Running head: Pre-trial bias and anchoring in jurors.

Verdictspotting: Investigating the effects of juror bias, evidence anchors, and verdict system in jurors.

Lee J. Curley[1]*, Jennifer Murray[2], Rory MacLean[3], James Munro[1], Martin Lages[4] Lara A. Frumkin[1], Phyllis Laybourn[3], & David Brown[5]

All authors declare no conflict of interests.

1. Faculty of Arts and Social Sciences, School of Psychology and Counselling, the Open University, Milton Keynes, England.

2. The School of Health and Social Care, Edinburgh Napier University, Edinburgh, Scotland.

3. School of Applied Sciences, Edinburgh Napier University, Edinburgh, Scotland.

4. College of Science and Engineering, The School of Psychology, The University of Glasgow, Glasgow, Scotland.

5. Not currently affiliated with a University.

Word count: 8707.

*For author enquiries please contact Dr Lee John Curley, at the Faculty of Arts and Social Sciences, School of Psychology and Counselling, the Open University, Milton Keynes, United Kingdom Email: Lee.Curley@open.ac.uk.

1

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the School of Applied Sciences Research Integrity Committee, Edinburgh Napier University and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent

Informed consent was obtained from all individual participants included in the study

Abstract

The unique Scottish 'bastard verdict' of *not proven* represents a second acquittal verdict which is, by design, not legally defined. Existing research into the influence of the *not proven* verdict on jury decision making is modest, despite the *not proven* verdict's regular appearance in Scottish media and policy discussion. The main aim of the current study was therefore to investigate the influence of verdict categories on juror decision making. The effect of pre-trial bias and evidence anchors on juror judgements were also examined. One-hundred and twenty-eight mock jurors listened to two homicide vignettes and were asked to rate their perception of guilt of the accused throughout two trials and to give a verdict at the end of each trial. The results suggest that pre-trial bias significantly affected both the verdict that was given and the final belief of guilt score. Evidence anchor was not a significant predictor of verdict given or the final belief of guilt score. In relation to verdict system, both *guilty* and *not guilty* verdicts were given with increased frequency in the two-verdict system when compared to the three-verdict system.

Keywords: Not proven verdict; Anchoring and adjustment; Pre-trial biases; Decision science; Juror decision-making; verdict systems; heuristics.

Verdictspotting: Investigating the effects of juror bias and evidence anchors in a two- and three-verdict system

The legal system's principles were developed based on logical, philosophical ideas (Basu, 2006), such as everyone deserving the right to a fair trial. However, research has demonstrated that pre-existing biases and intuitive decision processes may have an effect on how jurors make decisions (Chapman & Bornstein, 1996; Estrada-Reynolds, Gray, & Nunez, 2015); thus, jurors may not be rational, fair agents. The incorporation of such bias within juror decision processes may undermine the legal system's ability to provide justice to society. The current research therefore investigated the effects of pre-trial biases and initial evidence anchors (i.e., a piece of information that has a disproportionate influence on an individual's decision processes) on juror decisions across different verdict-systems, with perceived 'strong' (e.g., DNA, forensic science evidence) versus 'weak' (e.g., ambiguous information, secondary confessions) evidence anchors (Chapman & Bornstein, 1996).

In addition to legal principles, such as the right to a fair trial, two separate verdict systems have evolved in the UK. The Scottish legal system has three verdict options, allowing jurors to choose between giving either a *guilty*, a *not guilty* or a *not proven* verdict. The unique *not proven* verdict has not been legally defined (Duff, 1999), but has the same legal implications as a *not guilty* verdict. However, it has been suggested by Curley, MacLean, Murray, Laybourn and Brown (2019) that the *not proven* verdict represents a belief from a juror that the accused/defendant is probably *guilty* but there was not enough evidence to convict. Further, a large study examining 969 Scottish Jurors (see Ormston, Chalmers, Leverick, Munro, & Murray, 2019) demonstrated that there was confusion about how the *not proven* verdict differed from *not guilty*. The rest of the UK utilises a two-verdict system where only the verdicts of

guilty and *not guilty* are available to jurors. Jurisdictions in the United States of America, New Zealand and Australia also utilise the two-verdict system that originated in England. Such differences across the nations of the UK means that it is difficult to generalise previous research on the effects of pre-trial biases from the Anglo-American system to the Scottish system.

The current study therefore moves beyond individual juror characteristics and cognitive processes by also considering structural aspects of a trial: whether two-verdict versus three-verdict options would affect verdict decisions or considerations of guilt. The current study is, to the authors' best knowledge, the first paper that investigates the effects of individual differences, cognitive structures and varying legal jurisdictions in unison.

Pre-trial biases

Juror judgements can be biased by extra-legal factors, such as previous experience of fictional forensic/legal television shows, a victim's gender and a victim's attractiveness (Maeder, Yamamoto, & Saliba, 2015; Tyler, 2006; Williams, Demuth, & Holcomb, 2007). Previous research has found that pre-trial biases can be used to predict verdict choice in jurors (Lecci & Myers, 2008, 2009). Further, many models of juror decision making - such as the Information Integration Model (Ostrom, Werner, & Saks, 1978), Bayesian models (Marshall & Wise, 1975), and the Story Model (Pennington & Hastie, 1986, 1992) - have incorporated prior beliefs, since it is well documented that prior beliefs can have an impact on final juror judgements (Lecci & Myers, 2008).

The Story Model has been the most dominant model in the juror decision making literature for almost 30 years. This model suggests that jurors create narratives based on the evidence in a trial, pre-trial knowledge relating to similar crimes/trials, and heuristic-based knowledge of what makes a story complete (Pennington & Hastie, 1992; Willmott, Boduszek, Debowska, & Woodfield, 2018). In addition, pre-trial biases can influence the stories that jurors

create (Willmott et al., 2018). Jurors select a story based on certainty principles (coverage, coherence and uniqueness). Uniqueness will explain some of the juror's confidence in their story, with jurors being less confident in stories that are not unique. Verdicts that best align with a story are then chosen (Pennington & Hastie, 1986, 1992). Furthermore, pre-trial biases are likely to influence story construction and verdict outcome (Carlson & Russon, 2001; Pennington & Hastie, 1986, 1992, Willmott, 2018).

The adversarial system of law may increase the influence that pre-trial biases have on jurors. This is because in the adversarial system, two different narratives are presented to the jury, thus creating an ambiguous decision context, where bias dominates (Arkes, & Mellers, 2002; De La Fuente, De La Fuente, & García, 2003; Kaplan & Miller, 1978). Further, scientific evidence, legalese, the possibility of multiple charges (i.e., a defendant facing charges for both homicide and armed robbery), and different verdict systems makes the courtroom a complex place for a juror (Greene & Loftus, 1985; Schklar & Diamond, 1999; Severance & Loftus, 1982). This complexity may reduce the cognitive effort that jurors are able or willing to provide, which may increase the likelihood of jurors relying on pre-trial biases when making a decision.

One route to complexity within a jury trial lies in the comprehension and availability of different verdict choices. As previously mentioned, verdict options are not unified across the various legal systems in the UK, and the differences between seemingly similar verdict options are not always clear. Standard guidance for Scottish criminal procedure states that the *not proven* verdicts meaning should not be defined or described to juries (Gordon, Renton & Brown, 1996). However, the *not proven* verdict does have the same legal consequences as the *not guilty* verdict (Ormston et al., 2019). Ambiguity relating to the definition of a verdict or the differences between verdicts may also increase the influence that pre-trial biases play in the courtroom. Currently, the majority of the research conducted on the effects that pre-trial biases

have on juror verdicts have focussed on the two-verdict system. This means it is currently unknown how pre-trial biases influence juror decisions in the Scottish three-verdict system.

There are many possibly factors that could lead to pre-trial biases. Rather than attempting to summarise or investigate every possible pre-trial bias, the current research will only utilise the pre-trial bias score as measured by Lecci and Myers' (2008) Pre-trial Juror Attitude Questionnaire (PJAQ); see materials for discussion of the constructs in this questionnaire.

Anchoring and Adjustment

Tversky and Kahneman (1974) were the first to study the anchoring and adjustment heuristic. This heuristic occurs because human decision makers have limited cognitive capacity and tend to use initial pieces of information as anchors, particularly in complex environments (Tversky & Kahneman, 1974). However, decision makers do not sufficiently adjust from these anchors when integrating subsequent information in a decision-making process (Tversky & Kahneman, 1974). This means that decision makers who use the anchoring and adjustment heuristic deviate from normative models of decision making. Since the interpretation of novel cues is biased towards the initial anchor (so consequently, clues are not integrated independently) it is unlikely that a rational decision will be reached (Tversky & Kahneman, 1974).

Epley and Gilovich (2006) proposed that individuals use the anchoring and adjustment heuristic when the decision-making environment is uncertain. They suggested that in self-generated anchoring experiments, where the anchor is generated by the participants and not provided by the experimenter, individuals knew that the anchor they produced was incorrect, but they believed that their anchor was a good enough starting point (Epley & Gilovich, 2006). Epley and Gilovich (2006) found that individuals who had a low cognitive load, and were

therefore more able to process information efficiently, were able to adjust more suitably and to a greater extent from their existing anchors. This is because the less cognitive effort one is motivated or willing to expend on a certain task, the less likely one is to adjust rationally; thus, promoting heuristic and/or biased decision-making (Epley & Gilovich, 2006; Simon, 1956).

In a mock case where a woman sued her health-maintenance organisation for their potential involvement in her developing ovarian cancer, Chapman and Bornstein (1996) suggested that the anchoring and adjustment heuristic could be applied to juror decision making. They found that the strength of an evidence anchor (i.e., a piece of evidence used as an anchor) had an impact on liability judgements made during the case. Jurors who were shown strong evidence anchors (high chance of prescribed pill causing cancer) were more likely to perceive the defendant as liable than those shown weak evidence anchors (low chance of prescribed pill causing cancer). Jurors who were shown strong evidence anchors also perceived the defendant to be more to blame for the plaintiff's injuries. The anchoring and adjustment heuristic has also been shown to influence both judges and jurors when estimating financial penalties in civil cases (Greene & Bornstein, 2003, 2013; Guthrie, Rachlinski, & Wistrich, 2002; Orr & Guthrie, 2005). The current study therefore extends on the research of Chapman and Bornstein (1996) by testing whether evidence anchors (strong-first vs. weak-first) influence juror decision making in criminal trials.

Three versus two-verdict system.

Scotland's three-verdict system, with its additional verdict of *not proven*, originated in the 18th century (Barbato, 2004). This verdict system came to be as in the late 17th century Scottish jurors were only allowed to find individual factual allegations proven or not proven rather than to return a general verdict of guilty or not guilty, which was a question for the judge to answer based on those findings (Barbato, 2004). However, in the 18th century, jurors were

allowed, once more, to give verdicts that indicated the guilt (*guilty* and *not guilty*) of the accused. Despite the reintroduction of *guilty* and *not guilty* verdicts, the *not proven* verdict came to be used by jurors as a general verdict (Barbato, 2004), thus birthing the Scottish three-verdict system.

Notwithstanding the ambiguity of the *not proven* verdict, in 2015-2016, of the 7806 people acquitted in Scottish courts 1173 were given the not proven verdict (Scottish Government, 2021). Scepticism towards the *not proven* verdict has lasted for centuries, with Sir Walter Scott christening the *not proven* verdict as "the bastard verdict" because of its lack of legitimacy. More recently, *Miss M* (who successfully sued a man after a criminal trial found him *not proven* of rape) has been campaigning for the removal of the *not proven* verdict (BBC News, 2019). Her campaign was based on the fact that the *not proven* verdict is used disproportionately often in rape trials (Criminal Proceedings in Scotland, 2011). One reasoning for this may be that the availability of two acquittal verdicts may increase the chances of truly *guilty* individuals being acquitted (Rape Crisis Scotland, 2019).

Four previous publications/reports have investigated the effects that the *not proven* verdict may have on juror/jury decision making in an attempt to understand how jurors interpret said verdict. Research has shown that the availability of the not proven verdict is likely to influence the frequency by which *guilty* and *not guilty* verdicts are given. Smithson, Deady and Gracik (2007) found in murder trials that the introduction of the *not proven* verdict decreased the number of *not guilty* verdicts given. Since Smithson et al.'s (2007) study, Hope, Greene, Memon, Gavisk, and Houston (2008) and Curley et al. (2019) have found similar results in sexual assault and murder trials, respectively. Further, there seems to be consensus in the literature that the not proven verdict decreases the number of *not guilty* verdicts given by jurors. Ormston et al. (2019) found that the availability of the *not proven* verdict led to a significantly lower number of *guilty* verdicts being given by jurors in physical assault trials. Such effects

have not been seen as consistently, but the findings of Ormston et al. (2019) may have emerged because of the more realistic nature of the trial stimuli they used when compared to previous experimental investigations. Furthermore, research has shown that the structure of the legal environment (number of verdicts available) may influence the verdict choices of jurors (Ormston et al., 2019).

The way that jurors interpret the not proven verdict may be a key factor that influences juror decisions and consequently affects the frequency at which *not guilty* and *guilty* verdicts are made. Jurors seem to interpret the not proven verdict as a middling verdict, where they may perceive the accused as guilty but are not sure beyond reasonable doubt (Curley et al., 2019; Curley, Munro, Frumkin & Turner, 2021). Therefore, jurors who are just shy of reasonable doubt in two-verdict system may be forced to return a *not guilty* verdict, but in Scotland, they can show their scepticism relating to the innocence of the accused and the Crown's case through returning a not proven verdict (Jackson, 1998). This reasoning may, therefore, explain why the availability of the *not proven* verdict may reduce the number of *not guilty* verdicts jurors return. Further, research suggests that accused individuals who are given the *not proven* (over a *not guilty*) verdict are likely to face worse social sanctions in the community (Hope et al., 2008), highlighting the importance of testing the influence of the three-verdict system on *not guilty* verdict returns.

Legal professionals may view legal standards such as beyond reasonable doubt as objective benchmarks, meaning that conviction frequencies should not be influenced by the number of acquittal verdicts that are available (one versus two; Hope et al., 2008; Jackson, 1998). However, there are a number of reasons that can be given for the findings of Ormston et al. (2019) (i.e., fewer *guilty* verdicts in a three-verdict system). First, the option of *not proven* may direct the attention of jurors to the potential weaknesses of evidence presented (Hope et al.,2008; McKenzie, 1985); more focus on the proof provided rather than on the factual guilt

of the accused (Jackson, 1998). Second, due to the asymmetric dominance effect, jurors may give fewer guilty verdicts as the addition of another option (i.e., adding a *not proven* verdict to a *guilty* and *not guilty* verdict set) may "increase the proportion of alternative choices from the original set" (Hope et al., 2008, p.242). Finally, the three-verdict system may trigger a compromise effect (Hope et al., 2008; Simonson & Tversky 1992), whereby the existence of the *not proven* verdict makes jurors perceive *guilty* and *not guilty* verdicts as exaggerated extremes, and the not proven verdict as a realistic option in the middle (i.e., a compromise). This compromise effect may then decrease the chances of jurors choosing guilty or not guilty verdicts when compared to jurors who do not have the not proven verdict available to them (Hope et al., 2008). This potential reduction in convictions has been cited by Hope et al (2008) as a potential positive as it may decrease the chances of jurors picking incorrect guilty verdicts, and thus saving the Crown financially in relation to compensation claims.

In addition, Ormston et al. (2019) found that the availability of the *not proven* verdict did not significantly influence other factors such as 1) deliberation length, 2) juror participation, 3) accuracy of legal issue discussions, 4) and the number of issues raised relating to the evidence. Furthermore, the influence that the *not proven* verdict may have on juror decision-making may be context- and crime-type dependent.

Previous investigations on juror decision making in Scotland have focussed mostly on Scottish specific issues (e.g., *not proven* verdict) and largely ignored factors that influence jurors all over the world (e.g., pre-trial biases). Further, in previous mock juror studies conducted in Scotland, the Scottish *not proven* verdict and its effects on juror/jury decision processes has been either studied on its own or alongside other unique aspects of Scots law (e.g., 15-person jury). Commonly studied factors that may influence juror decision making (i.e., pre-trial biases) are therefore typically removed from experimental investigations of Scottish juror decisions. Therefore, the current study aims to add to the handful of publications

that exist in the literature by investigating how pre-trial bias influences jurors in both the two- and three-verdict system.

Research Aims and Questions

The aim of this research is to investigate how pre-trial biases, evidence anchors and verdict systems influence juror perceptions of the defendant/accused and decision outcomes. Juror perceptions of the defendant were measured by collecting likelihood of guilt scores after the presentation of each piece of information using a visual analogue scale (see figure.1). Decision outcomes were measured through asking participants to state their final verdict once they had heard all of the evidence.

The hypotheses were:

H_1. Pre-trial bias will be a significant predictor of the verdict that is given.

The more biased towards the prosecution jurors are, the more likely they will be to give a guilty verdict.

H_2. Evidence anchor will be a significant predictor of the verdict that is given.

Strong-first evidence anchors will lead to a higher frequency of guilty verdicts.

H_3. Verdict Systems will be a significant predictor of the verdict that is given.

The two-verdict system will lead to a higher frequency of not guilty verdicts. This hypothesis is derived from the majority of the literature suggesting that the availability of the not proven verdict significantly reduces the frequency by which the not guilty verdict is given (similar findings have not been as consistent with guilty verdicts). We

also believe it is of interest to examine how different acquittal verdicts are used given the different social sanctions associated with each one.

H4. Pre-trial bias will be a significant predictor of the final belief of guilt score. *The more biased towards the prosecution jurors are, the higher the final belief of guilt score will be.*

H5. Evidence anchor will be a significant predictor of the final belief of guilt score. *Strong-first evidence anchors will lead to higher final belief of guilt scores.*

No hypotheses were generated for potential interactions between verdict system, pre-trial bias, and evidence anchors, since no such interactions have been observed in previous research.

Method

Design

Two variables were manipulated within participants: 1) evidence anchor (strong-first versus weak-first); 2) verdict system (two- versus three-verdict systems). Total pre-trial bias score was used as a covariate in the analysis (this score is based on the PJAQ and is explained further in the materials section). Participants were also given two vignettes (vignette one and vignette two); thus, the vignette variable was included in the data analysis to ensure that this variable did not have an extraneous effect.

Two dependent variables were measured. The first dependent variable was the verdict given by the jurors. The verdict that could be given by jurors was dependent on the verdict system to which they were assigned: 1) 'guilty' and '*not guilty*'; or 2) 'guilty', '*not guilty*', and

'*not proven*'. The second dependent variable was the final belief of guilt score, which was measured on an accumulated scale from 0-14; with 0 representing no belief of guilt and 14 representing a total belief that the person is *guilty*; this measure will be explained more in the materials section.

In a pilot study (Appendix 1) it was determined which pieces of evidence to present in the weak-first and strong-first evidence anchor condition. The results of the pilot study suggested that evidence known as a secondary confession (where a third party states that the accused/defendant confessed their guilt; Neuschatz et al., 2012) was perceived as the weakest evidence anchor and DNA evidence was perceived as the strongest evidence anchor by participants. The weak-first evidence anchor was therefore secondary confessions and the strong-first evidence anchor was DNA evidence.

Participants

A total of 128 participants took part (female = 98; male = 30). The mean age of the participants was 24.93 years (SD = 8.02; range 18-61); and 114 participants were students, 13 were not students, and one participant did not record their occupation. To be included, participants had to be eligible to conduct jury duty in the UK. Participant demographics, including ethnicity, were not collected. While this is a potential limitation, these demographics are also not considered in the Scottish or the rest-of-UK jury selection process.

Information about the study and contact details were advertised on social media (Facebook and Twitter). Participants who wished to participate could then contact the researchers and an experimental time slot was allocated to them. Participants were asked to share the contact details for participating in the study with friends and family. The current study therefore used an opportunistic/snowball sampling scheme to recruit participants.

Ethical Approval

Standardised information sheets, consent forms, instruction sheets, debrief questionnaire, and debrief forms were given to all participants, and data were stored in anonymised form, following general BPS guidelines. Ethical approval for the study was granted from the host institution.

Materials

Homicide Vignettes. Two audio-recorded vignettes were used in the current research. Both described homicide trials and were limited to 962 spoken words. An audio playback device was used to play the vignettes to participants. Each of the vignettes followed the same structure: There was an opening statement from the judge, four pieces of prosecution evidence which were all followed by counterevidence from the defence, two closing statements from each of the respective stances, and, finally, the judge who told the jurors to only give a *guilty* verdict if they were sure beyond reasonable doubt. Participants heard one vignette which began with a strong evidence anchor (DNA evidence) and they also heard another that began with a weak evidence anchor (secondary confession); the presentation order of the evidence anchors was counterbalanced over the set of participants. All participants listened to both vignettes which were given by the same actor; variation was on the evidence order only. The other two pieces of prosecution evidence that were presented to the jurors in each of the vignettes (timepoint 4 and 6; see Figure. 2), which were not included as evidence anchors, were a piece of eyewitness testimony and a piece of expert testimony given by a psychologist; the presentation order of each of these pieces of evidence were also counterbalanced over the set of participants. For a more detailed discussion of these vignettes see Curley et al. (2019).

Pre-trial Juror Attitude Questionnaire. The PJAQ is a 29-item questionnaire which is composed of six constructs, each measuring a sperate pre-trial bias: racial bias (four items); innate criminality (four items); social injustice (four items); cynicism towards the defence (seven items); system confidence (six items); and conviction proneness (five items); system confidence and innate criminality share an item. In the current study, participants rated these items on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree; Mobely, 2007), with two items being reverse scored to reduce the risk of response bias. See Table 1 for the possible minimum and maximum scores that participants could give in each of the six constructs. The PJAQ has been used in number of recent juror decision making studies (Allison et al., 2014; Estrada-Reynolds et al., 2015 Lundrigan et al., 2016) and has been shown to outperform other measures, such as the Juror Bias Scale, when predicting verdicts (Lecci & Myers, 2008, 2009). The PJAQ has also been shown to have good internal validity (Lecci & Myers, 2008).

Table 1 about here

Three of the questions used in the PJAQ were slightly adapted in the current study to make the PJAQ more suitable for a Scottish/UK court setting. For example, the question: "If the defendant committed a victimless crime, like gambling or possession of marijuana, he should never be convicted" was changed to: "If the defendant committed a victimless crime, like possession of marijuana, he should never be convicted" since gambling is legal in the UK.

The PJAQ was has been used in previous research and it has consistently been found to have good predictive and internal validity. For example, when assessing the predictive validity of PJAQ, Lecci and Myers (2009) found that the questionnaire predicted: 1) 21% of pre-deliberation verdicts; 2) 15.1% of post-deliberation verdicts; and 3) 7.6% of the variance in

verdict change. Further, Lundrigan, Dhami, and Mueller-Johnson (2016) found that the PJAQ had the highest predictive ability compared to other pre-trial juror bias questionnaires (e.g., juror bias scale). Finally, the PJAQ has been utilised and validated in a number of other studies (Allison, Jung, Sweeney, & Culhane, 2014; Estrada-Reynolds, Gray, & Nuñez, 2015) and may therefore be considered as robust.

Belief in Guilt Scoring System. To measure participants' belief of guilt, a visual analogue scale was created (Figure. 1). This scale was printed on separate papers and was 14cm wide (0 = not guilt 14 = guilty). Participants were asked to mark their belief of guilt on this scale after hearing each piece of evidence (a total of 11 ratings per vignette, each time on a fresh visual analogue scale). Each belief of guilt score was recorded by measuring with a ruler from the arrow of *not guilty* (which symbolised 100% innocent) to the mark that the participants left on the visual analogue scale. The final belief of guilt score was used in the current analysis.

Figure 1 about here

The visual analogue method of recording belief of guilt scores was used because this should be an easy task to perform. Previous measures of perception of guilt have been numerical or likelihood based (Curley, Murray, MacLean, Laybourn, & Brown, 2018; Sommers & Ellsworth, 2000), but it is possible that statistically naïve individuals may have difficulties in understanding terms such as likelihood and probability (Gigerenzer & Hoffrage, 1995). Therefore, it was deemed appropriate to use a visual analogue scale to allow participants to return a belief of guilt judgement. Participants also gave a final verdict at the end of each vignette.

Procedure

Each participant took part in the study individually using a dedicated lab space. They read an information sheet and then completed a consent form and demographics questionnaire. The PJAQ was referred to as a 'legal questionnaire' to participants, as the real title may have indicated to participants that biases were being tested.

The experiment began by the researcher playing the judge's opening statement (here the judge provided the context of the case; who the defendant/accused and victim were, what the crime committed was and when the event occurred) for their first vignette. Then participants were asked to give an initial belief of guilt score using the visual analogue scale (Timepoint 1). Once the participant had marked their belief of guilt, the first piece of prosecution evidence was played on the audio playback device, with half of the participants hearing a weak evidence anchor first (*Mean age = 23.67 (SD = 6.83); Number of females = 51; Number of males = 13)* and half hearing the strong evidence anchor first (*Mean age = 26.17 (SD = 8.91); Number of females = 47; Number of males = 17*). Then, once again, participants marked their belief of guilt (Timepoint 2). After this, participants heard a rebuttal from the defence, who cross-examined the person currently giving evidence for the prosecution and marked their belief of guilt (Timepoint 3). Participants then continued this procedure of listening to information and marking their belief of guilt for a further six pieces of information (Timepoints 4, 6, and 8 = prosecution; 5, 7 and 9 = defence); if the first piece of prosecution evidence presented at Timepoint 2 was a strong anchor, then the last piece of prosecution evidence at Timepoint 8 would be a weak anchor, and vice versa.

Two closing statements (one from each stance) were presented to the participants once they had heard all of the evidence in the vignettes. After each closing statement, participants once again marked their belief of guilt (Timepoints 10 and 11). Before participants were asked

to give a verdict, they heard the following statement: "*Remember, to give a guilty verdict you must think that the defendant was guilty beyond reasonable doubt*".

Once the participants had heard all of the above information, they were asked to give a verdict. Half of the participants (*Mean age = 24.03 (SD = 9.31); Number of females = 49; Number of males = 15*) could give one of two verdicts (guilty or *not guilty*) in their first vignette, whereas the other half (*Mean age = 25.81 (SD = 6.39); Number of females = 49; Number of males = 15*) could give one of three verdicts (guilty, *not guilty* and *not proven*) in their first vignette. Participants who had two verdicts available to them in their first vignette, had three verdicts available to them in their second vignette; and vice versa. Once participants had completed one vignette and made a verdict, they were presented with a second vignette. Both of the vignettes presented to participants used the same procedure, and the presentation order of the two vignettes was counterbalanced across the participants.

Finally, participants read a debrief form and then confirmed whether or not they consented to their data being used. All participants gave their consent.

Results

Data Organisation and Analysis Plan

Descriptive statistics of each of the six PJAQ constructs are reported. Guilt scores averaged across participants are visualised across eleven timepoints of the strong-first and weak-first evidence anchor condition in Figure. 2. In the analyses, we used verdict given (categorical) and final belief of guilt score (continuous) as two dependent variables and verdict system (two-verdict system vs. three-verdict system), pre-trial bias (PJAQ score), evidence anchor (strong-first, weak-first), and vignette (1 and 2) as independent variables. Comprehensive model comparisons in terms of Akaike Information Criterion (AIC) indicated that mixed-effect models with by-subject random intercepts provided the most parsimonious

model fit for the linear regression on the final belief of guilt scores whereas standard models (without random by-subject intercepts) gave the most parsimonious fits for logistic regressions on the *guilty* and *not guilty* verdicts. Although we did not postulate any hypotheses about interactions, we included all interaction effects into the models.

Focussing on number of *guilty* verdicts we conducted a logistic regression by combining *not proven* verdicts with *not guilty* verdicts into a new category of *acquittal* verdicts. Similarly, focussing on the number of *not guilty* verdicts, we combined *not proven* verdicts with *guilty* verdicts into a new category of *other* verdicts. Hope et al. (2008) used the same categorisations when investigating verdict system (two and three categories) on number of verdicts given.

The *not proven* response category was only available in one of the two vignettes presented to each participant. The order of the two vignettes was counter-balanced across participants in terms of verdict system, vignettes, and evidence anchor in a 2 by 2 by 2 within-subjects design with repeated measures and total pre-trial bias PJAQ as a covariate.

Descriptive statistics for Pre-trial Bias constructs and Guilty Scores

Descriptive statistics were computed for each of the six constructs within the PJAQ (Table 2).

Table 2 about here

The scores from the six constructs were added to create a total pre-trial bias score ($M = 76.30$; $SD = 9.81$; range 54-101). The scoring of the PJAQ was conducted in accordance with previous research (Estrada-Reynolds et al., 2015; Lecci & Myers, 2008, 2009; Lundrigan et al., 2016). Racial bias was included in the total pre-trial bias score for two reasons. First, in other studies, which did not mention race, racial bias was also included in the PJAQ score

(Estrada-Reynolds et al., 2015). Second, audio-recorded vignettes were presented to participants that did not explicitly reveal the race of the defendant or accused.

In Figure 2 mean belief of guilt scores, averaged across vignettes and participants, are plotted for all eleven timepoints. There was considerable noise in the scores across participants. The lines connect the mean guilty scores for the weak-first (dashed line) and strong-first (solid line) evidence anchor condition. The values at Timepoint 1 reflect pre-trial bias (e.g., before any evidence was presented) and suggest a slightly higher belief of guilt score in the weak-first compared to the strong-first evidence anchor condition. The zig-zag lines reflect oscillations in belief of guilt after hearing first the prosecution and then the defence statement in alternation from Timepoints 2 and 3 to Timepoints 8 and 9. Both lines show a similar trend of increasing guilt scores up to Timepoint 8 and 9. In contrast to the experimental manipulation of evidence anchor the dashed line (weak-first) shows the largest peak at Timepoint 2 whereas the solid line (strong-first) shows the largest peak at Timepoint 8. Both evidence anchor conditions then assume a similar score just above 7 at Timepoint 9. Timepoints 10 and 11 reflect the effect of the summary statement of prosecution and defence on guilt scores. At Timepoint 11 the guilt scores show a similar pre-trial difference between conditions as at Timepoint 1. Please note that adding the mean guilt scores across timepoints leads to an overall larger total guilt score for the weak-first compared to the strong-first evidence anchor condition.

Figure 2 about here

The effect of pre-trial bias, evidence anchor and verdict system on juror decision making

We tested whether total pre-trial bias (PJAQ), evidence anchor and vignette predict final belief of guilt scores and whether total pre-trial bias (PJAQ), evidence anchor, vignette, and verdict system predict the categorical variable verdict. Linear regressions with mixed-effects and two logistic regression models (R-package lme4, Bates et al., 2014) were established to

predict final belief of guilt scores and the categorical verdicts, respectively. Descriptive statistics of dependent variables are presented in Tables 2 and 3.

For the final guilt scores, we conducted linear regressions with pre-trial bias (PJAQ) as a covariate and repeated measures for evidence anchor, and vignette. We also included by-subject intercepts as a random effect. The results of linear regressions with mixed effects (R-package lme4, Bates et al., 2014) are reported. The linear regression on the final guilt scores (Timepoint 11) gave the following findings in relation to the ability of the model to fit the data: marginal R^2 =0.07, conditional R^2 =0.41. In the following we report the results of the mixed-effect models for the final guilt scores as dependent variable, with reported *p*-values based on Satterswaithe's *t*-statistic with adjusted degrees of freedom.

For the final guilt scores (Range 0-14) the estimated intercept was 6.85. As in the analysis for total guilt score there were statistically significant fixed effects of PJAQ (*ß=0.0695, SE=0.025,* $t(124)$ =2.75, *p*<.007) and vignette (*ß=-0.505, SE=0.161,* $t(124)$ =-3.13, *p*=.002) but not for anchor (*ß=-0.165, SE=1.61,* $t(124)$ =-1.02, *p* =.308). No significant interactions were observed.

Table 3 about here

In addition, a logistic regression highlighted on guilty verdicts (vs other) revealed that the final guilt scores significantly predicted guilty verdicts (*ß=0.666, SE=0.128, z=4.802, p*<0.0001), highlighting that the measure had an adequate predictive ability.

Table 4 about here

The first logistic regression investigated the effect of total pre-trial bias (PJAQ), verdict system, vignette and evidence anchor on the probability of *guilty* vs *acquittal* verdicts. This model featured all interaction terms (Adjusted R^2 =0.21). The continuous predictor variable PJAQ was centred on the mean and the three categorical variables were deviation coded to facilitate the interpretation of coefficients and effects. In the following the reported *p*-values are based on Wald *z*-statistics for the logistic regressions (see Appendix 2 for details).

The significant intercept (z =-6.99; p <.0001) of the logistic regression model suggests that *guilty* verdicts occurred significantly less often than *acquittal* verdicts. The intercept of -1.20 log odds (odds ratio 0.30) suggests that *guilty* verdicts occurred with probability 0.23 and *acquittal* verdicts with (1-0.23) = 0.87. Verdict system had a significant effect on *guilty* verdicts (z =-2.33; p =.02). The coefficient of -0.80 log odds (odds ratio 0.45) suggests that on average 55% less *guilty* verdicts were given when the *not proven* option was available, provided all other variables remained constant. Total pre-trial bias (PJAQ) was also a significant predictor (z =2.82; p =.005). The coefficient of 0.053 log odds (odds ratio 1.052) suggested that with every unit on the total pre-trial bias score the likelihood for a *guilty* verdict increased by 5.2%, given all other predictor variables remained constant. Evidence anchor only approached statistical significance (z =-1.82; p =.07) and vignette was not significant (z=-1.21; p=0.23). Finally, the interaction between anchor and vignette was significant (z =-2.22; p =.03). No other significant interactions were observed.

The second logistic regression was conducted to evaluate the effects of total pre-trial bias (PJAQ), verdict system, and evidence anchor on *not-guilty* verdicts (*not-guilty* vs. *other*). The model featured all main effects and interaction terms (Adjusted R^2 =0.44). As in the previous analysis, the continuous predictor variable PJAQ was centred on the mean and the three categorical predictors were deviation coded. All reported *p*-values are based on Wald *z*-statistics.

The significant intercept (z = -2.85; p =.004) of the logistic regression model suggests that participants gave *not guilty* verdicts significantly less often than *other* verdicts. The intercept of -0.54 log odds (odds ratio 0.58) suggests that *not guilty* verdicts occurred with probability 0.37 and *other* verdicts with (1.0-0.37) 0.63. Verdict system had a significant effect on *not-guilty* verdicts (z =-7.10; p <.0001). A coefficient of -2.67 log odds (odds ratio 0.069) predicts that about 93.1% less *not guilty* verdicts were given when the *not proven* option was available, provided all other variables remained constant. Total pre-trial bias was also a significant predictor (z =-3.32; p =.001). The coefficient of -0.066 log odds (odds ratio 0.935) suggests that with every unit increase of the total pre-trial bias score the likelihood for a *not-guilty* verdict decreased by 6.5%, given all other predictor variables remain constant. The effect of evidence anchor (z =0.26; p =.799) and vignette (z =0.64; p =.869) were not statistically significant. However, the interaction between vignette and verdict system reached statistical significance (z =-2.42; p =.015). No other interactions were significant.

Discussion

The current research investigated the effects of pre-trial bias, anchoring, and different verdict systems on juror decision-making. It explored both belief in guilt and final verdict outcomes. To the authors' best knowledge, this is the first paper that investigates the effects of varying legal jurisdictions, pre-trial bias and evidence anchor in unison. Further, there are only a handful of empirical investigations into the effect of Scotland's unique *not proven* verdict on juror decisions. To this end, the current study adds significantly to the understanding of the influence of verdict options on juror decisions alongside the effects of pre-trial biases and evidence anchors on decisions in the court room.

The discussion is structured in line with the study design, with each of the factors being examined in turn. First, the effects of pre-trial biases on juror perceptions of guilt and the

verdicts given will be presented. A discussion surrounding the influence that evidence anchors had on juror perceptions of guilt and the verdicts given will then follow. Finally, the effects of different verdict systems on juror decision making will be outlined. General findings will then be presented, outlining how the current research can be interpreted alongside the Story Model. Future research will be outlined towards the end of this discussion.

Pre-trial bias effect on jurors

Pre-trial bias was found to be a significant predictor of the verdict that was given and the final belief of guilt score, supporting H_1 and H_4, respectively. The current study therefore provides evidence of pre-trial biases leading to pre-deliberation distortion (Carlson & Russo, 2001). This is because pre-trial biases in jurors may inform how jurors perceive the defendant/accused, which may then influence juror verdicts. Previous research has suggested similar effects (Hope, Memon, & McGeorge, 2004).

In a seminal paper by Carlson and Russo (2001) it was reported that mock jurors naturally developed leading verdicts (i.e., preferences towards either *guilty* or *not guilty*). These leading verdicts then distorted how jurors evaluated novel pieces of evidence. For example, mock jurors with a *guilty* leading verdict, would distort evidence from the defence in favour of their current leading verdict. Hope et al. (2004) found similar results, with negative pre-trial publicity causing jurors to favour the prosecution and consequently distort the evidence in favour of their preference. The results in the current study add to this body of literature, highlighting that pre-trial biases may lead to preferences towards a particular stance (prosecution vs. defence), which may then influence how guilty the accused/defendant is perceived to be throughout the trial, and these distorted perceptions may then influence verdict choice.

One implication of the current findings may be that in legal jurisdictions (e.g., the American two-verdict legal system) where voir dire (jury selection) is possible, the PJAQ might be a useful tool in allowing legal professionals to make accurate jury selection judgements (Lecci & Myers, 2009; Lundrigan et al., 2016). Under these jurisdictions, legal professionals select (directly or not) jurors according to their own beliefs and experiences (Lundrigan et al., 2016). However, these selection strategies are likely to be unreliable and invalid (Posey & Wrightsman, 2005). Therefore, the PJAQ may offer a more reliable and valid method of selecting jurors (Lecci & Myers, 2009; Lundrigan et al., 2016). In other jurisdictions, where voir dire is not permissible (e.g., the UK/Scotland), the current findings may suggest that the PJAQ could be used to filter out biased jurors, thus attenuating the influence that pre-trial biases have on juror judgements. The current research is the first of its kind to highlight that pre-trial biases are as relevant in Scottish three-verdict juror judgements as in two-verdict judgements.

Anchoring effects on jurors

Evidence anchors were found not to be a significant predictor for the final belief of guilt score (H5 not supported) or the verdict given (H2 can therefore be rejected). These findings demonstrate that evidence anchors do not influence juror verdict choices or juror perceptions of the defendant/accused.

Nevertheless, it should be mentioned here that the evidence anchors in this study had the opposite effect from what was expected. Strong-first evidence anchors (DNA evidence) led to lower guilt ratings at time-point two when compared to weak-first evidence anchors (Secondary confessions), despite the strength of these anchors being initially established in the pilot study. One explanation for this finding is that secondary confessions (used in the current study as a weak anchor) may give jurors some context surrounding the case (e.g., why or how

26

the defendant/accused committed the crime). DNA evidence (used in the current study as a strong anchor), on the other hand, may merely highlight that the defendant/accused has some connection with the victim. Further, DNA evidence may be "cold", whereas a secondary confession might be more emotive and thus more likely to bias jurors towards more guilty perceptions of the defendant/accused (Simon, Stenstrom, & Read, 2015). Future research is needed to further explore this potential explanation, however.

The findings of the current research also deviate from some of the findings by Chapman and Bornstein (1996), who found that evidence anchors could be used to anchor both causality judgements and liability ratings in civil court cases. In their study, individuals provided with strong anchors gave larger causality judgements and were more likely to perceive the defendant as liable in comparison to individuals who were shown weak anchors (Chapman & Bornstein, 1996). Although, the current study was not a direct replication of Chapman and Bornstein (1996), the measures used here did mirror the measures used by said authors. Liability ratings were categorical (liable, not liable) which is similar to the verdict given variable used in the current study (*guilty*, *not guilty*, *not proven*). Their causality judgements were also similar to the final belief of guilt measure used in the study as both responses measured the juror's perception of the degree of the defendant's/accused's role in the wrongdoing actioned upon the complainer/plaintiff. Furthermore, the current study's findings deviate from Chapman and Bornstein's (1996) in relation to the effects that evidence anchors may have on decision outcomes and perception of guilt.

An explanation for this difference in findings may relate to an interaction between cognitive mechanisms, evidence weights and environmental differences (e.g., different court dynamics and rules). Kaplan and Miller (1978) suggested that the impact that certain biases or evidence had in a trial was influenced by the weights of other pieces of evidence. In other words, a strong anchoring bias (either towards the prosecution or defence) may be attenuated

by a subsequent piece of evidence that was weighted strongly. Interestingly, in the weak-first evidence anchor condition, the largest spike in belief of guilt score was at Timepoint 2, when the secondary confession was presented. Likewise, in the strong-first evidence anchor condition, the largest spike in belief of guilt score was at Timepoint 8, when the secondary confession was presented. After Timepoint 8, the belief of guilt score was similar in both conditions.

These findings may highlight that secondary confessions increased perceptions of guilt as they provided context on the defendant/accused and the crime. However, the influence of evidence anchors (or particularly strong pieces of prosecution evidence) on jurors may be attenuated in a trial through the presentation of other pieces of evidence with strong weights (e.g. strong evidence that favours the defence) (see Figure. 2), meaning that by the time a juror is able to reach a verdict, the effects of evidence anchors of decision outcomes will be minimal and non-significant.

This interaction between anchoring and evidence weight described above may explain the difference in findings between the current research and that of Chapman and Bornstein (1996) when the environmental setting of each study is taken into account. The current research was based in the criminal court, where the beyond reasonable doubt standard is applied, whereas Chapman and Bornstein's (1996) study was conducted in the civil court using a lower standard of proof (i.e., the balance of probabilities). Furthermore, when strong evidence attenuates the effects of evidence anchors in the criminal court, the higher standard of proof may further reduce the effect that anchoring has on verdict outcomes. However, in the civil court, where the standard of proof is lower, early evidence anchors may be more likely to influence the final outcome. In other words, subtle anchoring effects may be more likely to influence decisions in civil court trials.

An alternative, yet related, explanation may relate to the fact that Chapman and Bornstein's (1996) study only presented participants with a one-sided piece of paper describing the trial, meaning that there was a limited space for subsequent pieces of information to attenuate the effects of the evidence anchor. In the current study four pieces of prosecution evidence (each with a rebuttal from the defence) and a closing statement from each stance were presented. This would have increased the chances for the respective pieces of evidence to dilute the effects of the evidence anchor on the decision outcomes of the jurors. Other cognitive fallacies and heuristics, such as the overconfidence effect, have also been found to disappear once participants are tested in a less artificial setting (Gigerenzer, 1991; Weiten & Diamond, 1979). In summary, variations in degree of ecological validity and/or type of trial may explain the differences found in the current research when compared with Chapman and Bornstein's (1996) study.

Verdict Systems

The verdict systems variable was a significant predictor of the verdict that was given, allowing H_3 to be accepted. Interestingly, the results highlighted that the availability of the *not proven* verdict led to a significant decrease in the number of *guilty* and *not guilty* verdicts given by jurors; this effect led to a larger reduction in the likelihood of giving a *not guilty* verdict, however.

Explanations as to why the availability of *not proven* verdict may decrease *not guilty* and *guilty* verdicts will now be discussed. First, jurors in a three-verdict system may be less likely to give a *not guilty* verdict in comparison to jurors in a two-verdict system as the not proven verdict may represent the mindset of most jurors more adequately (i.e., they do not think the crown proved their case, but they also do not think the accused is innocent; Jackson, 1998). Second, jurors in a three-verdict system may be less likely to give a *guilty* verdict in comparison

to jurors in a two-verdict system due to: 1) the availability of the *not proven* verdict focusing jurors to the weaknesses of the evidence; 2) the *not proven* verdict being used as a compromise by a juror when *guilty* and *not guilty* verdicts seem poles apart; 3) addition of the *not proven* verdict causing an asymmetric effect that decreases the proportion of both *guilty* and *not guilty* verdicts (Hope et al., 2008; McKenzie, 1985; Simonson & Tversky 1992). Further research would like to tease apart which explanation is most likely.

Practically the current findings highlight that a trial in Scotland will have a decreased probability of a guilty verdict being given in comparison to if the same trial was held in comparable jurisdictions with two verdicts (e.g., England or Wales). Similarly, truly innocent individuals will be less likely receive a *not guilty* verdict in Scotland compared to those tried in England and Wales. Despite the *not proven* verdict having the same legal outcomes as the *not guilty* verdict, individuals who receive the former may be more likely to face social sanctions (Hope et al., 2008). The implications of these findings are complex. For instance, the decreased *guilty* verdict frequency may decrease false positives in the conviction rate, equally though they may increase the rate of false negative acquittals. Further, a decrease in the likelihood of receiving a *not guilty* verdict may stain a truly innocent person's record, but may also cast doubt on the innocence of a truly *guilty* individual who would have been given a *not guilty* verdict if the *not proven* verdict was not available. We therefore urge the Scottish legal community to consider these findings and reflect upon the best possible course of action.

General discussion

The findings of the current study lend some support for the Story Model (Pennington & Hastie, 1986). As previously mentioned, in the Story Model, a narrative is created from three main aspects: 1) case specific information; 2) knowledge of how stories are created; and, 3) knowledge of similar events. Two of these three pieces of information that form the narrative

that jurors create have their origins in pre-trial information. Likewise, previous research has shown that evidence evaluation is influenced by preferences towards certain verdicts (Carlson & Russo, 2001; Hope et al., 2004). Therefore, pre-trial biases may also influence juror story construction, meaning that evidence distortion may be the consequence of evidence being interpreted in line with pre-trial biases in order for a coherent narrative to be produced (Carlson & Russo, 2001).

Evidence anchors (such as secondary confessions) may also set the scene and provide jurors with context surrounding the defendant/accused and the trial, which may then influence how novel information is interpreted. However, the impact of evidence anchors on verdict choices may be attenuated in a trial through the presentation of novel pieces of evidence with strong evidence weights. Therefore, for the creation of a narrative surrounding a trial, the presence, rather than the position, of a piece of evidence may be the important factor, as stories will be created through the integration of several pieces of evidence.

Nevertheless, once narratives have been created, they are then selected based on certainty principles (Pennington & Hastie, 1986). The verdict that best aligns with the selected story is then chosen. In the two-verdict system, the best fitting verdict (and thus the verdict chosen) will either be *guilty* or *not guilty*. However, in the three-verdict system, the best fitting verdict for some stories might be the *not proven* verdict, causing jurors to be less likely to select *guilty* and *not guilty* verdicts. This may then highlight that a binary verdict system does not give a full enough representation of guilt for jurors to match their stories onto. Therefore, the *not proven* verdict may allow jurors to communicate their belief of guilt more accurately to the courtroom (Curley et al., 2019). Nevertheless, if pre-trial biases (e.g. innate criminality, racial biases and rape myth; Lecci & Myers, 2009; Wimott, 2016) are informing how narratives are formed, the availability of the *not proven* verdict, as the best fitting verdict for a narrative, is problematic (Ormston et al., 2019). Because of this, the Scottish legal community should

review the current paper and make recommendations relating to the usefulness of the *not proven* verdict. Furthermore, the current study has highlighted how the Story Model can incorporate both different verdict systems and the influence of pre-trial biases within it.

Limitations

One potential limitation of the current study was that authors decided to utilise a within-subjects design for the evidence anchor and verdict system conditions. As the mock jurors in the current study saw two trials, their experience will be unlike real jurors who would have only witnessed one trial. The researchers believed that an initial study that used verdict system as a within-subjects element was needed, as individual difference surrounding the knowledge of the not proven verdict may have acted as an extraneous variable in a between-subjects design. This is due to the fact that the majority of TV court-based dramas aired within Scotland are set within English or American (two verdict) courtrooms. Thus, a significant proportion of the Scottish population do not know about the existence of the not proven verdict and/or do not fully understand its meaning (Hope et al., 2008; Ross, 2019). Future research should replicate the current study using more ecologically valid materials in a between-subjects design to establish if findings are consistent in more realistic settings.

A second potential limitation of the current study was that variables such as ethnicity were not collected. This was done for two reasons. First, previous studies (e.g., Hope et al., 2008) had not collected information on race. Second, the research team believed that it was unethical to ask participants further questions that we did not believe to be pertinent to the research question and which would not be analysed. However, future research should collect

participant ethnicity data to allow those descriptive data to be presented alongside other findings about demographics.

Future research

One line of enquiry from the current study relates to how other additional verdicts may influence verdict choice. The current study found that the inclusion of the *not proven* verdict significantly decreases the frequency of both *not guilty* and *guilty* verdicts. Future research could investigate how other additional verdicts, such as the *Guilty but Mentally Ill verdict* and the *diminished responsibility verdict* (Smithson et al., 2007), influence verdict choice. In addition, other jurisdictions have different legal proceedings. In Spain, for example, the judge presents a series of propositions and the jury declares each of the said propositions proven or not proven (Thaman, 1999). A final avenue for future research might be to add a condition to the factor of verdict system, in which participants could give either a proven and/or *not proven* verdict. This condition may more adequately reflect how legal professionals perceive the role of the jury in the courtroom (Curley et al., 2021; Jackson, 1998).

Conclusion

The current study presents a novel approach to investigating pre-trial biases, anchors, and verdict systems within juror decision-making. Three main findings were established. First, it was found that pre-trial bias had a significant effect on 1) final belief of guilt score; 2) *guilty* verdicts; and 3) *not guilty* verdicts. Second, evidence anchors did not significantly predict verdicts or the final belief of guilt score. One explanation for this is that evidence anchors may anchor beliefs of guilt, leading to guilt adjustments with each piece of novel information being influenced by the previous piece of information. However, as a trial progresses, and other

pieces of evidence that are weighted strongly are presented, the influence that evidence anchors have on jurors and their ensuing verdict will be reduced. Third, it was found that jurors were more likely to give both *guilty* and *not guilty* verdicts in the two-verdict system in comparison to the three-verdict system.

These findings can be interpreted through the Story Model. Pre-trial biases may set the scene for the narrative. Subsequent evidence, and associated perceptions of the defendant/accused, may then be interpreted in a manner that fits the opening narrative created by their pre-trial biases. The verdict system may then play a role in relation to which verdict best matches their constructed, and chosen, story. In a two-verdict system, the narratives can only lead to *guilty* or not-guilty verdicts. However, the addition of a third (non-legally defined) *not proven* verdict may better fit some constructed stories, meaning the *not proven* verdict may allow individuals to communicate their belief of guilt more accurately. However, when pre-trial beliefs, such as rape myths, influence chosen narratives, the availability of the *not proven* verdict is more problematic. The topic of the unique Scottish verdict is regularly discussed but research of its influence on jury decision making is in its infancy. Further research exploring the influence of different verdict systems on jury decision making is encouraged by the authors.

References

Allison, M., Jung, S., Sweeney, L., & Culhane, S. E. (2014). The impact of illegal alibi activities, corroborator involvement and corroborator certainty on mock juror perceptions. *Psychiatry, Psychology and Law*, *21*(2), 191-204.

Arkes, H. R., & Mellers, B. A. (2002). Do juries meet our expectations*? Law and Human Behavior*, 26(6), 625-639. doi:10.1023/A:1020929517312

Ask, K., Rebelius, A., & Granhag, P. A. (2008). The "elasticity" of criminal evidence: A Moderator of Investigator Bias. *Applied Cognitive Psychology*, *22* (1), 1245–1259. doi: 10.1002/acp.1432

Barbato, J. M. (2004). Scotland's Bastard Verdict: Intermediacy and the unique three-verdict system. *Ind. Int'l & Comp. L. Rev.*, *15*, 543-581.

Baron, R. M., & Kenny, D. A. (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.

Bates D, Mächler M, Bolker B, Walker S (2015). "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software*, **67**(1), 1–48. doi: 10.18637/jss.v067.i01.

BBC News (2019, May 1st). *Miss M case: Rape victim who sued attacker condemns bankruptcy move*. Retrieved from: https://www.bbc.co.uk/news/uk-scotland-edinburgh-east-fife-48117100

Borum, R., Otto, R., & Golding, S. (1993). Improving clinical judgement and decision making in forensic evaluation. *Journal of Psychiatry and Law*, 21, 35-76. doi:10.1177/009318539302100104

Carlson, K. A., & Russo, J. E. (2001). Biased interpretation of evidence by mock jurors. *Journal of Experimental Psychology: Applied*, *7*(2), 91- 103.

Chambliss, W. J., & Seidman, R. B. (1971). *Law, order, and power*. Reading, MA: Addison-Wesley.

Chapman, G. B., & Bornstein, B. H. (1996). The more you ask for, the more you get: Anchoring in personal injury verdicts. *Applied cognitive psychology*, *10* (6), 519-540. doi: 10.1002/(SICI)1099-0720(199612)10:6<519::AID-ACP417>3.0.CO;2-5

Criminal Justice System (2002). *Justice for All* (No. CM 5563). London, England: Criminal Justice System. Retrieved from: https://www.cps.gov.uk/publications/docs/jfawhitepaper.pdf

Curley, L. J., MacLean, R., Murray, J., Laybourn, P., & Brown, D. (2019). The bastard verdict and its influence on jurors. *Medicine, Science and the Law*, *59* (1), 26-35. doi: 10.1177%2F0025802418811740

Curley, L. J., Munro, J., Frumkin, L., Turner, J. (in press). Informing Reform: The views of legal professionals on the unique aspects of Scottish Law. *Medicine, Science and the law.*

Curley, L. J., Murray, J., MacLean, R., Laybourn, P., & Brown, D. (2018). Faith in thy threshold. *Medicine, Science and the Law*, *58*(4), 239-250.

Criminal Proceedings in Scotland (2011). Statistical Bulletin: Crime and Justice Series: Criminal Proceedings in Scotland, 2010-11. Retrieved from:

https://www.gov.scot/publications/statistical-bulletin-crime-justice-series-criminal-proceedings-scotland-2010-11/pages/5/

De La Fuente, L., De La Fuente, I., & García, J. (2003). Effects of pretrial juror bias, strength of evidence and deliberation process on juror decisions: new validity evidence of the Juror Bias Scale scores. *Psychology, Crime & Law, 9* (2), 197-209. doi: 10.1080/1068316031000116283.#

Duff, P. (1999). The Scottish criminal jury: A very peculiar institution. *Law and Contemp. Probs.*, *62*, 173-201.

Englich, B., Mussweiler, T., & Strack, F. (2005). The last word in court - A hidden disadvantage for the defence. *Law and Human Behavior*, *29*(6), 705-722. doi:10.1007/s10979-005-8380-7

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic why the adjustments are insufficient. *Psychological science*, *17*(4), 311-318. Retrieved from: http://journals.sagepub.com/home/pss

Estrada-Reynolds, V., Gray, J., & Nuñez, N. (2015). Information integration theory, juror bias, and sentence recommendations captured over time in a capital trial. *Applied Cognitive Psychology, 29* (7). 713–722. doi: 10.1002/acp.3155

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175-191. doi:10.3758/BF03193146

Field, A. (2013). *Discovering statistics using IBM SPSSS statistics* (4 ed.). London, England: Sage Publications.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases". *European Review of Social Psychology, 2*, 83-115.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency formats. *Psychological review*, *102*(4), 684-704.

Gliem, J. A., & Gliem, R. R. (2003). Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. *Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education, The Ohio State University.* 1-7. Retrieved from: http://hdl.handle.net/1805/344

Goldhill, O. (2014, September 4). When innocent men go to jail: Miscarriages of justice in Britain. *The Telegraph*, retrieved from: http://www.telegraph.co.uk/

GH Gordon, *Renton and Brown: Criminal Procedure according to the Law of Scotland,* 6th edn (1996) para 18-79.41.

Greene, E., & Bornstein, B. H. (2003). *Determining damages: The psychology of jury awards.* Washington: The American Psychological Association. doi: 10.1037/10464-000

Greene, E., & Bornstein, B. H. (2013). Nudging the justice system toward better decisions. *The Journal of Criminal Law and Criminology*, *103* (4), 1155-1170. Retrieved from: http://www.jstor.org/stable/43895390

Greene, E., & Loftus, E. F. (1985). When crimes are joined at trial. *Law and Human Behavior*, *9*(2), 193-2017. doi: 10.1007/BF01067051

Guthrie, C., Rachlinski, J. J., & Wistrich, A. J. (2002). Judging by heuristic-cognitive illusions in judicial decision making. *Judicature*, *86* (1), 44-50. Retrieved from: https://home.heinonline.org/

Heller, K. J. (2006). The cognitive psychology of circumstantial evidence. *Michigan Law Review*, *105* (2), 241-305. Retrieved from: http://www.jstor.org/stable/40041577

Hope, L., Greene, E., Memon, A., Gavisk, M., & Houston, K. (2008). A third verdict option: Exploring the impact of the Not Proven verdict on mock juror decision making. *Law and human behavior*, *32* (3), 241-252. doi:10.1007/s10979-007-9106-8

Hope, L., Memon, A., & McGeorge, P. (2004). Understanding pretrial publicity: Predecisional distortion of evidence by mock jurors. *Journal of Experimental Psychology: Applied*, *10*(2), 111-119. doi:10.1037/1076-898X.10.2.111

Jackson, B. S. (1998). Truth or proof?: The criminal verdict. *Revue internationale de semiotique juridique*, *11*(3), 227-273.

Kaplan, M. F., & Miller, L. E. (1978). Reducing the effects of juror bias. *Journal of Personality and Social Psychology*, *36*(12), 1443-1455. doi:10.1037/0022-3514.36.12.1443

Kassin, S. M., & Wrightsman, L. S. (1983). The construction and validation of a juror bias scale. *Journal of Research in Personality*, *17*(4), 423-442. doi:10.1016/0092-6566(83)90070-3

Lundrigan, S., Dhami, M. K., & Mueller-Johnson, K. (2016). Predicting verdicts using pre-trial attitudes and standard of proof. *Legal and Criminological Psychology*, *21*(1), 95-110.

Lecci, L., & Myers, B. (2008). Individual differences in attitudes relevant to juror decision making: development and validation of the pretrial juror attitude questionnaire (PJAQ) 1. *Journal of Applied Social Psychology*, *38*(8), 2010-2038. doi: 10.1111/j.1559-1816.2008.00378.x

Lecci, L., & Myers, B. (2009). Predicting guilt judgements and verdict change using a measure of pretrial bias in a videotaped mock trial with deliberating jurors. *Psychology, Crime & Law, 15* (7), 619-634. doi: 10.1080/10683160802477757

Lemieux, J. M. (2007). *The relationship between beliefs, strength of evidence, statistical presentation, and expert testimony on jury decision-making in DNA cases* (Doctoral dissertation, Carleton University).

Maeder, E., Yamamoto, S., & Saliba, P. (2015). The influence of defendant race and victim physical attractiveness on juror decision-making in a sexual assault trial. *Psychology, Crime & Law, 21* (1), 62-79, doi: 10.1080/1068316X.2014.915325

Marshall, C. R., & Wise, J. A. (1975). Juror decisions and the determination of guilt in capital punishment cases. In D. Wendt & C. Vlek (Eds.), *Utility, probability, and human decision making* (pp. 257–270). Dordrecht, The Netherlands: Reidel.

McKenzie, J. P. (1985). The Editorial Notebook; Between guilt and innocence. New York Times, A28.

Mobely, S. (2007). *Identifying juror bias: Using the pretrial juror attitude questionnaire for more effective voir dire.* Arlington, TX: ProQuest.

Nakagawa, S., Johnson, P.C.D., Schielzeth, H. (2017) The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *J. R. Soc. Interface* 14: 20170213.

Naughton, M. (2003). How big is the iceberg? A zemiological approach to quantifying miscarriages of justice. *Radical Statistics*, *81*(5), 5-17. Retrieved from: http://www.radstats.org.uk/no081/naughton.pdf

Neuschatz, J. S., Wilkinson, M. L., Goodsell, C. A., Wetmore, S. A., Quinlivan, D. S., & Jones, N. J. (2012). Secondary confessions, expert testimony, and unreliable testimony. *Journal of Police and Criminal Psychology*, *27*(2), 179-192.

Northcraft, G., & Neale, M. (1987). Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. *Organisational Behaviour and Human Decision Process, 84,* 87-93. doi:10.1016/0749-5978(87)90046-X

Ormston, R., Chalmers, J., Leverick, F., Munro, V., & Murray, L. (2019). Scottish Jury Research: findings from a large-scale mock jury study. *Crime and Justice*. Retrieved from: https://www.gov.scot/binaries/content/documents/govscot/publications/research-and-analysis/2019/10/scottish-jury-research-fingings-large-mock-jury-study/documents/scottish-jury-research-findings-large-scale-mock-jury-study/scottish-jury-research-findings-large-scale-mock-jury-study/govscot%3Adocument/scottish-jury-research-findings-large-scale-mock-jury-study.pdf

Orr, D., & Guthrie, C. (2005). Anchoring, information, expertise, and negotiation: New insights from meta-analysis. *Ohio St. J. on Disp. Resol.*, *21*, 597-628. Retreived from: https://home.heinonline.org/

Ostrom, T. M., Werner, C., & Saks, M. J. (1978). An integration theory analysis of jurors' presumption of guilt or innocence. *Journal of Personality and Social Psychology*, *36*, 436–450. doi:10.1037/0022-3514.36.4.436

Park, K. (2011). Estimating juror accuracy, juror ability, and the relationship between them. *Law and human behavior*, *35*(4), 288-305.

Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology*, *51*, 242– 258. doi:10.1037/0022-3514.51.2.242

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology*, *62*, 189–206.doi: 10.1037/0022-3514.62.2.189

Posey, A. J., & Wrightsman, L. S. (2005). *Trial consulting* (pp. 157–172). American Psychology-Law Society Series. New York: Oxford University Press.

Ramsey, R. J., & Frank, J. (2007). Wrongful conviction perceptions of criminal justice professionals regarding the frequency of wrongful conviction and the extent of system errors. *Crime & Delinquency*, *53*(3), 436-470. doi:10.1177/0011128706286554

Rape Crisis Scotland (2019). *End Not Proven*. Retrieved (13/11/2019) from: https://www.rapecrisisscotland.org.uk/not-proven/

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological review*, *111* (2), 333-267. doi: 10.1037/0033-295X.111.2.333

Ross, S. (2019). Sottish rape victim says most jurors 'don't' even know what the not proven verdict means. *The Scotsman*. Retrieved from: https://www.scotsman.com/news/uk-news/scottish-rape-victim-says-most-jurors-dont-even-know-what-not-proven-means-1405876

Runciman, V. (1993). Report of the royal commission on criminal justice. London, England: HMSO.

Shan, G., & Gerstenberger, S. (2017). Fisher's exact approach for post hoc analysis of a chi-squared test. *PloS one*, *12*(12), e0188709.

Schklar, J., & Diamond, S. S. (1999). Juror reactions to DNA evidence: Errors and expectancies. *Law and Human Behavior*, *23*(2), 159-184. doi: 10.1023/A:1022368801333

Scottish Government (2021). Not proven verdicts delivered from 2011 to 2016: FOI release [online]. Retrieved from: https://www.gov.scot/publications/foi-17-02213/

Severance, L. J., & Loftus, E. F. (1982). Improving the ability of jurors to comprehend and apply criminal jury instructions. *Law and Society Review*, 153-197. doi: 10.2307/3053535

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological review*, *63* (2), 1-11. doi: 10.1037/h0042769.

Simon, D., Stenstrom, D. M., & Read, S. J. (2015). The coherence effect: Blending cold and hot cognitions. *Journal of Personality and Social Psychology*, *109*(3), 369-394. doi: 10.1037/pspa0000029

Simonson, I., Tversky, A. (1992). Choice in context – tradeoff contrast and extremeness aversion. *Journal of Marketing Research*, *29*, 281–295.

Smith, L. & Bull. R. (2012). Identifying and measuring juror pre-trial bias for forensic evidence: Development and validation of the forensic evidence evaluation bias scale. *Psychology, Crime & Law, 18* (9), 797-815. doi: 10.1080/1068316X.2011.561800

Smith, L., & Bull, R. (2014). Validation of the factor structure and predictive validity of the forensic evidence evaluation bias scale for robbery and sexual assault trial scenarios. *Psychology, Crime & Law, 20* (5), 450-466. doi: 10.1080/1068316X.2013.793340

Smithson, M., Deady, S., & Gracik, L. (2007). Guilty, Not Guilty, or…? Multiple options in jury verdict choices. *Journal of Behavioral Decision Making*, *20* (5), 481-498. doi: 10.1002/bdm.572

Sommers, S. R., & Ellsworth, P. C. (2000). Race in the courtroom: Perceptions of guilt and dispositional attributions. *Personality and Social Psychology Bulletin*, *26*(11), 1367-1379.

Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology, 73* (3), 437-446. doi: 10.1037/0022-3514.73.3.437

Thaman, S. C. (1999). Europe's new jury systems: The cases of Spain and Russia. *Law and Contemporary Problems*, *62*(2), 233-259. doi: 10.2307/1192258

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science, New Series, 185* (9), 1124-1131. doi:10.1007/978-94-010-1834-0_8

Tyler, T. R. (2006). Viewing CSI and the threshold of guilt: Managing truth and justice in reality and fiction. *The Yale Law Journal*, 1050-1085. Retrieved from: https://home.heinonline.org/

Van der Wurff, A. W. G., Chan, Y. L., Van Straalen, N. M., & Schouten, J. (2000). TE-AFLP: combining rapidity and robustness in DNA fingerprinting. *Nucleic Acids Research*, *28*(24), 105-105. doi: 10.1093/nar/28.24.e105

Weiten, W., & Diamond, S. S. (1979). A critical review of the jury simulation paradigm. *Law and Human Behavior*, *3*(1-2), 71-93.

Wetmore, S., Neuschatz, J., & Gronlund, S. (2014). On the power of secondary confession evidence. *Psychology, Crime & Law, 20* (4), 339-357. doi:10.1080/1068316X.2013.777963

Williams, M. R., Demuth, S., & Holcomb, J. E. (2007). Understanding the influence of victim gender in death penalty cases: The importance of victim race, sex-related victimization, and jury decision making. *Criminology*, *45*(4), 865-891.doi:10.1111/j1745-9125.2007.00095.x

Willmott, D. (2016). Is Jury Bias Preventing Justice for Rape Victims. *The Conversation.* Retrieved from: https://theconversation.com/is-jury-bias-preventing-justice-for-rape-victims-60090

Willmott, D., Boduszek, D., Debowska, A., & Woodfield, R. (2018). Introduction and validation of the juror decision scale (JDS): an empirical investigation of the story model. *Journal of Criminal Justice*, *57*, 26-34.

Appendix 1 – Pilot

A pilot study was conducted to establish which types of evidence were perceived to be strongest and weakest. This data was used to develop anchors for the main study.

Pilot Rationale

The pilot was informed by a literature review on evidence strength, which investigated how researchers had previously defined evidence strength (Ask et al., 2008; Carlson & Russo, 2001; Estrada-Reynolds et al., 2015; Park, 2011; Smith & Bull, 2012; Van der Wurff, Chan, Van Straalen, & Schouten, 2000). The researchers used five difference types of measure: 1) evidence accuracy; 2) evidence robustness; 3) evidence strength; 4) evidence weight; and 5) decision accuracy based on evidence. The types of evidence included in the pilot were chosen based on this literature.

Pilot Method

Design

A within-subjects design was adopted for this inventory-based survey. The single independent variable was evidence type, which had four levels (DNA evidence; expert testimony; secondary confession; and eyewitness testimony). The dependent variable was perceived evidence strength (total of five evidence strength measures mentioned above).

Participants

In a power analysis (Faul et al., 2007) it was established that a sample of more than 70 participants would be required to achieve power of >.95 at a significance level of 5% (ANOVA with repeated measurements, effect size Cohen's d= 0.5). The pilot study received 78 completed questionnaires (female = 55; M age = 32.47, SD age = 12.19). Seventy-seven participants were from the UK and 1 from Poland. 64 of the participants were students. All participants were eligible to vote and were eligible to be on a jury. They were recruited online using social media (e.g. Craigslist, Gumtree, Call for Participants, Facebook and Twitter). Participants provided informed consent.

Materials

Each of the five measurement types highlighted in the rationale were converted into statements for each evidence type. Participants were therefore given 20 statements. An example of a statement used in the pilot questionnaire is: "*Eyewitness testimony would not allow me to reach an accurate decision*". The participant scored each statement on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). Five items were reverse-scored to reduce the risk of response bias. Participants were also asked to rank each of the evidence types from strongest (worth 4 points) to weakest (worth 1 point). Each evidence type could be, in total, rated a minimum of 6 and a maximum of 29.

Pilot Results

DNA evidence was rated as the strongest evidence type, followed by expert testimony, then eyewitness testimony and finally secondary confessions (table 5).

Table 5. Mean and standard deviation (SD) of strength of evidence for five different types of evidence.

Evidence Type	Mean (SD)
Eyewitness Testimony	17.83 (4.07)
DNA	24.77 (3.17)
Secondary Confession	13.88 (3.95)
Expert Testimony	19.64 (3.76)

Cronbach's Alpha tests were used to assess the internal consistency of each of the four different evidence types (table 6). All Alpha scores were greater than .8 which indicates that each construct (i.e., evidence type) had very good internal consistency (Gliem & Gliem, 2003).

Table 6. Cronbach's Alpha for each of the different types of evidence.

Evidence Type	Alpha	Number of Items
Eyewitness Testimony	0.81	6
DNA	0.85	6
Secondary Confession	0.86	6
Expert Testimony	0.83	6

A one-way within-subjects ANOVA was used to test whether the means across the evidence types differed significantly in terms of perceived evidence strength [$F_{(3, 231)} = 137.25$, $p < .001$, $\eta_p^2 = .64$]. Bonferroni post hoc comparisons showed that DNA evidence was rated as significantly stronger than expert testimony, eyewitness testimony and secondary confessions (all $p < .001$). Expert testimony was rated as significantly stronger than eyewitness testimony ($p = .01$) and secondary confessions ($p < .001$). Finally, eyewitness testimony was rated as significantly stronger than secondary confessions ($p < .001$). In summary, in the current study (for which this pilot was associated with) the strong anchor was always DNA evidence and weak anchor was always a secondary confession.

Tables

Table 1. Minimum and maximum scores for each of the constructs within the PJAQ.

PJAQ Construct	Minimum	Maximum
Racial Bias	4	20
Innate Criminality	4	20
Social Injustice	4	20
Cynicism Towards the Defence	7	35
System Confidence	6	30
Conviction Proneness	5	25
Total	30	150

Table 2. Descriptive statistics for each of the constructs within the PJAQ.

PJAQ Construct	Mean	SD
Racial Bias	8.57	2.12
Innate Criminality	7.53	2.16
Social Justice	12.58	2.30
Cynicism Toward the Defence	20.16	3.69
System Confidence	15.32	3.16
Conviction Proneness	13.99	3.23

Table 3. Final belief of guilt scores grouped by evidence anchor and vignette.

Factor	Final belief of guilt score (0-14)
Anchor	
Weak-first	
Mean (SD)	7.05 (3.44)
Minimum	0
Maximum	13.8
Strong-first	
Mean (SD)	6.7 (3.16)
Minimum	0.15
Maximum	13.4
Vignette One	
Mean (SD)	7.83 (3.49)
Minimum	0.25
Maximum	13.8
Vignette Two	
Mean (SD)	6.37 (3.03)
Minimum	0
Maximum	12.6

Table 4. Verdict frequency within the evidence anchor, verdict systems and vignette predictor variables.

	Verdict Frequency		
Factor	Guilty	Not Guilty	Not Proven
Anchor			
Weak-first	41	50	37
Strong-first	27	58	43
Verdict System			
Two	42	86	N/A
Three	26	22	80
Vignette			
Vignette One	41	50	37
Vignette Two	27	58	43

Figure Legends

Figure 1. Visual analogue scale measuring the belief in guilt score.
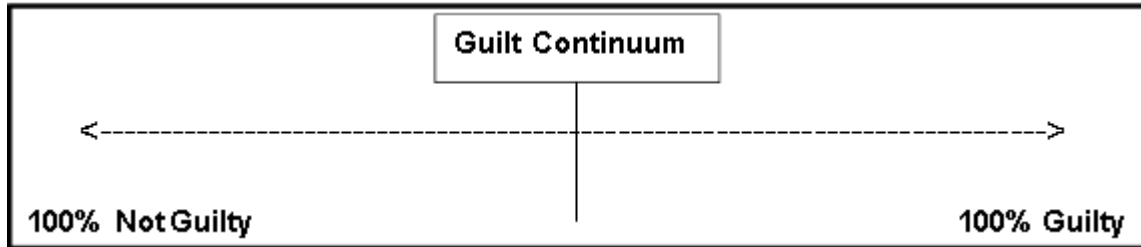
Figure 2. Line plots of average guilt scores at Timepoint 1 to 11 for weak-first (dashed line) and strong-first (solid line) evidence anchor condition. The corresponding shaded areas denote ±1 SE around the mean.