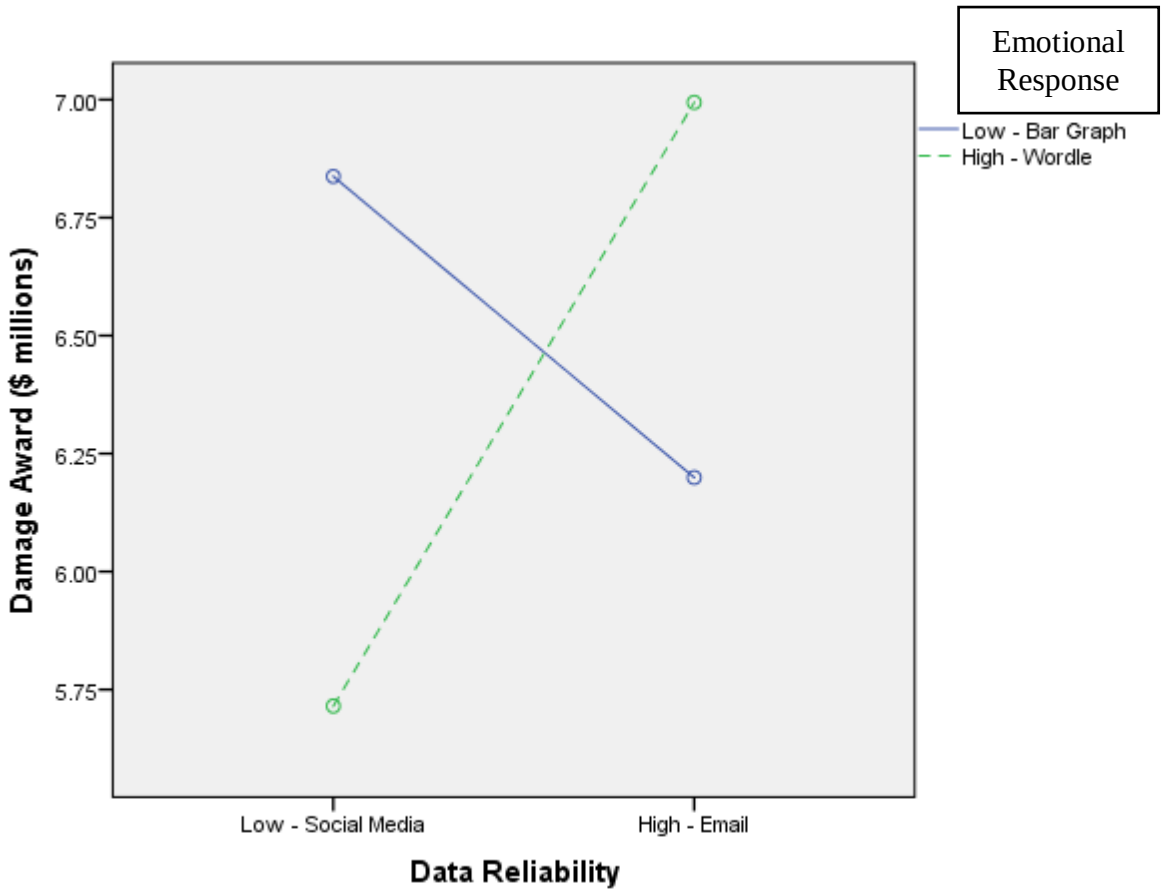


Figure 9 Interactive Effects of Data Reliability and Emotional Response on Juror Damage Awards

DV Damage Award = Amount of damages auditors should pay Nelson, if auditors were negligent ($0 to $10,000,000). Manipulation of Visualisation Type = Word Cloud visualisation (more emotional response) v. Stacked Bar Graph visualisation (less emotional response); Manipulation of Reliability of Data Source: Reliability of source underlying Big Data visualisation (Company email (more reliable) versus Social media (less reliable)).

Testing of simple effects further indicated that jurors care more about the reliability for more emotional visualisation, supporting H3. Further, when data are more reliable (mean = 6.97) than when they are less reliable (mean = 5.78) (see Table 17).

Table 17 Descriptive Statistics for Juror Assessment of Damage Award Means (Standard Deviation) [Number of Participants]

	More Emotional Visualisation	Less Emotional Visualisation	Total
More Reliable Evidence	6.73 (2.87) [35]	6.23 (2.22) [45]	6.45 (2.52) [80]
Less Reliable Evidence	5.78 (2.24) [36]	6.97 (2.83) [40]	6.41 (2.62) [76]
Total	6.25 (2.60) [71]	6.58 (2.54) [85]	6.43 (2.56) [156]

Debriefing Questions

In addition to the demographic questions relating to gender, age and occupation, participants were asked three debriefing questions. The questions and their purposes are described below.

In your opinion, how useful was the visualisation of employee sentiment? (0 to 100)?
I conjectured that the word cloud would clearly and emotionally express how the Omega workers felt. Therefore, participants pretending to serve on a jury should find the word cloud useful in making a decision regarding auditor negligence. Thus, the purpose of this debriefing question was to confirm whether this theory was correct.

In your opinion, how reliable were the data used to create the visualisation of employee sentiment? (0 to 100)?
I conjectured that jurors would find Omega's internal emails more credible than their social media posts. Therefore, the purpose of this question was to confirm this theory.

In your opinion, do you believe that this visualisation confirms or disconfirms Absolute's estimate that it is 90% likely to achieve the bonus for low employee turnover. (-5 to 5)?
The information displayed in the word cloud visualisation and stacked bar graph should disconfirm Absolute's above estimate. The purpose of this question was to discover if this

was the case and whether one visualisation was better at disconfirming Absolute's estimate than another. The results of these questions are presented below.

Table 18 Participants' Answers to Debriefing Questions

Treatment		N	Minimum	Maximum	Mean	Standard Deviation
1 (Bar Chart/Social Media)	Usefulness	101	0	100	63.21	24.172
	Reliability	101	0	100	54.96	24.594
	Likely to get bonus	101	–5	5	.82	2.165
2 (Bar Chart/Internal Emails)	Usefulness	100	0	100	59.21	24.554
	Reliability	100	0	100	54.96	23.634
	Likely to get bonus	100	–5	5	.66	2.090
3 (Word Cloud/Social Media)	Usefulness	111	1	100	62.42	22.202
	Reliability	111	0	100	57.27	22.635
	Likely to get bonus	111	–5	4	.94	1.983
4 (Word Cloud/Internal Emails)	Usefulness	98	9	100	62.33	25.183
	Reliability	98	0	100	54.26	25.798
	Likely to get bonus	98	–4	5	.98	2.168

Variable definitions: Usefulness = Answer to the question, 'In your opinion, how useful was the visualisation of employee sentiment?'; Reliability = Answer to the question, 'In your opinion, how reliable was the data that was used to create the visualisation of the employee survey?'; Likely to get bonus = Answer to the question, 'In your opinion, do you believe that this visualisation confirms or disconfirms Absolute's estimate that it is 90% likely to achieve the bonus for low employee turnover?'.

The subjects were not overly confident with the visualisations. The first two treatments consisted of the bar graph visualisation and the last two treatments consisted of the word cloud visualisation. For all treatments, the visualisation was found to be marginally useful, marginally reliable and marginally confirmatory of the auditor's assertions. Thus, using these particular visualisations in a jury trial may not be an overly persuasive means of evidence.

Mediation Analysis

This section evaluates whether perceived usefulness is a mediating variable. A mediating variable is the causal result of the independent variable and the casual antecedent of the dependent variable. That is, when the independent variable is combined with a mediating variable, it induces a causal result in the dependent variable (see Figure 10). Given that the previous analysis found that usefulness was an important covariate, it is possible that

usefulness is a mediating variable of the effect of visualization type and/or data reliability on the negligence scale and damage awards dependent variables.

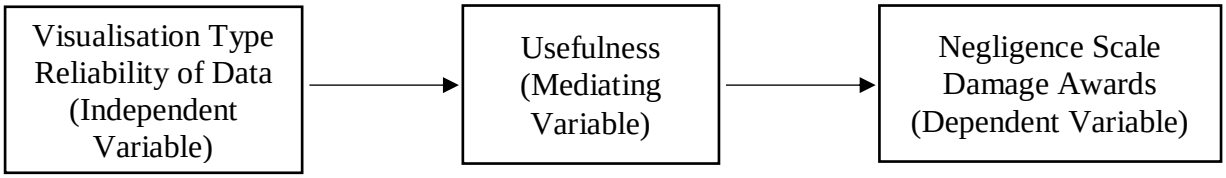


Figure 10 Mediating Variable on Dependent Variable

To explore usefulness as a possible mediating variable, a mediation analysis was conducted using the least-squares method following the approach outlined by Baron and Kenny (1986). The first step in this approach is to show that the causal variable is correlated with the outcome. This was accomplished by a simple regression test to investigate whether the independent variable affected the dependent variable (see Table 19).

Table 19 Does the IV (Visualisation Type/Reliability) Affect the DV (Negligent Scale)?

Model	Unstandardised B	Coefficients Std. Error	Standardised Coefficients Beta	t	Sig.
Constant	42.893	2.400		17.869	.000
Visualisation Type	-6.003	2.759	–.107	-2.176	.030
Reliability of Data Source	3.214	2.760	.057	1.165	.245

Variable definitions: DV Negligent Scale = Likelihood Smith and Larson was negligent (0 = Not Likely At All to 100 = Extremely Likely); Manipulation of Visualisation Type = Word Cloud visualisation (more emotional response) v. Stacked Bar Graph visualisation (less emotional response); Manipulation of Reliability of Data Source = Reliability of source underlying Big Data visualisation (Company email (more reliable) v. Social media (less reliable)).

As the results presented in Table 19 indicate, there is a statistically significant negative relationship between the visualisation variable (word cloud/bar graph) and the jury assessment of negligence variable ($r = –.107$, $p < .03$). This suggests that the more emotional the visualisation, the less likely it is that the auditor will be considered negligent.

The next step is to show that the causal variable (visualisation type variable) is correlated with the mediating variable (usefulness of visualisation). That is: does the

independent variable (visualisation type) affect the mediator (usefulness)? This requires a second regression (see Table 20).

Table 20 Does the IV (Visualisation Type/Reliability) affect the Mediator (Usefulness)?

Model	Unstandardised B	Coefficients Std. Error	Standardised Coefficients Beta	t	Sig.
Constant	59.366	2.247		26.416	.000
Visualisation Type	–12.087	2.583	–.226	–4.680	.000
Reliability of Data Source	2.414	2.584	.045	.934	.351

Dependent Variable: Juror assessment of usefulness of the visualisation type responses to the following question, 'In your opinion, how useful was the visualisation of the employee survey?' Responses are anchored on an 11-point scale ranging from 0 (Not at all useful) to 100 (very useful); Manipulation of Visualisation Type: Word Cloud visualisation (more emotional response) v. Stacked Bar Graph visualisation (less emotional response); Manipulation of Reliability of Data Source: Reliability of source underlying Big Data visualisation (Company email (more reliable) v. Social media (less reliable)).

Table 20 indicates that there is a statistically significant negative relationship between the visualisation type variable and its perceived usefulness ($r = –.226$, $p < .001$). Hence, the more emotional the visualisation, the lower its perceived extent of auditor negligence. These results disagree with the expected outcome and will be discussed further in this chapter.

The third step is to investigate whether the mediating variable (perceived usefulness of the visualisation) is significantly correlated with the outcome variable (extent of negligence). Table 21 reveals a statistically significant positive association between the perceived usefulness of the visualisation and perceived auditor negligence ($r = .414$, $p < .001$). The mediator (usefulness) affects the negligence scale ($p < .001$).

Table 21 Does the Mediator(Usefulness) Affect the DV (Negligence Scale)?

Model	Unstandardised B	Coefficients Std. Error	Standardised Coefficients Beta	t	Sig.
Constant	17.794	2.859		6.224	0.001
Usefulness of Visualisation	0.434	0.047	0.414	9.197	0.001

DV Negligent Scale = Likelihood Smith and Larson was negligent (0 = Not Likely At All to 100 = Extremely Likely); Usefulness of visualisation: Juror assessment of usefulness of the

visualisation type responses to the following question, 'In your opinion, how useful was the visualisation of the employee survey?' on an 11-point scale ranging from 0 (Not at all useful) to 100 (very useful).

The final step is to determine whether the mediating variable (perceived usefulness of the visualisation) reduces or eliminates the significant effect of the independent variable (visualisation type) on the dependent variable negligence scale. This would indicate full mediation and involves one final regression (see Table 22).

Table 22 Does Mediator (Usefulness) Reduce or Eliminate the Significant Effect of the Independent Variable (Visualisation Type) on the Dependent Variable (Negligence Scale)?

Model	Unstandardised B	Coefficients Std. Error	Standardised Coefficients Beta	t	Sig.
Constant	17.467	3.627		4.815	.000
Visualisation Type	–.827	2.598	–.015	–.318	.750
Reliability of Data Source	2.180	2.534	.039	.860	.390
Usefulness of Visualisation	.428	.049	.409	8.820	.000

DV Negligent Scale = Likelihood Smith and Larson was negligent (0 = Not Likely At All to 100 = Extremely Likely); Manipulation of Visualisation Type: Word Cloud visualisation (more emotional response) v. Stacked Bar Graph visualisation (less emotional response); Manipulation of Reliability of Data Source: Reliability of source underlying Big Data visualisation (Company email (more reliable) v. Social media (less reliable)); Usefulness of visualisation: Juror assessment of usefulness of the visualisation type as responses to the following question, 'In your opinion, how useful was the visualisation of the employee survey?' on an 11-point scale ranging from 0 (Not at all useful) to 100 (very useful).

Table 22 shows that the inclusion of the mediator causes the visualisation type variable to become non-significant ($p < 0.75$), which is evidence of full mediation. These mediation results suggest that the jurors' determination of whether the visualisation is useful for analysing sentiment is a key driver of the results. The results further suggest that jurors believe that the intention of the emotionally laden visualisation is to manipulate their feelings, rather than to display the information in a useful format. Therefore, they rely less on the information. The participants' interpretation of the intent of including an emotionally laden visualizations helps to explain why the results were the opposite of what was predicted.

Next, potential mediating effects of usefulness perceptions on the damage award determination were investigated using the same approach described by Baron and Kenney (1986), The first step is to check for significant correlation between the independent variables (visualisation type and reliability of the data) and the dependent variable damage award (see Table 23).

Table 23 Does the IV (Visualisation Type/Reliability) Affect the DV (Damage Scale)?

Model	Unstandardised B	Coefficients Std. Error	Standardised Coefficients Beta	t	Sig.
Constant	6.567	0.354		18.537	0.000
Visualisation Type	–0.332	0.414	–0.065	–0.803	0.423
Reliability of Data Source	0.028	0.412	0.005	0.067	0.947

DV Damage Award = Amount of damages auditors should pay Nelson, if auditors were negligent ($0 to $10,000,000); Manipulation of Visualisation Type: Word Cloud visualisation (more emotional response) v. Stacked Bar Graph visualisation (less emotional response); Manipulation of Reliability of Data Source: Reliability of source underlying Big Data visualisation (Company email (more reliable) v. Social media (less reliable)).

No mediating effects of the main effects were found with this method, but it is also possible that the mediator influences the moderating effect of reliability. That is, perceptions of usefulness could influence the interactive effect of visualization type and data reliability on damage awards. Therefore, a supplementary mediated moderation test was performed using the method outlined by Muller, Judd and Yzerbyt (2005). This complex, multistep process examines a series of regressions that combine mediation and moderation. Mediation involves the intervening mechanism that produces the treatment effect and moderation concerns factors that affect the magnitude of the treatment effect. To test for mediation of an interaction, also known as mediated moderation, Muller et al. (2005) propose the following model: $Y = \beta + \beta X + \beta Mo + \beta XMo + \beta Me + \beta MeMo$. In this model, Y is the damage award; X is the independent variable (visualisation type); Mo is the moderating variable (reliability), XMo is the product of the visualization type and reliability, Me is the mediating variable (usefulness), and MeMo represents the product of the mediator and moderator.

To begin the Muller et al. (2005) analysis, the first step is to establish the significance of the interaction (i.e., overall moderation). The results of this analysis in Table 24 show an interaction that is tended towards being statistically significant ($p < 0.085$, two-tailed). This

indicates a moderating effect of data reliability. Next, Muller et al. (2005) outline a procedure to determine whether the treatment effect on the mediator is moderated. Finally, their last step examines whether the mediator effects the dependent variable, while controlling for moderation and the interaction of the moderator and mediator (see Tables 24–26). The results do not provide evidence of mediated moderation.

Table 24 Test for Overall Moderation

Model	Unstandardised B	Coefficients Std Error	Standardised Coefficients Beta	t	Sig.
Constant	7.012	0.347		20.224	0.000
Visualisation Type	–0.951	0.503	–0.191	–1.889	0.060
Reliability of Data Source	–0.660	0.472	–0.133	–1.400	0.163
Interaction	1.213	0.701	0.204	1.729	0.085

DV Damage Award = Amount of damages auditors should pay Nelson, if auditors were negligent ($0 to $10,000,000); Manipulation of Visualisation Type: Word Cloud visualisation (more emotional response) v. Stacked Bar Graph visualisation (less emotional response); Manipulation of Reliability of Data Source: Reliability of source underlying Big Data visualisation (Company email (more reliable) v. Social media (less reliable))

Table 25 Mediated Interaction

Model	Unstandardised B	Coefficients Std Error	Standardised Coefficients Beta	t	Sig.
Constant	58.746	2.298		25.559	0.000
Visualisation Type	–11.052	3.226	–0.205	–3.426	0.001
Reliability of Data Source	4.164	3.226	0.077	1.291	0.197
Interaction	–3.240	4.558	–0.052	–0.711	0.478

DV Damage Award = Amount of damages auditors should pay Nelson, if auditors were negligent ($0 to $10,000,000); Manipulation of Visualisation Type: Word Cloud visualisation (more emotional response) v. Stacked Bar Graph visualisation (less emotional response); Manipulation of Reliability of Data Source: Reliability of source underlying Big Data visualisation (Company email (more reliable) v. Social media (less reliable))

Table 26 Mediation Moderation

Model	Unstandardised B	Coefficients Std Error	Standardised Coefficients Beta	t	Sig.
Constant	6.479	0.738		8.778	0.000
Visualisation Type	–0.916	0.500	–0.184	–1.832	0.068
Reliability of Data Source	–1.889	1.087	–0.381	–1.738	0.084
Interaction	1.188	0.695	0.200	1.709	0.089
Usefulness	0.008	0.010	0.078	0.815	0.416
Mediated Moderation	0.018	0.015	0.270	1.215	0.226

DV Damage Award = Amount of damages auditors should pay Nelson, if auditors were negligent ($0 to $10,000,000); Manipulation of Visualisation Type: Word Cloud visualisation (more emotional response) v. Stacked Bar Graph visualisation (less emotional response); Manipulation of Reliability of Data Source: Reliability of source underlying Big Data visualisation (Company email (more reliable) v. Social media (less reliable)); Usefulness of visualisation: Juror assessment of usefulness of the visualisation type as responses to the following question, 'In your opinion, how useful was the visualisation of the employee survey?' on an 11-point scale ranging from 0 (Not at all useful) to 100 (very useful).

Chapter Review

The chapter described the statistical tests used in this experiment and the data. It began by seeking outliers and found only one outlier in the experiment. Next, it examined normality and after conducting several tests, found that the results followed a normal distribution. Next, a MANOVA was used to determine if the experiment was effectively randomised across the population. The results show that the sample was effectively randomised, with the exception of age, which trended towards being significant. The chapter then described checks for covariates and found understanding to be a significant covariate in the experiment. The hypotheses were then tested. To further confirm that appropriate theories were used, debriefing questions were asked. Finally, mediation tests were conducted, and it was discovered that usefulness was a mediator of the relationship between visualization type and negligence judgments, which supported the conclusion that if a juror finds the visualisation useful, then they can be persuaded by more emotional visualisations.

Chapter 5
Discussion

Chapter Summary

This study provides evidence that auditors do not face an increased ligation risk with regard to emotional visualisations. Jurors have their own biases with regard to how audit evidence should be presented. Indeed, auditors are unlikely to increase their ligation risk by using less conventional visualisations. This study provides evidence that with regard to the reliability of data, jurors did not consider the source when making a negligence determination or deciding damage awards and, therefore, auditors should be aware that all audit evidence may be given equal weight.

It was initially predicted that visualisations with more emotional effect would cause jurors to find auditors more negligent and to award higher damages in audit negligence lawsuits. The results of this study do not confirm this prediction. In fact, the results are the opposite of what was expected. Jurors find auditors less negligent when visualisations are more emotional (mean = 38.40) than when they are less emotional (mean = 44.49). The reasons for these unexpected results are linked to perceptions of usefulness. People serving on a jury may find newer and more novel forms of displaying information less useful. Although convenient, a word cloud may not be a sophisticated or reliable means of displaying data. Therefore, when a jury sees a word cloud with emotional words and they view it as less useful, they are less likely to find the auditor negligent. They view it as unconvincing and unsophisticated.

To further explore the idea that a person must first find a visualisation useful to be persuaded by it, mediation analysis was conducted using the Barron and Kenny (1986) method of full mediation. The results suggest lower perceptions of audit negligence and slightly lower damage risks when the visualization is perceived as less useful.

In light of these unexpected results, a possible explanation for the phenomenon of people not trusting newer visualisations was sought in the literature; a possible explanation was found in psychology. This experiment used a heuristic approach (word cloud) to display information. Heuristic approaches are ‘mental shortcuts’ that allow one to make decisions quickly. Two heuristic approaches were used to display large amounts of information that was subsequently used as evidence in an audit negligence trial. One type of heuristic approach is known as a ‘representativeness heuristic’ (Tversky and Kahneman 1974). The

representativeness heuristic suggests that people express more confidence when what they are shown matches their stereotype of that object. For example, a group was told that the majority of people in a community are farmers and that there are only a few librarians. Despite that, upon hearing the description of a person who is quiet and introverted, people expressed more confidence that the person was a librarian (Tversky et al. 1974). It is likely that jurors have personal stereotypes that dictate how audit evidence should appear. It is a common stereotype within the accounting profession that accounting information is professional, technical, well-organised and, arguably, 'boring'. Therefore, information displayed in an unconventional, novel and emotional way in audit working papers, does not match a jury's expectation of audit evidence and thus, it is largely or completely disregarded as evidence.

Hypothesis 1b posited that more emotional visualisations will lead to higher damage awards. Once again, it was found that more emotional visualisations led to lower damage awards (more emotional, mean = 6.25; less emotional, mean = 6.58, Table 17). Also, a participant's perception of usefulness was not a mediating factor. Perhaps these findings indicate that participants should be encouraged to use more emotional visualisations because they will lead to slightly lower damage settlements and lower perceptions of audit negligence.

H2a predicted that jurors will find auditors negligent when the data used to create visualisations that contradict auditor decisions are more reliable. No effect on negligence determinations ($p = 0.446$, table 13) was found. This indicates that auditors should be careful about displaying company data because a juror will not consider the reliability of the data when making a negligence determination. That is, data from Facebook will be considered of equal reliability as data taken from sources, like internal email.

The next hypothesis (H2b) predicted that jurors who view visualisations that contradict the auditor's conclusions will pay larger damage awards when the data used to create the visualisation are more reliable. The ANCOVA model in Table 16 was used to test this hypothesis using the damage award as the dependent variable. No significant effect was found ($p = 0.431$); however, a significant disordinal interaction of emotion and reliability was found and suggests that main effects cannot be considered without evaluation of the interaction.

The final interaction hypothesis H3 predicted that jurors will be more influenced by the reliability of a visualisation's data source when the visualisation is more emotional and further, that the emotions of the visualisation will completely overwhelm the juror to the extent that they will not consider data reliability. No significant interactive effect was found between emotional sentiment and reliability on evaluations of negligence ($p = 0.666$; see Table 13). It appears that only emotional sentiment affects the negligence scale. However, with regard to the damage award determinations, the interaction term is significant ($p = 0.018$; see Table 16). Figure 9 shows that the interaction is disordinal, and the effects of reliability are different for the more emotional (word cloud) and less emotional (bar graph) visualisations. Tests of simple effects indicate that jurors respond more to reliability for the more emotional visualisations, which supports H3. When visualisations are more emotional, jurors give larger damage awards ($p = .034$) when data are more reliable (mean =6.97) than when data are less reliable (mean = 5.78). Conversely, when visualisations are less emotional, there is no significant difference in damage award decisions across levels of reliability ($p = .241$).

Conclusion

The results of the experiment were unexpected. This study found that jurors' negligence decisions were influenced by the type of visualisation; however, jurors found auditors more negligent when the visualisation created less emotional responses. It appears that jurors are sceptical of newer and more innovative means of displaying information. Further, the analysis found that juror negligence decisions were not influenced by the strong emotional responses created by visualisations; rather, they were affected by jurors' own determinations of the usefulness of different types of visualisations for evaluating the lawsuit. Their negligence decisions were ultimately driven by usefulness evaluations of the visualisation. Therefore, there is no indication that litigation risks to audit firms will be affected by adopting more emotional visualisations such as word clouds. This experiment does not find that audit firms using emotional visualisations will lead to more legal risk.

The experiment also investigated whether jurors considered the reliability of the data used to create visualisations when making decisions in auditor negligence lawsuits. The experiment found that jurors do not consider data reliability in their determinations of auditor negligence. Further, the result suggests that jurors' negligence judgments can be influenced by visualisations created from unreliable data sources. Even though jurors did not consider

data reliability when determining negligence, those jurors who found auditors negligent did consider data reliability when determining damage awards. Therefore, jurors made considerably larger damage awards when more emotional visualisations were made from more reliable data, compared to when an emotional visualisation was created from less reliable data. When visualisations were less emotional, data reliability did not affect damage award decisions. Thus there could be concerns for audit firms when emotional visualizations come from more reliable sources.

Limitations of the Experiment

Criticism of Amazon Mechanical Turk

Although AMT has been used in many accounting behavioural studies requiring non-professional subjects, such as jurors and non-professional investors, it has limitations. A recent study criticised the use of AMT because AMT participants can produce results that are different from samples obtained using other methods (Brink, Lee and Pyzoha 2018). Samples from the general US population and a population of AMT users were given the World Value Survey and the results showed that AMT values differed from those of the general population: 'For example, the M-Turk population is less religious, more willing to justify unethical behaviour, more politically left leaning, more trusting in others, and places lower importance on hard work and capitalistic values' (Brink, Lee and Pyzoha 2018, 4). However, Amazon Turk has been demonstrated to be a reliable method for examining jury decisions (Grenier et al. 2015).

Other criticism of AMT suggests that its samples are flawed because of the easily manoeuvrable barriers to participation, low pay rates and lack of experimental controls (Kraut et al. 2004). However, researchers of various subjects and backgrounds have validated the responses of AMT for consistency and psychometric properties (Buchheit et al. 2018). For example, it was found that AMT participants in a honesty study were no more dishonest than students and failed fewer attention checks in experimental research (Farrell, Grenier and Leiby 2007). Finally, Thibodeau, Williams, and Witte (2019)show that AMT is an effective tool for sourcing data as long as the correct controls are in place. These controls include using attention check questions to protect the integrity of the data. In my experiment I had such controls in place to protect the data. Furthermore, Thibodeau, et al, (2019) provide support for AMT as an acceptable means of sourcing data by noting that 41 articles were

published in top eight accounting journals between 2011 and 2016 that used on-line labour markets.

Other Limitations

Generalisability is another potential limitation of this experiment. This study was conducted entirely in the US and only examined litigation risk for auditors facing American juries. In less litigious societies or those with different legal systems, results could differ. It is possible that people serving on a jury in different cultures will react differently to the same visualisations. A potential future study could examine the effects of Big Data visualisations in the audit papers on juries cross-culturally.

In addition, as with any controlled experiment, subjects may behave differently in the real world. The act of reading a case and answering questions about how one would respond if serving on jury is a significantly different experience from actually serving on a jury. In serving, jurors are exposed to a host of other stimuli: hearing and seeing the defendant and learning how peers perceive the case. These additional factors could play a role in judgments of auditor negligence. However, to argue that nothing can be learned because people react differently in different contexts is also unfair, as many qualitative colleagues have argued. Making such a bold assertion would suggest that nothing can be learned from any controlled experiment.

Finally, I am aware that the two visualizations do not show the exact same information. The word cloud shows frequency and the bar graph shows time. However, it is important that the experiment mirrors reality and thus a commonly used bar chart visualization was chosen to prevent external validity problems. Future research could compare different frequency charts.

Contributions

This experiment gives auditors, regulators and academics a better understanding of how new evidence sources and new analysis and visualization techniques could influence litigation risk. This study tested the ability of neural affect decision theory and source credibility theory to explain the potential effects of different types of Big Data visualisations (bar graphs and word clouds) based on different data sources (social media and internal emails) on juror decisions. The study is important because the emotional effects that these

visualisations create in jurors have the potential to leave auditors vulnerable to negligence lawsuits. Ultimately, auditors should be aware of whether these new data analysis and visualisation techniques create new litigation risks. My findings suggest that visualisations that create powerful emotional responses are unlikely to increase litigation risk. Jurors are more influenced by their perceptions of the usefulness of visualisations than they are by their emotional effects. This is good news for audit firms because recent research found that the use of emotionally vivid visualisations has significant benefits for auditors and audit quality (Rose et al. 2019).

Potentially increased litigation risk created by new visualisation techniques is a significant concern for audit firms, and this study is the first to address this concern. The study also found that the perceived usefulness of the visualisation and jurors' understanding of the case were important factors in the determination of auditor negligence. Further, this study helps audit teams to build confidence because it shows that how data is presented will have very little influence in convincing a jury of whether or not the auditor is guilty of audit negligence and, therefore, auditor need not worry about data presentation techniques or the reliability of the source, but should focus on juries understanding the data they present.

Suggestions for Future Research

Reproductions of this study may benefit from having a sample group that is not presented with evidence as a control group. This will provide a better indication of the persuasiveness of the evidence used in the jury trial. An international study could also be conducted to see if results vary between different cultures and legal systems.

References

Alles, M. 2015. 'Drivers of the Use and Facilitators and Obstacles of the Evolution of Big Data by the Audit Profession'. *Accounting Horizons* 29 (2): 439–445.

Arel, B., M. Jennings, K. Pany and P. M. J. Reckers. 2012. 'Auditor liability: A comparison of judge and jury verdicts'. *Journal of Accounting and Public Policy* 31 (5): 516–532.

Backof, A. 2015. 'The impact of audit evidence documentation on jurors' negligence verdicts and damage awards'. *The Accounting Review* 90 (6): 2177–2204.

Baron, R. and D. Kenny. 1986. 'The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations'. *Journal of Personality and Social Psychology* 51 (6): 1173–1182.

Birnbaum, M. H. and S. E. Stegner. 1979. 'Source credibility in social judgment: Bias, expertise, and the judges point of view'. *Journal of Personality and Social Psychology* 37: 48–74.

Brasel, K, M. Doxey, J. Grenier and A. Reffett. 2016. 'Risk disclosure preceding negative outcomes: The effects of reporting critical audit matters on judgments of auditor liability'. *The Accounting Review* 91(5): 1345–1362.

Brink, W. L., L. Lee and J. Pyzoha. 2018. 'Values of Participants in Behavioral Accounting Research: A comparison of the M-Turk Population to a Nationally Representative Sample'. *Behavioral Research in Accounting.* In-Press.

Brown, J. O. and V. K. Popova. 2016. 'The Interplay of Client Management and the Audit Committee on Auditor Performance'. *Current Issues in Auditing* 10 (1): 11--17.

Brown-Liburd, H., H. Issa and D. Lombardi. 2015. 'Behavioral implications of Big Data's impact on audit judgment and decision making and future research direction'. *Accounting Horizons* 29 (2): 451–468.

Brown-Liburd, H. and M. A. Vasarhelyi. 2015. 'Big Data and audit evidence'. *Journal of Emerging Technologies in Accounting* 12 (1): 1–16.

Buchheit, H., M. Doxey, T. Pollard and S. Stinson. 2018. 'A Technical Guide to Amazon's Mechanical Turk in Behavioral Accounting Research'. *Behavioral Research in Accounting* 30 (1): 111–120.

Buhrmester, M., T. Kwang and S. D. Gosling. 2011. 'Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data?' *Perspectives on Psychological Science* 6 (1): 3–5.

Cao, M., R. Chychyla and T. Stewart. 2015. 'Big data analytics in financial statement audits'. *Accounting Horizons* 29 (2): 423–427.

Chaiken, S. and D. Maheswaran. 1994. 'Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment'. *Journal of Personality and Social Psychology* 66 (3): 460–473.

Chen, S. and S. Chaiken. 1999. *The heuristic-systematic model in its broader context. In Dual-Process Theories in Social Psychology*. New York, NY: The Guilford Press.

Cukier, K. and V. Mayer-Schoenberger. 2013. 'The rise of big data: how it's changing the way we think about the world'. *Foreign Affairs*, May–June 2003. *Academic OneFile*. http://link.galegroup.com/apps/doc/A329302643/AONE?u=vuw&sid=AONE&xid=7205b6fb.

Daly, T. M. and R. Nataraajan. 2015. 'Swapping bricks for clicks: Crowdsourcing longitudinal data on Amazon Turk'. *Journal of Business Research*, 68 (12): 2603—2609.

Dezoort, F., D. Hermanson and R. Houston. 2003. 'Audit committee member support for proposed audit adjustments: A source credibility perspective'. *Auditing: A Journal of Practice and Theory* 22 (2): 189–205.

Donelson, D. 2013. 'The potential for catastrophic auditor litigation'. *American Law and Economic Review* 15 (1): 333–380.

Eagly, A. H., W. Wood and S. Chaiken. 1978. 'Causal inferences about communicators and their effect on opinion change'. *Journal of Personality and Social Psychology* 36 (4): 424–435.

Eagly, A.H. and S. Chaiken. 1993. *The Psychology of Attitudes*. Fort Worth, TX: Harcourt Brace Jovanovich, Inc.

Farrell, A. J. and J. Leiby. 2017. 'Scoundrels or stars? Theory and evidence on the quality of workers in online labor markets'. *The Accounting Review* 92 (1): 93–114. https://doi.org/10.2308/accr-51447.

Field, A. 2009. *Discovering Statistics using SPSS*, 3rd ed. London: Sage.

Fiske, S. T. and S. E. Taylor. 1984. *Social Cognition*, 1st ed. Reading, MA: Addison-Wesley.

Franzel, J., J. Rose, J. Thibodeau and L. T. Williams. 2018, *The Audit of the Future: A Guide for Researchers and Regulators.* Unpublished paper, Bentley University.

Frasca, K. and M. Edwards. 2017. 'Web-based corporate, social and video recruitment media: Effects of media richness and source credibility on organizational attraction'. *International Journal of Selection and Assessment* 25(2): 125-137.

Gepp, A., M. Linnenluecke, J. Terrence and T. Smith. 2018. 'Big data techniques in auditing research and practice: Current trends and future opportunities', *Journal of Accounting Literature* 40: 102–115.

Grenier, G., B. Pomeroy and A. Reffett. 2012. 'Speak up or shut up? The moderating role of credibility on auditor remedial defence tactics'. *Auditing: A Journal of Practice & Theory* 31 (4): 65–83.

Grenier, J., J. Lowe, A. Reffett and R. Warne. 2015. 'The effects of independent expert recommendations on juror judgments of auditor negligence'. *Auditing: A Journal of Practice & Theory* 34 (4): 157–70.

Hastie, R. 1993. *Inside the Juror: The Psychology of Juror Decision Making*. *Cambridge Series on Judgment and Decision Making.* Cambridge, UK: Cambridge University Press.

Hayes, A. and K. Preacher. 2014 'Statistical mediation analysis with a multicategorical independent variable'. *British Journal of Mathematical and Statistical Psychology* 67(3): 451-470.

Herda, D. and J. Lavelle. 2011. 'The effects of organizational fairness and commitment on the extent of benefits big four alumni provide their former firm'. *Accounting, Organizations and Society* 36 (3): 156–166.

Herda, D. and J. Lavelle. 2012. 'The auditor-audit firm relationship and its effect on burnout and turnover intention'. *Accounting Horizons* 26 (4): 707–723.

Hertwig, R. T. Pachur and S. Kurzenhauser. 2005. 'Judgments of Risk Frequencies: Tests of Possible Cognitive Mechanisms'. *Journal of Experimental Psychology* 31 (4): 621—642.

Issa, H. and A. Kogan. 2014. 'A predictive ordered logistic regression model as a tool for quality review of control risk assessments'. *Journal of Information Systems* 28 (2): 209–229.

Kadous, K. 2000. 'The effects of audit quality and consequence severity on juror evaluations of auditor responsibility for plaintiff losses'. *The Accounting Review* 35 (1): 101–117.

Kadous. K. 2001. 'Improving jurors' evaluations of auditors in negligence cases'. *Contemporary Accounting Research* 18 (3): 425–444.

Kadous, K. and M. Mercer. 2012. 'Can Reporting Norms Create a Safe Harbour? Jury Verdicts against Auditors under Precise and Imprecise Accounting Standards'. *The Accounting Review* 87 (2): 565–587.

Kadous, K. and M. Mercer. 2016. 'Are juries more likely to second-guess auditors under imprecise accounting standards?' *Auditing: A Journal of Practice & Theory* 35 (1): 101–117.

Kaufman, D., M. Stasson and J. Hart. 1999. 'Are tabloids always wrong or is that just what we think? Need for cognition and perceptions of articles in print media'. *Journal of Applied Psychology* 29 (9): 1984–2000.

Keely, J. and M. Edwards. 2017. 'Web-based corporate, social and video recruitment media: Effects of media richness and source credibility on organizational attraction'. *International Journal of Selection and Assessment* 25: 125–137.

Keller, P. A. and L. Block. 1997. 'Vividness effects: A resource-matching perspective'. *Journal of Consumer Research* 24: 295–304.

Kraut, R., J. Olson, M. Banaji, A. Bruckman, J. Cohen and M. Couper. 2004. 'Psychological research online'. *The American Psychologist* 59 (2): 105–117. https://doi.org/10.1037/0003-066X.59.2.105.

Kruikemeier, S. and S. Lecheler. 2016. 'News consumer perceptions of new journalistic sourcing techniques'. *Journalism Studies* (June): 1–18.

Kung, F. N. Kwok, and D. Brown. 2018. 'Are Attention Check Questions a Threat to Scale validity?' *Applied Psychology: An International Review* 67 (2), 264--283.

Lichtenstein, S., P. Slovic, B. Fischhoff, M. Layman and B. Combs. 1978. 'Judged Frequency of Lethal events'. *Journal of Experimental Psychology: Human Learning and Memory* 4 (6): 551–578.

Litt, A., C. Eliasmith and P. Thagard. 2008. 'Neural affective decision theory: Choices, brains, and emotions'. *Cognitive Systems Research* 9(4): 252–273

Lyubimov, A., V. Arnold and S. Sutton. 2013. 'An examination of the legal liability associated with outsourcing and offshoring audit procedures'. *Auditing: A Journal of Practice and Theory* 32 (2): 97–118.

Martin, R. 2015. 'Jury Duty: Who Gets Called, And Who Actually Serves'. *NPR*. https://www.npr.org/2015/06/07/412633577/jury-duty-who-gets-called-and-who-actually-serves.

Mason, W. and S. Suri. 2012. 'Conducting Behavioral research on Amazon's Mechanical Turk'. *Behavior Research Methods* 44 (1):1–23.

McCroskey, J. 1966. 'Scales for the Measurement of Ethos'. *Speech Monographs* 33: 65--72.

McGinnies, E. and C. Ward. 1980. 'Better liked than right: Trustworthiness and expertise as factors in credibility'. *Personality and Social Psychology Bulletin* 6: 467–472.

Moffitt, K. and M. A. Vasarhelyi. 2013. 'AIS in an age of Big Data'. *Journal of Information Systems* 27 (2): 1–19.

Muller, D., M. Judd and V. Yzerbyt. 2005. 'When Moderation Is Mediated and Mediation is Moderated'. *Journal of Personality and Social Science* 89 (6): 852--863.

Nisbett, R. and L. Ross. 1980. *Human Inference Strategies and Shortcomings of Social Judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Palmrose, Z.-V. 2006. 'Maintaining the value and viability of independent auditors as gatekeepers under SOX: An auditing masters proposal'. In *Financial Gatekeepers: Can They Protect Investors*, edited by Y. Fuchita and R. Litan. Baltimore, MD: Brookings Institute Press and Nomura Institute of Capital Markets Research.

Pornpitakpan, C. 2004. 'The persuasiveness of source credibility: A critical review of five decades of evidence'. *Journal of Applied Social Psychology* 34 (3): 243–281.

O'Donnell, E. and J. D. Perkins. 2011. 'Assessing risk with analytical procedures: Do systems thinking tools help auditors focus on diagnostic patterns?' *Auditing: A Journal of Practice & Theory* 30 (4): 273–282.

Public Company Accounting Oversight Board (PCAOB). 2012. 'Consideration of Outreach and Research Regarding the Auditor's Approach to Detecting Fraud'. Standing Advisory Group Meeting, 15-16 November. http://pcaobus.org/News/Events/Documents/11152012_SAGMeeting/2012_11_15_%20SAG_BP_%20Fraud.pdf.

Rasso, J. 2014. 'Apology accepted: The benefits of an apology for a deficient audit following an audit failure'. *Auditing: A Journal of Practice & Theory* 33 (1): 161--176.

Reffett, A. B. 2010. 'Can identifying and investigating fraud risks increase auditors' liability?' *The Accounting Review* 85 (6): 2145-2167.

Reffett, A. B. 2011. 'No good deed goes unpunished? Recent evidence on the effects of identifying and investigating fraud risks on auditors' litigation exposure'. *Current Issues in Auditing* 5 (2): 1-8.

Reffett, A., B. Brester and B. Ballou. 2012. 'Comparing auditor versus non-auditor assessments of auditor liability: An experimental investigation of experts' versus

lay evaluators' judgments'. *Auditing: A Journal of Practice & Theory* 31 (3): 125--148.

Rose, A., J. Rose, K. Sanderson and J. Thibodeau. 2017. 'When should audit firms introduce analyses of big data into the audit process?' (Working paper, University of Waikato).

Russom, P. 2011. Big Data Analytics. *TDWI Best Practices Report*, Fourth Quarter.

Shaw, J. C., E. Wild and J. A. Colquitt. 2003. 'To justify or excuse? A meta-analytic review of the effects of explanations'. *Journal of Applied Psychology* 88 (3): 444--458.

Shedler, J. and M. Manis. 1986. 'Can the Availability Heuristic Explain Vividness Effects?' *Journal of Personality and Social Psychology* 39 (1): 2-21.

Syed, A., K. Gillela and C. Venugopal. 2013. 'The future revolution on Big Data'. *International Journal of Advanced Research in Computer and Communication Engineering* 2 (6): 2446–2451.

Stock, D. and C. Watson. 1984. 'Human judgment accuracy, multidimensional graphics, and human versus models'. *Journal of Accounting Research* 22 (1): 192-206.

Thibodeau, J., T. Williams and A. Witte. 2019. 'Point and Click Data: An Assessment of Editorial Perceptions and Recommendations for the Peer-Review Process in the New Data Frontier'. *Journal of Information Systems* 33(1): 129-144.

Tversky, A. and D. Kahnemen. 1974. 'Judgement under Uncertainty: Heuristics and Biases'. *Science* 185 (4157): 1124-1131.

Tversky, A. and D. Kahneman. 1981. 'The framing of decisions and the psychology of choice'. *Science* 211: 453–458.

Tversky, A. and D. Kahneman. 1986. 'Rational choice and the framing of decisions'. *Journal of Business* 59: 251–278.

Yoon, K., L. Hoogduin, L. Zhang. 2015. Big Data as complementary audit evidence. *Accounting Horizons* 29 (2): 431–438.

Appendix A
Auditor Negligence Literature

	Purpose	**Method**	**Independent Variable(s)**	**Dependent Variable(s)**	**Findings**	**Further Study**
Backof (2015)	To determine how documentation affects jurors' decisions	2 x 2 factorial design study	Alternative accounting treatments (no alternatives/ alternatives present) Risk-based approach (risks linked to procedures/ risks not linked to procedures)	The likelihood the audit firm was negligent (0–10) Was negligent (yes/no) Damages (0--13 million) Extent the audit firm caused the plaintiff loss (0--10) Whether the firm intended to conduct a risk-based audit (0--10)	Auditors' documentation of their accounting alternatives increases the likelihood that auditors are found negligent. When documentation that explicitly links the audit risks to the work performed to address each risk is furnished, jurors award lower damages because they perceive auditors' actions prior to the	Examine how differences in audit documentation affect auditors' ability to defend their work to different evaluators. Examine how tailoring expert witnesses' testimonies and attorneys' arguments to each experimental condition affects jurors' negligence verdicts and damage awards and could benefit auditors.

					negligent act as more compliant with the auditing standards.	
Grenier et al. (2015)	To examine how independent experts' recommendations affect jurors' judgments, and inform concerns that jurors will rely too heavily on independent experts' recommendations	3 x 2 between-participants experiment	Whether participants are provided with recommendations from an independent panel (control/ treatment) Specific case facts (the audit firm is negligent/not negligent) The auditors' use of a specialist (no specialist/specialist	Was Negligent (Yes/No) How negligent (0–100) The amount of compensatory damages (0--10 million) The amount of punitive damages (0--10 million)	The results support the utilisation of independent expert recommendations in cases of alleged professional negligence. Jurors will rubber stamp independent expert recommendations without sufficient consideration of specific case facts.	Examine how to ensure that independent experts appear to jurors to be unbiased and credible.
Grenier, Pomeroy and Reffett (2012)	To examine how perceptions of the credibility of remedial defence tactics affect jurors' assessments of	3 x 2 experiment	Remedial tactics (local tactic/ national tactic/ control) Client Importance (Low/High)	Compensatory damages (0--10 million) Punitive damages (0--10 million	Remedial tactics result in lower negligence assessments when such tactics are perceived as credible, but	Investigate other conditions beside the client being of high economic importance, which impair the national office's credibility.

	auditor negligence in cases of undetected fraud			Likelihood of negligence on an 11-point Likert scale Various process-related variables	backfire when perceived as not credible. Remedial tactics result in lower negligence assessments when client importance is low and the tactics are implemented by a national office—but result in higher negligence assessments when client importance is high, regardless of the tactic's source(local v. national).	Identify the effects of remedial tactics using unambiguous normative benchmarks. Investigate other independence threats, such as long auditor tenure and provision of non-audit services.
Kadous (2000)	To investigate whether the level of audit quality provided affects auditor's chances of being held liable for losses associated with	2 x 2 full-factorial experiment	Audit quality (High/Low) Consequence severity (moderate/severe)	Guilty of negligence (0--10) Innocent of negligence (0--10)	Participants serving in the role of juror assessed higher standards of care for auditors when the consequences of audit failure	Determine whether the results hold in conditions with extremely high audit quality. Incorporate additional factors expected to affect juror evaluation of auditors,

	subsequent audit failures				were more severe.	such as source of professional standards, the reputation of the auditor, whether the misstatement is caused by auditor error or client fraud, and, if by fraud, the type of fraud.
Kadous and Mercer (2012)	To provide empirical evidence of the effects of accounting standards precision on jury verdicts	2 x 2 x 2 between-participants experiment	Standard Precision (Precise/ Imprecise) Aggressiveness of client reporting (More Aggressive/ Less Aggressive) Accounting Norm (Consistent /Inconsistent)	The appropriateness of the auditor's action (nine-point scale) The quality of audit work (nine-point scale)	Under precise standards, juries follow the decision rule implicit in the standard. When standards are imprecise, juries turn to reporting norms for a decision. When the audit client's reports are less aggressive and comply with precise standards, few juries rule against the auditor.	None mentioned

					More juries rule against the auditor under the imprecise standard, but only when reporting is more aggressive than the reporting norm. When the client's accounting is more aggressive and violates the precise standard, jurors are less likely to return verdicts against auditors under imprecise standards than under precise standards.	
Kadous and Mercer (2016)	To understand if juries are more likely to second-guess auditors when accounting standards are not precise	2 x 2 between-participants experiment	Precision of the accounting guidance standard (Precise/ Imprecise) Aggressiveness of the client's	Was negligent (yes/no) Appropriateness of the audit firms' action	The results were mixed. When the client's conservative reporting complied with a precise standard,	Examine aggressive reporting that technically complies with audit standards.

			reporting choice (Aggressive/ Conservative)	(nine-point scale) Audit quality (nine-point scale) Firm's responsibility for investor losses (nine-point Likert scale)	most juries did not find the auditor negligent. However, half of the juries in the study found the auditor negligent for allowing the same conservative accounting reporting under an imprecise standard.	
Lyubimov, Arnold and Sutton (2013)	To examine the effect of outsourcing and offshoring on auditor legal liability	2 x 2 between-participants experiment	Sourcing (insourced to own firm/outsourced to a separate firm) Location (onshore /offshore)	Monetary assessments of compensatory and punitive damages to the plaintiff. (0--4 million)	Jurors award greater compensatory damages when work is outsourced. For punitive damages, outsourcing is not significant but the interaction between outsourcing and offshoring is significant with	Examining the effect of outsourcing and offshoring on the party perception of auditor professionalism and the potential related effects on the views of governing bodies on auditor self-regulation and other aspects of professionalism.

					the highest punitive damages awarded when the work is outsourced.	
Reffett (2010)	To provide empirical evidence that identifying and investigating fraud risks can increase auditors' exposure to litigation	Between-participant experiment	Fraud risk Investigation (No Investigation/ Low Investigation/ High Investigation)	Was Negligent (Yes/No) Probability of negligence (0--100) Damages awarded to the plaintiff (0--9 million)	Identifying fraud risk can increase an auditor's litigation exposure.	Examine the defence tactic of making jurors aware of their tendency to unintentionally punish auditors for having investigated fraud.
Reffett, Brewster, and Ballou (2012)	To provide empirical evidence that the judgments of lay evaluators differ from those of auditor evaluators	2 x 2 full-factorial experiment	Audit Quality (High/ Low) Professional Background (auditor evaluator/ lay evaluators)	Auditor negligence (0--100) Negligent (yes/no) Damage (0--9 million)	Auditor evaluators provide lower assessment of auditor liability than lay auditors, regardless of the audit quality.	Examine the costs and benefits of utilising panels of court-appointed auditors and lay evaluators. Examine whether mixed groups of lay auditors and audit experts make better judgments of auditor liability.

Appendix B
Full Instrument for the Experiment

RESEARCH PARTICIPANT CONSENT

TITLE OF THE STUDY: Perceptions of Auditor Negligence

DESCRIPTION OF THE RESEARCH
You are invited to take part in a research study that examines the thoughts of the public about financial statement auditor negligence. You are invited to take part because you have similar skills and experiences as people who serve on juries in the United States. Completion of the study will require about 20 minutes of your time during one day.

WHAT WILL MY PARTICIPATION INVOLVE?
You will be asked to complete a short survey to find out whether you qualify to take part in this study. If you qualify and you choose to take part, you will be asked to pretend you are a juror in a legal case where an audit firm is sued for being negligent in performing a financial statement audit. Our primary interest in this study is to obtain your thoughts about whether the auditors were negligent in doing their job. There are no right or wrong answers.

ARE THERE ANY RISKS TO ME?
We do not anticipate any risks to you from participation in this study.

VOLUNTARY NATURE OF PARTICIPATION
Your participation is completely voluntary. If you decide not to participate or to withdraw from the study, there is no penalty.

HOW WILL MY CONFIDENTIALITY BE PROTECTED?
Data collected in this study is confidential. We do not gather any personally identifiable information. The findings will be summarized and reported in group form.

NOTE: Participation in this research indicates your consent and agreement with the aforementioned description of this study.

Study Qualification Questions

Individuals must answer these questions to determine their eligibility to participate in this research study. Only eligible participants will be invited to complete the research study. Please read the questions and select the most suitable option for each question below.

1. Are you eligible to serve on a jury in the United States? To be eligible you must:
   - be a US citizen, **AND**
   - be at least 18 years old, **AND**
   - have not been convicted of a felony.

   [ ] Yes [ ] No

2. Library is to book as book is to?
   a. page
   b. copy
   c. binding
   d. cover

3. How many college level accounting courses have you completed?
   - ❑ 0
   - ❑ 1 - 2
   - ❑ 3 – 5
   - ❑ 6+

INFORMATION ABOUT FINANCIAL STATEMENTS

Many people, such as lenders (e.g., banks) and investors, use the information in a company's financial statements to make decisions about the company. For example, potential lenders use the company's financial statements to assess the financial 'health' of the company when deciding whether or not to issue a loan.

Financial statements give information about a company's assets and liabilities at a particular point in time. Assets are items of value that a company owns. For example, cash, inventory, property and equipment are all assets. Liabilities are debts that the company owes to other parties. For example, if a company borrowed $1 million from a bank, the company would have a liability for $1 million until it repaid the money to the bank.

A company that increases the total amount of assets owned relative to the amount of liabilities owed improves its financial position and decreases the risk that it will become bankrupt. Alternatively, a company that increases its liabilities relative to assets owned increases its risk of bankruptcy. In other words, the financial 'health' or value of a company increases as the assets increase above and beyond its liabilities. Generally, the lower the risk of bankruptcy the more attractive a company is to lenders and investors. Therefore, it is very important for a company's financial statements to reasonably portray the company's financial situation.

REVIEW QUESTIONS
TO ANSWER THESE QUESTIONS, YOU MAY REVIEW THE INFORMATION IN THE PREVIOUS PARAGRAPHS.
Check the correct answer.

1. Lenders and investors rely on the information in a company's financial statements when deciding whether to loan money to, or invest in that company.

   - ❑ True
   - ❑ False

2. If a company's financial statements do not portray the true health or value of the company, it can cause lenders and investors to make poor decisions.

   - ❑ True
   - ❑ False

THE ROLE OF EXTERNAL AUDITORS

Companies hire external auditors (i.e., accounting professionals from outside of the company) to report on the accuracy of their financial statements. Having external auditors report on the accuracy of financial statements increases the credibility of the financial statements, and the confidence that users (such as lenders and investors) have in relying on the financial statements.

Financial statements can be materially misstated. A material misstatement is one that is large enough to change the decisions of financial statement users such as lenders. Material misstatements occur when the amounts reported in the financial statements are inaccurate. Material misstatements can result from unintentional errors and/or intentional (fraudulent) misrepresentations made by the company's management.

External auditors test for material misstatements by examining the financial records for the company's assets and liabilities to determine if the information in the financial statements is materially accurate. Auditors issue a report, called an opinion, which states whether a company's financial statements are materially accurate. Auditors issue unqualified or 'clean' audit opinions when they judge the company's financial statements to be materially accurate. An unqualified opinion gives lenders, investors, and other users of the financial statements reasonable assurance or increased confidence that the financial statements are free from material misstatements. Auditors issue adverse opinions when they believe the financial statements have material misstatements. 'Audit failures' occur when auditors issue a 'clean' opinion for financial statements that have material misstatements.

Reasonable assurance is not the same as absolute assurance. So, even after auditors perform a good audit that does not reveal a material misstatement, a company's financial statements could still contain a material misstatement. In other words, 'perfect' audits, do not exist. So, a good audit is one performed by auditors who exercise due professional care. Auditors exercise due professional care by being as careful and competent as any other auditor would in following the auditing guidance issued by regulators and audit standard setters. Exercising due professional care reduces the probability of an audit failure but no auditor can get the risk of an audit failure down to a zero probability.

REVIEW QUESTIONS
YOU MAY REVIEW THE INFORMATION ABOVE WHEN ANSWERING THESE QUESTIONS.
Check the correct answer.

3. External auditors are hired by the company to perform tests and provide reasonable assurance (confidence) that a company's financial statements are free from material misstatements.

   - ❑ True
   - ❑ False

4. The probability of not catching material misstatements in the financial statements will be HIGHER if the auditor exercises due professional care.

   - ❑ True
   - ❑ False

PART 1

Please assume that you have been selected to serve on a jury that is evaluating a case related to the audit of Absolute Corporation. Please read the following information carefully and then answer the questions that follow

BACKGROUND INFORMATION
Absolute is a publicly traded corporation that provides management services to various not-for-profit entities, such as health clinics and counseling, education, and social welfare organizations. Absolute's internal staff provides a wide range of services for its clients. Absolute's revenue consists primarily of management fees. Absolute has sales offices in most major cities in the United States, and its fiscal year end is December 31.

COMPLAINT
The plaintiff, Nelson Inc., alleges that the defendant, the accounting firm known as Smith and Larson LLC, was negligent in performing its audit of the 2016 financial statements of Absolute, Inc.

ANSWER
The defendant, Smith and Larson LLC, responds that it complied with professional auditing standards and was not negligent.

PLAINTIFF'S ATTORNEY OPENING STATEMENT
This case is about auditor negligence. The defendant, Smith and Larson LLC, audited Absolute's 2016 financial statements. My client, Nelson Inc, received and relied on Absolute's audited 2016 financial statements, but later found out that those audited financial statements were misstated. Smith and Larson's negligence in the conduct of its audit of Absolute cost Nelson Inc. $10 million dollars in a loan to Absolute.

It is my job to prove to you, on behalf of Nelson, Inc., that Smith and Larson LLC was negligent in its performance of the audit of Absolute's 2016 financial statements. Smith and Larson reported that the 2016 financial statements of Absolute were not materially misstated. In other words, Smith and Larson gave Absolute a 'clean' report. However, after the audited financial statements were released, it came to light that Absolute's financial statements listed revenues that were $4 million too high. Smith and Larson LLC failed to find this inaccuracy because the auditors did not perform an audit of sufficient quality; that is, they did not exercise the same degree of care in their conduct of the audit that other auditors in their position would have used. The $4 million overstatement of revenue hid Absolute's financial problems from Nelson Inc., and from others as well. Nelson Inc. relied on the misstated financial statements when it decided to loan Absolute $10 million.

Absolute's financial problems came to light, Absolute declared bankruptcy. The company closed and all 115 employees lost their jobs. Because of the terrible job market, many of them are still unemployed now, and others had to accept far less attractive jobs. The stockholders of Absolute suffered large losses when the company declared bankruptcy, and my client has received nothing in return for its loan to Absolute. Nelson, Inc. feels that the

auditor, who negligently allowed the overstatement of revenue to occur, should reimburse it for its loss.

After I present my case, the defense will present its case. The defense will claim Smith and Larson satisfied professional auditing standards with the work that it did and that Smith and Larson's judgments were reasonable given the facts available at the time of the audit. Consider carefully whether you believe this to be true and whether the auditors performed their duties in an appropriate manner in this particular case. I am confident that after weighing the evidence you will find for the plaintiff as Smith and Larson LLC was negligent in performing its audit of the 2016 financial statements of Absolute and my client suffered as a result of Smith and Larson's negligence.

DEFENSE'S ATTORNEY OPENING STATEMENT

The plaintiff has alleged that my client, Smith and Larson LLC, was negligent in its audit of Absolute's 2016 financial statements. The plaintiff makes a point of mentioning the loss of his client, Nelson, Inc. That loss is not relevant in determining whether Smith and Larson LLC was negligent in performing its audit of Absolute's 2016 financial statements. Only the actions and decisions made by Smith and Larson, as compared with those that would have been made by other competent Certified Public Accountants (CPAs) in similar circumstances, are relevant. Further, the losses of parties other than Nelson, Inc. are not relevant to this trial.

Negligence can be established only when an auditor fails to exercise the usual judgment, care, skill, and diligence employed by other CPAs in the community. CPAs use the guidance provided in the professional auditing standards to plan and perform their audit work, but the professional auditing standards also require auditors to use their professional judgment throughout an audit. According to the professional auditing standards, auditors must plan and conduct an audit so that they can provide reasonable assurance that the financial statements are free of material misstatements. In other words, an audit conducted in accordance with the professional auditing standards reduces, but cannot completely eliminate, the chance that people receive misstated financial statements. It is the defense's position that if an auditor complies with professional auditing standards and makes reasonable professional judgments given the facts available at the time of the audit, he or she has not been negligent. It is my job to prove to you that Smith and Larson LLC did just that. I will present evidence that proves that Smith and Larson conducted a quality audit in accordance with the auditing standards and used good professional judgment when evaluating the facts available at the time of the audit.

The plaintiff must prove its allegations that Smith and Larson LLC was negligent by a preponderance of the evidence. This means that it must show that the charges are more probably true than not true. The plaintiff cannot do so. The audit evidence proves that Smith and Larson conducted a quality audit, made a reasonable professional judgment, and in no way violated professional auditing standards. Smith and Larson LLC is a competent, esteemed accounting firm, and I am confident that you will find in its favor.

EVIDENCE PRESENTED BY PLAINTIFF'S ATTORNEY

Effective April 1, 2016, Absolute signed a five-year contract to provide management services to Omega. For this contract, the services include human resources, accounting and administration, and sales and marketing. Since Omega is providing Absolute with significant authority to take action in these service areas when needed, the contract with Omega provided substantial performance incentives to Absolute. This incentive compensation is in addition to the established annual management fee from Omega to Absolute.

According to the current accounting standards, Absolute should record revenue in the amount it expects to receive for satisfying the obligations under its contract at the end of each quarter. To do so, Absolute must estimate the amount of revenue it is most likely to realize for performance incentives.

According to the contract, Absolute can earn up to $4 million each quarter if employee turnover rates remain below certain thresholds at the end of each quarter. That is, Absolute could earn a very large bonus if Omega employees were content and remained with Omega. Using historical results for Omega as well as current expectations, Absolute estimated the chances of achieving this performance incentive as follows:

Employee Turnover Rates	Performance Bonus	% Chance of Achieving
Below 5%	$4 million	90%
Between 5% and 6%	$2 million	7.5%
6% or above	$0	2.5%

Based on these expectations, Absolute recorded $4 million in performance incentive revenue for the fourth quarter.

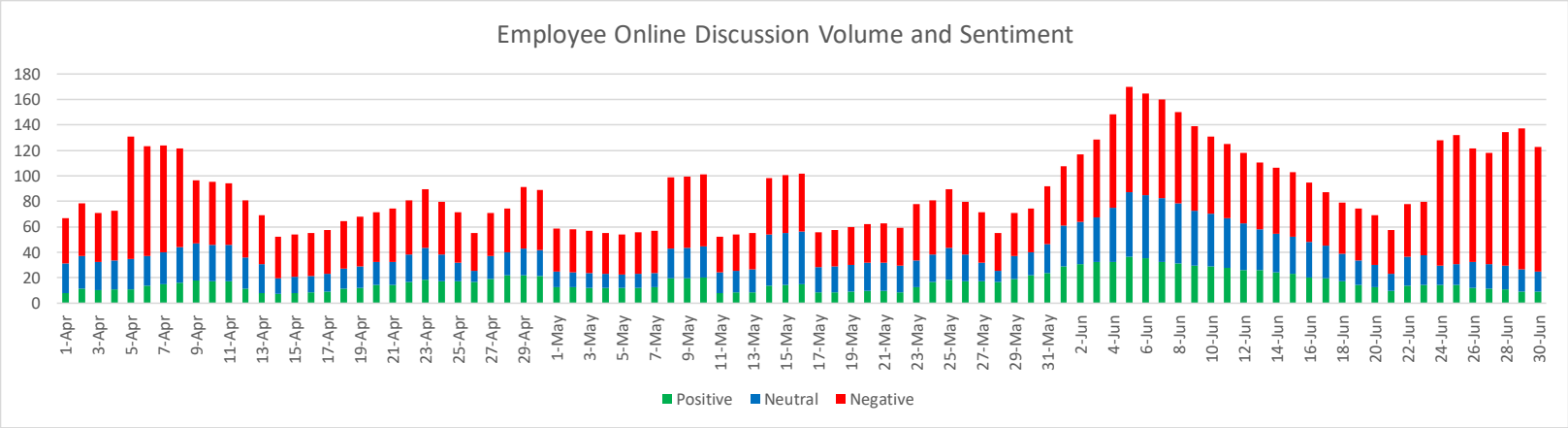
The defendant, Smith and Larson LLC, analyzed data from social networking sites in order to seek information that would confirm or disconfirm Absolute's estimate of revenue. One of the most critical visualisations is presented below. What is very clear in this graphic is that employee morale was low, and not high as Absolute claimed when it estimated a 90% chance that employee turnover would be extremely low. The evidence collected and prepared by Smith and Larson clearly shows that they should have questioned the recording of $4 million in revenue and demonstrates that Smith and Larson LLC were negligent and failed to satisfy professional auditing standards when it ignored this evidence. As a result of this negligence, Nelson Inc. lost $10 million in a loan to Absolute.

For reviewers: The experimental manipulations appear on the following pages. Manipulation 1 is the type of visualisation (graph versus word cloud). This manipulation involves the emotional impact of the visualisation. The word cloud has a stronger emotional impact. Manipulation 2 is the source of information for the visualisation (social media or the client's email). The visualisations of sentiment provide evidence that disconfirms the assumption by Absolute that morale is high and turnover is expected to be low. The visualisations represent two ways to display

the same data, and we use the same underlying sentiment data to create the two different visualisations.

Treatment 1 [Bar Chart / social media]

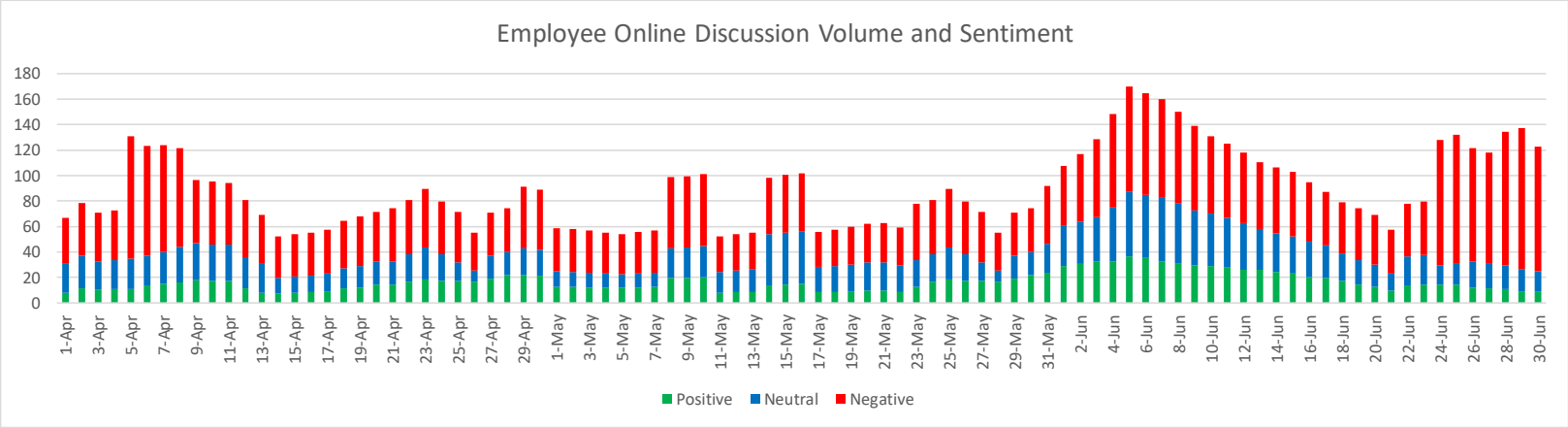
Smith and Larson LLC analyzed social media postings that are believed to have been made by Omega employees and conducted text sentiment analyses on these postings to determine whether the words in the postings represented a positive attitude towards Absolute, a negative attitude towards Absolute, or a neutral attitude. The visualisation presents the volume and sentiment of online discussions related to Absolute during the fourth quarter on social media sites such as Twitter, Facebook, and Google+.



Source	Mentions	Positive	Positive%	Neutral	Neutral%	Negative	Negative%	Net Sentiment
Total	8117	1496	18%	2130	26%	4492	55%	-37%

Treatment 2 [Bar Chart / email]

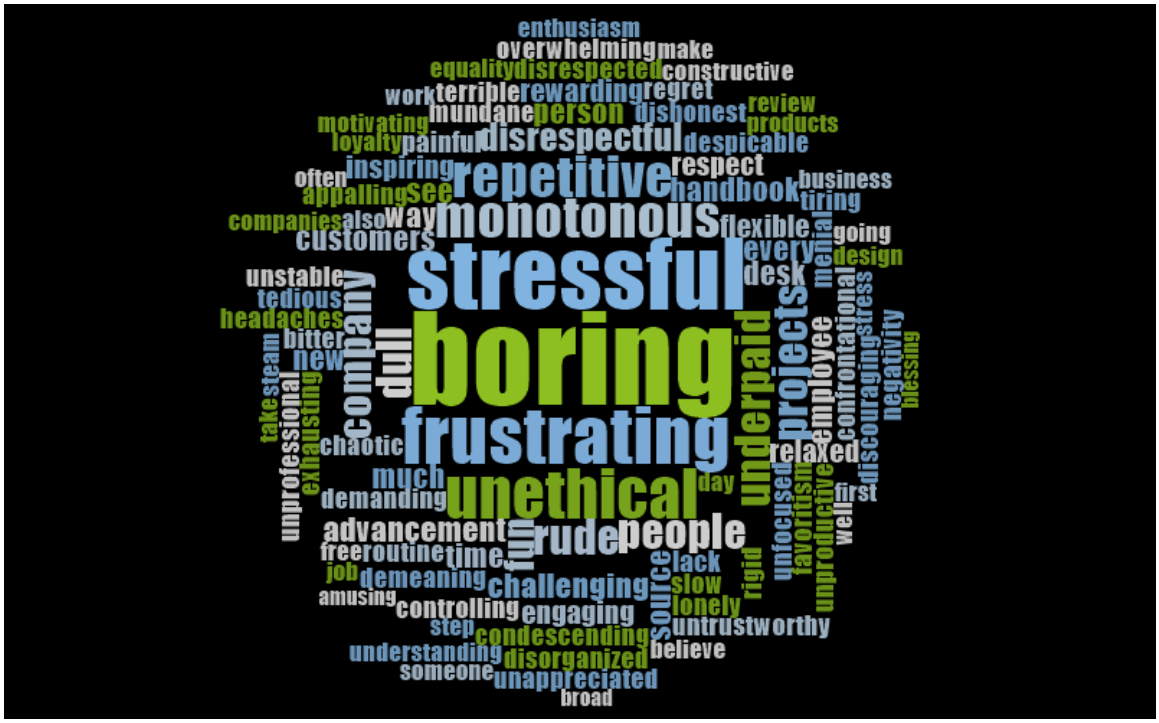
Smith and Larson LLC analyzed emails sent by Omega employees to the Omega Human Resources department and upper management and conducted text sentiment analyses on these emails to determine whether the words in the emails represented a positive attitude towards Absolute, a negative attitude towards Absolute, or a neutral attitude. The visualisation presents the volume and sentiment of emails related to Absolute during the fourth quarter.



Source	Mentions	Positive	Positive%	Neutral	Neutral%	Negative	Negative%	Net Sentiment
Total	8117	1496	18%	2130	26%	4492	55%	-37%

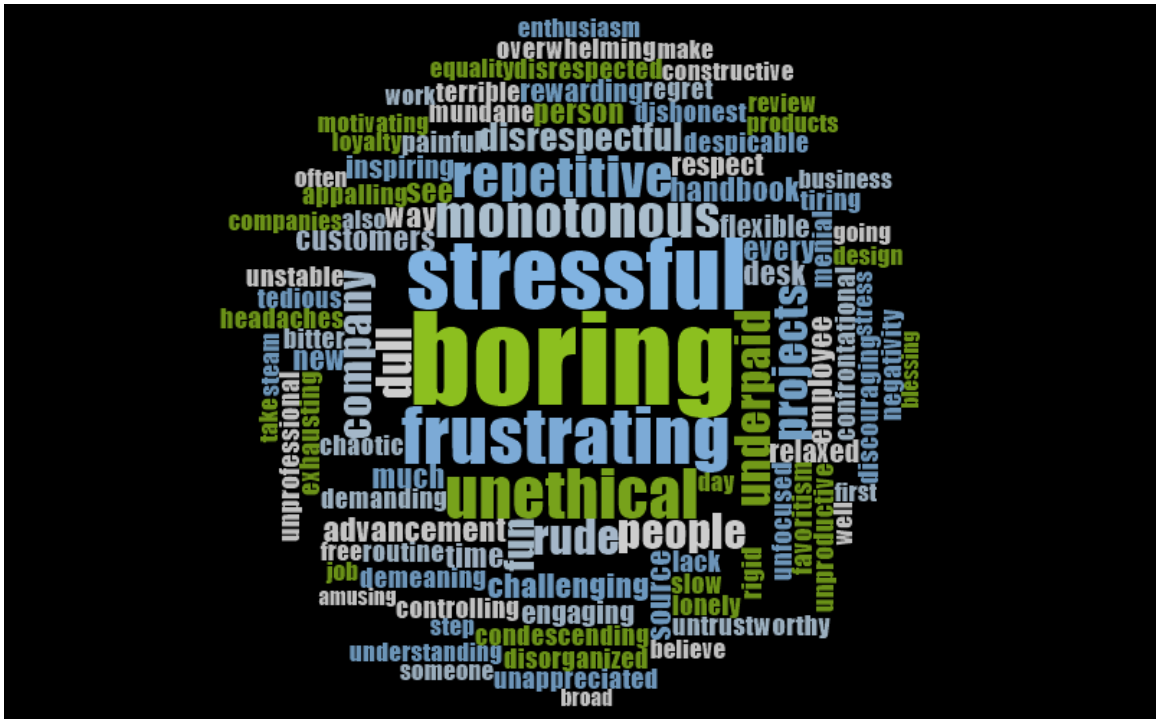
Treatment 3 [word cloud / social media]

Smith and Larson LLC analyzed social media postings that are believed to have been made by Omega employees and conducted text sentiment analyses on these postings to determine whether the words in the postings represented a positive attitude towards Absolute, a negative attitude towards Absolute, or a neutral attitude. The visualisation presents the words most commonly used on social media social media sites such as Twitter, Facebook, and Google+. Words that appear more often are larger, and words often used together are closer to each other.



Treatment 4 [word cloud / email]

Smith and Larson LLC analyzed emails sent by Omega employees to the Omega Human Resources department and upper management and conducted text sentiment analyses on these emails to determine whether the words in the emails represented a positive attitude towards Absolute, a negative attitude towards Absolute, or a neutral attitude. The visualisation presents the words most commonly used in the emails. Words that appear more often are larger, and words often used together are closer to each other.



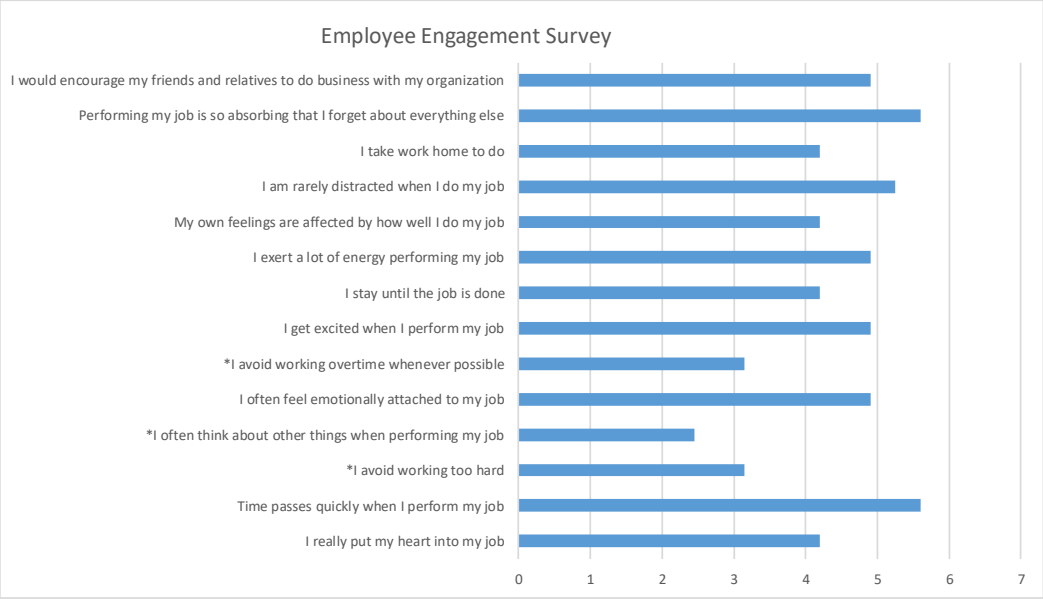
EVIDENCE PRESENTED BY DEFENSE'S ATTORNEY

Smith and Larson LLC evaluated the estimation methods employed by Absolute to estimate Omega's likely employee turnover ratios and found that they were consistent with models used in the industry. As described by the defense, Smith and Larson also reviewed information available from their data analytics and visualisation team to determine whether the assumptions underlying management estimates and the 2016 revenue recognized for the new Omega contract were appropriate. This demonstrates that Smith and Larson LLC went above and beyond the professional standards because examination of social media is not required by the standards, and this was done as additional testing.

Importantly, the plaintiff, Nelson Inc., failed to mention that the visualisation of sentiment was only one of many pieces of evidence examined. Absolute based its estimate of the likelihood of turnover on measures of employee satisfaction. The engagement survey below was administered by Absolute during the third quarter of 2016 to investigate employee morale at Omega. Approximately 35% of the employee workforce responded to the survey. Smith and Larson LLC examined these survey results as part of the 2016 financial audit.

Employee Survey Results

A response of 0 = Definitely Not True
A response of 7 = Definitely True



Absolute's survey of Omega's employees clearly shows that employee morale was extremely high, and employee turnover was very unlikely. This supports the amount of revenue that was recorded by Absolute. The visualisation of employee sentiment involves far less reliable and less verifiable data than the direct survey measure of employee satisfaction. Smith and Larson LLC made the same determination that all auditors would make under these circumstances and followed the more reliable audit evidence. There are

often conflicts between different sources of evidence, and auditors must use their professional judgment when deciding which evidence is more reliable and meaningful. Smith and Larson LLC made the best decision given the information that was available, and they followed and even exceeded all professional standards.

Nelson, Inc. v. Smith and Larson LLC
Case Summary

Complaint: The plaintiff, Nelson, Inc., alleges that the defendant, Smith and Larson LLC, was negligent in performing its audit of the 2016 financial statements of Absolute, Inc.

Answer: The defendant, Smith and Larson LLC, responds that it complied with auditing standards and that therefore it was not negligent.

Summary of Jury Instructions: It is your responsibility to determine whether or not Smith and Larson LLC were negligent based on the evidence presented to you during the trial. Auditors are required to use the same judgment, care, skill, and diligence employed by other CPAs in the community. Auditors comply with this standard of care by adhering to the professional auditing standards. Therefore, you should consider whether the defendant complied with professional auditing standards in making your evaluation. If you believe that the evidence suggests that for the most part Smith and Larson LLC acted as other CPAs would have given the same circumstances, then you should conclude that Smith and Larson is not guilty of negligence. On the other hand, if you decide that the majority of evidence suggests that Smith and Larson did not act as other CPAs would have given the same circumstances, you should conclude that Smith and Larson LLC is guilty of negligence.

QUESTIONS

How likely it is that Smith and Larson LLC was negligent in performing the audit? (circle a number on the scale to indicate our response)

0	10	20	30	40	50	60	70	80	90	100
Not Likely At All										Extremely Likely

You will now cast a vote for the verdict. If the jury took a poll before deliberations, how would you vote? (check one response to indicate your response)

_____ Smith and Larson LLC was **negligent**

_____ Smith and Larson LLC was **not negligent**

If you voted negligent, indicate the amount of damages (from $0 to $13 million) that you would require Smith and Larson LLC to pay to Nelson, Inc., the plaintiff. Please write an amount in the range of $0 to $13,000,000 to indicate your response. If you voted not negligent, please leave this line blank.

$_____

PART 2

How do you feel about the defendant (the auditing firm Smith and Larson LLC)? (circle a number on the scale to indicate your response)

-5 -4 -3 -2 -1 0 +1 +2 +3 +4 +5

Very Negative Feelings — Very Positive Feelings

How do you feel about the plaintiff, Nelson Inc.? (circle a number on the scale to indicate our response)

-5 -4 -3 -2 -1 0 +1 +2 +3 +4 +5

Very Negative Feelings — Very Positive Feelings

How do you feel about Absolute Corporation? (circle a number on the scale to indicate our response)

-5 -4 -3 -2 -1 0 +1 +2 +3 +4 +5

Very Negative Feelings — Very Positive Feelings

Did you understand the case and questionnaire?

0 50 100

Not At All — Completely

1) **Do you believe that the Omega employees were happy?**

0 50 100

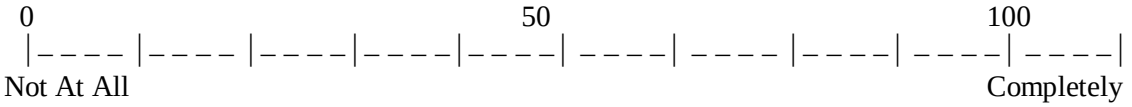
Not At All — Completely

2) **Do you believe that the Omega employees were discouraged?**

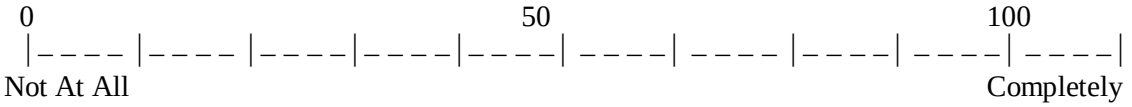
0 50 100

Not At All — Completely

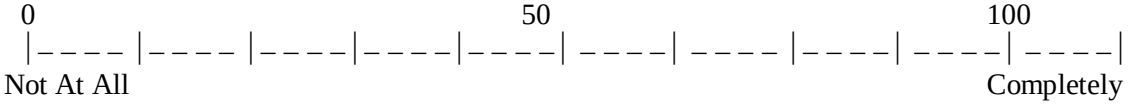
3) **Do you believe that the Omega employees were angry?**



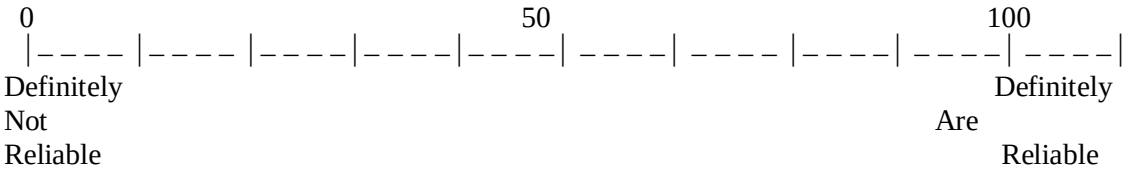
4) **Do you believe that the Omega employees were frustrated?**



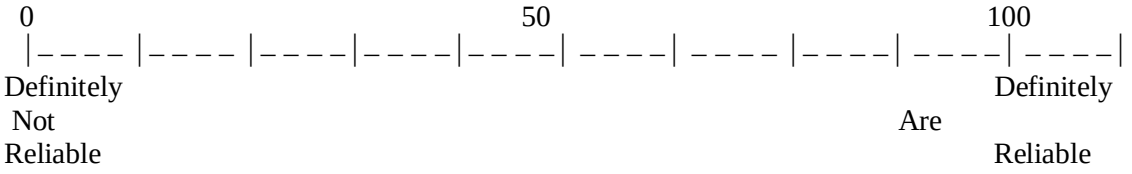
5) **Do you believe that the Omega employees were depressed?**



6) **Do you believe that social media postings about Absolute are reliable sources of information about employees' intentions to remain with the firm?**



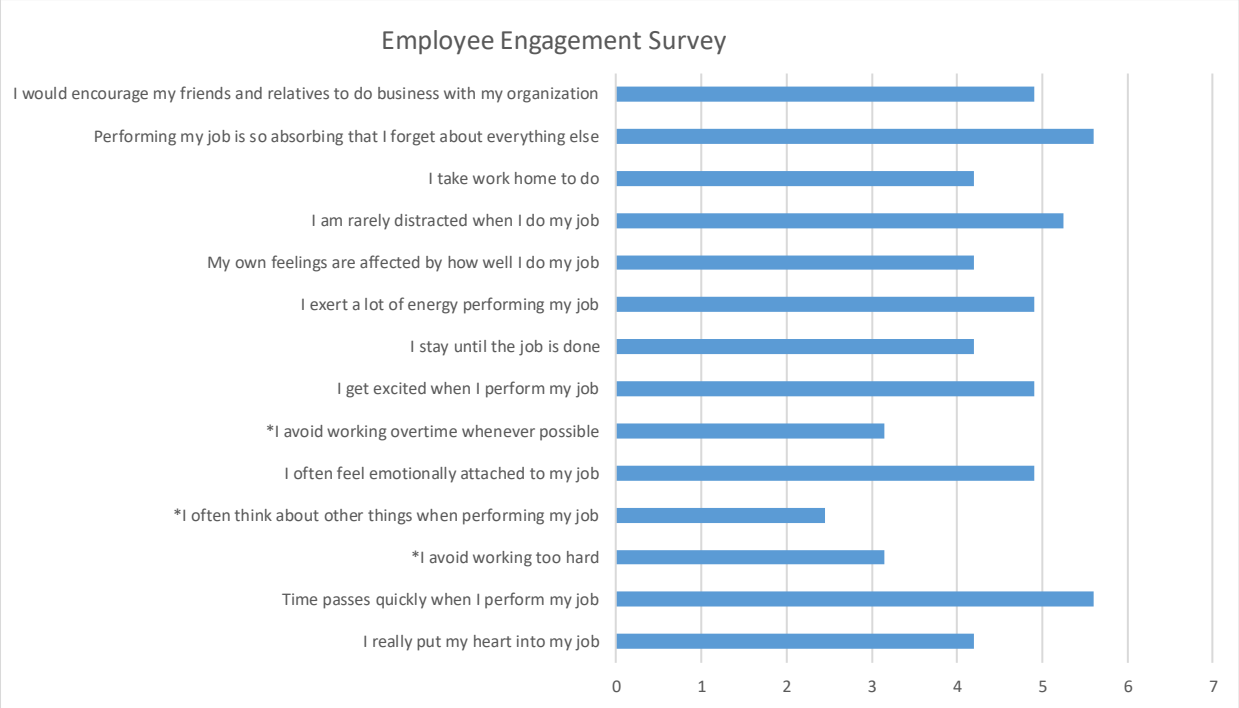
7) **Do you believe that emails sent by Omega employees to the Human Resources department and management of Absolute are reliable sources of information about employees' intentions to remain with the firm?**



PART 3

Below are the visualisations that you examined previously. For each visualisation, please answer the questions related to each visualisation.

This engagement survey was administered by Absolute during the third quarter of the current year to investigate employee morale at Omega. Approximately 35% of the employee workforce responded to the survey. A response of 0 = Definitely Not True
A response of 7 = Definitely True



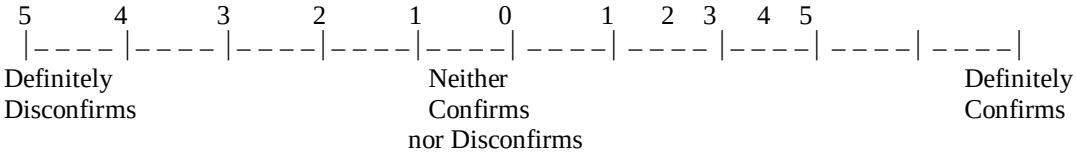
1) In your opinion, how useful was the visualisation of the employee survey?

0% 50% 100%
|– – – – |– – – – |– – – –|– – – –|– – – –| – – – –| – – – – |– – – –| – – – –| – – – –|
Not At All Useful — Very Useful

2) In your opinion, how reliable was the data that was used to create the visualisation of the employee survey?

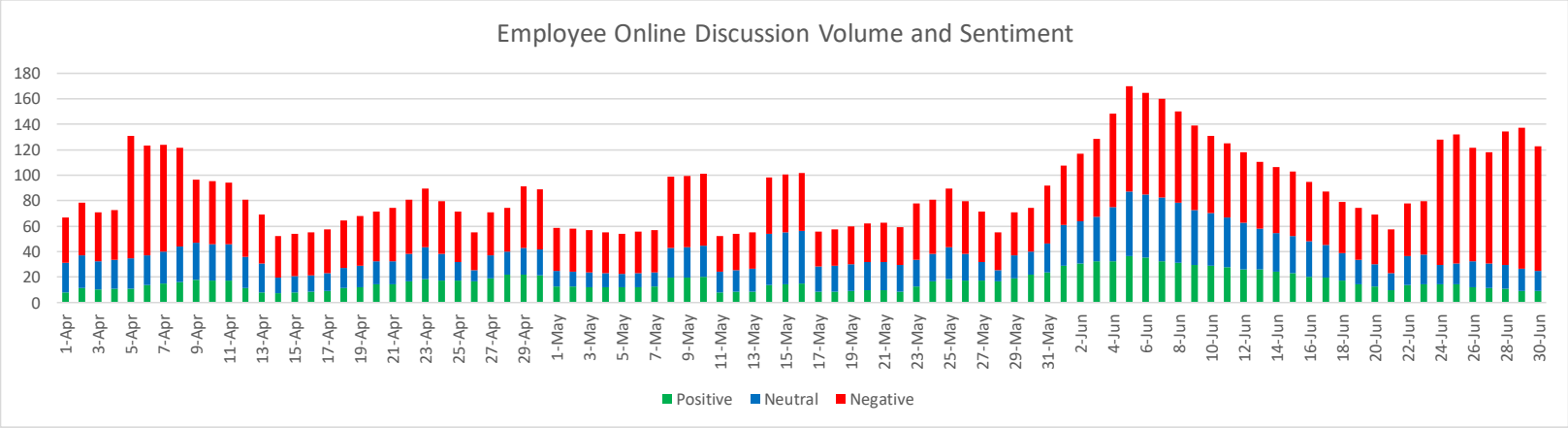
0% 50% 100%
|– – – – |– – – – |– – – –|– – – –|– – – –| – – – –| – – – – |– – – –| – – – –| – – – –|
Not At All Reliable — Completely Reliable

3) In your opinion, do you believe that this visualisation confirms or disconfirms Absolute's estimate that it is 90% likely to achieve the bonus for low employee turnover?



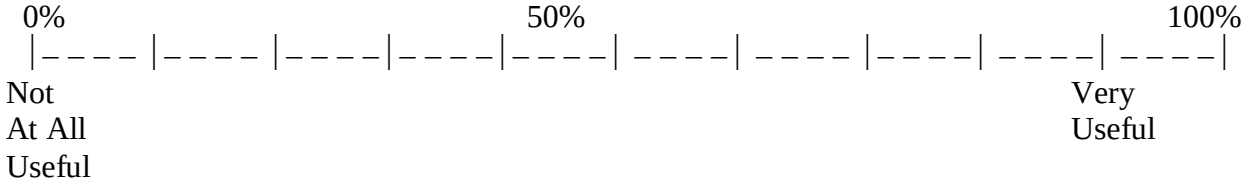
Note that each participant will see the same visualisation here that was presented earlier (i.e., the visualisation here matches the treatment condition). I have only included one, but each version of the instrument would have the appropriate visual here.

Smith and Larson LLC analyzed social media postings that are believed to have been made by Omega employees and conducted text sentiment analyses on these postings to determine whether the words in the postings represented a positive attitude towards Absolute, a negative attitude towards Absolute, or a neutral attitude. The visualisation presents the volume and sentiment of online discussions related to Absolute during the fourth quarter on social media sites such as Twitter, Facebook, and Google+.

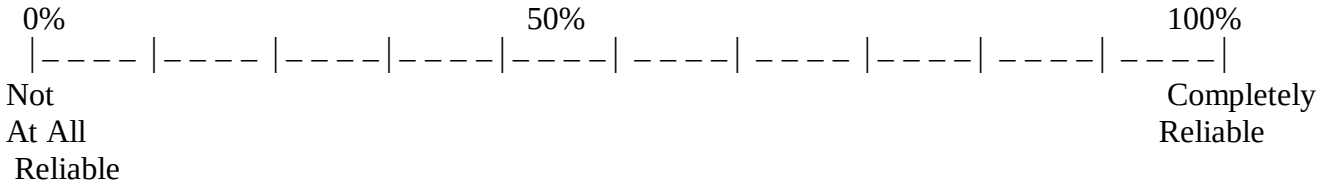


Source	Mentions	Positive	Positive%	Neutral	Neutral%	Negative	Negative%	Net Sentiment
Total	8117	1496	18%	2130	26%	4492	55%	-37%

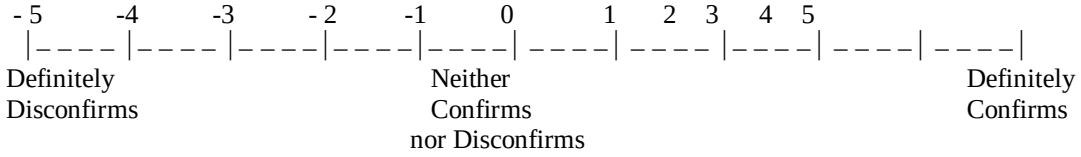
1) **In your opinion, how useful was the visualisation of employee sentiment?**



2) **In your opinion, how reliable was the data that was used to create the visualisation of employee sentiment?**



3) **In your opinion, do you believe that this visualisation confirms or disconfirms Absolute's estimate that it is 90% likely to achieve the bonus for low employee turnover?**



4) **Have you previously seen visualisations that display words of varying sizes (these are often called word clouds):**

Yes_____
No_____

The final 4 questions relate to you and your background:

5) What is your gender? _____Male _____Female

6) Are you currently a student? _____ No _____Yes

7) What is your age _____

8) What is your primary occupation? _____

9) Are you registered to vote? _____ No _____Yes

10) What is the highest level of Education you have completed? _____some High School _____High School graduate____ Some College ____College Graduate___Post Graduate Degree

Thank you for completing the questionnaire.