Error vs. accuracy rates: Evaluating forensic expert credibility.

Jonica Koodrin

This thesis is submitted in partial fulfilment of the Honours Degree of Bachelor of Psychological Science

School of Psychology

The University of Adelaide

Submitted on the 29th of September, 2020

Word count: 9,127

Table of Contents

List of Figures

Abstract

In a criminal trial, judges and jurors need to be able to quickly, easily, and confidently interpret forensic expert data. It is also important for them to determine the credibility of forensic experts in order to properly evaluate their opinion on legal evidence. Communicating forensic experts' opinions on legal evidence has been in the spotlight for researchers. More recently, focus has begun to shift towards the importance of understanding and communicating forensic expert performance data. In a 2 x 2 x 2 fully between-subjects design three fictitious expert reports (extracted from Martire, et al., 2020) were used to communicate forensic expert performance data to mock jurors (*N* = 143). The framing of the data - (error rates vs. accuracy rates), presentation format (data presented as a mean value vs. individual data points on a scatterplot) and colour of the data (colour vs. greyscale) - were manipulated. To measure judgements of credibility in forensic experts the traits 'reliable', 'accurate', and 'trustworthy' were used. Negatively framed forensic expert performance data (i.e. error rates) lead to lower credibility ratings compared to positively framed forensic expert performance data (i.e. accuracy rates). Presentation format had no significant effect on participants' credibility ratings. Finally, participants' ratings of expert credibility were significantly greater when data was presented in colour compared with greyscale. The outcome of this study extends research on message framing to a legal decision making context looking at the communication of forensic data.

Key words: Loss aversion, science communication, forensic science, psychology-law, decision-making.

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University, and, to the best of my knowledge, this thesis contains no material previously published except where due reference is made. I give permission for the digital version of this thesis to be made available on the web, via the University of Adelaide's digital thesis repository, the Library Search and through web search engines, unless permission has been granted by the School to restrict access for a period of time.

Signed,

Jonica Koodrin.

September 2020

Contribution Statement

In creating and writing this thesis, Dr Rachel Searston and I collaborated to generate the research question of interest and design the appropriate methodology. I conducted the literature search and worked with my supervisor to complete the ethics application. The expert reports were created by my supervisor and we worked together to write the Qualtrics Survey. I was responsible for all participant recruitment and data collection. The R script utilised in this study was written by my supervisor. I ran all of the reported analyses and plots in R, with guidance from my supervisor on the construction on the reproducible R Notebook. I was responsible for the thesis write-up with my supervisor providing constructive and general feedback.

Acknowledgements

I would like to acknowledge and thank my supervisor, Dr Rachel Searston, for her support and encouragement throughout this challenging year. Your passion towards this thesis and your own research motivated me to work hard to produce the best piece of work I am capable of. I would also like to thank the expert cognition lab for the group support throughout the first stage of this year. Knowing that I wasn't alone in my struggles and fears and having a group to share this with helped me remain calm to be able to continue moving forward. To Claire and Jade, this year has been so much brighter with Social Support Wednesdays. Thank you for the laughter, lunchbreaks and comradery throughout this stressful year.

Thank you to all of my family for putting up with my moodiness and listening to all my rants about the ins and outs of my thesis topic. To my oldest friend Maddi, thank you for the endless supply of memes and encouragement to take breaks for self-care. You have always been a pillar of support throughout my life, I am grateful for your friendship.

Finally, to Josh – you came back into my life at the beginning of the year and have given support and encouragement throughout. Thank you for being there for me, for making me tea, and for providing love and laughter when I needed it most.

Chapter 1: Introduction

1.1 Forensic Expert Impact on the Court

Forensic evidence is used in the criminal justice system to provide judges and jurors with information about the legal evidence of a case at hand (National Research Council, 2009). For example, a fingerprint left on a firearm found at the scene of a crime may be used to determine whether or not a suspect was at the scene. Forensic evidence plays a vital role in the decision-making process of jurors, who rely on the source conclusions of forensic experts, in order to make a decision to prosecute or set free the defendant in a particular case (National Research Council, 2009; PCAST, 2016). A source conclusion is the final decision that experts make and present to the jury after reviewing legal evidence. Forensic science has been used for decades in police investigations and the courtroom, however there has been a rising concern about the reliability and validity of forensic methods (PCAST, 2016).

In 2009, in *R v Jama*, Farah Jama was charged with rape based solely on DNA evidence. Mr. Jama served 15 months before it was discovered that he was innocent, and that in addition to problems that arose from poor interprofessional communication, the DNA evidence was mishandled (Vincent, 2010). Cases like *R v Jama* (2009) are called exonerations, and in the United States alone 45% of exonerations included a misapplication of forensic methods (Innocence Project, 2020). Between 1922 and 2015 in Australia, there has been an estimated 71 high profile cases where the accused was wrongly convicted (Dioso-Villa, 2015). These exoneration cases prompted critical appraisal of forensic science in the criminal justice system (Doyle, 2010). It has been highlighted that communication standards of forensic performance in the courtroom are lacking, in that they do not prevent experts from presenting misleading and inaccurate representations of their discipline (PCAST, 2016). For example, it has been noted that some forensic scientists claim to have an error rate that is “essentially zero” when presenting in the courtroom, which is an inaccurate representation of their actual performance and is misleading to the jury (PCAST, 2016). As such, several peak scientific bodies have recommended clearer standards for the communication of forensic expert performance data, as increased transparency will lead to more valid, reliable, and scientifically established methods (National Research Council, 2009; PCAST, 2016).

Forensic experts are increasingly expected to openly report scientific estimates of their accuracy to improve the communication of legal evidence (McAndrew & Houck, 2020; Searston & Chin, 2019). For example, experts may present their performance error rates alongside their conclusions about the source of the legal evidence so that jurors are aware of the limits of their expertise. As an indicator of performance error rates, it has been suggested that experts may report identification and exclusion rates specific to their domain (PCAST, 2016). Presenting such performance error rates is important as it will provide a clear understanding of the probative value of the expert’s evidence. Probative value can be described as the strength of the evidence used to prove a fact in a criminal trial. Reporting the probative value (e.g. in the form of expert performance error rates) of forensic expert evidence may provide judges and jurors with better insight into the reliability and validity of the expert’s conclusions. This is crucial because jurors may not have the expertise to evaluate forensic evidence, the time to discuss or research their decision and the power to dispute evidence (Koehler, Schweitzer, Saks, & McQuiston, 2016; M. B. Thompson, Tangen, & McCarthy, 2013).

Despite the push for forensic scientists to present their performance error rates (as a means of providing transparent information to jurors) there is little research regarding the effects of, and methods used to present this kind of data (Neumann, Kaye, Jackson, Reyna, & Ranadive, 2016; PCAST, 2016; Thompson, Vuille, Taroni, & Bidermann, 2018). This is because in the past it was not common practice for some forensic experts (e.g. firearm and bitemark examiners) to report the limits of their discipline or validation of scientific methods in criminal trails (PCAST, 2016). This thesis therefore investigates the extent to which the framing of forensic expert performance data might influence peoples' judgments of an expert's credibility.

1.2 The Credibility of Forensic Experts

Source credibility stems from the term *ethos*, which can be described as a means of persuasion or the intent of the speaker to appear credible to an audience (McCroskey & Young, 1981). Previously, credibility was thought to have two key factors: trustworthiness and expertise (Hovland, Janis, & Kelley, 1953 as cited in Hurwitz, Miron, & Johnson, 1992). However, subsequent research on the nature of credibility, and the variety of factors that impact it argued that credibility is a continuous variable which is subjectively perceived by those receiving information (McCroskey, 1966). This means that the term credibility will be viewed differently by different people depending on the context of the message being received.

Due to concern about the credibility of forensic expert opinions (National Research Council, 2009), some researchers have argued that forensic experts should report both broad and narrow assessments of accuracy, reliability, and validity (Edmond et al., 2016; National Research Council, 2009; PCAST, 2016). Narrow assessments of validity focus on the witness (juror) level and how these witnesses evaluate the credibility of an experts' opinion (Edmond et al., 2016). Some researchers have been exploring factors that affect perceptions of expert credibility at the witness level, in order to improve the assessment of forensic expert opinion (Brodsky, Griffin, & Cramer, 2010; Hurwitz, et al., 1992; Martire, Edmond, & Navarro, 2020).

Brodsky et al., (2010) addressed the gap in the assessment of expert opinions by developing and validating a scale that can be used to measure juror perceptions of forensic experts' credibility. The Witness Credibility Scale (WCS) was created through a series experiments using a variety of participants (judges, jurors and undergraduate students). The final version of the WCS includes 20 adjectives within four subscales (Confidence, Likeability, Knowledge and Trustworthiness) that together are thought to capture peoples' beliefs about an expert's credibility. However, in the development and validation studies reported, participant responses were measured based on an example capital murder trial. The researchers pointed out that further studies may wish to explore other types of testimony (and case examples) to validate that these factors of credibility are generalisable to other contexts (Brodsky et al., 2010).

More recently, Martire et al., (2020) explored the Expert Persuasion Expectancy (ExPEx) framework. They allocated 437 participants to one of nine groups (eight were attributes of credibility such as Ability and Consistency and one was a control group). Presented with strong and weak versions of the ExPEx, participants were asked to rate the persuasiveness of the expert report. The results found that participants were strongly influenced by Ability, Consistency, and Trustworthiness (Martire et al., 2020). That is, if these factors are strong in an expert report it will improve jurors' assessments of expert opinions. However, if they are weak, then jurors will have a harder time assessing the credibility of an expert's opinion (Martire et al., 2020). This suggests that juror perceptions of expert opinion can be swayed by the availability of information regarding the credibility of the forensic expert, particularly Ability, Consistency, and Trustworthiness.

Inspired by the results in Brodsky et al., (2010) and Martire et al., (2020), this thesis aims to test the impact that framing effects have on judgements of expert credibility at a witness level. A better understanding of the factors that affect perceptions of an experts' credibility may help forensic experts to more effectively communicate their conclusions (Carl I. Hovland & Weiss, 1951). While there are many factors that are important to include when measuring juror perceptions of forensic expert credibility (Brodsky et al., 2010; Martire et al., 2020; PCAST, 2016), this study included the three factors that were most consistent across studies; Reliability, Accuracy and Trustworthiness (based primarily on the results of Martire et al., 2020).

Based off prior findings in studies looking at forensic expert credibility (Brodsky et al., 2010; Martire et al., 2020), ratings of expert credibility as they are affected by framed expert performance data will be measured. Comparing the ratings of expert's credibility will indicate how effective the framed data was in influencing juror decisions. Understanding how different presentations of expert performance data affect assessments of expert credibility will increase our understanding of how juror decisions can be influenced by performance data. Further, the findings from this thesis may provide empirical evidence to encourage openness of forensic performance in the courtroom (McAndrew & Houck, 2020; Searston & Chin, 2019).

1.3 Communicating Forensic Science

Communication of forensic expert evidence has tended to focus on how experts present their source conclusions to the courtroom, noting that source conclusion is the final decision an expert presents to the jury. For example, DNA experts may use Random Match Probabilities (RMPs) to express the likelihood that an unrelated person, randomly sampled from the general population, would match the DNA sample found at the crime scene (Eldridge, 2019; Thompson, Grady, Lai, & Stern, 2018). Firearm and fingerprint experts, on the other hand, tend to make categorical statements about the source of trace evidence (Thompson et al., 2018). For example, a fingerprint examiner will conclude a fingerprint as being identified or individuated as a particular person or possession of a person (Thompson et al., 2018).

Recently, researchers have begun questioning the accuracy of the science behind the methods that forensic experts use to testify with in the courtroom (e.g. the likelihood that the defendants fingerprint matches prints left at the scene of the crime; Thompson, 2018). Further, people are likely to misunderstand and overweigh RMP values as well as categorical statements such as “match” usually because they have misconceptions about forensic disciplines (McQuiston-Surrett & Saks, 2009; Spellman, 2018).

Some studies have begun to look at the effects of presenting the limitations of forensic science alongside expert source conclusions (Kovera & McAuliff, 2000; McQuiston-Surrett & Saks, 2009). Forensic scientific limitations were presented in two forms: a cross examination by the defendant’s attorney and instructions given by the judge (McQuiston-Surrett & Saks, 2009). Responses were collected from 350 current jury members in order to gain actual juror opinions, rather than creating a mock jury situation (McQuiston-Surrett & Saks, 2009). Responses were compared based on whether their decision to find the defendant guilty or not guilty differed after hearing the limitations of the forensic method that was used in the mock case report (McQuiston-Surrett & Saks, 2009).

This study did not find a significant difference between groups that were presented with the limitations and groups that were presented without. This means that the jury members’ decisions were not influenced by the knowledge of limitations within the field (McQuiston-Surrett & Saks, 2009). Jurors’ decisions not being affected by the presentation of scientific limitations suggests that jurors may not be able to determine the difference between sound or flawed forensic methods used to report conclusions on legal evidence (Kovera & McAuliff, 2000; McQuiston-Surrett & Saks, 2009). Both researchers and peak scientific bodies have recognised that there is a gap in the literature regarding the presentation of forensic limitations via performance error rates (National Research Council, 2009; Neumann et al., 2016; PCAST, 2016; Thompson, 2018).

Prior research has focused on the communication of forensic scientific source attribution conclusions to jurors (e.g. RMPs and categorical statements; Eldridge, 2019; Thompson et al., 2018), while some studies have begun to look at the effects of reporting the scientific limitations of forensic methods (Kovera & McAuliff, 2000; McQuiston-Surrett & Saks, 2009). However, a form of communication that is yet to be researched is the use of forensic expert performance error rates (Thompson, 2018; PCAST, 2016). Reporting performance error rates may provide jurors with a better understanding of the credibility of an expert's opinion (PCAST, 2016). Further, it has been suggested that presenting expert performance rates alongside source conclusions will increase the transparency of forensic science methods and close the gap in assisting jurors with the assessment of expert opinions (National Research Council, 2009; PCAST, 2016). This thesis will be the first to investigate how the framing of forensic expert performance data can affect perceptions of forensic expert credibility. It is expected that an understanding of how forensic expert performance data influences perceptions of expert credibility may shed light on how juror decisions are affected in the courtroom.

1.4 How Jurors Evaluate Forensic Experts

Prior studies have investigated the factors that influence juror judgements of forensic expert evidence and why jurors often under evaluate expert conclusions (Faigman & Baglioni, 1988; Goodman, 1992; Nance & Morris, 2005; B. C. Smith, Penrod, Otto, & Park, 1996; L. L. Smith, Bull, & Holliday, 2011). Research has stated that when evaluating the credibility of evidence given in a trial, jurors' are often influenced by cognitive bias and heuristics (Curley, Munro, & Lages, 2020; Smith et al., 2011). Cognitive bias can be described as a person's general knowledge of objects or events that deviate from normal or intended perceptions, usually due to preconceived ideas and feelings that a person has about the object or event (Curley et al., 2020; Haselton, Nettle, & Murray, 2015; Smith et al., 2011).

Cognitive biases can cause jurors to misinterpret expert conclusions and misuse evidence when making a decision (McQuiston-Surrett & Saks, 2009). For example, jurors' weigh forensic evidence differently compared to experts. Categorical statements such as "match" are viewed as strong evidence by jurors but as weaker evidence by forensic experts (Eldridge, 2019). Additionally, people lacking skills in mathematics may misinterpret RMPs (Peters & Levin, 2008). Jurors may also misinterpret the RMP indicative of the defendant being innocent, and not as a probability of guilt (also known as the prosecutor's fallacy; Thompson, 2018).

Further, cognitive biases may occur when jury members look for information to validate their own values, beliefs and logical reasoning about forensic experts (known as confirmation bias; Eldridge, 2019). For example, people may believe that forensic expert methods are well researched as they often appraise trustworthiness by considering an expert's consistency in performance and use of positive language (Haley & Sidanius, 2006). Additionally, people tend to ignore information that may otherwise contradict their beliefs (Haley & Sidanius, 2006).

Finally, jurors may also rely on easily understood cues of credibility such as expert experience (Brodsky et al., 2010; Koehler et al., 2016). Researchers focussing on factors that affected the persuasiveness (perceived credibility) of an experts' source conclusion, found that perceptions of stronger persuasiveness were associated with higher credentials (Cooper, Bennett, & Sukel, 1996; Jurs, 2016). For example, experience in a specialised field (e.g. firearm or fingerprint examiner) being rated higher than university education (Wilcox & NicDaeid, 2018). While it is important for jurors to know the credentials of forensic experts (Wilcox & NicDaeid, 2018), there are arguably more useful aspects of an expert's credibility that can be presented to the courtroom (PCAST, 2016). Performance error rates, for example, offer a more transparent window into expert methods. Openness in the communication of forensic performance methods may allow jurors to assess expert conclusions with greater understanding of the accuracy and reliability of their technique (Koehler, 2011).

1.5 Framing Techniques

One technique that may affect perceptions of expert credibility is the framing of forensic expert performance data in terms of losses versus gains (Kahneman & Tversky, 1979). This thesis is the first to investigate framing effects in the presentation of forensic expert performance data. The effect of different computer graphical techniques (Chien, 2011; Eldridge, 2019) on judgments of forensic expert credibility will also be explored.

1.5.1 Framed forensic data. Presenting the same data as a gain rather than a loss, can influence and change the outcome of people's decisions. Prospect theory is a theory of decision-making which suggests that people perceive potential gains or losses relative to their specific situation, rather than in absolute terms (Kahneman & Tversky, 1979). A key prediction of prospect theory is that people will make decisions that preference avoiding losses more than acquiring the equivalent gains. This phenomenon, known as 'loss aversion', has been demonstrated in a wide range of decision-making contexts. The most common examples of loss aversion focus on judgements of monetary value and human life, elements demonstrated by the "Asian disease problem" (Tversky & Kahneman, 1981).

Briefly described, the Asian disease problem offers a story with two problems proposed (one positively framed; the other negative). In both of these problems, there are two options; one offers certainty and the other a probability. When presented with the positive frame, viewers pick the certain option; however when presented with the negative frame, they are more likely to chose the probability (the risky option; Steiger & Kühberger, 2018; Takemura, 1993; Tversky & Kahneman, 1981). That is, equal gains or losses are not always deemed psychologically equal depending on how the message is framed.

To our knowledge, there is no literature that looks at loss aversion within a legal context. However, several studies have focussed on framing effects within a health context. A meta-analysis looking at 165 studies on the effectiveness of persuasion in gain and loss framed messages (in the context of disease prevention) found moderately significant results in favour of gain-framed messages (O'Keefe & Jensen, 2006, 2007). This means that when given a message about disease prevention, people are slightly more likely to listen and take action when they view a gain frame over a loss frame. However, when presented with gain and loss framed messages (in the context of disease detection), there was no difference between the two message frames (O'Keefe & Jensen, 2006, 2009). This may suggest that the type of message being conveyed also matters when choosing to frame it as either a gain or a loss.

The current research draws on these previous demonstrations of loss aversion to investigate the impact of framing effects in the communication of forensic expert performance rates. Forensic expert performance tends to be reported as an error rate (PCAST, 2016), which has a distinct loss frame. This thesis explores the extent to which framing forensic expert performance in terms of losses (e.g. erroneous decisions) versus gains (e.g. correct decisions) influences perceptions of forensic expert credibility. In line with previous demonstrations of loss aversion, it is hypothesised that presentations of forensic expert performance data that emphasize error (loss framing) will result in judgments of forensic expert credibility that reflect an avoidance of risky courtroom decisions that may be incorrect (i.e. lower ratings of credibility). Understanding how the framing of forensic expert performance data affects peoples' assessments of forensic expert opinions is important for future research. This is because it will inform the development of evidence-based methods for communicating forensic scientific data (National Research Council, 2009; PCAST, 2016). This study lays the foundation for future researchers to explore the benefits, and potential problems, of reporting forensic expert performance as an error rate.

1.5.2 Presentation format. Research on numerical cognition has shown that people tend to focus on mean information and neglect variance information (i.e. looking at the mean and neglecting other information such as sample size) when considering how trustworthy a new finding is (Kramer, Telfer, & Towler, 2017). In the context of communicating forensic expert performance data to a jury, this tendency to neglect variance information may cause the jury to overestimate the reliability of summary statistics such as mean error rates. Visual displays of data (e.g. graphs that use simple visual features to represent different data points along a common axis) have been shown to improve comprehension of variability information when evaluating differences between data subsets (Kramer et al., 2017).

There has been an abundance of research looking at how to make data visualisation memorable and easily understandable to laypeople. Visualisation types such as diagrams and line and bar graphs are popular tools for experts to use to communicate quantitative information (Borkin et al., 2013; Newman & Scholl, 2012; Weissgerber, Milic, Winham, & Garovic, 2015). While these types of data displays are common (if not overused) in physiology, psychology, and medicine journals (Lane & Sándor, 2009), they can be problematic for a number of reasons. Often they fail to show outliers, normal distributions, and trends in changes across individuals (Weissgerber et al., 2015). On the other hand, univariate scatterplots, box plots and estimation plots clearly show data distribution and comparisons between two means, which is an important feature that gives power to the viewer (Ho, Tumkaya, Aryal, Choi, & Claridge-Chang, 2019).

Thus, a secondary aim of this thesis is to explore the extent to which displays of variance information affect judgments of expert credibility. It is hypothesised that more transparent displays of variability in expert performance (e.g. scatterplots versus standalone mean values) will result in lower ratings of expert credibility as viewing the raw data will reduce people's confidence in the credibility of forensic expert opinions.

1.5.3 Colour framing effects. Colour is highly subjective and is dependent on assumptions made by both the researcher preparing to present information and the viewer receiving information (Jonauskaite et al., 2016; Shah & Hoeffner, 2002). Despite obvious limitations, colour is useful for grouping elements on a graph (Shah & Hoeffner, 2002), and it can also be used to prime framed messages to create stronger responses from viewers (Chien, 2011; Voss Jr, Corser, McCormick, & Jasper, 2018).

There is little research regarding the use of colour in forensic expert reports. However, some research has been conducted on colour effects on message framing in health communications. Using a message describing the benefits of sunscreen to test out colour influence, one study found that peoples' intentions to wear sunscreen were influenced by message framing and the use of colour (Voss Jr et al., 2018). In particular, health messages that were framed positively and in short wavelength colours (i.e. blue and purple) increased intentions to wear sunscreen (Voss Jr et al., 2018). However, messages framed negatively in long wavelength colours (i.e. red and orange) did not produce significant intentions to wear sunscreen (Voss Jr et al., 2018). Another health communication study looking at willingness to receive vaccinations, found significant interactions with colour and message framing

(Chien, 2011). These results suggest that a priming effect occurs, such that short wave-length colours strengthen gain-framed messages and long wavelength colours strengthen loss-framed messages, resulting in a larger framing effect.

These findings pair nicely with other studies looking at most and least preferred hues. For example, cooler hues such as blue, green-blue, and purple tend to be preferred over warmer hues (Bakker, Voordt, Vink, Boon, & Bazley, 2015; Granger, 1955; Jonauskaite et al., 2016; Yu et al., 2018). As mentioned above, an obvious limitation when measuring the effects of colour is the subjectiveness that may stem from factors, such as positive and negative experiences and assumptions (Jonauskaite et al., 2016; Shah & Hoeffner, 2002). It is recognise that within these studies some warmer colours were also preferred (e.g. the colour red in Jonauskaite et al., 2016, and orange and yellow in Eysenck, 1941). However, the present study adopts the general trend that cool hues are viewed more positively and warmer hues are viewed more negatively.

Thus, a third hypothesis is that the use of colours generally associated with positive and negative affect, paired with the positive and negative framing of forensic expert performance data will increase the magnitude of the framing effect. This means that participants will give stronger opinions of credibility when in colour conditions as opposed to those in greyscale conditions.

1.6 Current Study

This thesis extends on prior loss aversion studies, and forensic communication research, to examine how different presentations of forensic expert performance data influences lay evaluations of expert credibility. The aim of this project is to better understand how the framing and presentation of empirical data can sway perceptions of forensic expert opinions.

Based on the literature reviewed and hypotheses details above, the predictions of this thesis can be summarised as follows:

- Ratings of the credibility of forensic experts' opinions will be lower for participants who view negatively framed forensic expert performance data compared with those who view positively framed forensic expert performance data;
- This 'loss aversion' effect will be significantly more pronounced for participants who view the gain-framed forensic expert performance data presented in short wavelength colours (i.e. purples and blues) and the loss-framed forensic expert performance data in long wavelength colours (i.e. reds and oranges), compared with participants who view the same data in greyscale;
- Ratings of forensic expert credibility will be lower for participants who are shown how experts vary in their performance (i.e. individual points on a graph) compared with those who are only shown summary statistical information (i.e. mean values).

Chapter 2: Method

2.1 Ethics

This study was approved by the *University of Adelaide Human Research Ethics Subcommittee*, approval number 20/27. Participation was voluntary and anonymous for all participants coming from the University of Adelaide and the general public. Informed consent was obtained from participants prior to their commencement in the study, this was indicated by ticking 'continue' at the bottom of the information and consent page (Appendix A). University students that participated in exchange for course credit were assured anonymity through identification numbers. Participants were informed that the de-identified data from the study would be made available on the Open Science Framework website.

2.2 Participants

A total of 210 participants completed the online survey. Participants recruited were all over the age of 18, fluent in English and had normal to corrected-to-normal vision. Participants who failed to complete the survey were removed, leaving a total sample of 145 participants. As the experiment was conducted online it was necessary to ensure participants engaged with the survey. Twenty-six of the remaining participants who completed the experiment, incorrectly recalled one of the forensic domains. The question asked participants what the respective fields were of the three forensic experts used in the survey. However, these participants were retained in the sample to avoid losing power in the analysis and because genuine jury samples may also include members that do not easily or accurately recall specific details about experts. For example, people in a jury may experience lapses in attention or pay more attention to the overall gist of an expert's conclusions to a greater extent than the specific details (Eldridge, 2019).

An a priori power analysis indicated that at least 144 participants (18 per condition) were needed to achieve 80% power to detect an effect size as small as *d* = .02. The smallest framing effect of interest in this experiment was based on prior estimates, ranging from η^2 = 0.17 in McQuiston-Surrett & Saks (2009) to η^2_p= .06 – .19 in Chien (2011).

A study in a similar field of research has provided support for the use of student populations in mock-juror research (Brodsky et al., 2010). As a result, undergraduate psychology students at the University of Adelaide (*n*= 108) were recruited via the Research Participation Pool. They completed the study in exchange for course credit (0.5 credits). Further, the study recruited participants via social media and word of mouth (*n* = 37). The general public participants were not given a reward in exchange for their time. The study consisted of 41 males, 102 females, 1 non-binary participant and 1 participant who did not wish to disclose their gender. Participant ages ranged from 18 – 62 with an average of 23 years. The majority of the participants had graduated high school or had completed some college or university studies (n= 104). The remaining participants had reported less than high school (n= 4), two, three or four year degree (n= 32) and professional degree (n= 5). There was a variety of educational backgrounds with most participants coming from psychology (n= 46), biological science (n= 19), chemistry (n= 15), mathematics (n= 12), economics (n= 12), criminology (n= 11), arts and design (n= 9), education (n= 7), and law (n= 6). Many other fields of education were recorded however they were not as popular as the fields stated above.

2.3 Design

In a 2 (Framing: Error vs. Accuracy) x 2 (Presentation Format: Scatterplot vs. Summary Statistics) x 2 (Colour: Colour vs. Greyscale) between-subjects design three fictitious expert reports (based on Martire et al. 2020) were used to communicate forensic expert performance data. The message frame, presentation format and colour of the data were manipulated in order to measure and compare participants' ratings of expert credibility (see Figure 1).

The purpose of this thesis was to examine how different presentations of forensic expert performance rates influenced peoples' perceptions of expert credibility, as such there were no control variables as comparisons were made between groups. Further, although in the past it was not common practice for some forensic experts to report the limits of their discipline (PCAST, 2016), it is now increasingly expected (Searston et al., 2019). Therefore, adding a condition where expert performance was not reported would be irrelevant in relation to expected forensic practice in the courtroom. Additionally, comparing the effects of error vs. accuracy rates is central to the hypotheses, therefore adding a control group would be pointless and outside the scope of this thesis.

2.4 Procedure

Participants accessed the survey via a Qualtrics[tm] web address and were randomly allocated to one of the eight experimental conditions (Figure 1). Participants first read the information and consent form and then gave consent to have their de-identified data used in this study. The purpose of the study was not fully disclosed, rather it was broadly stated that the aim was to assess the credibility of experts based on performance data. Prior to commencing the experiment, University students were asked to provide their student code in order to receive course credits in exchange for their participation. Participants rated the field-wide credibility of six different forensic domains (fingerprint, faces, firearms, handwriting, blood pattern and bitemarks) before and after reading the three mock expert reports. The pre and post responses of credibility were measured in an exploratory analysis to see if there was a generalised effect of the three framed expert reports on all forensic domains. The mock expert reports consisted of a firearms examiner, fingerprint examiner, and facial examiner and participants rated the credibility of these experts on perceived ‘Reliability’, ‘Accuracy’, and ‘Trustworthiness’. Participants then filled out a series of demographic questions related to their educational background and experience with interpreting statistical graphs. Participants were also asked to recall the forensic domains of the experts in each of the case reports at the end of the experiment to see how well they remembered details about the expert reports.

2.5 Expert Reports

The three expert reports gave a general description about the forensic evidence in a mock case and a specific description of the expert’s opinion about the legal evidence in the case. A detailed description of a graph depicting simulated data on the performance of forensic examiners from the same field as the expert in the case was also provided. The three fictitious female experts (Smith, Jones, and Wilson) represent experts from three forensic disciplines (firearm examination, fingerprint examination, and facial examination). The reports are identical across all 8 conditions except for systematic changes to the graph based on the performance data corresponding to the three forensic disciplines. There were a total of 24 framed expert reports created. Credibility ratings were measured using a slider scale ranging from 0 to 100, with corresponding scale anchors (e.g. *‘Very low’, Moderate’* and *‘Very high’*) to assist participants.

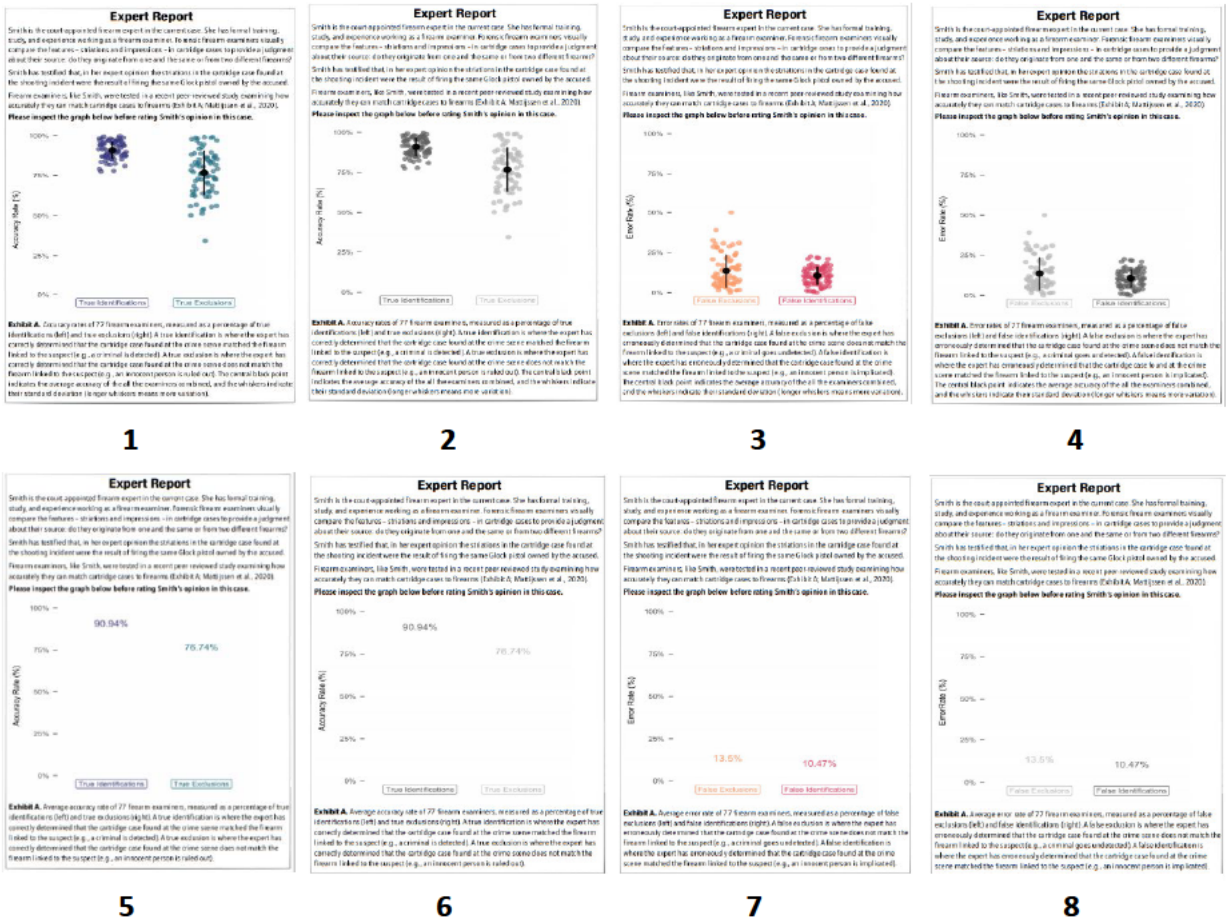

Figure 1. Mock expert reports displaying framed forensic performance data.

As seen from left to right – 1) positive, scatterplot, colour; 2) positive, scatterplot, greyscale; 3) negative, scatterplot, colour; 4) negative, scatterplot, greyscale; 5) positive, mean value, colour; 6) positive, mean value, greyscale; 7) negative, mean value, colour; 8) negative,

Chapter 3: Results

3.1 Main Analysis

Firstly, the data was cleaned and de-identified for analysis in R, all analyses were parametric tests. The hypotheses are directional for the main analysis and two-tailed exact statistics are reported. Summary statistics were run to check for homogeneity and to see preliminary effects, then plots were developed to visualise the data to check for trends. Finally, we conducted three one-way between-subjects Analysis of Variances (ANOVA) to test each of the primary predictions.

Hypothesis 1: Framed forensic data. It was hypothesised that credibility ratings of forensic experts' opinions would be lower for participants who view the negatively framed forensic expert performance data compared with those who view the positively framed forensic expert performance data. The descriptive statistics revealed that the experts' were perceived as less credible with the negatively framed performance data (*M*= 64.97, *SD*= 18.56) compared with the positively framed performance data (*M*= 75.99, *SD*= 13.65). The scatterplot in Figure 2 gives a visual illustration of the difference in mean credibility ratings between the two framing conditions.

A 2 (Framing: Positive vs. Negative) x 2 (Presentation Format: Mean values vs. Scatterplots) x 2 (Colour: Colour vs. Greyscale) between-subjects ANOVA revealed that message framing had a significant effect on participants' mean expert credibility ratings, with participants' in the negative framing (loss-frame) condition providing significantly lower credibility ratings compared with the participants in the positive framing (gain-frame) condition, $F(136) = 18.02$, $p < .001$, $\eta^2_G = 0.117$.

Hypothesis 2: Presentation format. Further, it was predicted that perceptions of the expert's credibility would be lower for participants who viewed the data as individual data points (e.g. scatterplot), compared with those who viewed the data as a single aggregated numerical value (e.g. mean accuracy or error rates). However there was no significant difference in the credibility rating provided by participants who were presented with the scatterplots (M= 69.48, SD= 18.27) compared with participants who were presented with the mean performance rates (M= 71.67, SD= 15.95), $F(136) = 1.22$, $p = 0.27$, $\eta^2_G = 0.008$.

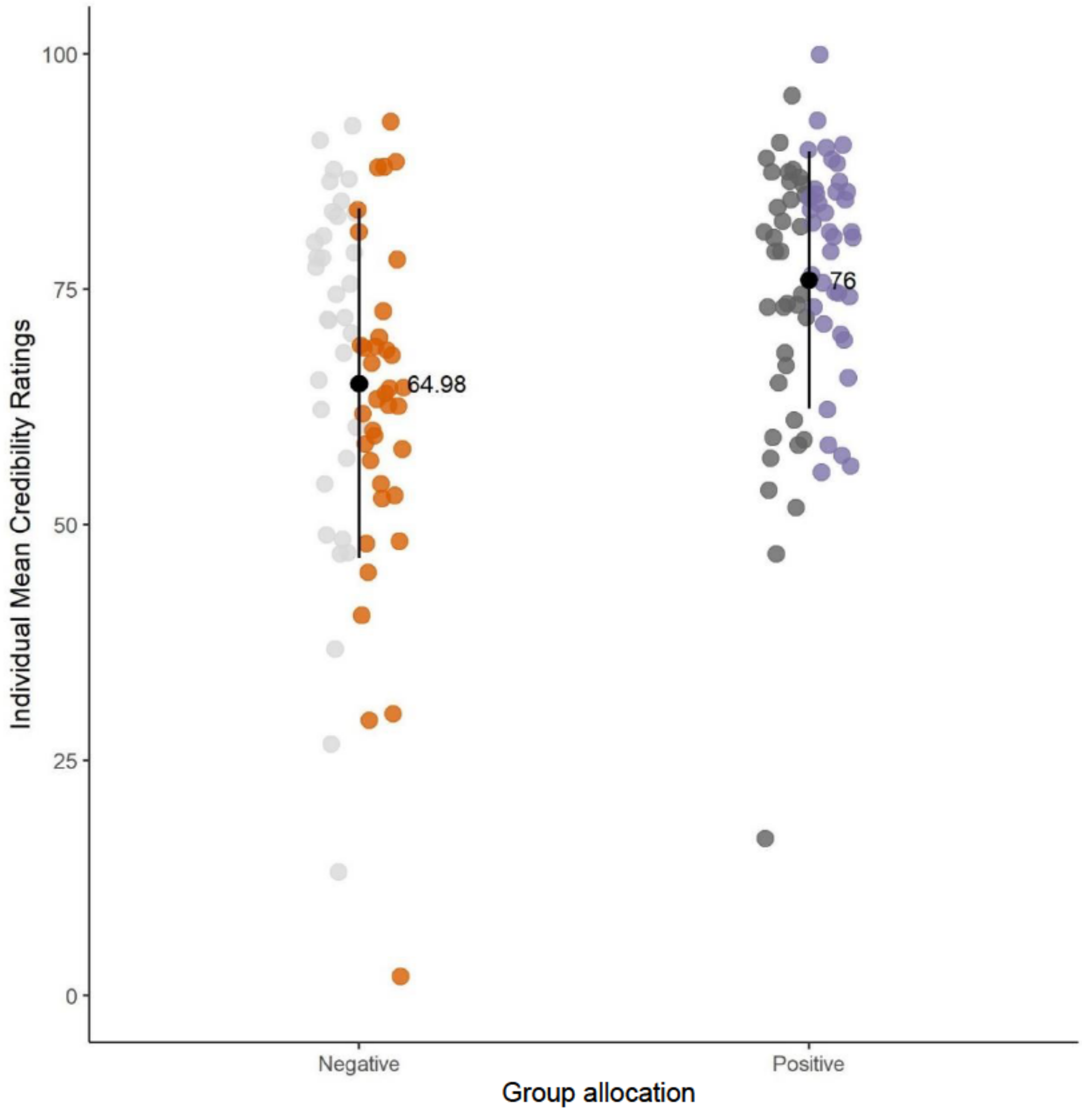

Figure 2. Mean credibility ratings for positive vs. negative and colour vs. greyscale conditions.

Hypothesis 3: Colour framing effect. It was also predicted that the message frame would be significantly more pronounced for participants who view the gain frame in short wavelength colours (blue and purple) and the loss frame in long wavelength colours (red and orange), compared to participants who viewed the data in greyscale.

The descriptive statistics showed that the negative framing condition (error rates) presented in long wavelength colours (*M*= 61.93, *SD*= 18.00) indeed appeared to result in lower credibility, compared with the negative framing condition presented in greyscale (*M*= 68.29, *SD*= 18.85). Similarly, the positive framing condition (accuracy rates) presented in short wavelength colours (*M*= 78.60, *SD*= 10.77) seemed to have an increased positive effect on expert credibility ratings, compared with a positive frame presented in greyscale (*M*= 73.24, *SD*= 15.84). Demonstrated in Figure 2, there is a difference in mean expert credibility ratings within gain frame and loss frame conditions when colour is introduced.

This interaction was also significant, $F(136) = 4.17$, $p < .05$, $\eta^2_G = 0.03$. That is, the effect of framing on participants' credibility ratings was significantly more pronounced when the data were presented in colours congruent with their framing condition (e.g. long-wave length colours paired with error rates, and short-wave length colours paired with accuracy rates).

3.2 Exploratory Analysis

Additional, an exploratory analysis was conducted to see if the pre and post (referred to as *time* in the analysis) measures of participants' perceptions of the credibility of the different forensic domain was affected by the expert reports they encountered during the experiment. For the analysis, the expert domains that were used in the mock expert reports (i.e. firearm examiner, fingerprint examiner and facial examiner) were labelled "target" domains and those not used (i.e. bitemark examiner, handwriting examiner and blood pattern examiner) were labelled “distractors” domains. Two-tailed statistics are reported.

An within-subjects ANOVA was conducted revealing a main effect of domain, $F(286) = 19.44$, $p < .05$, $\eta^2_G = 0.05$, and time, $F(286) = 7.48$, $p < .05$, $\eta^2_G = 0.004$, on participants’ judgements of domain credibility. That is, participants rated the forensic domains as less credible after reading the expert reports in the experiment overall and that this effect was strongest for the domains that were not encountered in the expert reports.

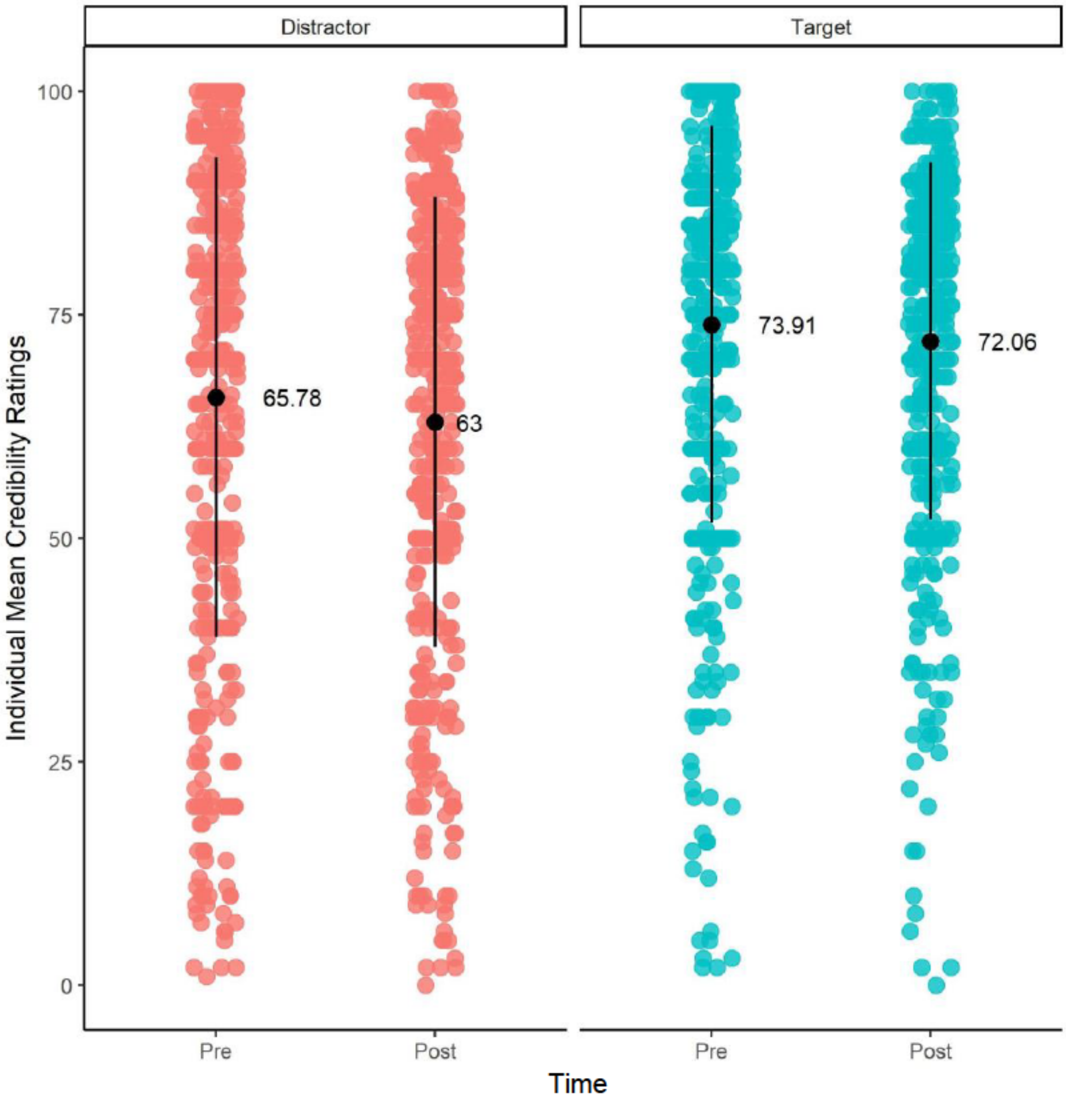

Figure 3. Pre and post mean credibility ratings for target and distractor forensic domains.

Post-hoc comparisons were conducted to determine the decrease in field credibility from pre to post remained significant for the target domains. For the target domains, the scores from the pre-test (*M*= 73.91, *SD*= 22.20) and post-test (*M*= 72.06, *SD*= 19.95) in the credibility rating task indicated that there was no significant effect, $t(434) = 1.88$, $p = .06$, meaning these forensic domains were rated consistently across both time points. As seen in Figure 3, the mean credibility rating (black dot) and the standard deviation (black lines) for the target domains show little variation between pre and post times. On the other hand, the credibility rating task for the distractor domains indicated that scores from the pre-test (*M*= 65.77, *SD*= 26.82) and post-test (*M*= 63, *SD*= 25.17) had a significant decline, $t(434) = 3.55$, $p = <.001$. Figure 3 illustrates the significant effect between pre and post time points for the distractor domains. To summarise, after viewing the framed expert reports participants did not change their perceptions of credibility for the target disciplines, however they gave lower credibility ratings for the distractor disciplines.

Chapter 4: Discussion

4.1 Overview

This thesis investigated the effects of framed forensic expert performance rates on peoples' perceptions of expert credibility. The results suggest that some communication techniques are more effective than others at increasing credibility ratings of forensic experts. However, these findings should be interpreted with caution as there are limitations that may have affected the outcome of this study.

4.2 Framed Forensic Data

The present study found that ratings of forensic credibility were lower for participants who viewed negatively framed forensic expert performance data, supporting the first hypothesis. Results also found that ratings of credibility were higher for participants who viewed positively framed expert performance data. These results were in line with past research, and support prospect theory within a legal context (Kahneman & Tversky, 1979). As such, there is an effect of loss aversion where message framing impacts decision making for participants regarding the credibility of forensic experts (Steiger & Kühberger, 2018; Tversky & Kahneman, 1981).

The findings from this study illustrate that peoples' perceptions of forensic expert credibility can be swayed using message framing techniques. However, this study only measured the effects of framed expert performance rates on perceptions of credibility. Future research may wish to explore the effects that loss aversion and expert performance rates have on juror decisions to convict or set free the defendant in a criminal trial. Further, future work focussing on mock criminal court cases may explore if message framing and expert performance rates have a direct effect on juror decisions in a criminal trial, causing them to make more conservative decisions. On the other hand, there may be a mediating effect on the outcome of the trial through the juries' assessment of an expert's credibility. Furthering research may help to understand the extents to which message framing influences juror decisions in criminal courts.

Further, this thesis shows that presenting framed expert performance rates has a significant effect on expert credibility ratings. However, this study did not measure the reliability or validity of this form of presentation. Future researchers may wish to conduct comparison studies between this new design and other forms of communication that express forensic limitations (e.g. cross examinations or a judge giving instructions to a jury; McQuiston-Surrett & Saks, 2009). Comparison studies will help to increase experts understanding of the reliability and validity of this new design. Further, researchers may explore the strengths and weaknesses of the different forms of communication and the varying effects they may have on jurors' decisions in the courtroom.

Due to the relative infancy of this topic, there was little prior knowledge available to build on. Therefore, knowledge was acquired from other domains of research and adapted various methods to achieve this study's design. Adapting the loss aversion effect from prospect theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1981) and message framing in a health context (O'Keefe & Jensen, 2006, 2007, 2009), revealed strength in the generalisability of these theories. The findings from this study not only show that peoples' perceptions of expert credibility can be swayed, it also contributes to the literature as it supports robust theories from past research within a legal context, which has not previously been studied.

4.2.1 The Added Effect of Colour. The current study also revealed that colour acted as a prime that magnified the effects of loss aversion, supporting hypothesis three. Specifically, in the loss-frame groups, participants in colour conditions gave lower ratings of expert credibility, compared to those in the loss-frame greyscale conditions. A similar effect occurred for the gain-frame messages where participants gave higher ratings of credibility when in colour conditions. The findings correspond with previous colour studies in a health context describing this priming effect enhancing the frame of a message (Chien, 2011; Voss Jr et al., 2017). This study contributes to the literature by showing that loss aversion and colour priming are generalisable theories that have significant effects within a legal context.

This thesis found a significant effect of colour that supports research on colours that are best suited to specific message frames (Bakker et al., 2015; Chien, 2011; Granger, 1955; Voss Jr et al., 2018; Yu et al., 2018). However, an obvious limitation of this study includes the subjectiveness of colours. Factors such as positive and negative experiences and assumptions have the potential to cause cognitive bias among participants (Jonauskaite et al., 2016; Shah & Hoeffner 2002).

It is worth noted that experts should carefully consider the impact they want to have on jurors when choosing to pair colour with a message frame. Our findings suggest that short wavelength colours (blue/purple) enhance gain-frames, increasing credibility ratings. On the other hand long wavelength colours (red/orange) enhance loss-frames, decreasing credibility ratings. Enhancing messages through the use of colour may assist people to make more informed assessments of forensic expert opinions. If forensic experts are aware of the ways that colour priming can enhance framed performance reports of accuracy and error, this may assist them in increasing open communication in the courtroom.

4.4 Presentation Format

Presenting data in scatterplots did not cause significantly lower ratings of credibility, which means that hypothesis two was not supported. What our study found instead was that the data presented as a mean value and data presented as individual points on a scatterplot produced approximately the same expert credibility ratings. This is surprising as the literature suggests that graphs which show data distribution and comparisons between two means are more effective in communicating quantitative data (i.e. the expert performance data; Ho et al., 2018; Kramer et al., 2017). Our prediction was made based on the assumption that presenting individual performance rates of forensic experts in a given domain may reduce people's confidence in the credibility of forensic experts' opinions.

A possible explanation as to why no significant effects were present in presentation format may be because participants may not have understood either of the two formats used in this study (i.e. mean values and individual data points scatterplots). This may be due to a lack of numeracy skills or lack of experience with interpreting data on a graph (Peters & Levin, 2008). Numeracy skills refer to a person's ability to understand and use mathematical concepts. As such, a lack of numeracy skills may look like a person struggling to calculate the difference between a regular price and a sale price using a calculator (Peters & Levin, 2008). To help increase juror understanding of expert methods and performance levels (e. g. rates of true identification and true exclusions of a fingerprint examiner), Rousselet, Pernet, & Wilcox (2017) have suggested to use subgroups to organise data into certain limit groups (above, within or below standard forensic practice). In the context of forensic experts reporting their performance error rates in a courtroom, scatterplots may be presented with additional subgroup lines to further clarify the standards of forensic performance in a specific domain. Future researchers may run this study again, but with added subgroups describing standard levels of practice for the forensic domain. This may help less numerate people to understand forensic expert performance data as research has shown an increase in understanding when non-numerical sources of information are presented (Peters & Levin, 2008).

Further, to avoid confusion with interpretation, and to help the less numerate, Neumann et al. (2016) suggests presenting data in both verbal and numerical form. Presenting data in multiple formats may assist people when interpreting and understanding source conclusions. For example, researchers may present the current studies expert performance reports with both the mean value and individual data points together on a graph. Together they may provide greater clarity to participants about the reliability, accuracy and trustworthiness of forensic experts' methods.

Additionally, the findings from this project are particularly important as jurors often under evaluate forensic expert source conclusions, due to cognitive bias, the prosecutor's fallacy, and lacking numeracy skills (Curley et al., 2020; Peters & Levin, 2008; Smith et al., 2011). The present study has identified a number of ways that experts might shape the presentation of their performance rates, in order to reduce the impact if these erroneous negative evaluation of credibility. These are, presenting mean value rates alongside individual data points (verbal and numerical; Neumann et al., 2016), or using forensic standard practice indicators on a scatterplot graph to further inform jurors of expert performance rates (Rousselet, Pernet & Wilcox, 2017), while assisting the less numerate (Peters & Levin, 2008).

As mentioned before, transparency of expert credibility may help jurors to make more informed decisions in the courtroom. That is, they will be made aware of the reliability and validity of scientific methods that were used by the expert to come their conclusion about the legal evidence in a criminal trial. Additionally, exploring different presentation formats of forensic performance data may help to understand why jurors often under evaluate forensic expert conclusions (Smith et al., 2011). Furthering empirical research on forensic expert performance error rates may lead to normalising open communication about forensic limitations in the courtroom.

4.5. Exploratory Analysis

As a part of the survey we asked participants to rate the credibility of six forensic experts' domains (firearms, fingerprint, facial, handwriting, bitemark and blood pattern) before and after reading the framed expert reports. There was an interest to see if there was a change in credibility rating scores across time (pre and post) and in the grouped forensic domains; target (firearm, fingerprint and facial) and distractor (handwriting, blood pattern and bitemarks). The analysis found that the pre and post credibility scores for all forensic domains decreased significantly, after participants viewed the framed expert reports. Further, it was found that pre credibility ratings for the target domains did not change significantly after reading the framed expert reports, but credibility ratings for the distractor domains decreased significantly. These are interesting finding as it suggests that when participants view performance rates of forensic experts, their perception of forensic credibility decreases as a whole. This may be due to participants' misconceptions about forensic performance and prior beliefs that forensic methods have "essentially zero" error rates (McQuiston-Surrett & Saks, 2009; PCAST, 2016). Being presented with data that contradicts their prior beliefs about the accuracy of forensic science methods may have caused participants to give lower credibility ratings.

The effects found in the exploratory analysis may be explained by the CSI effect (Crime Scene Investigation effect; Hayes-Smith, 2009; Hui & Lo, 2017).This theory has been explored by researchers as a potential factor that may encourage jurors to misinterpret forensic performance and legal evidence, due to altered expectations (Hayes-Smith, 2009; Hui & Lo, 2017; Kinsey, 2011). The CSI effect describes the influence that forensic science crime drama television series (e.g. CSI or Bones) has on juror perceptions of legal evidence and decision making (Kinsey, 2012). Hui and Lo (2017) found that perceptions of legal evidence (i.e. Participants' ratings of the reliability of forensic evidence) was significant influenced by the CSI effect. Similarly, Schweitzer and Saks (2007) found that those who watched more forensic dramas had a greater expectations for the quality of forensic science than what is often presented in the courtroom. The CSI effect may explain why the participants in this study gave lower credibility ratings after viewing the expert reports that revealed the performance accuracy and error rates of forensic expert methods.

The CSI effect is an appealing theory to provide an explanation for the findings in this study. However, it is important to note that there are debates about the reliability of the CSI effect in the literature regarding whether it helps or hinders the criminal justice system (Kinsey, 2012). Future researches may be interested to explore how prior perceptions of forensic experts change when presented with expert performance rates. These studies may lead to a greater understanding of how peoples' perceptions of forensic experts can be influenced by information that opposes their beliefs. Additionally, researchers may explore the CSI effect as a factor influencing juror expectations of the reliability and validity of forensic experts' performance, rather than something that directly affects their decisions in a criminal trial (Hui & Lo, 2017; Schweitzer & Saks, 2007).

Cognitive bias may be another possible answer for what might be occurring in the data. Participants' prior perceptions (or rather misconceptions) of forensic expert performance may have caused them to rate all fields of forensics less credible, rather than specific disciplines. Future research may expand on our initial findings where mean credibility ratings of all forensic domains decreased significantly after viewing the framed expert reports. The findings from the exploratory analysis are interesting however, we are limited in our knowledge about potential effects that might be occurring as it is beyond the scope of this thesis. Further, due to time constraints the current study was not able to explore these findings further. However, these findings are interesting as it suggests that perceptions of credibility for one forensic discipline may affect the image of the entire field.

4.6 Limitations

Firstly, the study was conducted online, which meant that it was not possible to control the environment of the participants while they completed the survey. Participants were asked to follow instructions and complete the survey correctly (e.g. not multitasking or rushing through the questions), however there is no guarantee that they complied with these expectations.

Majority of the sample (*n*= 108) was collected from the first-year research participation pool. Although a similar study utilised student participation and yielded significant results (Brodsky et al., 2010), there are other data bases such as such as Amazon Mechanical Turk which would give the study a more robust sample that would better represents a jury in a courtroom. Further, other studies have used actual jury members as participants in their study which provides an insight into how previous or current members of a jury have made decisions (McQuiston-Surrett & Saks, 2009). Therefore, future research could look at the effects within a larger more diverse sample, or a sample drawn from current or previous jury members. This may provide a deeper insight into the effects that presenting forensic expert performance rates has on expert credibility.

Lastly, the thesis was conducted under strict time constraints. Time constraints limited the ability to run appropriate pilot tests before fully launching the survey online. This meant that more effort was required in the development of the survey, with the researchers' pilot testing the survey themselves to check for consistency in language, ease of use on mobile devices and other additional queries. Further, there were interesting discoveries in the exploratory analysis that could have been explored further. However, in order to conduct additional studies to explore these findings, a longer time frame is needed.

4.7 Implications for Forensic Science

Inquiries by several peak scientific bodies recognised that forensic scientists need to improve the reliability and validity of their scientific methods and communication of performance data (National Research Council, 2009; PCAST; 2016). Further, there has been a lack of research regarding the communication of forensic expert performance error rates (Neumann, 2016; Thompson, 2018; PCAST, 2016). As such any knowledge on the effects of performance error rates is relevant for the legal system.

This thesis offers an insight into a relatively new area of research regarding the presentation of framed forensic expert performance data, and the influence it has on peoples' perceptions of expert credibility. It has been highlighted that message framing and colour priming have the ability to sway peoples' opinions of forensic expert credibility. The results contribute to the literature looking at how jurors understand and evaluate forensic expert data. Such that, peoples' opinions of forensic credibility may be swayed when they view framed expert performance data. In the context of a criminal trial, the presentation of forensic limitations may assist jurors' in making more informed decisions. This is because jurors' will know the accuracy and error rates of the scientific methods the forensic expert used in order to reach their conclusion about the legal evidence.

Additionally, forensic experts are expected to report the scientific estimates of their accuracy to improve communication of legal evidence and transparency of their methods (McAndrew & Houck, 2020; Searston & Chin, 2019). This thesis addresses the issue of transparency in communication by measuring the effects of forensic expert performance rates. Presenting performance data to inform participants about forensic experts' reliability and validity seemed to be an effective way of achieving greater openness in forensic expert communication. This is evidenced by the fact that the forensic expert reports, using performance rates presented in a gain-frame or loss-frame and colour, significantly influenced how people rated forensic expert credibility. Presenting forensic data that is easily understood by everyone is important so that jurors have clear expectations about expert methods and how reliable their conclusions are in terms of standard practice.

Compared with the findings from previous studies, the results from this thesis do not reach so far as to show that transparency of forensic performance can improve the communication of legal evidence or affect the outcome of a trial. However, this research demonstrate that forensic expert performance data can sway peoples' perceptions of forensic expert credibility. Knowing this, forensic experts may increase the effectiveness of their communication of scientific limitations.

Further, this thesis provides the foundations for future research to explore the effects of expert performance error rates on the outcome of a criminal trial. It is hoped that increasing a juries understanding of forensic limitations will assist in making better informed decisions, thus decreasing the rate of wrongful convictions that directly link to the miscommunication of forensic evidence.

4.9 Concluding Remarks

Forensic experts are introducing their findings in criminal courts without presenting scientific validation, further they are overstating the probative value of their error rates (National Research Council, 2009; PCAST, 2016). This thesis is the first to explore how the presentation of forensic expert performance error rates influence expert credibility. New findings between message framing and colour priming were established within a legal context. These findings suggest that message framing influences peoples' perceptions of forensic expert credibility. Further, that colour acts as a primer for the message frame, magnifying expert credibility ratings. Although this thesis did not find a significant effect in the presentation of mean values or individual data points on a graph, suggestions were made about how future researchers may explore the best way to present expert performance error rates to better inform judges and jurors of the reliability and validity of forensic expert opinions.

As it stands, the findings from this thesis shed light on a new area of research that has the potential to increase the effectiveness of forensic communication in the courtroom. Given that expert performance rates influence peoples' confidence in expert credibility, it is only a matter of time before presenting forensic limitations in the form of performance rates is used in the criminal justice system.

References

Bakker, I., Voordt, T. van der, Vink, P., Boon, J. de, & Bazley, C. (2015). Color preferences for different topics in connection to personal characteristics. *Color Research & Application*, *40*(1), 62–71.

Borkin, M. A., Vo, A. A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., & Pfister, H. (2013). What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics*, *19*(12), 2306–2315.

Brodsky, S. L., Griffin, M. P., & Cramer, R. J. (2010). The Witness Credibility Scale: An outcome measure for expert witness research. *Behavioural Sciences & the Law*, *28*(6), 892–907.

Chien, Y.-H. (2011). Message framing and color combination in the perception of medical information. *Psychological Reports*, *108*(2), 667–672.

Cooper, J., Bennett, E. A., & Sukel, H. L. (1996). Complex Scientific Testimony: How Do Jurors Make Decisions? *Law and Human Behavior*, *20*(4), 379–394.

Curley, L. J., Munro, J., & Lages, M. (2020). An inconvenient truth: More rigorous and ecologically valid research is needed to properly understand cognitive bias in forensic decisions. *Forensic Science International: Synergy*, *2*, 107–109.

Dioso-Villa, R. (2015). A repository of wrongful convictions in Australia: First steps toward estimating prevalence and causal contributing factors. *41*.

Doyle, J. M. (2010). Learning from error in American criminal justice. *The Journal of Criminal Law and Criminology*, *100*(1), 109–148.

Edmond, G., Found, B., Martire, K., Ballantyne, K., Hamer, D., Searston, R., … Roberts, A. (2016). Model forensic science. *Australian Journal of Forensic Sciences*, *48*(5), 496–537.

Eldridge, H. (2019). Juror comprehension of forensic expert testimony: A literature review and gap analysis. *Forensic Science International: Synergy*.

Eysenck, H. J. (1941). A Critical and Experimental Study of Colour Preferences. *The American Journal of Psychology, 54*(3), 385–394.

Faigman, D. L., & Baglioni, A. J. (1988). Bayes' theorem in the trial process. *Law and Human Behavior*, *12*(1), 1–17.

Goodman, J. (1992). Jurors' Comprehension and Assessment of Probabilistic Evidence. *American Journal of Trial Advocacy*, *16*(2), 361–390.

Granger, G. W. (1955). An experimental study of color preferences. *Journal of General Psychology; Provincetown, Mass., Etc.*, *52*, 3–20.

Haley, H., & Sidanius, J. (2006). The positive and negative framing of affirmative action: A group dominance perspective. *Personality and Social Psychology Bulletin*, *32*(5), 656–668.

Haselton, M. G., Nettle, D., & Murray, D. R. (2015). The Evolution of Cognitive Bias. In *The Handbook of Evolutionary Psychology* (pp. 1–20). American Cancer Society.

Hayes-Smith, R. (2009). *The CSI effect: Jurors' perceptions and trial decisions* (Ph.D., University of Florida). University of Florida, United States -- Florida.

Ho, J., Tumkaya, T., Aryal, S., Choi, H., & Claridge-Chang, A. (2019). Moving beyond P values: Everyday data analysis with estimation plots. *BioRxiv*, 377978.

Hovland, Carl I., & Weiss, W. (1951). The Influence of Source Credibility on Communication Effectiveness. *Public Opinion Quarterly*, *15*(4), 635–650.

Hovland, C.I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion*. New Haven, CT, US: Yale University Press.

Hui, C. Y. T., & Lo, T. W. (2017). Examination of the “CSI Effect” on Perceptions of Scientific and Testimonial Evidence in a Hong Kong Chinese Sample. *International Journal of Offender Therapy and Comparative Criminology*, *61*(7), 819–833.

Hurwitz, S. D., Miron, M. S., & Johnson, B. T. (1992). Source credibility and the language of expert testimony 1. *Journal of Applied Social Psychology*, *22*(24), 1909–1939.

Innocence Project. (2020). Overturning Wrongful Convictions Involving Misapplied Forensics. Retrieved August 31, 2020, from Innocence Project website: https://www.innocenceproject.org/overturning-wrongful-convictions-involving-flawed-forensics/

Jonauskaite, D., Mohr, C., Antonietti, J.-P., Spiers, P. M., Althaus, B., Anil, S., & Dael, N. (2016). Most and least preferred colours differ according to object context: New insights from an unrestricted colour range. *PloS One*, *11*(3).

Jurs, A. W. (2016). Expert Prevalence, Persuasion, and Price: What Trial Participants Really Think about Experts. *Indiana Law Journal*, *91*(2), 353–392.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–292.

Kinsey, C. L. (2011). CSI: From the Television to the Courtroom. *Virginia Sports and Entertainment Law Journal*, *11*(2), 313–361.

Koehler, J. J. (2011). If the Shoe Fits They Might Acquit: The Value of Forensic Science Testimony. *Journal of Empirical Legal Studies*, *8*(s1), 21–48.

Koehler, J. J., Schweitzer, N. J., Saks, M. J., & McQuiston, D. E. (2016). Science, technology, or the expert witness: What influences jurors' judgments about forensic science testimony? *Psychology, Public Policy, and Law*, *22*(4), 401–413.

Kovera, M. B., & McAuliff, B. D. (2000). The effects of peer review and evidence quality on judge evaluations of psychological science: Are judges effective gatekeepers? *Journal of Applied Psychology*, *85*(4), 574.

Kramer, R. S. S., Telfer, C. G. R., & Towler, A. (2017). Visual comparison of two data sets: Do people use the means and the variability? *Journal of Numerical Cognition*, *3*(1), 97–111.

Lane, D. M., & Sándor, A. (2009). Designing better graphs by including distributional information and integrating words, numbers, and images. *Psychological Methods*, *14*(3), 239–257.

Martire, K. A., Edmond, G., & Navarro, D. (2020). Exploring juror evaluations of expert opinions using the Expert Persuasion Expectancy framework. *Legal and Criminological Psychology*.

McAndrew, W. P., & Houck, M. M. (2020). Interpol review of forensic science management literature 2016–2019. *Forensic Science International: Synergy*.

McCroskey, J. C. (1966). Scales for the measurement of ethos. *Speech Monographs*, *33*(1), 65–72.

McCroskey, J. C., & Young, T. J. (1981). Ethos and credibility: The construct and its measurement after three decades. *Central States Speech Journal*, *32*(1), 24–34.

McQuiston-Surrett, D., & Saks, M. J. (2009). The testimony of forensic identification science: What expert witnesses say and what factfinders hear. *Law and Human Behavior, 33*(5), 436.

Nance, D. A., & Morris, S. B. (2005). Juror Understanding of DNA Evidence: An Empirical Assessment of Presentation Formats for Trace Evidence with a Relatively Small Random-Match Probability. *Journal of Legal Studies*, *34*(2), 395–444.

National Research Council. (2009). *Strengthening forensic science in the United States: A path forward*. Washington, DC: The National Academies Press.

Neumann, C., Kaye, D., Jackson, G., Reyna, V., & Ranadive, A. (2016). Presenting quantitative and qualitative information on forensic science Evidence in the courtroom. *Chance*, *29*(1), 37–43.

Newman, G. E., & Scholl, B. J. (2012). Bar graphs depicting averages are perceptually misinterpreted: The within-the-bar bias. *Psychonomic Bulletin & Review*, *19*(4), 601–607.

O'Keefe, D. J., & Jensen, J. D. (2006). The Advantages of Compliance or the Disadvantages of Noncompliance? A Meta-Analytic Review of the Relative Persuasive Effectiveness of Gain-Framed and Loss-Framed Messages. *Annals of the International Communication Association, 30*(1), 1–43.

O'Keefe, D. J., & Jensen, J. D. (2007). The Relative Persuasiveness of Gain-Framed Loss-Framed Messages for Encouraging Disease Prevention Behaviors: A Meta-Analytic Review. *Journal of Health Communication*, *12*(7), 623–644.

O'Keefe, D. J., & Jensen, J. D. (2009). The Relative Persuasiveness of Gain-Framed and Loss-Framed Messages for Encouraging Disease Detection Behaviors: A Meta-Analytic Review. *Journal of Communication*, *59*(2), 296–316.

PCAST. (2016). *Report to the president, forensic science in criminal courts: Ensuring scientific validity of feature-comparison methods*. Executive Office of the President of the United States, President's Council of Advisors on Science and Technology.

Peters, E., & Levin, I. P. (2008). *Dissecting the risky-choice framing effect: Numeracy as an individual-difference factor in weighting risky and riskless options*.

Rousselet, G. A., Pernet, C. R., & Wilcox, R. R. (2017). Beyond differences in means: Robust graphical methods to compare two groups in neuroscience. *European Journal of Neuroscience*, *46*(2), 1738–1748.

Schweitzer, N. J., & Saks, M. J. (2007). The CSI Effect: Popular Fiction about Forensic Science Affects the Public's Expectations about Real Forensic Science. *Jurimetrics*, *47*(3), 357–364.

Searston, R., & Chin, J. (2019). The legal and scientific challenge of black box expertise. *Searston, RA, and J. Chin,'The Legal and Scientific Challenge of Black Box Expertise', University of Queensland Law Journal (Forthcoming)*.

Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review*, *14*(1), 47–69.

Smith, B. C., Penrod, S. D., Otto, A. L., & Park, R. C. (1996). Jurors' Use of Probabilistic Evidence. *Law and Human Behavior*, *20*(1), 49–82. JSTOR.

Smith, L. L., Bull, R., & Holliday, R. (2011). Understanding Juror Perceptions of Forensic Evidence: Investigating the Impact of Case Context on Perceptions of Forensic Evidence Strength. *Journal of Forensic Sciences*, *56*(2), 409–414.

Spellman, B. (2018). Communicating Forensic Evidence: Lessons from Psychological Science. *Seton Hall L. Rev.*, *48*(827), 14.

Steiger, A., & Kühberger, A. (2018). A meta-analytic re-appraisal of the framing effect. *Zeitschrift Für Psychologie*.

Takemura, K. (1993). The effect of decision frame and decision justification on risky choice. *Japanese Psychological Research*, *35*(1), 36–40.

Thompson, M. B., Tangen, J. M., & McCarthy, D. J. (2013). Expertise in Fingerprint Identification. *Journal of Forensic Sciences*, *58*(6), 1519–1530.

Thompson, W. C. (2018). How Should Forensic Scientists Present Source Conclusions. *Seton Hall Law Review*, *48*(3), 773–814.

Thompson, W. C., Grady, R. H., Lai, E., & Stern, H. S. (2018). Perceived strength of forensic scientists' reporting statements about source conclusions. *Law, Probability and Risk*, *17*(2), 133–155.

Thompson, W. C., Vuille, J., Taroni, F., & Bidermann, A. (2018). After uniqueness: The evolution of forensic science opinions. *Judicature*, *102*, 18.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 263–292.

Vincent, F. H. (2010). *Report: Inquiry into the circumstances that led to the conviction of Mr Farah Abdulkadir Jama*. Government Printer, South Africa.

Voss Jr, R. P., Corser, R., McCormick, M., & Jasper, J. D. (2018). Influencing health decision-making: A study of colour and message framing. *Psychology & Health*, *33*(7), 941–954.

Weissgerber, T. L., Milic, N. M., Winham, S. J., & Garovic, V. D. (2015). Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. *PLoS Biology*, *13*(4).

Wilcox, M. A., & NicDaeid, N. (2018). Jurors' perceptions of forensic science expert witnesses: Experience, qualifications, testimony style and credibility. *Forensic Science International*, *291*, 100–108.

Yu, L., Westland, S., Li, Z., Pan, Q., Shin, M. J., & Won, S. (2018). The role of individual colour preferences in consumer purchase decisions. *Color Research & Application*, *43*(2), 258–267.

Appendix A.

Participant Information and consent form.

PERCEPTIONS OF FORENSIC EXPERTS

Participant Information

PROJECT TITLE: Perceptions of Forensic Experts
SCHOOL OF PSYCHOLOGY HUMAN RESEARCH ETHICS SUB-COMMITTEE APPROVAL NUMBER: 20/27
PRINCIPAL INVESTIGATOR: Dr Rachel Searston
STUDENT RESEARCHER: Jonica Koodrin
STUDENT'S DEGREE: Honours

Hello,

My name is Jonica Koodrin and I'm an Honours student at The University of Adelaide. Thank you for your interest in participating in this study on how people perceive forensic experts.

Who can participate?
You are able to participate if you are an adult over the age of 18 years who speaks fluent English and has normal or corrected to normal vision.

What is involved?
This study involves a short survey about your perception of forensic experts. In this survey, you will be asked to read reports about three different forensic experts, and answer some questions about their performance. You will also be asked some demographic questions.

Are there any risks associated with participating?
The risks associated with this project are no greater than that of everyday living.

Can I withdraw from the project?
Participation in this project is completely voluntary. If you agree to participate, you can withdraw from the study at any time by exiting out of the browser you are using to take the survey. However, once the survey has been submitted, withdrawal is no longer possible.

What will happen to my information?
All data gathered will be stored on password protected computers and servers. Deidentified data will be made available on the Open Science Framework for other researchers to access. There are no plans to destroy the deidentified data.

Who do I contact if I have questions about the project?
Should you have any questions, you can contact the following people:

[redacted]

If you wish to speak with an independent person regarding concerns or a complaint, the University's policy on research involving human participants, or your rights as a participant, please contact the Convenor, School of Psychology Human Research Ethics Sub-Committee at: paul.delfabbro@adelaide.edu.au. Any complaint or concern will be treated in confidence and fully investigated and you will be informed of the outcome.

Participant Consent Sheet

This study has been approved by the School of Psychology Human Research Ethics Sub-Committee at the University of Adelaide (approval number 20/27). This research project will be conducted according to the NHMRC National Statement on Ethical Conduct in Human Research 2007 (Updated 2018).

If I want to participate, what do I do?
1. I freely consent to participate in the project named above, which is for research purposes.
2. I have had the project and the potential risks and burdens fully explained to my satisfaction by the researchers.
3. I agree to participate in the activities outlined in the participant information sheet.
4. I understand that as my participation is anonymous, I can withdraw any time up until submission of my data for analysis.
5. I have been informed that the information gained in the project may be published in a thesis, conference presentation, website, report, book or journal.
6. I have been informed that in the published materials I will not be identified and my personal results will not be divulged.
7. I consent to any data gathered from this participation to be used for future research purposes and to the data being stored in an online public repository (e.g., Open Science Framework).
8. I understand my information will only be disclosed according to the consent provided, except where disclosure is required by law.

If you understand and agree to the statements above, and freely consent to participating in this study, click "next" to begin the survey.

Yours sincerely,
Jonica