Evaluating Hypothesis-Inconsistent Evidence: The Effects of Background Knowledge, Order, and Strength of Arguments

Elizabeth Mackenzie BPsy(Hons)BHlth

Supervisor: A/Prof. Colin Wastell

Macquarie University

Empirical thesis submitted in partial fulfilment of the requirements for the degree of Master of Research (Human Sciences), Macquarie University, 2017.

Table of Contents

List of Tables

List of Figures

Declaration of Originality

I hereby confirm that all material contained in this project are my original authorship and ideas, except where the work of others has been acknowledged or referenced. I also confirm that the work has not been submitted for a higher degree to any other university or institution. The research project was approved by the Macquarie University Human Research Ethics Committee (Approval Number: 5201700100).

Elizabeth Mackenzie

Acknowledgements

First, I would like to extend sincere thanks to my supervisor, Associate Professor Colin Wastell. Through the entire research project, his guidance, support, and encouragement have been unwavering and invaluable. I would also like to thank Dr Matt Roberts for his help with the technical side of the experiments. From programming through to data retrieval, Matt was very supportive and helpful with any problems that arose. Finally, I would like to thank to my wonderful family and friends who have been incredibly understanding and supportive over this past year, with special thanks to my brother Edward for helping and encouraging me in the final stages of editing.

Abstract

Coherence-based models of reasoning suggest that the ways in which individuals incorporate information into a decision representation varies dependent upon whether it is hypothesis-consistent or hypothesis-inconsistent. This is particularly relevant in a legal context as in order to produce a judgment or decision, legal decision makers must decide which pieces of information are relevant to their preferred explanation of the case and what information will be filtered out. Background knowledge, order of information presentation, and strength of arguments have been highlighted as important to decision making. However, these elements have not yet been explored together to determine whether there are any interaction effects upon decisions made in the face of hypothesis-inconsistent information. Through the use of fictional murder and robbery scenarios, two experiments were conducted in order to determine the effects of incriminating and non-incriminating evidence on decision making in conditions of varying suspect-related background information and order of evidence presentation. Additionally, each experiment looked at different strengths of hypothesis-inconsistent arguments. In both experiments, the relationship between decision making and order of non-incriminating evidence presentation was different dependent upon background information, however, this relationship also varied between the murder and robbery scenarios. These experiments have added to the literature which suggest that the preconceptions of investigators can bias their evaluation and perception of subsequent information, and that order of evidence presentation impacts subsequent judgments of guilt. Additional research is required to determine whether these findings may generalise to more complex investigative representations.

Introduction

Legal decision making has been the subject of many experiments examining complex decision situations. In legal contexts, judgments, decisions, and hypotheses are informed by the available information, such as evidence, as well as ones' prior knowledge (Resnikoff, Ribaux, Baylon, Jendly, & Rossy, 2015; Shadlen & Kiani, 2013). By assembling and making sense of assorted pieces of evidence, legal decision makers, such as investigators or members of a jury, aim to reach a judgment (e.g., guilty or innocent; Lagnado, 2011). The purpose of amassing a body of evidence (BoE) is ultimately to provide support for a hypothesis or to refute it (Jamieson, 2003). However, in situations where ones' aim is to explain an event, there can be an abundance of information (Englich, Mussweiler, & Strack, 2006). In order to form a judgment and produce a decision, legal decision makers must determine which pieces of information are relevant to their preferred explanation of the case; they must determine what information will impact their decision and what information will be filtered out (Koslowski, Marasia, Chelenza, & Dublin, 2008; Shadlen & Kiani, 2013). During this process, different explanations may be evaluated and compared in competition with one another, and alternative explanations may be ignored or discounted (Ask & Granhag, 2007; Koslowski et al., 2008).

According to Ask and Granhag (2007), it is commonly recognized that individuals treat information differently based on whether it is consistent or inconsistent with their prior knowledge or beliefs. Individuals have exhibited a tendency to favour information that supports their hypothesis; this has been referred to as confirmation bias, positive test strategy, or myside bias in different contexts (Ask & Granhag, 2005; McKenzie, 2005; Mercier, 2016; Navarro & Perfors, 2011; Rassin, Eerland, & Kuijpers, 2010). Individuals have been shown to "seek and interpret information in ways that are partial toward existing beliefs… [and] avoid information that would contradict those beliefs and support alternative possibilities."

(Ask & Granhag, 2005, p. 45). For example, in a fictional murder investigation, Wastell, Weeks, Wearing, and Duncan (2012b) found that participants displayed hypothesis-confirmation behaviour when acquiring additional case-related information.

Individuals may also seek to discredit or reinterpret evidence that is contrary to, or not in support of, their hypothesis (Garrison & Hoskisson, 1989; Mercier & Sperber, 2009). Information that is not consistent with an individuals' emerging decision may be deliberately suppressed, or be subjected to a higher level of scrutiny, scepticism, and cognitive analysis, compared to hypothesis-consistent information (Ask & Granhag, 2007; Simon, Snow, & Read, 2004). Lagnado (2011) noted that these tendencies appear to be reinforced by the constraints of legal decision making. With pressures to reach a final verdict, individuals are compelled to construct a comprehensive, yet *one-sided*, explanation of a case that leaves no room for alternate versions (Lagnado, 2011). Indeed, Ask and Granhag (2005) found, in a fictional murder investigation with a sample of police investigators, that an increase in hypothesis-confirmation seeking behaviour was related to an increased need for case closure.

Understanding the underlying mechanisms and strategies for decision making in complex decision environments are important, as Lagnado (2011) noted, it is unlikely that individuals would be able to form judgments and make decisions from complex, interrelated bodies of evidence without employing strategies and methods for organising their mental representation. Coherence-based models of reasoning and the argumentative theory of reasoning have endeavoured to understand possible mental representations of decision situations and factors that may be relevant to reasoning and decision making. These models suggest that individuals aim to produce a coherent, consistent representation of the evidence they are presented with in order to make a decision that will stand up to argument (Mercier, 2016). In this process, the mind reportedly rejects incoherent representations, in place of coherent ones (Simon et al., 2004). This type of information processing facilitates decision

making in complex or ambiguous situations that would otherwise have been overwhelming (Simon et al., 2004).

Background knowledge, order of information presentation, and strength of arguments have been highlighted as important to the process of forming judgments and decision making in the face of inconsistent or conflicting information (Ask & Granhag, 2007; Charman, Carbone, Kekessie, & Villalba, 2015; Keppens & Schafer, 2006; Price & Dahl, 2013; Resnikoff et al., 2015). As each of these elements are said to impact the production of judgments and decisions, it is important to understand how they interact with one another. However, background knowledge, order of information presentation, and strength of arguments have not yet been explored in the context of each other to determine whether there are any *interaction* effects upon judgments made in the face of hypothesis-inconsistent information. Using a garden-path argument methodology, two experiments were conducted in order to determine the effects of incriminating and non-incriminating evidence in conditions of varying background information and varying order of presentation. Additionally, each experiment looked at varied strengths of hypothesis-inconsistent arguments. It is apparent that legal decision representations are complex and, due to the potentially serious impacts of related decisions, it is important to understand how various manipulations to available information and evidence may alter the outcomes.

Coherence-Based Models of Reasoning

Often, evidence is viewed as *positive* or *negative* in the way it relates to a particular hypothesis, however, simply summing evidence in this way does not account for the complex interrelations between various pieces of evidence (Lagnado, 2011). Coherence-based models of reasoning emphasise these complex relationships and suggest that the mind attempts to produce a coherent mental representation of the decision problem (Simon et al., 2004; Simon,

Pham, Le, & Holyoak, 2001). It has been suggested that, when forming coherent representations, pieces of evidence are grouped together if they follow the same direction (e.g. whether they support or do not support ones' hypothesis; Lagnado, 2011; Simon et al., 2004). This grouping is important and may assist in overcoming limitations of memory and information processing, as individuals can reportedly only manage around three to four sentences in working memory (Baddeley, 2010; Botvinick & Plaut, 2006; Lagnado, 2011; Ryan, 1969). However, elements that cohere also reportedly tend to be accepted or rejected from the mental representation together (Lagnado, 2011). In this way, coherent representations may discount some of the evidence which does not fit with their hypothesis. According to Simon et al. (2001), from the outset decision makers actively seek coherence, thereby reducing "costs associated with post-decisional regret" (p. 28). Incoherent representations are rejected in favour of coherent representations, which then inform judgments and decisions (Simon et al., 2004; Frigotto & Rossi, 2015).

According to Simon et al. (2004), the mechanism for processing this network of information and defining highly coherent variables can be found in constraint satisfaction models (see Simon & Holyoak, 2002). It is said that the representation of a decision problem is a continuum, and the emerging decision moves toward a decision alternative at either end of the continuum (Simon et al., 2004). Information that supports the emerging decision is highly endorsed, whereas information that supports the other decision alternative is suppressed by the 'constraint satisfaction mechanism' (Simon et al., 2004). Furthermore, it is not necessary for the non-coherent information to be in direct opposition to a hypothesis (e.g. exonerating evidence in a criminal trial), instead it may have implications for a competing explanation (e.g. incriminating an alternate suspect; Simon et al., 2001). As a result, this adaptive mechanism can *distort* the evidence, in that a coherent representation may be produced in the presence of an inconsistent pattern of information (Simon et al., 2004). The

mechanisms that appear to allow individuals to confidently form judgments and then make decisions from potentially ambiguous evidence also reportedly allow decisions to be made based on 'insufficient evidence' (Simon et al., 2004). Indeed, individuals may only consider a subset of the available information (Hernandez & Preston, 2013). As Simon et al. (2004) have noted, "it is not hard to see how weak and ambiguous evidence can be constructed to seem incriminating beyond a reasonable doubt" (p. 830). The outcomes of coherence-driven decision making could be very serious, for example, through false convictions during criminal trials.

Arguments and disfluency. The argumentative theory of reasoning has also emphasised coherence-based decision making. When integrating new information, individuals are said to perform 'coherence checking', whereby they compare new information to previously held beliefs and background knowledge (Mercier & Sperber, 2011). During this process, individuals monitor for any inconsistencies, which must be resolved by choosing between two decision alternatives (Mercier & Sperber, 2011). This theory also emphasises the importance of arguments, as they enable individuals to get their message across to others, and allow ideas and beliefs to be communicated (Mercier, 2016). Exchanging arguments is useful, both in testimony and joint decision making, for example: between officers investigating a case or during jury deliberations (Mercier, 2016). In such cases, multiple arguments may be in competition and an individual may prefer one argument over another (Mercier & Sperber, 2011). Along these lines, individuals are not necessarily seeking *truth*, but are seeking to support their own views with the information available (Mercier & Sperber, 2011). Individuals typically fail to seek information at odds with their beliefs; instead, they seek evidence that supports their own position, merely entertaining counterarguments if they anticipate the need to refute them (Mercier & Sperber, 2011). Whilst reasoners can objectively evaluate the arguments of others, especially arguments

contrary to their views, they can be biased when producing or evaluating their own argument (Mercier, 2016).

It has been suggested that 'processing fluency', the perceived ease of information processing, is highly important to decision making and biases, particularly as it has been suggested that legal decision makers have been influenced by the most fluent, coherent story (Alter & Oppenheimer, 2009). According to Hernandez and Preston (2013), *disfluency* of information and arguments can assist individuals in "overcoming biases that can often distort reasoning" (p. 181). They reported that participants provided 'less biased' assessments of guilt when information was presented disfluently; however, this only occurred in the condition that was not time pressured and did not have an additional memory task, suggesting that cognitive resources were important (Hernandez & Preston, 2013). Through their findings, Hernandez and Preston (2013) suggested the disfluent presentation of information may offer individuals the opportunity to form better judgments, and reduce premature discounting of alternate perspectives that may have otherwise been overlooked. It is important to note that the relationship between disfluency and decision making has been criticised and some findings that suggest a relationship have not been replicated (see Meyer et al., 2015), however these criticisms and findings are related to *disfluent fonts* (i.e., fonts that are difficult to read) and not to the argument-related disfluency discussed in this research.

Conflicting arguments have been explored in a number of studies through the use of the garden-path method (Feeney, Coley, & Crisp, 2010; Mackenzie, Chalmers, Wastell, Duncan, & Roberts, 2017; Wastell, Feeney, Coley, & Weeks, 2012a). Garden-path arguments endeavour to reduce processing fluency and increase cognitive effort by presenting the reasoner with information that does not support their pre-existing hypothesis (Feeney et al., 2010). When presented with a garden-path argument, reasoners are said to hypothesise a certain way based on early information; however, they have been 'led down a garden path',

as their hypothesis is not supported by later information (Feeney et al., 2010). Even beyond the content of the information, the fluency of its' presentation can impact judgments (Alter & Oppenheimer, 2009). The garden-path method is important in the context of coherence seeking, as it confronts the reasoner with hypothesis-inconsistent information. Once an unexpected piece of information is encountered, an error signal reportedly prompts the reconsideration of the initial hypothesis (Solska & Rojcsk, 2015).

Feeney et al. (2010) reported that participants in their garden-path argument experiments showed an increase in cognitive effort when presented with hypothesis-inconsistent information, or information that led them to produce new hypotheses, compared to hypothesis-consistent information. Furthermore, marked reductions in perceived argument strength occurred with the addition of hypothesis-inconsistent information (Feeney et al., 2010). Additionally, in a simulated crime investigation task, Wastell et al. (2012a) found that some level of belief revision occurred for participants upon being presented with information that was inconsistent with suspect guilt and with information previously presented. This was evident in reductions to subsequent ratings of guilt following the presentation of the hypothesis-inconsistent information (Wastell et al., 2012a).

In contrast, in a simulated murder investigation where participants were presented with an incriminating BoE followed by an additional piece of evidence, Mackenzie et al. (2017) found that resulting judgments regarding the suspects' level of guilt did not significantly differ based on whether the additional piece of evidence was hypothesis-consistent or hypothesis-inconsistent. Furthermore, Christianson, Hollingworth, Halliwell, and Ferreira (2001) noted that, more frequently than previously expected, garden-path mental representations do not accurately represent the content of the input. Instead, individuals may be slow to register the related error signal, spend less time attempting the integration of new information, as well as fail to revise their initial representation (Christianson, Luke, Hussey,

& Wochna, 2017). This indicates that garden-path arguments may not always prompt individuals to reconsider their representation of the decision situation, or the resulting decisions made.

When conflicts are generated by arguments, they may not always be strong enough to promote belief change. Mercier (2016) noted that change is more likely to occur where initial attitudes are relatively weak or the argument is particularly strong. Altering ones' beliefs can be an effortful and time consuming process; for some, it may be easier to avoid modifying existing beliefs by ignoring or discounting alternate arguments or perspectives (Hernandez & Preston, 2013). Inaccurate belief updating in response to hypothesis-inconsistent information aligns with the notion that individuals can cognitively justify discounting and even ignoring evidence that does not support their existing hypothesis (Tetlock, 2005; Tetlock & Gardener, 2015). Hypothesis-inconsistent information may also be subject to a higher level of scrutiny, scepticism and cognitive analysis, compared to hypothesis-consistent evidence (Ask & Granhag, 2007). This is important, as Simon et al. (2004) noted that information that is not consistent with an emerging decision may be deliberately suppressed to maintain coherence.

Background Knowledge

In many situations decision makers are presented with, or can access from prior knowledge, two types of information: general (base rate) information about 'how things tend to be in similar situations'; and more specific individuating information about "how things appear to be in the particular situation" (Bar-Hillel, 1980, p. 1). According to Evans (2006), it is normal, as well as adaptive, to utilise all knowledge and beliefs relevant to the decision problem, and to draw inferences from these. Indeed, for many individuals there is a tendency to provide default responses based on prior knowledge (Evans, 2006; Wright & Goodwin, 2009). Legal decision makers' background knowledge has been shown to be very important

in the decision making process. Investigators typically utilise available evidence to inform judgments, decision making, and their course of investigative action, in addition to their previous experiences and knowledge (Resnikoff et al., 2015). Whilst there is little research into how this way of conceptualising background knowledge applies to a legal context, Mackenzie et al. (2017) found that the *combination* of base rate and individuating information had the greatest impact on decision making, rather than either type of information in opposition with the other; participants reported significantly higher guilt ratings when presented with both general and individuating information that enhanced the likelihood a suspect was responsible for a fictitious murder, compared to those presented with mixed information.

Crime intelligence, which encompasses knowledge related to the crime environment and underlying processes, can also be helpful in identifying patterns, crime typicality and commonalities between cases (Resnikoff et al., 2015). Furthermore, the use of background knowledge in criminal investigations is routinely accepted, as this enables investigators to identify whether the particular case is connected to another case, and allows organisations to manage high volumes of crimes (Resnikoff et al., 2015). In this way, crime intelligence acts as an orienting frame upon which investigators recognise typical patterns and search for new information in the situation (Resnikoff et al., 2015). Medin, Coley, Storms, and Hayes (2003) too noted that an individuals' knowledge may impact their expectations about what additional information may be relevant. This process reportedly allows legal decision makers to target their investigation and collect and base their decisions upon information deemed relevant and worth collecting (Resnikoff et al., 2015).

In addition to influencing judgments and decision making, background knowledge may also impact how arguments are perceived. This is particularly important when considering the conclusion of an argument, as ones' background knowledge may affect their

perception of the arguments' plausibility (Mercier, 2016). Judgments of plausibility are made in many varying situations, including during the evaluation of criminal evidence by investigators and juries (Canter, Grieve, Nicol, & Benneworth, 2003). Plausibility is said to be determined partly by base rate, whereby occurrences with higher base rates are judged to be more likely to occur (Mackenzie et al., 2017); for example: suspecting the spouse in a murder investigation. Plausibility is also said to be determined partly by individuating information; the presence of factors linked with an increase in a particular occurrence, for example that the spouse showed physically violent tendencies or knew of infidelity, is said to increase the perceived plausibility that the spouse may be responsible for the murder (Mackenzie et al., 2017). In certain legal decision making situations, individuals may be presented with multiple arguments that aim to present the most likely account of events. Individuals in these contexts likely choose the one they deem most plausible (Baudet, Jhean-Larose, & Legros, 1994).

However, individuals also appear to display a preference for, and a 'want to believe', information that is consistent with the hypothesis under consideration; this is in comparison to information that does not cohere, which is met with increased scepticism, more attempts of refutation, and increased cognitive analysis (Ask & Granhag, 2007). According to Ask and Granhag (2007), "the most effective means to avoid adjusting ones' belief is indeed to simply ignore subsequent information because critical examination of the information might lead to the detection of disturbing inconsistencies" (p. 580). Individuals are generally inclined to hold on to their beliefs quite strongly; it becomes quite difficult to change them, especially where they are reinforced by ones' background knowledge (Hernandez & Preston, 2013). Taken together, these tendencies may *blind* investigators to other possibilities within their investigation (Keppens & Schafer, 2006). Whilst some arguments may be strong enough to

overcome this, it is more likely to occur where the individual had a relatively weak attitude initially or where there is less opposition to the arguments' conclusion (Mercier, 2016).

Moreover, the preconceptions of investigators can also bias their evaluation and perception of subsequent information (Ask & Granhag, 2007). Being overly dependent on crime intelligence and background knowledge can result in biased judgments and decision making (Resnikoff et al., 2015). In an experiment where participants rated the likely guilt of a suspect in a fictional murder scenario, those presented with background information that enhanced the likelihood that a suspect was responsible reported significantly greater initial guilt ratings compared with those presented with low plausibility information (Mackenzie et al., 2017). This finding is particularly important, as early judgments can have strong consequences for how an investigation may then proceed (Ask & Granhag, 2007). Furthermore, studies have found that police investigators tend to target their investigations *against* suspects identified very early on in the investigation process (Keppens & Schafer, 2006; Sedley, 1993).

Evidence Presentation

In general, jurors and criminal investigators learn about, and become exposed to pieces of evidence in a gradual, sequential way (Charman et al., 2015). Thus, it is also important to understand how evidence presentation may impact judgments and decision making. The process of 'evidence integration' involves combining various pieces of evidence into an integrated mental representation, in order to form a global assessment of a suspects' guilt (Charman et al., 2015). With a focus on investigator decision making, Ask, Granhag, and Rebelius (2011) noted that a step-by-step belief updating process could be utilised due to the complexities of criminal cases. However, the sequential retrieval of evidence may also exacerbate judgment and decision making biases (Ask & Granhag, 2007). Whilst legal

decision making should be informed by the evidence, studies have shown that additional processes may be involved (Canter et al., 2003). The order of evidence presentation and the amount of evidence which supports one argument or another have been examined in the context of criminal investigations, and have been shown to be important to decision making outcomes (Charman et al., 2015; Hogarth & Einhorn, 1992; Lagnado, 2011; Simon et al., 2004).

Order of evidence. When integrating evidence into a mental representation, a broad range of experiments have found that the last piece of evidence presented to participants had the greatest impact upon an assessment of suspect guilt, indicating a *recency* effect (e.g., Charman et al., 2015; Dahl, Brimacombe, & Lindsay, 2009; Price & Dahl, 2013). However, findings related to order effects tend to vary, dependent upon types of evidence presented, strength of evidence, initial beliefs of guilt, and more. Lagnado and Harvey (2008) found that order of evidence presentation impacted undergraduate student participants' judgments in relation to various crime scenarios. When discrediting evidence was presented last, a final judgment of guilt was significantly lower, compared to when the discrediting evidence was presented between two pieces of incriminating evidence (Lagnado & Harvey, 2008). Whilst the impact of discrediting evidence extended to unrelated items that supported the same hypothesis (e.g., similarly incriminating evidence), the effect was dependent upon the order of its' presentation (Lagnado & Harvey, 2008).

Price and Dahl (2013) demonstrated recency effects in a mock robbery investigation scenario with evidence that was either incriminating or exonerating. Additionally, the level of influence was enhanced where the recently presented evidence was strong. Price and Dahl noted that evidence evaluation occurs within the context of the other pieces of evidence; their findings supported the idea that the order in which information is encountered could disproportionately influence evidence evaluation and the judgments made regarding a

suspects' level of guilt. Similarly, in a study that assessed the impact of eye-witness evidence related to a fictional robbery, Dahl et al. (2009) found that evidence "presentation order can dramatically affect some decisions… and influence how that information is weighed" (p. 373). In a sample of university students, the order of evidence presentation impacted suspect guilt ratings, evaluations of evidence (including perceived evidence credibility), as well as willingness to arrest a suspect (Dahl et al., 2009). Recency effects were observed, indicating that the last piece of evidence had a greater impact on decision making, but only where that information was 'highly contradictory' (Dahl et al., 2009).

In a study of undergraduate university students who listened to fictional murder case testimony, Charman et al. (2015) demonstrated similar order effects but with evidence that was ambiguous (i.e., evidence that did not clearly incriminate nor exonerate a suspect). The last piece of evidence presented to participants had the greatest impact on an overall assessment of suspect guilt (Charman et al., 2015). However, Charman et al. also found that earlier evidence influenced the evaluation of subsequent pieces of evidence. It is important to note that they did not find an equivalent retroactive influence of later evidence on early evidence evaluation, indicating that order is also relevant for this process (Charman et al., 2015). Taken together, these order effect findings indicate that participants do not simply add up individual pieces of evidence to form a coherent mental representation of a crime situation. Evidence appears to be integrated with an emphasis on *recency*, however, early evidence may impact how later evidence is evaluated (Charman et al., 2015; Dahl et al., 2009; Price & Dahl, 2013). These effects are particularly important where the accumulated evidence may provide support for varying hypotheses; as Dahl et al. (2009) noted, where investigators are exposed to contradictory evidence, "the order in which they receive this information could affect how they pursue such an investigation" (p. 370).

Strength of argument. In addition to order effects, variations in the amount or strength of evidence which supports one argument or another has been highlighted as important to decision making. For Price and Dahl (2013), strength of evidence impacted judgments of guilt, whereby the recency effects they observed were enhanced where the evidence presented most recently was particularly strong, or where it was in conflict with previous strong evidence. Furthermore, Dahl et al. (2009) found that the impact of strong evidence varied depending on the order it was encountered, yet, weaker alibis did not have as much effect on judgments made. According to Mercier (2016), when evaluating arguments individuals tend to respond appropriately to variations of strength, and regardless of ones' prior attitudes or beliefs, strong arguments tend to be effective. That is, even when arguments challenge ones' beliefs, individuals are typically able to be objective in their evaluations of argument strength (Mercier, 2016). Through the presentation of strong evidence, Price and Dahl noted that tendencies for confirmation bias might be reduced. Argument strength, however, is not solely related to the *content* of the evidence. Verheij (2014) suggested that the *addition* of information can also alter the perceived strength of an argument, either strengthening or weakening it. A hypothesis that may have been supported initially, may no longer be supported following the presentation of additional information (Verheij, 2014).

Argument strength has also been linked to issues in the subsequent processing of evidence. Charman et al. (2015) found that when strongly incriminating evidence was presented first, subsequently presented information appeared to be processed on a *superficial* level. However, in the absence of strong evidence that provided support for guilt or innocence judgments, evidence was processed more deeply (Charman et al., 2015). This finding is important, according to Charman et al., as participants tended to evaluate evidence 'fairly thoroughly' where there was no strong belief in the guilt of a suspect; whereas, if a strong guilt rating was produced, context effects were activated and a superficial evaluation of

subsequent evidence followed. According to Ask et al. (2011), superficial processing can cause context effects, typically impacting upon the assessment of evidence that follows the establishment of a belief in suspect's guilt. Finally, commitment to a belief of *guilt* appeared to produce stronger constraints, compared to beliefs of *innocence*, to the subsequent evaluation of evidence (Charman et al., 2015).

Additional Considerations

Reasoning and decision making are complex and it is evident that many factors are involved. When attempting to study the formation of judgments, it is also important to consider elements related to reasoning and decision making, such as evidence re-examination, confidence, and potential response style biases.

Evidence re-examination. Wastell et al. (2012a) found a link between evidence re-examination and belief change, specifically that the relationship between hypothesis-inconsistent information and belief revision was mediated by evidence re-examination. In legal decision making, it is expected that at some point the information collated may be reviewed, for example: when a jury meets to deliberate a verdict (Klevorick & Rothschild, 1979; Silberger, Van Wezemael, Paisiou, & Strebel, 2010). Whilst additional processes may be involved in evidence re-examination, such as group decision making, it is important to understand how revising information may impact the production of a final decision, especially in the context of order effects. The garden-path argument methodology is well suited to examine this process; it allows the measurement of incremental judgments following the presentation of each piece of information, which can then be compared to the judgments made by participants once they have had the opportunity to re-examine the evidence.

Confidence. Level of confidence is also viewed as important in the decision making process, especially when confronted with conflicting information. Decisions and judgments

are typically accompanied by some level of confidence (Kiani, Corthell, & Shadlen, 2014). When individuals reason on their own, over confidence and decision polarisation may occur as a result of the accumulation of arguments in support of one's own opinion (Mercier, 2016). Furthermore, any failure to adjust their level of confidence may be explained by an increased motivation to maintain ones' initial belief (Ask & Granhag, 2007). Decision confidence is also important to confirmation bias, as individuals become less likely to seek additional information or attempt to incorporate it into their existing mental representation (Phillips, Prybutok, & Peak, 2014). Decision confidence can reportedly impact the interpretation and use of information, as well as the decision outcome (Phillips et al., 2014). Finally, when a high level of confirmation bias is present and a high level of confidence exists, individuals are not easily persuaded to alter their decision (Phillips et al., 2014).

Social desirability. Social desirability bias is the tendency for respondents to make selections based on what they deem to be more socially accepted, compared to responses that reflect their thoughts and attitudes (Grimm, 2010). Response biases are very important to take into account when collecting data that attempts to capture judgments and decision making, as respondents may behave in ways to make themselves 'look good' and misrepresent their actual judgments (Norwood & Lusk, 2011). In regards to *socially sensitive issues*, socially desirable responses may 'become an issue', and is considered a response style that should be monitored by researchers (Ganster, Hennessey, & Luthans, 1983; Grimm, 2010). Respondents may be unwilling to accurately report their attitudes, beliefs, intentions, or behaviours regarding sensitive topics, such as murder (Fisher, 1993).

Additionally, socially desirable response styles can produce spurious relationships between variables and negatively impact data quality (Ganster et al., 1983; Krumpal, 2013). Individuals who yield high scores on a social desirability scale reportedly have an increased need for social approval and are more likely to portray themselves positively (King &

Brunner, 2000). Furthermore, social desirability bias scores indicate an individuals' level of engagement in 'positive impression management' (Edens, Buffington, Tominic, & Riley, 2001). Whilst previous garden-path research has not included this variable, the present research determined its' importance as being relevant to the inclusion of background information which has not previously been explored in this context. For investigators, biases may "entail serious consequences for the entire judicial process." (Ask & Granhag, 2007, p. 562). Thus, it is important to gather true attitudes and decisions rather than socially desirable ones.

Rationale and Aims

It is apparent from the reviewed literature that background knowledge, order of information presentation, and strength of arguments are important to judgments and decision making, and more specifically to legal decision making. It is evident that background information can be drawn from to assist in decision making. Background knowledge may be used by investigators as an orienting frame to better target their investigation, and may also be used to determine the plausibility of arguments they form or are presented with (Mercier, 2016; Resnikoff et al., 2015). It is also evident from the literature that background knowledge may impact how hypothesis-inconsistent information is integrated into a decision situation mental representation (Ask & Granhag, 2007; Hernandez & Preston, 2013).

With regards to order of evidence presentation, research has found that order is particularly important where the accumulated evidence may provide support for varying hypotheses; where investigators are exposed to contradictory evidence, the order they receive information might influence how the investigation is pursued (Dahl et al., 2009). In the literature there has been an emphasis on recency effects, whereby a broad range of studies have found that the last piece of evidence had the greatest impact upon an assessment of

suspect guilt (e.g., Charman et al., 2015; Dahl et al., 2009; Price & Dahl, 2013). However, background knowledge may impact how order effects are manifested, as the preconceptions of investigators, as well as early evidence, might influence how later evidence is evaluated (Ask & Granhag, 2007; Charman et al., 2015).

In the reviewed literature, it has also been noted that the strength of evidence which supports one argument or another may be related to the formation of judgments, decision making, and issues surrounding the subsequent processing of evidence. Studies have found that strong arguments may lead subsequent evidence to be processed superficially (Ask et al., 2011; Charman et al., 2015). Furthermore, commitment to a belief of *guilt* appeared to produce stronger constraints to the subsequent evaluation of evidence compared to beliefs of *innocence* (Charman et al., 2015). However, research also suggests that individuals are typically able to be objective in their evaluations of argument strength (Mercier, 2016). That is, when evaluating arguments, individuals tend to respond appropriately to variations of strength, and regardless of ones' prior attitudes or beliefs, strong arguments tend to be effective (Mercier, 2016).

Each of these elements have been shown to influence judgments and decision making yet, as previously noted, they have not been explored together to determine whether there are any interaction effects upon judgments made in the face of hypothesis-inconsistent information. Through the use of a garden-path argument methodology, two experiments were conducted with the aim of providing a more thorough understanding of the relationship between background knowledge, order of evidence presentation, and judgments. These were conducted using two fictional crime scenarios, murder and robbery. Additionally, each experiment looked at different strengths of hypothesis-inconsistent arguments; participants in Experiment 1 were presented with one piece of non-incriminating evidence compared to three pieces of incriminating evidence, whereas participants in Experiment 2 were presented with

an equal number of non-incriminating and incriminating pieces of evidence. The hypotheses are presented with each experiment. It is evident that factors surrounding legal decision making and judgments are complex. As the impacts of related decisions may be very serious, it is important that we better understand how background knowledge, order of evidence presentation, and argument strengths may alter judgments made.

General Method

Overview

This research, which consisted of a pilot study and two experiments, was approved by the Macquarie University Human Research Ethics Committee (reference number: HREC 5201700100; see Appendix A). The pilot study was conducted using the materials that would be presented to participants in the main experiments. As previously stated, Experiment 1 focused on the effects of one piece of non-incriminating evidence on guilt ratings in varying conditions of background information, order of presentation, and crimes. Whereas, Experiment 2 focused on the effects of a non-incriminating BoE on guilt ratings in varying conditions of background information, order of presentation, and crimes.

Participants

In total, this research included 525 participants: 25 participated in the pilot study, 250 participated in Experiment 1, and 250 participated in Experiment 2. Participants were recruited from the Macquarie University participation pool website and received course credit in exchange for their participation. Additionally, they were only able to participate in the pilot study *or one* of the two experiments. For the pilot study, the sample consisted of four males and 21 females whose ages ranged from 17 to 27 years (M = 19.60, SD = 2.363). For Experiment 1, the sample consisted of 34 males, 214 females, one other, and one not specified, whose ages ranged from 17 to 59 years (M = 19.82, SD = 4.041). Ninety-six

participants were enrolled in a course that required internal participation and 153 participated externally: internal participants completed their participation, which was scheduled and supervised by the researcher, using university facilities; whereas, external participation was not supervised and involved the use of participants' own devices. For Experiment 2, the sample consisted of 43 males, 205 females, and two not specified, whose ages ranged from 17 to 46 years (M = 19.65, SD = 3.159). Sixty-six participants were enrolled in a course that required internal participation and 184 participated externally.

Apparatus and Materials

Materials for the pilot study were presented to participants through the survey website Qualtrics (2017), and for the experiments through both Qualtrics and the Analysis Simulation Project (ASP; adapted from Weeks, Wastell, Taylor, Wearing, & Duncan, 2012). The ASP recorded participants' information accessing activity, which included: guilt ratings and confidence ratings, time spent accessing each piece of evidence, and whether and for how long each piece of evidence was revisited.

Crime scenarios. The primary materials were two fictitious crime scenarios, a murder and a robbery, which contained a victim, two suspects who had been questioned in relation to the crime, and general information relating to the crime (see Appendix B). The crime scenarios were adapted from previous research conducted by Wastell et al. (2012a) and Mackenzie et al. (2017). Participants were instructed to read through the scenarios from the perspective of an investigator who had just 'arrived on scene' and were also instructed that they would be presented with witness statements collected prior to their arrival on scene.

Adapted crime scenarios. Within the crime scenarios presented in Experiments 1 and 2, participants were also provided with a *prime suspect,* for example: "At present, the prime suspect is *Jessica Salt*." They were also notified that they would be required to indicate the

extent they believed the prime suspect was guilty, through the statement: “You will also be required to indicate to what extent you believe [*Jessica*] to be guilty”.

General and specific information intending to draw on participants’ background knowledge was also provided in the crime scenarios presented in Experiments 1 and 2. General background information presented to participants was related to the base-rate occurance, for example: “in cases with similar circumstances, most often the victim’s spouse is responsible”. This background information was either related to the prime suspect (high general plausibility) or related to the alternate suspect (low general plausibility). Specific background information presented to participants was related to the prime suspects’ individual circumstances, for example: “[prime suspect] is (not) known to have current financial problems, a factor associated with robbery.” This information indicated that an individuating factor associated with being responsible for the crime was either present (high specific plausibility) or not present (low specific plausibility) for the prime suspect.

Plausibility manipulation check. The Plausibility Manipulation Check, adapted from Mackenzie et al. (2017), consisted of five true or false questions that participants were instructed to answer “based on the information presented to you in the [crime scenario]”. These questions related to the plausibility aspects of each scenario (see Appendix C). For example: “in similar circumstances, it is more common that someone other than the spouse is responsible for the murder”. The maximum score on this scale was five and a higher score was associated with a greater ability to report on the plausibility related aspects for each crime scenario.

Witness statements. For each crime scenario there were six related pieces of evidence, in the form of witness statements (see Appendix D). Witness statements were used as, according to Ask and Alison (2010), they are ‘*by far’* the most common form of information used by investigators. Each piece of evidence included the witness’ name,

relationship to the suspect, and their statement, for example: “Lillian Simon. Status: Emily’s close friend. Emily found out about her husband’s affair a couple of days ago, she was devastated and told me she was considering confronting him about it”. Three pieces of evidence were related to each of the two people questioned in relation to the crime. Evidence was considered ‘incriminating’ if it was related to the prime suspect, and ‘non-incriminating’ if it was related to the alternate suspect. A BoE was defined as three pieces of evidence relating to the same suspect.

Pilot study questions. For the pilot study, accompanying the evidence were four questions: “How clear (easy to understand) is this piece of evidence?”; “How relevant do you think this piece of evidence is in relation to the case details?”; “If you were asked to determine the level of guilt for either one of the suspects, how important would this piece of evidence be in your decision making process?”; “How likely is it that this piece of evidence would impact upon your decision, in the form of a guilt rating, for either one of the suspects?”. Each of the questions were rated on a seven-point Likert scale with semantic anchors associated with each answer alternative, for example, for question one: 1 = *extremely clear*, 2 = *moderately clear*, 3 = *slightly clear*, 4 = *neither clear nor unclear*, 5 = *slightly unclear*, 6 = *moderately unclear*, 7 = *extremely unclear* (see Appendix E). The questions were operationalised as being related to each piece of evidence’s level of clarity, perceived relevance, importance, and impact, respectively. A lower score indicated a higher level of clarity, perceived relevance, importance, or impact.

Guilt rating and confidence rating. Guilt ratings were collected on an 11-point Likert scale anchored at -5 (*most probably not guilty*), 0 (*not sure*), and 5 (*most probably guilty*). Guilt ratings were in response to the question: “Given the information so far, please rate the likelihood that the prime suspect, [name], is guilty of the [crime] on a scale from -5

(Most Probably Not Guilty) to 5 (Most Probably Guilty)" (see Appendix F). A high, positive score indicated a higher level of perceived guilt for the prime suspect.

Confidence ratings were also collected on an 11-point Likert scale anchored at 0% (*no confidence*), 50% (*neutral*), and 100% (*complete confidence*). Confidence ratings were in response to the question: "Given the information so far, please rate your level of confidence in the rating you provided for prime suspect, [name], being guilty of the [crime] on a scale from 0% (No Confidence) to 100% (Complete Confidence)" (see Appendix G). A higher score indicated a higher level of confidence in their corresponding guilt rating. The guilt and confidence rating scales were displayed on the same screen, with guilt rating presented first, followed by confidence rating.

Social desirability scale. The M-C Form C, taken from Reynolds (1982), is a 13 question short form of the 33 item Marlowe-Crown Social Desirability scale (see Appendix H). Individuals who yield high scores on a social desirability scale are reported to have an increased need for social approval and are more likely to portray themselves positively (King & Brunner, 2000). The reliability of M-C Form C with the original Marlowe-Crowne scale is high ($r = 0.93$, $p < 0.001$; Reynolds, 1982). Responses were recorded on a two point true/false scale. Edens et al. (2001) proposed that scores greater than one and a half standard deviations above the mean are considered socially desirable responses.

Pilot Study

Evidence has been shown to be very important in legal decision making, as noted by Lagnado (2011), "there are various different ways in which the evidence can exert its influence on a hypothesis" (p. 185). Furthermore, judgments and decisions are likely to be heavily impacted by the content of evidence (Price & Dahl, 2013). It was the aim of the main experiments to determine the impact of hypothesis-inconsistent information on judgments in

the form of guilt ratings in contexts of varying background information, order of presentation, and amount of non-incriminating compared to incriminating information. In order to fulfil those aims, first, the materials to be presented in the experiments were piloted to ensure a relatively consistent level of clarity, perceived relevance, importance, and likelihood that they would have a similar level of impact upon judgments and decision making. These aspects are important as, if not controlled for, they may unduly effect judgments over and above the variables of interest.

Procedure

As previously stated, participants were recruited from the Macquarie University participation pool website and received course credit in exchange for their participation. Upon signing up for the pilot study, participants were directed to complete the survey online. Participants were first presented with an information and consent form (see Appendix I) and demographic questions (see Appendix J). Upon consent, participants were randomly presented with one of the two crime scenarios, either murder or robbery. Next, in random order participants were sequentially presented with the six pieces of evidence related to the crime and were asked to answer the four pilot study questions with each piece of evidence. This process was then repeated for the second crime scenario.

Results and Discussion

Descriptive statistics are displayed in Table 1. Furthermore, for each BoE, average ratings for clarity, relevance, importance, and impact were calculated.

Table 1.

Pilot Study Descriptive Statistics

Crime	Suspect	Evidence	Clarity M(SD)	Relevance M(SD)	Importance M(SD)	Impact M(SD)
Murder	Emily	Mary	2.32(1.145)	2.76(1.128)	3.12(1.236)	3.36(1.630)
		Hillary	1.48(0.653)	2.72(1.429)	3.08(1.352)	2.92(1.288)
		Lillian	1.24(0.523)	2.04(1.060)	2.60(1.190)	2.80(1.225)
		Average for BoE	1.68(0.597)	2.51(0.806)	2.93(0.770)	3.03(1.032)
	Ashley	Jane	1.64(0.907)	1.76(0.663)	2.44(1.502)	2.32(0.988)
		Bob	1.48(0.653)	2.16(1.068)	3.08(1.656)	2.64(1.319)
		John	2.04(0.841)	2.36(0.810)	2.84(0.850)	3.20(1.291)
		Average for BoE	1.72(0.468)	2.09(0.532)	2.79(0.937)	2.72(0.875)
Robbery	Luke	Clarissa	2.36(1.287)	3.00(1.323)	3.08(1.382)	3.56(1.417)
		Mitchell	1.20(0.500)	2.12(1.333)	2.28(1.208)	2.56(1.387)
		Christian	1.56(0.712)	1.72(0.737)	2.24(1.363)	2.48(1.122)
		Average for BoE	1.71(0.669)	2.28(0.762)	2.53(1.179)	2.87(1.009)
	Jessica	Jacinta	1.44(0.507)	2.16(1.214)	2.60(1.472)	2.72(1.208)
		Isla	1.72(0.614)	1.96(1.020)	2.44(1.193)	2.52(1.085)
		Elias	1.44(1.083)	1.56(1.158)	1.92(1.552)	1.84(1.143)
		Average for BoE	1.53(0.553)	1.89(0.843)	2.32(1.211)	2.36(0.793)

Note. N = 25. BoE = body of evidence. Rating scales were from 1 to 7.

For both murder and robbery, on average, participants' responses regarding the clarity of the evidence ranged from moderately to extremely clear. Furthermore, average ratings regarding the perceived relevance of each piece of evidence ranged from slightly to moderately relevant. Likewise, on average, participants' rated the evidence as ranging from slightly to moderately important to the decision making process. Finally, when asked to rate the likelihood that each piece of evidence would impact upon decision making in the form of a guilt rating, average ratings ranged from slightly to moderately likely.

Within each crime, the bodies of evidence were compared through paired samples t-tests. Due to the exploratory nature of the pilot study, statistical significance was defined as *p*

< .01 (Field, 2013). For murder, there was no significant difference between each BoE in terms of clarity, $t(24) = -0.350$, $p = .730$, $d = 0.075$, importance, $t(24) = 0.891$, $p = .382$, $d = 0.163$, or impact, $t(24) = 1.729$, $p = .097$, $d = 0.324$. The comparison for relevance approached significance, $t(24) = 2.744$, $p = .011$, $d = 0.615$, with a medium effect size, indicating that the BoE implicating Ashley was close to being considered more *relevant* to the crime scenario compared to the BoE implicating Emily. However, as both ratings were around 'moderately relevant', and the other comparisons produced very small effect sizes and were not significantly different, the bodies of evidence were not considered to be significantly different from one another. For robbery, there was no significant difference between each BoE in terms of clarity, $t(24) = 1.459$, $p = .158$, $d = 0.293$, relevance, $t(24) = 2.669$, $p = .013$, $d = 0.485$, importance, $t(24) = 1.933$, $p = .065$, $d = 0.176$, or impact, $t(24) = 2.392$, $p = .025$, $d = 0.562$.

For Experiment 1, within each crime it was also necessary to select one piece of evidence that would become the non-incriminating piece of evidence presented in sequence with the alternate, incriminating BoE. Upon viewing the descriptive statistics, for murder: 'Jane' and 'Lillian' were selected, and for robbery: 'Christian' and 'Elias' were selected. Overall, they appeared to be rated most clear, relevant, important, and impactful compared to the other pieces of evidence that implicated the same suspect. Furthermore, paired samples t-tests determined that there were no significant differences between 'Jane' and 'Lillian' in terms of clarity, $t(24) = 2.191$, $p = .038$, $d = 0.540$, relevance, $t(24) = -1.071$, $p = .295$, $d = 0.317$, importance, $t(24) = -0.558$, $p = .582$, $d = 0.118$, or impact, $t(24) = -1.365$, $p = .185$, $d = 0.431$. Similarly, paired samples t-tests determined there were no significant differences between 'Elias' and 'Christian' in terms of clarity, $t(24) = -0.461$, $p = .649$, $d = 0.131$, relevance, $t(24) = -0.582$, $p = .566$, $d = 0.165$, importance, $t(24) = -1.138$, $p = .266$, $d = 0.219$, or impact, $t(24) = -2.028$, $p = .054$, $d = 0.565$.

A comparison of potential difference in responses by gender was not possible due to the small sample of males. However, it was not predicted that males and females would respond differently, or that they would utilise the evidence to inform their judgments or decisions differently. O'Donnell and Safer (2017) found that while gender was related to judgments surrounding the 'likelihood that the defendant killed the victim' in a simulated murder trial, these effects did not extend to the verdicts produced. Furthermore, whilst gender differences in guilt ratings have been found when respondents were presented with crime types such as child sexual abuse cases (e.g., Pozzulo, Dempsey, Maeder, & Allen, 2010), a meta-analysis revealed 'little if any' link between the gender of jurors and guilt judgments for crimes that were classified as 'violent', 'property-related', or 'homicides' (Devine & Caughlin, 2014).

As has been noted, evidence is very important to legal decision making. By piloting the evidence that was to be presented in the main experiments, the pilot study fulfilled its' aims to ensure a relatively consistent level of clarity, perceived relevance, importance, and likelihood that each piece of evidence would have a similar level of impact upon judgments and decision making.

Experiment 1

Hypotheses

Overall analysis. It was predicted that across the various guilt ratings, there would be an interaction between plausibility and order of non-incriminating evidence presentation. As participants rated guilt in two different and complex scenarios, differences between ratings in the murder and robbery scenarios were also monitored. Whilst it appears the majority of research has been conducted using murder scenarios, studies have included other crime types

as well (e.g., Dahl et al., 2009; Lagnado & Harvey, 2008; Wastell et al., 2012a). It was not predicted that there would be any differences between crimes.

Initial guilt rating. In line with previous research (Mackenzie et al., 2017), it was predicted that those presented with high plausibility background information would report significantly higher initial guilt ratings compared to those presented with low plausibility background information.

Non-incriminating evidence and order. For those in the low plausibility condition, it was predicted that guilt ratings would significantly decrease upon the presentation of non-incriminating evidence (Wastell et al., 2012a). It was also predicted that those presented with non-incriminating evidence *last* would subsequently report significantly lower guilt ratings, compared to those presented with non-incriminating evidence *first*. This is in accordance with previous research that have found recency effects (e.g., Charman et al., 2015; Dahl et al., 2009; Price & Dahl, 2013).

For those in the high plausibility condition however, the same effects were not expected upon the presentation of non-incriminating evidence, as previous research has suggested that the preconceptions of investigators might influence how later evidence is evaluated, and that individuals may seek to reinterpret evidence that does not support their hypothesis (Ask & Granhag, 2007; Charman et al., 2015; Garrison & Hoskisson, 1989; Mercier & Sperber, 2009). As previously noted, hypothesis-inconsistent information may be deliberately suppressed, or be subjected to a higher level of scrutiny, scepticism and cognitive analysis, compared to hypothesis-consistent information (Ask & Granhag, 2007; Simon et al., 2004). Furthermore, it becomes particularly difficult to change ones' beliefs where they are reinforced by ones' background knowledge (Hernandez & Preston, 2013). It was also predicted that, following the presentation of non-incriminating evidence, those in the low

plausibility conditions would report significantly lower guilt ratings compared to those in the high plausibility conditions.

Incriminating body of evidence. Based on the findings of Mackenzie et al. (2017) and Wastell et al. (2012a), it was predicted that following the presentation of an incriminating BoE guilt ratings would increase in each plausibility, order, and crime condition.

Final guilt rating. In an exploratory comparison, this research also aimed to determine the effects of evidence re-examination. Whilst there is little research into how the effects of re-examining evidence may predict belief change, Wastell et al. (2012a) found a relationship between belief change and evidence re-examination. Furthermore, if a summary of evidence is viewed before providing a final guilt rating, it is important to know how this may impact a recency order effect. As all participants were presented with a summary of the evidence they had viewed following the last piece of evidence, a significant change from the penultimate guilt rating to the final guilt rating would indicate that the evidence summary impacted their final judgment.

Design

The dependent variable for Experiment 1 was guilt rating, which was collected at six intervals per crime. There were also three between-subjects independent variables, each of which had two-levels: plausibility (low or high), order (non-incriminating evidence first or last), and crime (murder or robbery). Participants were randomly allocated to a plausibility condition, which remained constant for the duration of the experiment. Whereas, participants were presented with both crimes and the order of the non-incriminating evidence differed in each crime. The order of crime presentation and which order of evidence participants were presented with was also randomised.

Procedure

As with the pilot study, participants were recruited from the Macquarie University participation pool website and received course credit in exchange for their participation. As previously stated, some participants were enrolled in a course that required internal participation and some participated externally: internal participants completed their participation using university facilities and were supervised by the researcher; external participation was not supervised and involved the use of participants' own devices. Instructions for these participants were the same and responses were monitored for any differences. In groups of approximately five, internal participants were placed at individual computers shielded from one another and were instructed to start at staggered time intervals. Upon being directed to the online materials, participants provided informed consent and answered demographics questions.

Next, participants were randomly allocated to a plausibility condition and were presented with one of the two crime scenarios. Immediately following the crime scenario, participants were asked to complete the corresponding Plausibility Manipulation Check. Following this, participants' reported initial guilt and confidence ratings. Next, participants were sequentially presented with four pieces of evidence (see Figure 1). Whether participants were presented with an incriminating BoE or a non-incriminating piece of evidence *first* was determined by random allocation to an order condition, which varied for each participant between crimes. Furthermore, the incriminating BoE was presented in a quasi-random sequence. Immediately after the presentation of each of the pieces of evidence, participants reported guilt and confidence ratings.

Once all of the evidence had been viewed, participants were presented with an evidence summary that included descriptors of the evidence previously displayed (e.g.,

"Lillian Simon. Status: Emily's close friend"; see Appendix K). On this screen participants were also able to open a full description of the evidence. There was no limit to how many pieces of evidence could be re-viewed, for how long, or how many times. Following this, participants gave their final guilt and confidence ratings. This process was then repeated for the second crime scenario. Finally, participants were presented with and completed the social desirability scale (M-C Form C).

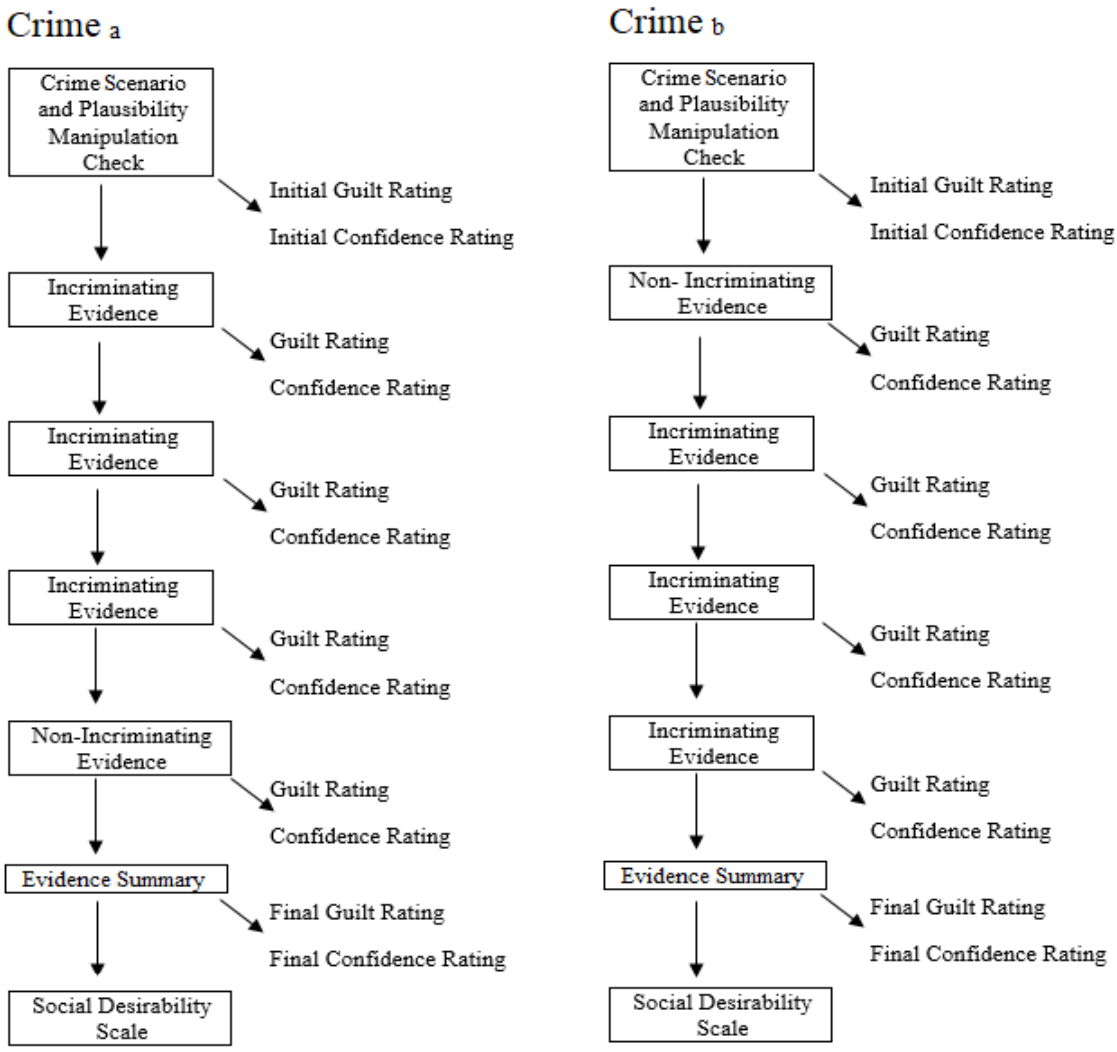


Figure 1. Experiment 1 Procedure.

Results

Inclusion criteria. Analyses were conducted using statistical package SPSS version 20 (IBM, 2011). Statistical significance for Experiment 1 was defined as $p < .05$, unless

otherwise specified. Descriptive statistics for each guilt rating by plausibility, order, and crime scenario were produced; guilt ratings that were more than two standard deviations from the mean were excluded from the analyses, as this were determined to be 'outside of normality' (Field, 2013; see Table 2 in Appendix L for description of removed cases). Additionally, social desirability scores were assessed and participants whose scores were more than Edens et al.'s (2001) criteria of one and a half standard deviations above the mean were considered to be displaying socially desirable responding and were removed. Analyses were conducted using 215 of the 250 participants; the sample consisted of 30 males, 183 females, one other, and one not specified (M_{age} = 19.75, SD_{age} = 4.116). Seventy-six participated internally and 139 participated externally.

Descriptive statistics. For each condition, the means and standard deviations for guilt and confidence ratings were produced and can be found in Table 3 and Table 4 (see Appendix M). Guilt ratings are also displayed in Figure 2 and Figure 3, for murder and robbery respectively. The Plausibility Manipulation Check scores were also calculated and neared the maximum of five for both plausibility conditions in the murder (high plausibility: M = 4.42, SD = 0.967; low plausibility: M = 4.18, SD = 1.100) and robbery scenarios (high plausibility: M = 4.61, SD = 0.719; low plausibility: M = 4.49, SD = 0.921).

Analyses of Variance (ANOVAs) revealed that there were no significant differences in responses between males and females, $F(1) = 2.051$, $p = .153$, $\eta_p^2 = .006$ (not specified or other genders were not included due to the small sample size), nor were there significant differences in responses between internal and external participants, $F(1) = .073$, $p = .787$, $\eta_p^2 < .0005$. Furthermore, there was no effect relating to the quasi-randomisation of order of evidence presentation within a BoE, $F(2) = 0.387$, $p = .679$, $\eta_p^2 = .004$.

Analyses. A Greenhouse-Geisser corrected mixed ANOVA was conducted to determine the effects of background information, order of non-incriminating evidence presentation, and crime on guilt ratings. There was a significant interaction, indicating that the guilt ratings collected throughout the experiment varied dependent upon the level of plausibility, order of non-incriminating evidence presentation, and which crime scenario was presented, $F(2.767) = 4.527$, $p = .005$, $\eta_p^2 = .012$ (see Figure 2 for murder and Figure 3 for robbery). Due to this significant interaction, the main effects and lower level interactions were not interpreted (see Table 5 in Appendix N for a comprehensive list of effects); pairwise contrasts comparing the different levels of the independent variables were conducted in order to address the research questions. Furthermore, for ease of interpretation, the two crimes will be viewed separately.

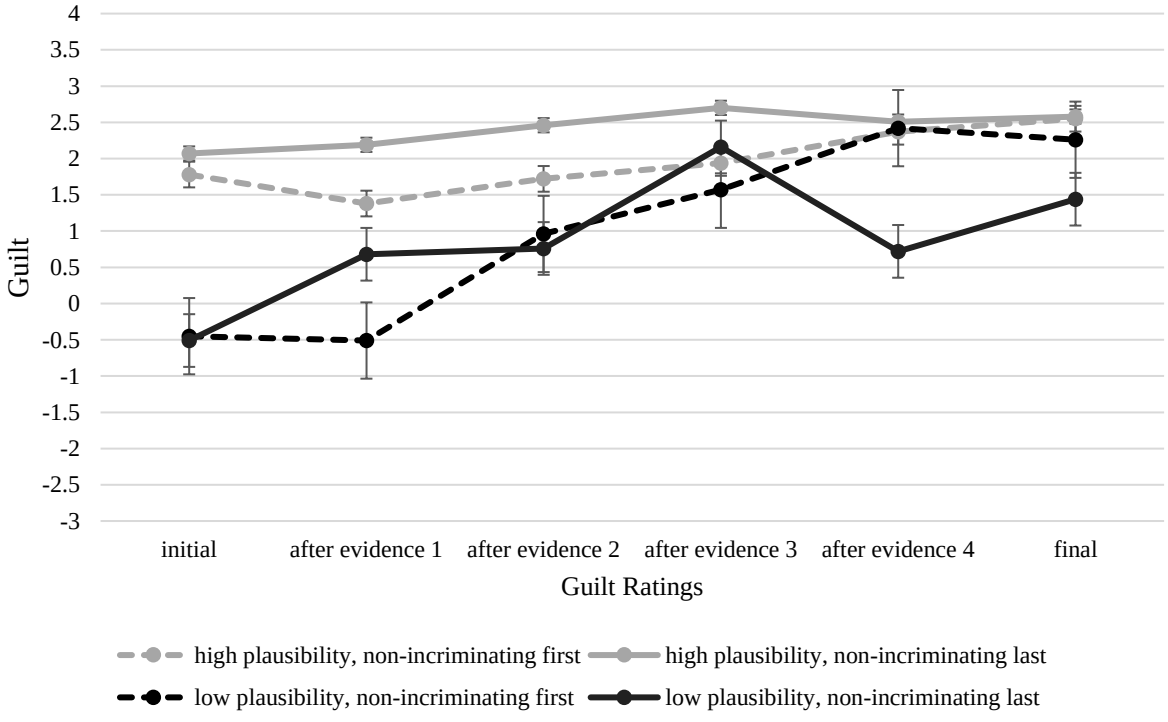


Figure 2. Guilt ratings by plausibility and order of non-incriminating piece of evidence for murder. Error bars represent standard errors. For the non-incriminating first conditions, 'evidence 1' was non-incriminating whereas the last three were incriminating. For the non-incriminating last conditions, the first three pieces of evidence were incriminating and 'evidence 4' was non-incriminating.

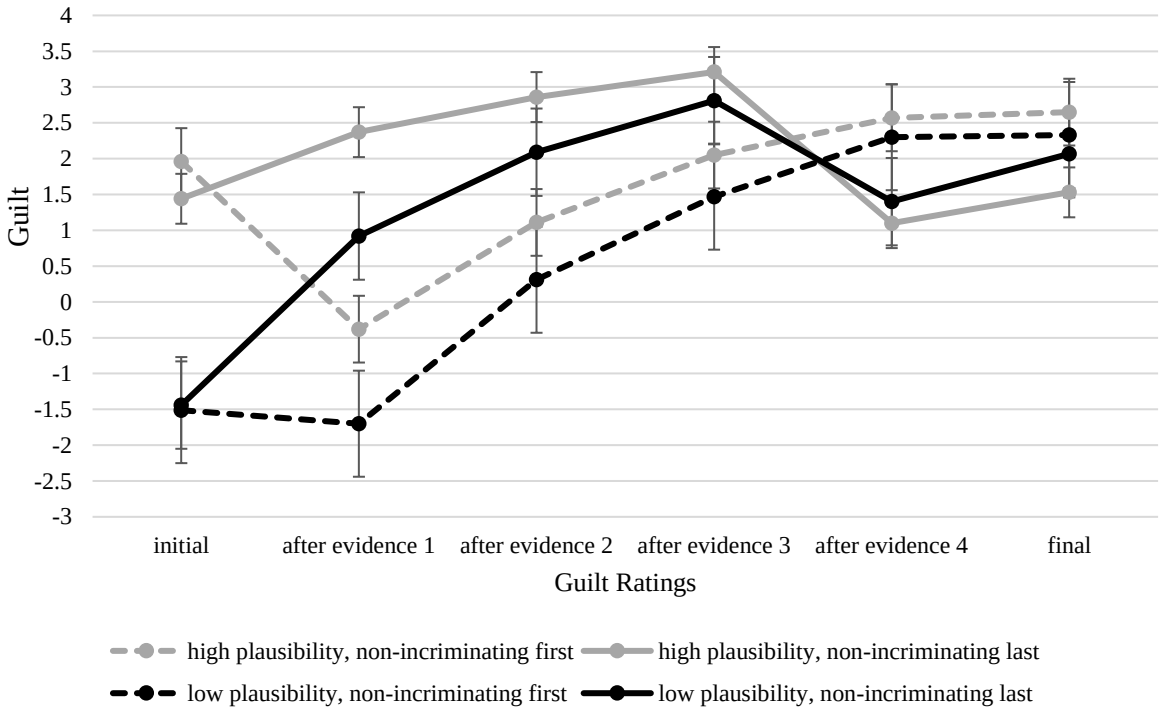


Figure 3. Guilt ratings by plausibility and order of non-incriminating piece of evidence for robbery. Error bars represent standard errors. For the non-incriminating first conditions, 'evidence 1' was non-incriminating whereas the last three were incriminating. For the non-incriminating last conditions, the first three pieces of evidence were incriminating and 'evidence 4' was non-incriminating.

Confidence. When comparing participants' level of confidence in each guilt rating between plausibility and order conditions there was only one significant ANOVA, which was for the confidence rating following the final piece of evidence, $F(7) = 3.449$, $p = .001$, $\eta_p^2 = .051$; for the robbery scenario, in the high plausibility condition, those presented with non-incriminating evidence last reported a confidence rating 15.01 lower compared to those presented with non-incriminating evidence first ($p = .006$, $d = 0.694$). Confidence ratings related to the other guilt ratings were not significantly different between conditions, with significance values ranging from $p = .055$ to .562.

Between-groups pairwise comparisons for murder.

Plausibility.

Initial guilt rating. For the initial guilt rating, high plausibility participants reported significantly higher guilt ratings compared to low plausibility participants for both those allocated to the non-incriminating first condition, by 2.20 ($p < .0005$, $d = 1.071$), and the non-incriminating last condition, by 2.59 ($p < .0005$, $d = 1.460$; see Figure 2, initial differences between the grey and black lines).

Incriminating body of evidence. For the guilt rating following the incriminating BoE, there were no significant differences between high and low plausibility conditions for those presented with non-incriminating evidence last ($p = .450$, $d = 0.392$), or for those presented with non-incriminating evidence first ($p > .9999$, $d = 0.030$).

Non-incriminating evidence. For the guilt rating following the non-incriminating piece of evidence, those in the low plausibility conditions' guilt ratings were significantly lower compared to those in the high plausibility conditions, by 1.91 when the non-incriminating evidence was presented first ($p < .0005$, $d = 0.908$; see Figure 2, after evidence 1 difference between the grey and black broken lines), and by 1.80 when the non-incriminating evidence was presented last ($p < .0005$, $d = 0.869$; see Figure 2, after evidence 4 difference between the grey and black unbroken lines).

Final guilt rating. For those presented with non-incriminating evidence last, there was a significant difference in final guilt rating, where those in the low plausibility condition reported a significantly lower guilt rating compared to those in the high plausibility condition, by 1.22 ($p = .003$, $d = 0.580$). However, there was no significant difference in final guilt rating between high and low plausibility conditions for those presented with non-incriminating evidence first ($p > .9999$, $d = 0.175$).

Order. For the initial guilt rating, there were no significant differences between those presented with non-incriminating evidence first or last in either plausibility condition (low plausibility: $p > .9999$, $d = 0.028$; high plausibility: $p > .9999$, $d = 0.171$). For the guilt rating immediately following the *first piece of evidence*, those in the low plausibility condition who were presented with non-incriminating evidence reported a significantly lower guilt rating compared to those presented with incriminating evidence, by 1.25 ($p = .005$, $d = 0.543$). However there was no significant difference for those in the high plausibility conditions ($p = .110$, $d = 0.455$). When comparing the guilt rating immediately following the *last piece of evidence* for those in the low plausibility conditions, participants presented with a non-incriminating piece of evidence last reported a guilt rating 1.59 lower than those who had been presented with the non-incriminating evidence first ($p < .0005$, $d = 0.816$; see Figure 2, evidence 4 difference between black broken and unbroken lines). Again however, there was no significant difference for those in the high plausibility conditions ($p > .9999$, $d = 0.086$). Furthermore, there were no significant differences in the final guilt ratings between order conditions for either the low ($p = .213$ $d = 0.420$) or high plausibility conditions ($p > .9999$, $d = 0.018$).

Summary of between-group effects in the murder scenario.

Initial guilt rating. The between-groups analyses revealed that background information had a large, significant effect, whereby those in the high plausibility conditions reported significantly higher initial guilt ratings compared to those in the low plausibility conditions.

Non-incriminating evidence and order. When participants were presented with non-incriminating evidence, either first or last, those in the low plausibility conditions reported significantly lower guilt ratings compared to those in the high plausibility conditions, with

large effect sizes. Additionally, for those in the low plausibility conditions, upon being presented with non-incriminating evidence their subsequent guilt rating was significantly lower compared to those in the low plausibility condition who had been presented with incriminating evidence. This occurred in both order conditions. However, this was not the case for those in the high plausibility conditions, as there were no significant differences.

Incriminating body of evidence. In both order conditions, guilt ratings were not significantly different between high and low plausibility conditions following the presentation of an incriminating BoE.

Final guilt rating. When comparing the difference in final guilt ratings between plausibility and order conditions, the between-groups analyses revealed one significant difference: this was between the high and low plausibility conditions for those presented with non-incriminating evidence *last*. The final guilt rating of those in the low plausibility, non-incriminating evidence last condition was significantly lower compared to those in the high plausibility, non-incriminating evidence last condition. There were no other significant differences in final guilt ratings.

Repeated-measures pairwise comparisons for murder. For the following repeated-measures analyses, through a Bonferroni adjustment, statistical significance was defined as $p < .017$.

Low plausibility.

Non-incriminating evidence first. From the initial guilt rating to that following the non-incriminating piece of evidence, there was no significant change ($p = .642$, $d = 0.036$). Following the non-incriminating evidence, upon being presented with an incriminating BoE there was a significant increase in guilt rating of 2.72 ($p < .0005$, $d = 1.422$). However, from

the guilt rating following the incriminating BoE to the final guilt rating, there was no significant change (p = .350, d = 0.083; see Figure 2, broken black line).

Non-incriminating evidence last. From the initial guilt rating to that following an incriminating BoE, there was an increase of 2.57 (p < .0005, d = 1.573). Additionally, there was a significant reduction of 1.10 from the guilt rating following the incriminating BoE to that following the non-incriminating piece of evidence (p < .0005, d = 0.625). From the guilt rating following the non-incriminating piece of evidence to the final guilt rating there was a significant increase of 0.51 (p = .003, d = 0.230; see Figure 2, unbroken black line).

High plausibility.

Non-incriminating evidence first. From the initial guilt rating to that following the non-incriminating piece of evidence, there was no significant change (p = .050, d = 0.191). Following the non-incriminating evidence, upon being presented with an incriminating BoE there was a significant increase in guilt rating of 1.11 (p < .0005, d = 0.627). Furthermore, from the guilt rating following the incriminating BoE to the final guilt rating, there was no significant change (p = .086, d = 0.110; see Figure 2, broken grey line).

Non-incriminating evidence last. From the initial guilt rating to that following an incriminating BoE, there was an increase of 0.71 (p < .0005, d = 0.485). However, there was no significant change from the guilt rating following the incriminating BoE to that following the non-incriminating piece of evidence (p = .350, d = 0.107). Likewise, from the guilt rating following the non-incriminating piece of evidence to the final guilt rating there was no significant change (p = .761, d = 0.018; see Figure 2, unbroken grey line).

Summary of repeated-measures effects in the murder scenario. The analyses revealed that upon being presented with an incriminating BoE, significant increases in guilt ratings occurred in each of the plausibility and order conditions. The presentation of non-

incriminating evidence resulted in a significant, moderate reduction in guilt rating but *only* for those in the low plausibility, non-incriminating evidence last condition. Furthermore, they also reported a small, yet significant increase from the guilt rating following the non-incriminating piece of evidence to the final guilt rating, whereas, there were no changes in the other conditions.

Between-groups pairwise comparisons for robbery.

Plausibility.

Initial guilt rating. For the initial guilt rating, the high plausibility condition compared to the low plausibility condition reported significantly higher guilt ratings for those allocated to the non-incriminating first condition, by 3.48 ($p < .0005$, $d = 1.723$), and the non-incriminating last condition, by 2.89 ($p < .0005$, $d = 1.576$; see Figure 3, initial differences between grey and lines).

Incriminating body of evidence. Following the presentation of an incriminating BoE, there was no significant difference in guilt ratings between the high and low plausibility conditions for either those presented with non-incriminating evidence last ($p > .9999$, $d = 0.287$), or those presented with non-incriminating evidence first ($p > .9999$, $d = 0.030$).

Non-incriminating evidence. For those in the non-incriminating evidence first condition, the guilt rating following the non-incriminating evidence was 1.33 lower in the low plausibility condition compared to the high plausibility condition ($p < .0005$, $d = 0.570$; see Figure 3, after evidence 1 difference between grey and black broken lines). However, for those in the non-incriminating evidence last condition, there was no significant difference in guilt ratings between plausibility conditions following the non-incriminating evidence ($p > .9999$, $d = 0.140$).

Final guilt rating. For the final guilt rating there were no significant differences between high and low plausibility for those presented with non-incriminating evidence last ($p > .9999$, $d = 0.274$), or for those presented with non-incriminating evidence first ($p = .422$, $d = 0.193$).

Order. For the initial guilt rating, there were no significant differences between those presented with non-incriminating evidence first or last (low plausibility: $p > .9999$, $d = 0.037$; high plausibility: $p = .670$, $d = 0.290$). For the guilt rating immediately following the *first piece of evidence*, those presented with non-incriminating evidence reported significantly lower guilt ratings compared to those presented with incriminating evidence, by 2.70 for the low plausibility condition ($p < .0005$, $d = 1.394$), and by 2.83 for the high plausibility condition ($p < .0005$, $d = 1.720$). For the guilt rating immediately following the *last piece of evidence*, there was no significant difference between low plausibility order conditions ($p = .096$, $d = 0.502$), however, in the high plausibility condition, those presented with non-incriminating evidence last reported a guilt rating 1.54 lower than those presented with non-incriminating evidence first ($p < .0005$, $d = 0.766$). For the final guilt rating, there were no significant differences between low plausibility order conditions ($p > .9999$, $d = 0.146$), however, in the highly plausibility condition, those presented with non-incriminating evidence last reported a guilt rating 1.18 lower compared to those presented with non-incriminating evidence first ($p = .005$, $d = 0.600$; see Figure 3, after evidence 4 difference between grey broken and unbroken lines).

Summary of between-group effects in the robbery scenario.

Initial guilt rating. Similar to the murder scenario, the between-groups analyses revealed that background information had a significant effect, as those in the high plausibility

conditions reported significantly higher initial guilt ratings compared to those in the low plausibility conditions, with large effect sizes.

Non-incriminating evidence and order. When participants were presented with non-incriminating evidence, those in the low plausibility condition reported a moderately lower guilt rating compared to those in the high plausibility condition, but only where the non-incriminating evidence was presented *first*. Additionally, in both plausibility conditions, for the guilt rating immediately following the *first* piece of evidence, those presented with non-incriminating evidence reported significantly lower guilt ratings compared to those presented with incriminating evidence.

There was also a difference between order conditions in guilt ratings following the final piece of evidence for those in the high plausibility condition, but *not* for those in the low plausibility condition. In the high plausibility condition, those presented with non-incriminating evidence *last* reported moderately lower guilt ratings compared to those presented with non-incriminating evidence *first.* This difference was also associated with a difference in confidence ratings, as those presented with non-incriminating evidence last reported a moderately lower confidence rating compared to those presented with non-incriminating evidence first.

Incriminating body of evidence. As occurred in the murder scenario, in both order conditions guilt ratings were not significantly different between high and low plausibility conditions following the presentation of an incriminating BoE.

Final guilt rating. Furthermore, for those in the high plausibility conditions, the difference in guilt ratings between order conditions following the final piece of evidence remained for the final guilt rating. Yet there were no other differences in final guilt ratings between plausibility or order conditions.

Repeated-measures pairwise comparisons for robbery. As with the murder scenario, for the following repeated-measures analyses statistical significance was defined as $p < .017$.

Low plausibility.

Non-incriminating evidence first. From the initial guilt rating to the guilt rating following the non-incriminating piece of evidence, there was no significant change ($p = .412$, $d = 0.091$). Following the non-incriminating evidence, upon being presented with an incriminating BoE there was a significant increase in guilt rating of 3.86 ($p < .0005$, $d = 2.154$). However, from the guilt rating following the incriminating BoE to the final guilt rating, there was no significant change ($p = .344$, $d = 0.018$; see Figure 3, broken black line).

Non-incriminating evidence last. When comparing the initial guilt rating to that following an incriminating BoE, there was an increase of 4.05 ($p < .0005$, $d = 2.689$). There was also a significant reduction of 1.18 from the guilt rating following the incriminating BoE to that following the non-incriminating piece of evidence ($p < .0005$, $d = 0.805$). Additionally, from the guilt rating following the non-incriminating piece of evidence to the final guilt rating there was a significant increase of 0.39 ($p = .008$, $d = 0.355$; see Figure 3, unbroken black line).

High plausibility.

Non-incriminating evidence first. From the initial guilt rating to the guilt rating following the non-incriminating piece of evidence, there was a significant reduction of 2.30 ($p < .0005$, $d = 1.352$). Following the non-incriminating piece of evidence, upon being presented with an incriminating BoE there was a significant increase in guilt rating of 2.86 ($p < .0005$, $d = 1.780$). From the guilt rating following the incriminating BoE to the final guilt rating, there was no significant change ($p > .9999$, $d = 0.054$; see Figure 3, broken grey line).

Non-incriminating evidence last. When comparing the initial guilt rating to that following an incriminating BoE, there was an increase of 1.50 ($p < .0005$, $d = 1.060$). Additionally, there was a significant reduction of 1.73 from the guilt rating following the incriminating BoE to that following the non-incriminating piece of evidence ($p < .0005$, $d = 1.128$). Finally, from the guilt rating following the non-incriminating piece of evidence to the final guilt rating there was no significant change ($p = .049$, $d = 0.193$; see Figure 3, unbroken grey line).

Summary of repeated-measures effects in the robbery scenario. As occurred in the murder scenario, the repeated-measures analyses revealed that significant increases in guilt ratings in each plausibility and order condition occurred following the presentation of an incriminating BoE. Upon being presented with non-incriminating evidence, there were large significant reductions in guilt ratings in all but the low plausibility, non-incriminating first condition. From the guilt rating following the last piece of evidence to the final guilt rating there was only one significant change which occurred in the low plausibility, non-incriminating evidence last condition. They reported a small, yet significant increase whereas there were no significant changes in the other conditions.

Experiment 1 Discussion

The present study aimed to determine the effects of incriminating and non-incriminating evidence on judgments in the form of guilt ratings, and predicted that there would be an interaction between guilt ratings, background information plausibility level, and order of non-incriminating evidence presentation. The findings provided some support for the hypothesised interaction, however, there was also an effect of crime which indicated that the relationship between guilt ratings, plausibility, and order of non-incriminating evidence presentation was different based on the type of crime scenario.

Initial guilt rating. It was predicted that those presented with high plausibility background information would report significantly higher initial guilt ratings compared to those presented with low plausibility background information. There was strong support for this hypothesis in both the murder and robbery scenarios, demonstrated through the large significant differences in initial guilt ratings between plausibility conditions. Furthermore, the findings were in line with previous research conducted by Mackenzie et al. (2017).

Incriminating body of evidence. Upon being presented with the incriminating BoE, it was predicted that guilt ratings would increase in all conditions. There was also strong support for this hypothesis, as there were moderate to large increases in guilt ratings in each of the plausibility and order conditions, in both crime scenarios. These findings indicate that the incriminating BoE enhanced participants' belief that the suspect was responsible for either the murder or the robbery, and are in accordance with previous research that found similar effects of incriminating evidence on guilt rating judgments (Mackenzie et al., 2017; Wastell et al., 2012a).

Non-incriminating evidence. Next, as part of the interaction, it was important to understand how the initial impact of background information may influence judgments made following the presentation of non-incriminating evidence. Based on the previous literature, it was predicted that guilt ratings for those in the low plausibility conditions would significantly *decrease* upon the presentation of non-incriminating evidence; whereas, for those in the high plausibility conditions, the same effects were not expected upon the presentation of non-incriminating evidence (Ask & Granhag, 2007; Simon et al., 2004; Wastell et al., 2012a). Based on the previous literature, it was also predicted that those in the low plausibility conditions would report significantly *lower* guilt ratings following the presentation of the non-incriminating evidence compared to those in the high plausibility conditions (Ask & Granhag, 2007; Mercier & Sperber, 2009; Simon et al., 2004).

The pattern of results predicted were largely supported by the murder scenario findings, whereby a moderate reduction in guilt rating occurred following the presentation of non-incriminating evidence for those in the low plausibility, non-incriminating last condition. Whereas, there were no reductions in guilt ratings for those in the high plausibility conditions or the low plausibility, non-incriminating first condition. Whilst the lack of reduction for those in the low plausibility, non-incriminating first condition was surprising, their guilt rating following the non-incriminating evidence remained significantly lower than those in the high plausibility, non-incriminating first condition, with a large effect size. Following the non-incriminating evidence, guilt ratings were also significantly lower for the low plausibility condition compared to the high plausibility condition where the non-incriminating evidence was presented last, with a large effect size also. Taken together, these findings support the idea that the ways in which judgments and decisions are impacted by hypothesis-inconsistent evidence differ based on background information.

However, the pattern of results predicted were not supported by the robbery scenario findings. Upon being presented with non-incriminating evidence, in all but the low plausibility, non-incriminating first condition, there were large significant reductions in guilt ratings. This finding was not in line with the reviewed literature as, upon being presented with non-incriminating evidence *first* or *last*, those in the high plausibility conditions' guilt ratings reduced significantly. Furthermore, these results contrasted with the murder scenario findings, where neither of the high plausibility conditions' guilt ratings reduced upon being presented with a non-incriminating piece of evidence. Surprisingly, in the robbery scenario, following the non-incriminating evidence there was only a significant difference between plausibility conditions where the non-incriminating evidence was presented *first*, and not where it was presented *last*.

Order. With regards to the order of non-incriminating evidence presentation, it was also predicted for low plausibility participants that those presented with non-incriminating evidence *last* would report significantly lower subsequent guilt ratings compared to those presented with non-incriminating evidence *first*. This is in accordance with previous research that have found recency effects in both murder and robbery scenarios (Charman et al., 2015; Dahl et al., 2009; Price & Dahl, 2013). Again, the murder scenario findings provided support for this hypothesis as, for those in the low plausibility conditions, the guilt rating following the final piece of evidence was significantly lower for those presented with non-incriminating evidence last compared to those presented with non-incriminating evidence first. However, in the robbery scenario, this same effect occurred for those in the high plausibility condition, but not the low plausibility condition. The difference in guilt ratings following the final piece of evidence between high plausibility order conditions was also associated with a lower level of confidence for those in the non-incriminating last condition. This difference in confidence may be related to the significant finding, as decision confidence can also reportedly impact how information is used, as well as the decision outcome (Phillips et al., 2014). However, this does not explain why there was no significant difference between the low plausibility order conditions.

Final guilt rating. In an exploratory comparison, this research also aimed to determine the effects of evidence re-examination on judgments made. Notably, Wastell et al. (2012a) found a relationship between belief change and evidence re-examination. This is important as, it is expected that the information collated in relation to an investigation or trial may at some point be reviewed (Klevorick & Rothschild, 1979; Silberger et al., 2010). Furthermore, if a summary of evidence is viewed before providing a final guilt rating, it is important to know how this may impact recency order effects.

In both crime scenarios, from the guilt rating following the last piece of evidence to the final guilt rating, significant changes occurred in the low plausibility, non-incriminating evidence last condition, whereas there were no significant changes in the other conditions. Participants in the low plausibility, non-incriminating last conditions reported small increases in guilt ratings, indicating that the evidence summary impacted their final rating. Additionally, in the murder scenario where there was a significant difference in guilt ratings between low plausibility order conditions following the final piece of evidence, there was no longer a difference following the evidence summary. This finding indicates that the recency effect observed in the murder scenario was impacted by the evidence summary.

Conclusion. The findings relating to the initial guilt rating, the incriminating BoE, and the final guilt rating were largely the same for the murder and robbery scenarios. However, based on the findings related to the non-incriminating evidence and its order of presentation, it is evident that there were important differences between the murder and robbery scenarios. Where the murder findings were mostly in agreement with the previous literature and with the predictions made, the robbery findings were not. It will be important to determine whether these differences are also present in Experiment 2.

Furthermore, previous studies have noted that variations in the amount or strength of evidence which supports one argument or another is also important to judgments and decision making (e.g., Mercier, 2016; Price & Dahl, 2013). Through the presentation of strong evidence, Price and Dahl (2013) noted that tendencies for confirmation bias might be reduced. Perhaps it is the case that one piece of evidence was not considered a ‘strong enough argument’, for example: in the murder scenario, where there was no significant change in guilt rating following the non-incriminating evidence for those in the high plausibility condition. With this consideration in mind, following Experiment 1, it was important that Experiment 2 was conducted with the aim of determining the impact of a non-incriminating

BoE on judgments made in varying conditions of background information and order of presentation in the context of a stronger non-incriminating argument. As Verheij (2014) suggested the *addition* of evidence can alter the perceived strength of an argument, through the presentation of a non-incriminating BoE rather than single piece of non-incriminating evidence Experiment 2 aimed to determine whether the impact of non-incriminating evidence may change.

Experiment 2

Hypotheses

Overall analysis. As previously mentioned, as an extension of Experiment 1, Experiment 2 aimed to determine the impact of a non-incriminating BoE on judgments made in varying conditions of background information and order of presentation. It was predicted that both background information plausibility and order of the non-incriminating BoE would impact guilt ratings. Additionally, as Experiment 1 found a significant differences between crimes, Experiment 2 also monitored for potential differences between the murder and robbery scenarios to determine whether this effect would be replicated.

As Verheij (2014) suggested the *addition* of evidence can alter the perceived strength of an argument, it was expected that through presenting a non-incriminating BoE rather than one piece of non-incriminating evidence, the hypothesis-inconsistent argument would be strengthened. It was also predicted that the change in argument strength might alter the relationship between plausibility and order found in Experiment 1 as, even when arguments challenge ones' beliefs, individuals are typically able to be objective in their evaluations of argument strength and respond appropriately (Mercier, 2016). Price and Dahl (2013) noted that, through the presentation of strong evidence, tendencies to favour hypothesis-consistent information might be reduced; furthermore, if such tendencies are reduced, individuals may

also no longer seek to discredit, suppress, or be more sceptical of information that is hypothesis-inconsistent. Instead, individuals may be more inclined to attempt to incorporate this information into their decision representation.

Initial guilt rating. In line with previous research conducted by Mackenzie et al. (2017) and the findings of Experiment 1, it was predicted that those presented with high plausibility background information would report significantly higher initial guilt ratings compared to those presented with low plausibility background information.

Non-incriminating body of evidence and order. With the increase in hypothesis-inconsistent information, for those in the low plausibility condition it was predicted that guilt ratings would significantly decrease upon the presentation of a non-incriminating BoE. Furthermore, in accordance with previous research, such as Charman et al. (2015) and Dahl et al. (2009), and in accordance with the murder scenario findings in Experiment 1, it was also predicted that those presented with the non-incriminating BoE *last* would report significantly lower guilt ratings following the last piece of evidence, compared to those presented with the non-incriminating BoE *first*.

As previously noted, individuals are said to respond appropriately to variations of argument strength regardless of their prior beliefs, and the *addition* of evidence can reportedly strengthen or weaken an argument (Mercier, 2016; Verheij, 2014). Through the presentation of a non-incriminating BoE instead of a non-incriminating piece of evidence, it was predicted that those in the high plausibility conditions would also report decreases in guilt ratings following the presentation of the non-incriminating BoE in both order conditions. Additionally, in Experiment 1, a recency effect was found for those in the high plausibility condition, in the robbery scenario; thus, Experiment 2 also monitored for these differences and predicted a recency order effect for those in the high plausibility conditions.

Incriminating body of evidence. Based on the findings of Mackenzie et al. (2017), Wastell et al. (2012a), as well as the findings of Experiment 1, it was predicted that following the presentation of an incriminating BoE guilt ratings would increase in all conditions.

Final guilt rating. As with Experiment 1, in an exploratory comparison this research aimed to determine the effects of evidence re-examination on judgments. Experiment 1 found that in both crime scenarios, those in the low plausibility, non-incriminating last condition reported small, significant increases in guilt ratings after being presented with the evidence summary, whereas there was no change in any of the other conditions. Experiment 2 aimed to determine whether the same pattern of effects would occur in the context of a non-incriminating BoE.

Design

The design for Experiment 2 was largely the same as Experiment 1, except that the dependent variable: guilt rating, was collected at eight intervals per crime, instead of six.

Procedure

The procedure for Experiment 2 was also largely the same as Experiment 1, except that participants were presented with a non-incriminating BoE instead of one piece of non-incriminating evidence (see Figure 4).

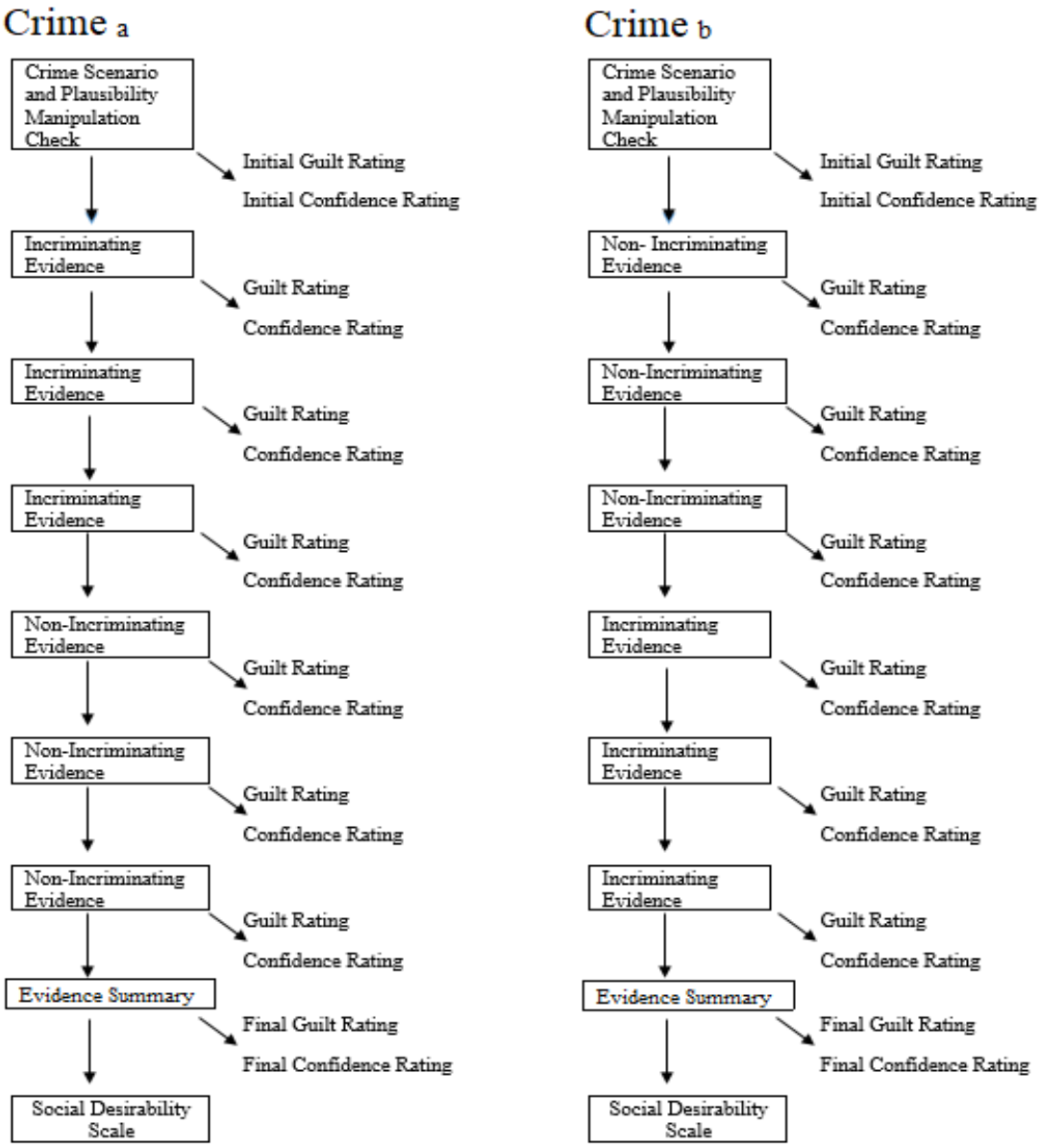


Figure 4. Experiment 2 Procedure.

Thus, participants were presented with six pieces of evidence instead of four, and were also asked to report two additional guilt and confidence ratings. As with the incriminating BoE, the non-incriminating BoE was presented to participants in a quasi-random sequence.

Results

Inclusion criteria. Analyses were conducted using statistical package SPSS version 20 (IBM, 2011). As with Experiment 1, statistical significance was defined as $p < .05$, unless

otherwise specified. Adopting the same inclusion criteria as Experiment 1, analyses were conducted using 204 of the 250 participants; the sample consisted of 36 males, 167 females, and one not specified (M_{age} = 19.74, SD_{age} = 3.097; see Table 6 in Appendix L for description of removed cases). Fifty-seven participated internally and 147 participated externally.

Descriptive statistics. For each condition, the means and standard deviations for guilt and confidence ratings were calculated and can be found in Table 7 and Table 8 (see Appendix O). Guilt ratings are also displayed in Figure 5 and Figure 6. The Plausibility Manipulation Check scores neared the maximum of five for both plausibility conditions in the murder (high plausibility: M = 4.48, SD = 0.955; low plausibility: M = 4.47, SD = 0.841) and robbery scenarios (high plausibility: M = 4.56, SD = 0.740; low plausibility: M = 4.57, SD = 0.980). For the main analyses, ANOVAs revealed no difference in responses between males and females, $F(1) = 0.032$, $p = .858$, $\eta_p^2 < .0005$ (gender not specified was not included due to the small sample size). There were also no differences between those who participated internally compared to those who participated externally, $F(1) = 0.921$, $p = .338$, $\eta_p^2 = .003$. Furthermore, an ANOVA determined there was no effect related to the quasi-random order of evidence presentation within a BoE on responses, $F(2) = 2.218$, $p = .111$, $\eta_p^2 = .016$.

Analyses

A Greenhouse-Geisser corrected mixed ANOVA was conducted to determine the effects of plausibility of background information, order of non-incriminating evidence presentation, and crime on guilt ratings. There was a significant interaction between the guilt ratings, level of plausibility, order of non-incriminating BoE presentation, and which crime scenario was presented, $F(3.560) = 2.861$, $p = .028$, $\eta_p^2 = .008$ (see Figure 5 and Figure 6). Due to this significant interaction the main effects and lower level interactions were not interpreted (see Table 9 in Appendix N for a comprehensive list of effects) and, as with

Experiment 1, pairwise contrasts comparing the different levels of the independent variables were conducted and will be the focus of interpretation. Furthermore, for ease of interpretation, the two crimes will be viewed separately.

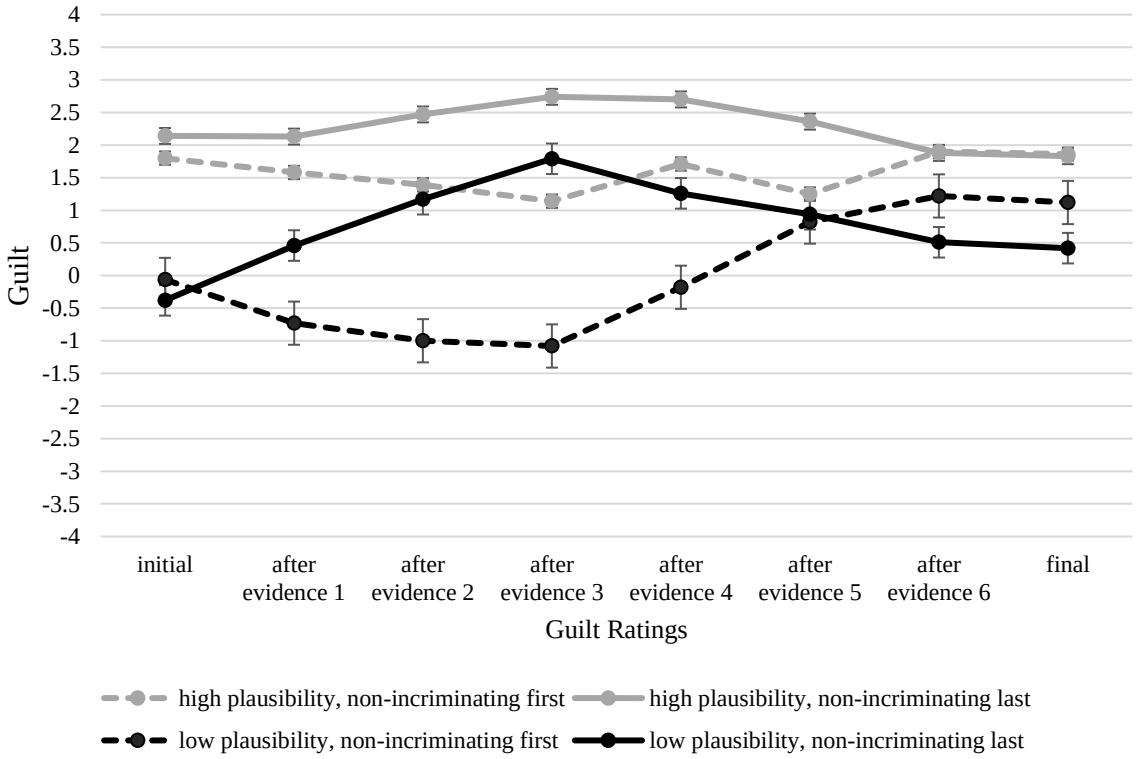


Figure 5. Guilt ratings by plausibility and order of non-incriminating body of evidence for murder. Error bars represent standard errors. For the non-incriminating first conditions, the first three pieces of evidence were non-incriminating and the last three were incriminating. For the non-incriminating last conditions, the first three pieces of evidence were incriminating and the last three were non-incriminating.

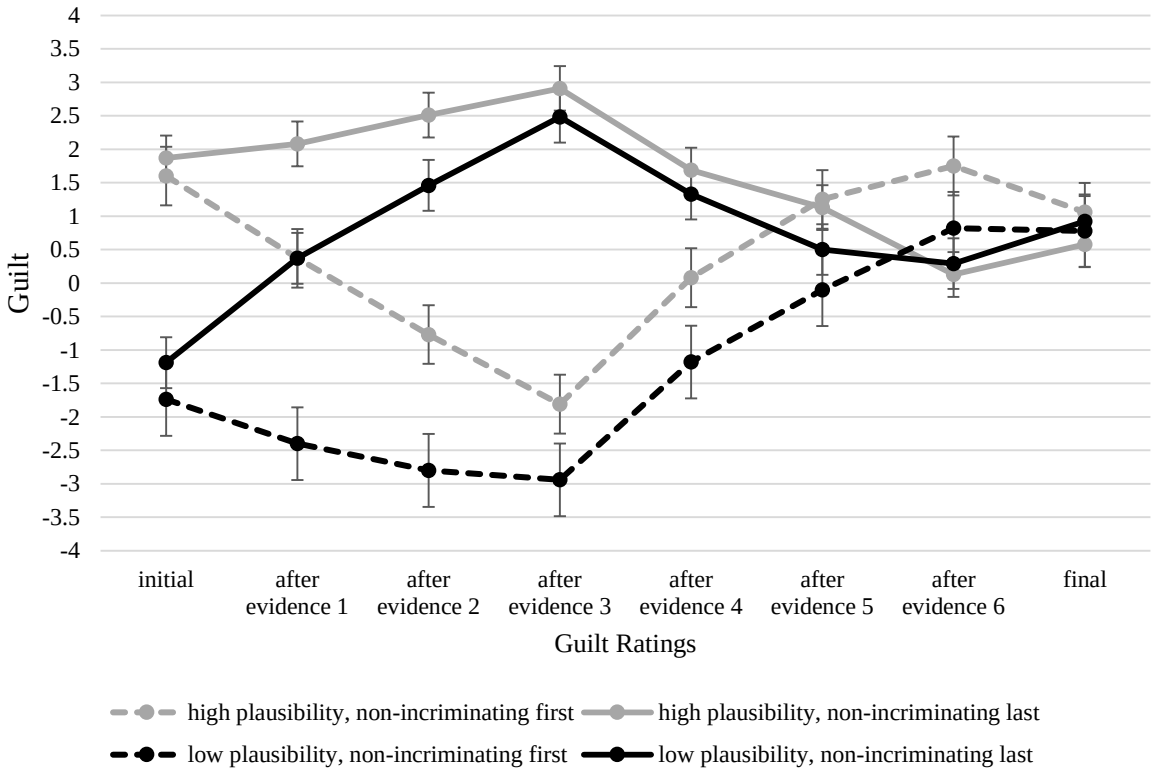


Figure 6. Guilt ratings by plausibility and order of non-incriminating body of evidence for robbery. Error bars represent standard errors. For the non-incriminating first conditions, the first three pieces of evidence were non-incriminating and the last three were incriminating. For the non-incriminating last conditions, the first three pieces of evidence were incriminating and the last three were non-incriminating.

Confidence. When comparing participants' level of confidence in each guilt rating between conditions, there was only one significant ANOVA which was for the confidence rating following the first BoE, $F(7) = 3.936$, $p < .0005$, $\eta_p^2 = .068$. However, with follow-up contrasts it was determined that the significant difference was between the high plausibility, non-incriminating first condition and the low plausibility, non-incriminating last condition in the robbery scenario; as these conditions were not compared in any of the pairwise analyses, this difference in confidence was not considered to be relevant. Confidence ratings related to the other guilt ratings were compared between each of the conditions and were not significantly different, with significance values ranging from $p = .110$ to .288.

Between-groups pairwise comparisons for murder.

***Plausibility*.**

Initial guilt rating. Initial guilt ratings were significantly higher for the high plausibility conditions compared to the low plausibility conditions, by 1.86 in the non-incriminating first condition ($p < .0005$, $d = 0.958$), and by 2.52 in the non-incriminating last condition ($p < .0005$, $d = 1.315$; see Figure 5, initial differences between grey and black lines).

Incriminating body of evidence. Guilt ratings following the incriminating BoE were not significantly different between the high and low plausibility conditions for either those presented with the non-incriminating BoE last ($p = .159$, $d = 0.603$), or the non-incriminating BoE first ($p = .718$, $d = 0.305$).

Non-incriminating body of evidence. In both order conditions, the guilt rating immediately following the non-incriminating BoE was significantly lower for those in the low plausibility conditions compared to the high plausibility conditions, by 2.21 for those presented with the non-incriminating BoE first ($p < .0005$, $d = 0.925$; see Figure 5, after evidence 3 difference between grey and black broken lines), and by 1.36 for those presented with the non-incriminating BoE last ($p = .015$, $d = 0.632$; see Figure 5, after evidence 6 difference between grey and black unbroken lines).

Final guilt rating. For the final guilt rating there was a significant difference between plausibility conditions for those presented with the non-incriminating BoE last, whereby those in the low plausibility condition's final guilt rating was 1.41 lower than those in the high plausibility condition ($p = .033$, $d = 0.531$). However, there was no significant difference in final guilt rating between plausibility conditions for those presented with the non-incriminating BoE first ($p = .833$, $d = 0.321$).

Order. For the initial guilt rating, there was no significant difference between order conditions for either the low ($p > .9999$, $d = 0.146$) or high plausibility conditions ($p > .9999$, $d = 0.209$). For the guilt rating immediately following the first BoE, those presented with non-incriminating evidence reported significantly lower guilt ratings compared to those presented with incriminating evidence, by 2.86 for the low plausibility condition ($p < .0005$, $d = 1.317$), and by 1.60 for the high plausibility condition ($p = .001$, $d = 0.856$). For the guilt rating immediately following the final BoE, there were no significant differences between order conditions in either the low ($p = .663$, $d = 0.306$) or high plausibility conditions ($p > .9999$, $d = 0.010$). Furthermore, there were no significant differences in final guilt ratings between order conditions for either the low ($p = .962$, $d = 0.264$) or high plausibility conditions ($p > .9999$, $d = 0.013$).

Summary of between-groups effects in the murder scenario.

Initial guilt rating. The between-groups analyses revealed that those in the high plausibility conditions reported significantly higher initial guilt ratings compared to those in the low plausibility conditions.

Non-incriminating body of evidence and order. When participants were presented with a non-incriminating BoE either first or last, those in the low plausibility conditions reported significantly lower guilt ratings compared to those in the high plausibility conditions. The analyses also revealed that in both plausibility conditions, following the presentation of the *first* BoE, those who were presented with a non-incriminating BoE reported significantly lower guilt ratings compared to those presented with an incriminating BoE. However, following the presentation of the *final* BoE, there were no significant differences in guilt ratings between order conditions for either plausibility condition.

Incriminating body of evidence. In both order conditions, following the presentation of an incriminating BoE there were no significant differences in guilt ratings between plausibility conditions.

Final guilt rating. Furthermore, Experiment 2 found a significant difference between plausibility conditions for those presented with the non-incriminating BoE last, where the final guilt rating was significantly lower for those in the low plausibility condition compared to the high plausibility condition. However, there were no other significant differences in final guilt ratings.

Repeated-measures pairwise comparisons for murder. As was the case in Experiment 1, for the repeated-measures analyses statistical significance was defined as $p < .017$.

Low plausibility.

Non-incriminating evidence first. From the initial guilt rating to the guilt rating following the non-incriminating BoE, there was a significant reduction of 0.94 ($p = .002$, $d = 0.445$). Following the non-incriminating BoE, upon being presented with an incriminating BoE there was a significant increase in guilt rating of 2.28 ($p < .0005$, $d = 0.916$). Though, from the guilt rating following the incriminating BoE to the final guilt rating, there was no significant change ($p = .684$, $d = 0.039$; see Figure 5, black broken line).

Non-incriminating evidence last. There was a significant increase of 1.98 from the initial guilt rating to the guilt rating following the incriminating BoE ($p < .0005$, $d = 1.046$). Additionally, there was a significant reduction in guilt rating of 1.19 from the guilt rating following the incriminating BoE to that following the non-incriminating BoE ($p < .0005$, $d = 0.653$). However, from the guilt rating following the non-incriminating BoE to the final guilt

rating there was no significant change (p = .931, d = 0.037; see Figure 5, black unbroken line).

High plausibility.

Non-incriminating evidence first. From the initial guilt rating to that following the non-incriminating BoE, there was no significant change (p = .078, d = 0.319). Whereas, there was a significant increase of 0.61 from the guilt rating following the non-incriminating BoE to the guilt rating following the incriminating BoE (p = .005, d = 0.362). There was, however, no change in guilt rating from that following the incriminating BoE to the final guilt rating (p > .9999, d = 0.021; see Figure 5, grey broken line).

Non-incriminating evidence last. There was a significant increase of 0.46 from the initial guilt rating to that following an incriminating BoE (p = .001, d = 0.243). Though, there was no significant change from the guilt rating following the incriminating BoE to that following the non-incriminating BoE (p = .033, d = 0.323), nor was there a significant change from the guilt rating following the non-incriminating BoE to the final guilt rating (p = .776, d = 0.021; see Figure 5, grey unbroken line).

Summary of repeated-measures effects in the murder scenario. The repeated-measures analyses revealed that the presentation of an incriminating BoE resulted in significant increases in guilt ratings for all of the conditions, with small and large effect sizes in the high and low plausibility conditions respectively. Following the presentation of the non-incriminating BoE, significant reductions in guilt ratings occurred for both low plausibility order conditions. However, there were no significant changes in guilt ratings following the presentation of the non-incriminating BoE for those in the high plausibility conditions. Furthermore, in each plausibility and order condition, there were no significant changes from the penultimate to the final guilt rating.

Between-groups pairwise comparisons for robbery.

Plausibility.

Initial guilt rating. Those in the high plausibility conditions reported significantly higher initial guilt ratings compared to those in the low plausibility conditions, by 3.59 for the non-incriminating first condition ($p < .0005$, $d = 2.106$), and by 3.15 for the non-incriminating last condition ($p < .0005$, $d = 1.741$; see Figure 6, initial differences between grey and black lines).

Incriminating body of evidence. The guilt rating following the incriminating BoE was not significantly different between plausibility conditions for those presented with the non-incriminating BoE last ($p > .9999$, $d = .225$), or those presented with the non-incriminating BoE first ($p = .249$, $d = 0.484$).

Non-incriminating body of evidence. For those in the non-incriminating first condition, the guilt rating immediately following the non-incriminating BoE was significantly lower in the low plausibility condition compared to the high plausibility condition, by 1.69 ($p < .0005$, $d = 0.751$; see Figure 6, after evidence 3 difference between grey and black broken lines). However, when the non-incriminating BoE was presented last, there were no significant differences in the guilt rating following the non-incriminating BoE between plausibility conditions ($p > .9999$, $d = 0.098$).

Final guilt rating. The final guilt rating was not significantly different between the plausibility conditions for those who were presented with the non-incriminating BoE first ($p > .9999$, $d = 0.317$), or those presented with the non-incriminating BoE last ($p > .9999$, $d = 0.090$).

Order. For the initial guilt rating, there was no significant difference between order conditions (low plausibility: $p = .318$, $d = 0.335$; high plausibility: $p = > .9999$, $d = 0.179$).

For the guilt rating immediately following the *first* BoE, those presented with a non-incriminating BoE reported significantly lower guilt ratings compared to those presented with an incriminating BoE, by 6.46 for the low plausibility condition ($p < .0005$, $d = 4.439$), and by 5.08 for the high plausibility condition ($p < .0005$, $d = 2.320$). For the guilt rating following the *last* BoE, in the high plausibility condition, those presented with a non-incriminating BoE last reported a significantly lower guilt rating compared to those presented with an incriminating BoE, by 1.82 ($p = .001$. $d = 0.774$; see Figure 6, after evidence 6 difference between grey broken and unbroken lines); however, there was no difference between the low plausibility order conditions ($p > .9999$, $d = 0.227$). Furthermore, there were no significant differences in final guilt ratings between order conditions for either the low ($p > .9999$, $d = 0.008$) or high plausibility conditions ($p = .431$, $d = 0.395$).

Summary of between-groups effects in the robbery scenario.

Initial guilt rating. The between-groups analyses revealed that background information had a significant effect on initial guilt rating, whereby those in the high plausibility conditions reported significantly higher initial guilt ratings compared to those in the low plausibility conditions, with large effect sizes.

Non-incriminating body of evidence and order. When participants were presented with the non-incriminating BoE, those in the low plausibility condition reported moderately lower guilt ratings compared to those in the high plausibility condition but only where the non-incriminating BoE was presented *first*. This was different to the murder scenario, as in the murder scenario this effect also occurred for those presented with non-incriminating evidence last.

The between-subjects analyses also revealed that in both plausibility conditions, for the guilt rating immediately following the *first* BoE, those presented with a non-incriminating

BoE reported significantly lower guilt ratings compared to those presented with an incriminating BoE. There was a significant difference in guilt rating between the order conditions for those in the high plausibility condition following the *last* BoE: those presented with a non-incriminating BoE last reported moderately lower guilt ratings compared to those presented with a non-incriminating BoE first. However, there was no difference between low plausibility order conditions.

Incriminating body of evidence. In both order conditions, guilt ratings were not significantly different between plausibility conditions following the presentation of an incriminating BoE.

Final guilt rating. There were no significant differences in final guilt ratings between plausibility or order conditions.

Repeated-measures pairwise comparisons for robbery. As per the murder scenario and Experiment 1, statistical significance for the following repeated-measures analyses was defined as $p < .017$.

Low plausibility.

Non-incriminating evidence first. From the initial guilt rating to the guilt rating following the non-incriminating BoE, there was a significant reduction of 1.66 ($p < .0005$, $d = 0.930$). Following the non-incriminating BoE, upon being presented with an incriminating BoE there was a significant increase in guilt rating of 4.30 ($p < .0005$, $d = 2.152$). However, from the guilt rating following the incriminating BoE to the final guilt rating, there was no significant change ($p = .799$, $d = 0.016$; see Figure 6, broken black line).

Non-incriminating evidence last. From the initial guilt rating to that following an incriminating BoE, there was a significant increase of 3.90 ($p < .0005$, $d = 2.404$). There was a significant reduction in guilt rating of 2.19 from the guilt rating following the incriminating

BoE to that following the non-incriminating BoE ($p < .0005$, $d = 1.299$). Whereas, from the guilt rating following the non-incriminating BoE to the final guilt rating there was no significant change ($p = .030$, $d = 0.218$; see Figure 6, unbroken black line).

High plausibility.

Non-incriminating evidence first. There was a significant reduction of 3.69 from the initial guilt rating to that following the non-incriminating BoE ($p < .0005$, $d = 1.764$). Additionally, following the non-incriminating BoE, upon being presented with an incriminating BoE there was a significant increase in guilt rating of 3.73 ($p < .0005$, $d = 1.767$). Yet, from the guilt rating following the incriminating BoE to the final guilt rating there was no significant change ($p = .033$, $d = 0.219$; see Figure 6, broken grey line).

Non-incriminating evidence last. From the initial guilt rating to that following an incriminating BoE, there was a significant increase of 1.22 ($p < .0005$, $d = 0.665$). There was also a significant reduction of 2.82 from the guilt rating following the incriminating BoE to that following the non-incriminating BoE ($p < .0005$, $d = 1.211$). Finally, from the guilt rating following the non-incriminating BoE to the final guilt rating there was no significant change ($p = .065$, $d = 0.202$; see Figure 6, unbroken grey line).

Summary of repeated-measures effects in the robbery scenario. The repeated-measures analyses revealed that significant increases in guilt ratings occurred following the presentation of an incriminating BoE in each of the plausibility and order conditions, with moderate to large effect sizes. Upon being presented with a non-incriminating BoE, there were also large significant reductions in guilt ratings in each plausibility and order condition. Finally, from the penultimate to the final guilt rating there were no significant changes.

Experiment 2 Discussion

It was the aim of Experiment 2 to determine the impact of an incriminating and a non-incriminating BoE on judgments made in varying conditions of background information and order of presentation. Through presenting a non-incriminating BoE rather than single piece of non-incriminating evidence, it was suggested that the hypothesis-inconsistent argument would be strengthened, tendencies to favour hypothesis-consistent information might be reduced, and the hypothesis-inconsistent evidence may be more likely to impact judgments made (Mercier, 2016; Price & Dahl, 2013; Verheij, 2014). It was predicted that the change in argument strength would alter the relationship between plausibility and order found in Experiment 1, however, there was a significant interaction with a small effect size: The impact of the order of non-incriminating evidence on the various guilt ratings differed based on background information plausibility level and, as occurred in Experiment 1, this interaction also varied by crime.

Initial guilt rating. It was predicted that those presented with high plausibility background information would report significantly higher initial guilt ratings compared to those presented with low plausibility background information. There was strong support for this hypothesis in both the murder and robbery scenarios; when provided with high plausibility background information, participants' initial guilt ratings were significantly higher compared to those presented with low plausibility information. These findings provide support for the idea that background information influences initial judgments, in the form of guilt ratings, and are in line with the findings of Experiment 1 and previous research conducted by Mackenzie et al. (2017).

Incriminating body of evidence. Upon being presented with the incriminating BoE, it was predicted that guilt ratings would increase in all conditions. The findings of

Experiment 2 provided strong support for this hypothesis, as there were moderate to large increases in guilt ratings in each of the plausibility and order conditions, in both the murder and robbery scenarios. These findings indicate that the incriminating BoE enhanced participants' belief that the suspect was responsible for the crime, and are in agreement with previous research (Mackenzie et al., 2017; Wastell et al., 2012a).

Non-incriminating body of evidence. As previously noted, individuals are said to respond appropriately to variations of argument strength regardless of their prior beliefs and the *addition* of evidence can reportedly strengthen or weaken an argument (Mercier, 2016; Verheij, 2014). Accordingly, it was predicted that following the presentation of a non-incriminating BoE, both those in the high plausibility and low plausibility conditions would report decreases in guilt ratings. The findings were mixed, however, and did not fully align with the predictions made from the literature. For the murder scenario, following the presentation of the non-incriminating evidence, where Experiment 1 found a reduction in guilt rating *only* for those in the low plausibility, non-incriminating evidence last condition, Experiment 2 found small to moderate, significant reductions in guilt ratings for both low plausibility order conditions. Surprisingly however, there were no significant changes in guilt ratings for those in the high plausibility conditions following the presentation of the non-incriminating BoE. In contrast, for the robbery scenario there were large significant reductions in guilt ratings in each plausibility and order condition following the presentation of the non-incriminating BoE. This finding was consistent with the predictions made and was mostly consistent with the robbery findings from Experiment 1, where significant reductions in guilt ratings occurred following the non-incriminating piece of evidence for all but those in the low plausibility, non-incriminating first condition.

Order. In accordance with previous research, such as Charman et al. (2015) and Dahl et al. (2009), it was also predicted that those presented with the non-incriminating BoE *last*

would report significantly lower guilt ratings following the last piece of evidence, compared to those presented with the non-incriminating BoE *first*. In the murder scenario, following the presentation of the final BoE, there were no significant differences in guilt ratings between order conditions for either plausibility condition. These findings were surprising as it appears the recency effects observed for the low plausibility condition in Experiment 1 were reduced with the additional non-incriminating evidence. In contrast, for the robbery scenario there was partial support for this hypothesis; those in the high plausibility, non-incriminating BoE last condition reported moderately lower guilt ratings following the last BoE compared to those in the high plausibility, non-incriminating BoE first condition. However, there was no difference between low plausibility order conditions. In the robbery scenario, is unclear why the recency effect occurred only for those in the high plausibility conditions, especially as there were no significant differences between plausibility conditions following the last BoE.

Final guilt rating. As with Experiment 1, in an exploratory comparison, this research also aimed to determine the effects of evidence re-examination. Experiment 1 found a significant increase from the penultimate to the final guilt rating for those in the low plausibility, non-incriminating evidence last condition in both crime scenarios. However, there were no significant changes in either of the crime scenarios in Experiment 2.

Conclusion. As previously stated, Experiment 2 aimed to determine the impact of an incriminating and non-incriminating BoE on judgments made in varying conditions of background information and order of presentation. In line with Experiment 1, the findings relating to the initial guilt rating, the incriminating BoE, and the final guilt rating were largely the same for the murder and robbery scenarios. However, based on the findings related to the non-incriminating evidence and its' order of presentation, it is evident that there were important differences between the murder and robbery scenarios. These findings will be discussed further in the General Discussion.

General Discussion

It is evident that background knowledge, order of information presentation, and strength of arguments are important to decision making. Each of these elements have been shown to influence judgments and decision making yet, as previously noted, they had not been explored in the context of each other to determine their impacts upon judgments made in the face of hypothesis-inconsistent information. Two experiments with different hypothesis-inconsistent argument strengths were conducted with the aim of providing a more thorough understanding of the relationship between background knowledge, order of evidence presentation, and judgments made in both a murder and a robbery scenario. The findings of Experiments 1 and 2 were mixed in relation to the hypotheses. Largely, the findings relating to the initial guilt ratings, the guilt ratings following the incriminating BoE, and the final guilt ratings were similar across crimes and experiments. However, the findings related to ratings made following the non-incriminating evidence were more varied.

Initial Guilt Rating and Incriminating Body of Evidence

It was clear from both experiments that background information influenced initial judgments, in the form of guilt ratings. In both crime scenarios in both experiments, those who were provided with background information that enhanced the likelihood the prime suspect was responsible for the crime reported significantly higher initial guilt ratings compared to those presented with low plausibility background information. These findings are important as they indicate that mock-investigators' judgments were influenced by background information. This was consistent with previous literature that have looked at both mock and real-world investigators' judgments and decision making (Mackenzie et al., 2017; Resnikoff et al., 2015). Furthermore, consistent with previous literature conducted by Mackenzie et al. (2017) and Wastell et al. (2012a), both experiments provided support for the

idea that guilt ratings increase following the presentation of an incriminating BoE; this increase occurred in each of the experiments, crime scenarios, plausibility, and order conditions. These findings were important as they demonstrate that participants were not simply reliant on background information to inform their judgments, instead they appeared to integrate the hypothesis-consistent information into their mental representation in order to form a judgment regarding the prime suspects' guilt (Charman et al., 2015).

Final Guilt Rating

The final guilt rating findings differed between experiments, but within each experiment the findings were the same for the different crime scenarios. Whilst it is unclear how the change in strength of non-incriminating evidence would impact whether participants' guilt ratings changed following the evidence summary, one possible explanation could be related to the changes in order effects. In the murder scenario in Experiment 1, following the changes in guilt rating that occurred after viewing the evidence summary, there was no longer a difference in guilt ratings between order conditions for those in the low plausibility condition. As this effect did not occur in Experiment 2, it is possible that there was no longer a cause for participants to adjust their guilt ratings after being reminded of the evidence they had been presented with. The findings are not conclusive regarding how evidence re-examination may impact judgments or decision making and warrants further exploration as, in legal decision making, it is expected that at some point the collated information may be reviewed (Klevorick & Rothschild, 1979; Silberger et al., 2010).

Non-Incriminating Evidence

With a focus on the murder scenario findings, both experiments found that background information appeared to influence how hypothesis-inconsistent information effected judgments made at each level of the variable: order. Whether it was a non-

incriminating piece of evidence or BoE, only those provided with low plausibility background information reported significant reductions in guilt ratings. Whereas, ratings made by those provided with high plausibility background information did not significantly change upon being presented with either a non-incriminating piece of evidence or BoE. This interaction was expected in Experiment 1, as previous research has suggested that the preconceptions of investigators might influence how later evidence is evaluated and that individuals may seek to reinterpret evidence that does not support their hypothesis (Ask & Granhag, 2007; Charman et al., 2015; Garrison & Hoskisson, 1989; Mercier & Sperber, 2009). However, the interaction was not expected in Experiment 2, as Mercier (2016) noted that even when arguments challenge ones' beliefs, individuals are typically able to respond appropriately to *strong* arguments.

One possible explanation for these findings are that the non-incriminating evidence did not represent a strong argument for those primed with high plausibility background information. However, this explanation is not favoured given the similarity in pilot study ratings for the different bodies of evidence, and that significant changes in guilt ratings occurred in the low plausibility condition when this same evidence was presented as the incriminating BoE. Furthermore, Simon et al. (2004) noted that it is not necessary for the hypothesis-inconsistent information to be in direct opposition to a hypothesis, such as exonerating evidence, instead it may have implications for a competing explanation. Another possible explanation for these findings was that those presented with high plausibility background information subsequently processed evidence on a superficial level, whereas for those in the low plausibility conditions, in the absence of strong evidence that provided support for guilt or innocence judgments, evidence was processed more deeply (Charman et al., 2015). However, as guilt ratings increased following the presentation of incriminating

evidence for those in the high plausibility conditions it is unlikely that all subsequent information was processed *superficially*.

The favoured explanation was derived from coherence-based models of reasoning. When forming a coherent representation of the decision problem, both Lagnado (2011) and Simon et al. (2004) noted that pieces of evidence are reportedly *grouped* together if they follow the same direction. This could explain why there was no significant impact of the additional non-incriminating evidence on judgments made for those in the high plausibility condition, as the decision representation still only included two *groups* of information. Furthermore, as elements that cohere also reportedly tend to be accepted or rejected from the mental representation together (Lagnado, 2011), the findings of Experiments 1 and 2 could suggest that the hypothesis-inconsistent information was accepted or rejected based on background knowledge. That is, it was accepted for those in the low plausibility conditions, demonstrated through their reductions in guilt ratings following its' presentation, but was rejected for those in the high plausibility conditions, as their guilt ratings did not significantly change and were significantly higher in each order condition compared to those in the low plausibility conditions following the non-incriminating evidence.

Order. Consistent with previous research, where differences between order conditions were found, they indicated recency effects. Following the final piece of evidence, there were significant differences between the high plausibility order conditions in both experiments' robbery scenarios, and significant differences between the low plausibility order conditions in the murder scenario in Experiment 1. Largely however, in the presence of high or low plausibility background information, there was less of an effect of non-incriminating evidence presentation order than expected. As Dahl et al. (2009) noted, the last piece of evidence had a greater impact on decision making, but only where that information was

'highly contradictory'; it is possible that the non-incriminating evidence was not considered to be 'highly contradictory' and thus, had less of an impact.

Crime difference. Interestingly, the murder scenario findings were mostly in agreement with the previous literature and with the predictions made, however, the robbery findings were not. This was surprising as it was not expected that there would be any differences between crimes. Both murder are robbery are considered to be violent, person related crimes, and have many similarities in comparison to other crime types (Ask & Alison, 2010; Bynum et al., 1982; Yechiam et al., 2008). Furthermore, in other judgment and decision making research that have utilised multiple crime types, significant differences between crimes were not reported (e.g., Lagnado & Harvey, 2008; Wastell et al., 2012a). The findings related to order effects have been viewed in both murder and robbery scenarios (e.g., Charman et al., 2015; Dahl et al., 2009; Price & Dahl, 2013). Whereas, the impact of background knowledge on decision making and decision making in the face of hypothesis-inconsistent information have largely been viewed in relation to murder and homicide scenarios, or to crime in *general*, compared to robbery (Ask & Granhag, 2007; Ask & Granhag, 2005; Hernandez & Preston, 2013; Keppens & Schafer, 2006; Mackenzie et al., 2017; Resnikoff et al., 2015). Ask and Alison (2010) noted that in relation to generating and testing hypotheses, different crime types may prompt different responses. Whilst there is little research looking into the differences in judgments and decision making between crime types, the differences found in this research warrant further investigation.

Limitations and Future Research

The findings of these experiments should be considered in the context of some limitations. Investigations are highly complex and there are still many differences between real-world investigations and the fictional investigations presented to participants. Firstly,

participants in this study were mock-investigators, who did not have investigative experience or expertise. Ultimately, it is not known whether responses would differ in a sample of police officers, however, according to Dahl et al. (2009), "prior research provides little support for the idea that professionals differ qualitatively from lay people in their reliance on heuristics [or] susceptibility to biases" (p. 370). Additionally, investigations typically take place over a time span of weeks or longer, whereas participants' judgments were made in one sitting. It should be noted that this research was also limited as only one type of evidence was presented to participants, whereas other types of evidence are used in investigative decision making as well, for example: forensic evidence. Future research could utilise real-world cases and investigators to determine whether these findings have broader generalisability. Additionally, the present research focused on *judgments* made by participants compared to the process of actioning *decisions*. It will be important for future research to explore how guilt ratings relate to decisions made, such as the decision to arrest a suspect.

Furthermore, as Dahl et al. (2009) found recency effects where the final piece of information was 'highly contradictory', it is possible that the *non-incriminating* evidence was not considered 'highly contradictory', and that effects may differ for other types of hypothesis-inconsistent information. It is suggested that future research look at the impacts of *exonerating* evidence in the presence of high and low plausibility background information, to determine whether the findings differ to those found in this research. It is also recommended that future research look at other types of crimes and include multiple, varying scenarios of each crime type to further explore the differences in judgments and decisions made in relation to hypothesis-inconsistent evidence that were found in these experiments.

Conclusion

Legal decision making has been the subject of many experiments examining complex decision situations. Furthermore, background knowledge, order of information presentation,

and strength of arguments have been highlighted as important to judgments and decision making in the face of hypothesis-inconsistent information (Ask & Granhag, 2007; Charman et al., 2015; Keppens & Schafer, 2006; Price & Dahl, 2013; Resnikoff et al., 2015). However, these elements had not yet been explored in the context of each other to determine the possible interaction effects upon judgments made in the face of hypothesis-inconsistent information. The present research found that the impact of the presentation order of non-incriminating evidence on the various guilt ratings differed based on background information plausibility level, indicating that background information impacted the ways in which participants incorporated non-incriminating evidence into their mental decision representation. However, it was also found that this relationship varied based on the type of crime scenario. Importantly, this interaction occurred whether participants were presented with a non-incriminating piece of evidence or a non-incriminating BoE, indicating a likelihood that participants grouped hypothesis-consistent information and hypothesis-inconsistent information, either accepting or rejecting the hypothesis-inconsistent evidence from their mental representation as a group.

The experiments aimed to provide a more thorough understanding of the relationship between background knowledge, order of evidence presentation, and investigative judgments made in the face of hypothesis-inconsistent information. It is apparent that legal decision representations are complex and, due to the potentially serious impacts of related decisions, it is important that we continue to better our understanding of how background knowledge, order of evidence presentation, argument strengths, and crime types may alter judgments and decision outcomes.

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review 13*(1), 219-235. doi:10.1177/1088868309341564

Ask, K., & Alison, L. J. (2010). Investigators' decision making. *Forensic psychology in context: Nordic and international perspectives*, 35-55. Retrieved from https://www.researchgate.net/profile/Karl_Ask/publication/316241525_Investigators'_decision_making/links/58f749150f7e9b67a34dd6ad/Investigators-decision-making.pdf

Ask, K., & Granhag, P. A. (2005). Motivational sources of confirmation bias in criminal investigations: The need for cognitive closure. *Journal of Investigative Psychology and Offender Profiling*, *2*(1), 43-63. doi:10.1002/jip.19

Ask, K., & Granhag, P. A. (2007). Motivational bias in criminal investigators' judgments of witness reliability1. *Journal of Applied Social Psychology*, *37*(3), 561-591. doi:10.1111/j.1559-1816.2007.00175.x

Ask, K., Rebelius, A., & Granhag, P. A. (2008). The 'elasticity' of criminal evidence: A moderator of investigator bias. *Applied Cognitive Psychology*, 22, 1245–1259. doi:10.1002/acp.1432

Baddeley, A. (2010). Working memory. *Current biology*, *20*(4), 136-R140. doi:10.1016/j.cub.2009.12.014

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, *44*(3), 211-233. doi:10.1016/0001-6918(80)90046-3

Baudet, S., Jhean-Larose, S., & Legros, D. (1994). Coherence and truth: a cognitive model of propositional truth attribution. *International Journal of Psychology, 29*, 219-350. doi:10.1080/00207599408246547

Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order: a recurrent neural network model. *Psychological review, 113*(2), 201. doi: 10.1037/0033-295X.113.2.201

Bynum, T. S., Cordner, G. W., & Greene, J. R. (1982). Victim and offense characteristics: Impact on police investigative decision-making. *Criminology, 20*(3-4), 301-318. doi:10.1111/j.1745-9125.1982.tb00463.x

Canter, D. V., Grieve, N., Nicol, C., & Benneworth, K. (2003). Narrative plausibility: The impact of sequence and anchoring. *Behavioral sciences & the law, 21*(2), 251-267. doi:10.1002/bsl.528

Charman, S. D., Carbone, J., Kekessie, S., & Villalba, D. K. (2015). Evidence evaluation and evidence integration in legal decision-making: order of evidence presentation as a moderator of context effects. *Applied Cognitive Psychology*. doi:10.1002/acp.3181

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology, 42*, 368-407. doi:10.1006/cogp.2001.0752

Christianson, K., Luke, S. G., Hussey, E. K., & Wochna, K. L. (2017). Why reread? Evidence from garden-path and local coherence structures. *The Quarterly Journal of Experimental Psychology, 70*(7), 1380-1405. doi:10.1080/17470218.2016.1186200

Dahl, L. C., Brimacombe, C. A., & Lindsay, D. S. (2009). Investigating investigators: how order influences participant-investigators' interpretations of eyewitness identification and alibi evidence. *Law and Human Behaviour, 33*(5), 368-380. doi:10.1007/s10979-008-9151-y

Devine, D. J., Clayton, L. D., Dunford, B. B., Seying, R., & Pryce, J. (2001). Jury decision making: 45 years of empirical research on deliberating groups. *Psychology, Public Policy, and Law, 7*(3), 622-631. doi:10.1037/1076-8971.7.3.622

Edens, J. F., Buffington, J. K., Tominic, T. L., & Riley, B. D. (2001). Effects of postivie impression management on the psychopathic personality inventory. *Law Hum Behav, 25*(3), 235-56. doi:10.1023/A:1010793810896

Englich, B., Mussweiler, T., & Strack, F. (2006). Playing dice with criminal sentences: the influence of irrelevant anchors on experts' judicial decision making. *Pers Soc Psychol Bull, 32*(2), 188-200. doi:10.1177/0146167205282152

Evans, J. ST. B. T. (2006). The heuristic-analytic theory of reasoning: extension and evaluation. *Psychonomic Bulletin & Review, 13*(3), 378-395. doi:10.3758/BF03193858

Feeney, A., Coley, J. D., & Crisp, A. (2010). The relevance framework for category-based induction: Evidence from garden-path arguments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(4), 906-911. doi: 10.1037/a0019762

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). LA: Sage.

Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of consumer research*, *20*(2), 303-315. doi:10.1086/209351

Frigotto, M. L., & Rossi, A. (2015). An explanatory coherence model of decision making in ill-structured problems. *Mind Soc, 14*, 35-55. doi:10.1007/s11299-014-0158-4.

Ganster, D. C., Hennessey, H. W., & Luthans, F. (1983). Social desirability response effects: Three alternative models. *Academy of Management Journal*, *26*(2), 321-331. doi:10.2307/255979

Garrison, J. W., & Hoskisson, K. (1989). Confirmation bias in predictive reading. *The Reading Teacher*, *42*(7), 482-486. Retreived from: http://www.jstor.org/stable/20200195?seq=1#page_scan_tab_contents

Grimm, P. (2010). Social desirability bias. *Wiley International Encyclopedia of Marketing*. doi:10.1002/9781444316568.wiem02057

Hernandez, I. & Preston, J. L. (2013). Disfluency disrupts the confirmation bias. *Journal of Experimental Social Psychology, 49*, 178-182. doi: 10.1016/j.jesp.2012.08.010.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive psychology*, *24*(1), 1-55. doi:10.1016/0010-0285(92)90002-J

Keppens, J., & Schafer, B. (2006). Knowledge based crime scenario modelling. *Expert Systems with Applications*, *30*(2), 203-222. doi:10.1016/j.eswa.2005.07.011

Kiani, R., Corthell, L., & Shadlen, M. N. (2014). Choice certainty is informed by both evidence and decision time. *Neuron, 84*(6), 1329-1342. doi:10.1016/j.neuron.2014.12.015

King, M. F., & Brunner, G. C. (2000). Social desirability bias: a neglected aspect of validity testing. *Psychology & Marketing*, *17*(2), 79-103. doi: 10.1002/(SICI)1520-6793(200002)17:2<79::AID-MAR2>3.0.CO;2-0

Klevorick, A. K., & Rothschild, M. (1979). A model of the jury decision process. *The Journal of Legal Studies*, *8*(1), 141-164. doi:10.1086/467604

Koslowski, B., Marasia, J., Chelenza, M., & Dublin, R. (2008). Information becomes evidence when an explanation can incorporate it into a causal framework. *Cognitive Development*, *23*(4), 472-487. doi: 10.1016/j.cogdev.2008.09.007

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity*, *47*(4), 2025-2047.doi:10.1007/s11135-011-9640-9

Lagnado, D. (2011). Thinking about Evidence1. In *Proceedings of the British Academy* 171, 183-223. Retrieved from http://www.ucl.ac.uk

Lagnado, D. A., & Harvey, N. (2008). The impact of discredited evidence. *Psychonomic Bulletin & Review*, *15*(6), 1166-1173. doi:10.3758/PBR.15.6.1166

Mackenzie, E., Chalmers, E., Wastell,C. A., Duncan, P., & Roberts, M. (2017). *Choice and Background Knowledge: How do individuals evaluate accumulating evidence in a murder scenario?*. Manuscript submitted for publication.

McKenzie, C. R. M. (2005). Judgment and decision making. In R. L. G. Koen Lamberts (Ed.), *Handbook of cognition* (pp. 321-338). Sage.

Medin, D. L., Coley, J. D., Storms, G., & Hayes, B. L. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review*, *10*(3), 517-532. doi:10.3758/BF03196515

Mercier, H. (2016). The argumentative theory: predictions and empirical evidence. *Trends in Cognitive Science, 20*(9). doi:10.1016/j.tics.2016.07.001

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34*(2), 57-74. doi:10.1017/S0140525X10000968

Meyer, A., Frederick, S., Burnham, T. C., Guevara Pinto, J. D., Boyer, T. W., Ball, L. J., ... & Schuldt, J. P. (2015). Disfluent fonts don't help people solve math problems. *Journal of Experimental Psychology: General*, *144*(2), e16. doi:10.1037/xge0000049

Navarro, D. J., & Perfors, A. F. (2010). Hypothesis generation, sparse categories, and the positive test strategy. *Psychology Review, 118*(1), 120-134. doi:10.1037/a0021110

Norwood, F. B., & Lusk, J. L. (2011). Social desirability bias in real, hypothetical, and inferred valuation experiments. *American Journal of Agricultural Economics*, *93*(2), 528-534. doi: 10.1093/ajae/aaq142

O'Donnell, C. M., & Safer, M. A. (2017). Jury instructions and mock-juror sensitivity to confession evidence in a simulated criminal case. *Psychology, Crime & Law, 23*(10). doi:10.1080/1068316X.2017.1351965

Phillips, B., Prybutok, V. R., & Peak, D. A. (2014). Decision confidence, information usefulness, and information seeking intention in the presence of disconfirming information. *Informing Science: the International Journal of an Emerging Transdiscipline*, *17*, 1-25. Retrieved from http://www.inform.nu/Articles/Vol17/ISJv17p001- 024Phillips386.pdf

Pozzulo, J. D., Dempsey, J., Maeder, E., & Allen, L. (2010). The effects of victim gender, defendant gender, and defendant age on juror decision making. Criminal Justice and Behavior, 37, 47–63. doi:10.1177/0093854809344173

Price, H. L., & Dahl, L. C. (2013). Order and strength matter for evaluation of alibi and eyewitness evidence. *Applied Cognitive Psychology, 28,* 143-150. doi:10.1002/acp.2983

Rassin, E., Eerland, A., & Kuijpers, I. (2010). Let's find the evidence: an analogue study of confirmation bias in criminal investigations. *Journal of Investigative Psychology and Offender Profiling, 7*, 231-246. doi:10.1002/jip.126

Resnikoff, T., Ribaux, O., Baylon, A., Jendly, M., & Rossy, Q. (2015). The polymorphism of crime scene investigation: An exploratory analysis of the influence of crime and forensic intelligence on decisions made by crime scene examiners. *Forensic Science International, 257,* 425-434. doi:10.1016/j.forsciint.2015.10.022

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale. *Journal of clinical psychology*, *38*(1), 119-125. doi: 10.1002/1097-4679(198201)38:1<119::AID-JCLP2270380118>3.0.CO;2-I

Ryan, J. (1969). Temporal grouping, rehearsal and short-term memory. *The Quarterly Journal of Experimental Psychology*, *21*(2), 148-155. doi:10.1080/14640746908400207

Sedley, S. (1993). Whose Justice? In *London Review of Books, page 6, 23 June 1993.*

Shadlen, M. N., & Kiani, R. (2013). Decision making as a window on cognition. *Neuron, 80,* 791-806. doi:10.1016/j.neuron.2013.10.047

Silberberger, J., Van Wezemael, J., Paisiou, S., & Strebel, I. (2010). Spaces of knowledge creation: tracing'knowing in action'in jury-based decision-making processes in Switzerland. *International Journal of Knowledge-Based Development*, *1*(4), 287-302. doi:10.1504/IJKBD.2010.038040

Simon, D., & Holyoak, K. J. (2002). Structural dynamics of cognition: From consistency theories to constraint satisfaction. *Personality and social psychology review*, *6*(4), 283-294. doi: 10.1207/S15327957PSPR0604_03

Simon, D., Pham, L. B., Le, Q. A., & Holyoak, K. J. (2001). The emergence of coherence over the course of decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(5), 1250-1260. doi: 10.1037/0278-7393.27.5.1250

Simon, D., Snow, C. J., & Read, S. J. (2004). The redux of cognitive consistency theories: evidence judgments by constraint satisfaction. *Journal of Personality and Social Psychology*, *86*(6), 814-819. doi.org/10.1037/0022-3514.86.6.814

Solska, A., & Rojcsyk, A. (2015). Exploring interpretation and misinterpretation of garden-path sentences in polish. *International Review of Pragmatics, 7,* 98-127. doi:10.1163/18773109-01802100

Tetlock, P. E. (2005). *Expert political judgment.* Princeton: Princeton University Press.

Tetlock, P. E., & Gardner, D. (2015). *Superforecasting: The art and science of prediction*. New York: Random House.

Verheij, B. (2014). To catch a thief with and without numbers: arguments, scenarios and probabilities in evidential reasoning. *Law, Probability and Risk, 13*(3-4), 307-325. doi:10.1093/lpr/mgu011

Wastell, C., Feeney, A., Coley, J. & Weeks, N. (2012a). *Processing disfluency relates to belief revision via a process of evidence reappraisal*. Paper presented at the International Conference on Thinking, London.

Wastell, C. A., Weeks, N., Wearing, A., & Duncan, P. (2012b). Identifying hypothesis confirmation behaviors in a simulated murder investigation: Implications for practice. *Journal of Investigative Psychology and Offender Profiling.* 9, 184-198. doi:10.1002/jip.1362

Weeks, N. J., Wastell, C. A., Taylor, A. J., Wearing, A. J., & Duncan, P. (2012). Tracing decision processes in complex, ambiguous, information-rich environments. *International Journal of Psychological Studies, 4*(1), 158-173. Retrieved from http://proquest.com

Wright, G., & Goodwin, P. (2009). Decision making and planning under low levels of predictability: enhancing the scenario method. *International Journal of Forecasting, 25*, 813-825. doi:10.1016/j.ijforecast.2009.06.019

Yechiam, E., Kanz, J. E., Bechara, A., Stout, J. C., Busemeyer, J. R., Altmaier, E. M., & Paulsen, J. S. (2008). Neurocognitive deficits related to poor decision making in people behind bars. *Psychonomic Bulletin & Review, 15*(1), 44-51. doi: 10.3758/PBR.15.1.44

Appendix A of this thesis has been removed as it may contain sensitive/confidential content

Appendix B

Crime Scenarios

B1 – Pilot Study Murder Scenario

As a private investigator, you have been commissioned to investigate the murder of James Smith. Listed below are some of the case details.

Date: Saturday the 10th of February 2018
Estimated time of death: Between 2:00pm to 4:00pm
Location: 4 High Street, Rydel

Cause of death report: James Smith suffered a fatal gunshot wound to the head. Crime scene investigators reported finding gunshot residue on the victims face suggesting that the shooting occurred at close range. The bullet was extracted and examined, yet experts were unable to specify the exact type of weapon the bullet came from. Initial reports indicate that James Smith was having an affair with Ashley Buddle, a close neighbour of the Smiths.

Both Ashley Buddle and the victim's wife, Emily Smith, have been questioned in relation to the murder. Both Ashley and Emily are suspects in this case.

The following sections will present witness statements collected prior to your arrival on scene related to each suspect. You will also be required to indicate to what extent each piece of evidence may impact upon your decisions regarding each suspect's level of guilt

B2 – Pilot Study Robbery Scenario

As a private investigator, you have been commissioned to investigate the robbery of Alex Higgins. Listed below are some of the case details.

Date: Friday the 6th of June, 2018
Estimated time of robbery: Between 11:30pm and 12:00am
Location: Clyde Alley, Placton

Case report: Alex Higgins was found blindfolded, hands bound with rope in Clyde Alley late Friday night. An initial search determined that his wallet, mobile phone and expensive gold wrist watch had been taken. Alex Higgins reported that his assailant approached from behind and threatened him with a knife but could not provide any further details.

Initial reports indicate that two individuals, Jessica Salt and Luke Garrow, were seen near the alley around the time of the assault. Both Jessica and Luke are suspects in this case.

The following sections will present witness statements collected prior to your arrival on scene related to each suspect. You will also be required to indicate to what extent each piece of evidence may impact upon your decisions regarding each suspect's level of guilt

B3 – Experiment 1 and 2 Murder Scenario: High Plausibility

As a private investigator, you have been commissioned to investigate the murder of James Smith. Listed below are some of the case details.

Date: Saturday the 10th of February 2018
Estimated time of death: Between 2:00pm to 4:00pm
Location: 4 High Street, Rydel

Cause of death report: James Smith suffered a fatal gunshot wound to the head. Crime scene investigators reported finding gunshot residue on the victims face suggesting that the shooting occurred at close range. The bullet was extracted and examined, yet experts were unable to specify the exact type of weapon the bullet came from.
Initial reports indicate that James Smith was having an affair with Ashley Buddle, a close neighbour of the Smiths. Australian Bureau of Crime Statistics data suggests that, in cases with similar circumstances, most often the victim's spouse is responsible. Both Ashley Buddle and the victim's wife, Emily Smith, have been questioned in relation to the murder. Emily is known to have a history of physically violent tendencies, a factor associated with murder.

At present the prime suspect is *Emily Smith*.

The following pages will present witness statements collected prior to your arrival on scene. You will also be required to indicate to what extent you believe *Emily* to be guilty.

B4 – Experiment 1 and 2 Murder Scenario: Low Plausibility

As a private investigator, you have been commissioned to investigate the murder of James Smith. Listed below are some of the case details.

Date: Saturday the 10th of February 2018
Estimated time of death: Between 2:00pm to 4:00pm
Location: 4 High Street, Rydel

Cause of death report: James Smith suffered a fatal gunshot wound to the head. Crime scene investigators reported finding gunshot residue on the victims face suggesting that the shooting occurred at close range. The bullet was extracted and examined, yet experts were unable to specify the exact type of weapon the bullet came from.
Initial reports indicate that James Smith was having an affair with Ashley Buddle, a close neighbour of the Smiths. Australian Bureau of Crime Statistics data suggests that, in cases with similar circumstances, most often the victim's spouse is responsible. Both Ashley Buddle and the victim's wife, Emily Smith, have been questioned in relation to the murder. Ashley is not known to have a history of physically violent tendencies, a factor associated with murder.

At present the prime suspect is *Ashley Buddle*.

The following pages will present witness statements collected prior to your arrival on scene. You will also be required to indicate to what extent you believe *Ashley* to be guilty.

B5 – Experiment 1 and 2 Robbery Scenario: High Plausibility

As a private investigator, you have been commissioned to investigate the robbery of Alex Higgins. Listed below are some of the case details.

Date: Friday the 6th of June, 2018
Estimated time of robbery: Between 11:30pm and 12:00am
Location: Clyde Alley, Placton

Case report: Alex Higgins was found blindfolded, hands bound with rope in Clyde Alley late Friday night. An initial search determined that his wallet, mobile phone and expensive gold wrist watch had been taken. Alex Higgins reported that his assailant approached from behind and threatened him with a knife but could not provide any further details.
Australian Bureau of Crime Statistics data suggests that, in cases with similar circumstances, most often a male is responsible. Initial reports indicate that two individuals, Jessica Salt and Luke Garrow, were seen near the alley around the time of the assault. Luke Garrow is known to have current financial problems, a factor associated with robbery.

At present the prime suspect is *Luke Garrow*.

The following pages will present witness statements collected prior to your arrival on scene. You will also be required to indicate to what extent you believe *Luke* to be guilty.

B6 – Experiment 1 and 2 Robbery Scenario: Low Plausibility

As a private investigator, you have been commissioned to investigate the robbery of Alex Higgins. Listed below are some of the case details.

Date: Friday the 6th of June, 2018
Estimated time of robbery: Between 11:30pm and 12:00am
Location: Clyde Alley, Placton

Case report: Alex Higgins was found blindfolded, hands bound with rope in Clyde Alley late Friday night. An initial search determined that his wallet, mobile phone and wrist watch had been taken. Alex Higgins reported that his assailant approached from behind and threatened him with a knife. Placton is a busy neighbourhood and many businesses were open in the hours leading up to and following the incident.
Australian Bureau of Crime Statistics data suggests that, in cases with similar circumstances, most often a male is responsible. Initial reports indicate that two individuals, Jessica Salt and Luke Garrow, were seen near the alley around the time of the assault. Jessica Salt is not known to have any current financial problems, a factor associated with robbery.

At present the prime suspect is *Jessica Salt.*

The following pages will present witness statements collected prior to your arrival on scene. You will also be required to indicate to what extent you believe *Jessica* to be guilty.

Appendix C

Plausibility Manipulation Checks

C1 – Murder Plausibility Manipulation Check

Based on the information presented to you in the previous section please answer the following True or False questions:

In similar circumstances, it is more common that the spouse is responsible for the murder

- True
- False

In similar circumstances, it is more common that someone other than the spouse is responsible for the murder

- True
- False

People who are responsible for murder are more likely to have physically violent tendencies

- True
- False

[Prime suspect] has a history of physically violent tendencies

- True
- False

[Prime suspect] has no history of physically violent tendencies

- True
- False

C2 – Robbery Plausibility Manipulation Check

Based on the information presented to you in the previous section please answer the following true or false questions:

In similar circumstances, it is more common that a male is responsible for the robbery

- True
- False

In similar circumstances, it is more common that a female is responsible for the robbery

- True
- False

People who are responsible for robberies are more likely to have current financial problems

- True
- False

[Prime suspect] has current financial problems

- True
- False

[Prime suspect] does not have any current financial problems

- True
- False

Appendix D

Witness Statements

D1 – Murder Witness Statements

Witness: Mary Ladle
Status: Emily's co-worker

Emily is always yelling at someone on the phone. I guess it must be her husband. This morning, yeah, it sounded pretty bad.

Witness: Hillary Thompson
Status: Emily's boss

Emily has had some problems at work lately. I had to give her a warning the other day. Afterwards, others heard her yelling at her husband on the phone

---

Witness: Lillian Simon
Status: Emily's close friend

Emily found out about her husband's affair a couple of days ago, she was devastated and told me she was considering confronting him about it

---

Witness: Bob Newton
Status: Ashley's close friend

Yeah, Ashley and I go to the gun range all the time. We're avid enthusiasts. This morning, sure, we always go to the range on Saturday

---

Witness: Jane Morslee
Status: Neighbour

Ashley was definitely at the house around that time. I was out for a walk and saw her go inside. What time, I guess it was around 2pm.

Witness: John Buddle
Status: Ashley's Brother

Ashley seemed really tense the last couple of days. She was muttering, something about being in a bad situation… something she needed to get out of

Note. Quasi-randomisation of evidence was: a, b, c; b, c, a; c, a, b

D2 – Robbery Witness Statements

Witness: Clarissa Johnston
Status: Nearby neighbour

I'm sure I saw a man hurrying out of the alley right around midnight … I suppose he fits the description of Luke.

Witness: Mitchell Gray
Status: 'Gray and Sons Hardware' Store Clerk

Luke's one of our regular customers, he was just in here yesterday. I remember he bought rope because I had to get more from the storeroom to restock the shelves afterwards.

Witness: Christian Murray
Status: Luke's co-worker

I was in a club across the street from the alley, I went outside for some air at about 11.50pm or midnight when I saw Luke. It looked like he'd just come out of the alley.

Witness: Jacinta Morris
Status: Jessica's roommate

Jessica has a knife collection that she keeps in her room: she's got all kinds of knives, vintage ones, decorative ones- she really loves that collection.

Witness: Isla Williams
Status: Nearby neighbour

Anyone hanging around the alley? Come to think of it, I did see someone going into the alley, might've been 11:45pm. She looked a lot like one of my neighbours, Jessica.

Witness: Elias Harrison
Status: Jessica's boyfriend

Jessica was supposed to meet me for a drink around 11:30pm. She was running really late which made me quite mad, well, that is until she surprised me with a gold wrist watch for my birthday

Note. Quasi-randomisation of evidence was: a, b, c; b, c, a; c, a, b

Appendix E

Pilot Study Questions

How clear (easy to understand) is this piece of evidence? ○ Extremely clear ○ Moderately clear ○ Slightly clear ○ Neither clear nor unclear ○ Slightly unclear ○ Moderately unclear ○ Extremely unclear	If you were asked to determine the level of guilt for either one of the suspects, how important would this piece of evidence be in your decision making process? ○ Extremely important ○ Moderately important ○ Slightly important ○ Neither important nor unimportant ○ Slightly unimportant ○ Moderately unimportant ○ Extremely unimportant
How relevant do you think this piece of evidence is in relation to the case details? ○ Extremely relevant ○ Moderately relevant ○ Slightly relevant ○ Neither relevant nor irrelevant ○ Slightly irrelevant ○ Moderately irrelevant ○ Extremely irrelevant	How likely is it that this piece of evidence would impact upon your decision, in the form of a guilt rating, for either one of the suspects? ○ Extremely likely ○ Moderately likely ○ Slightly likely ○ Neither likely nor unlikely ○ Slightly unlikely ○ Moderately unlikely ○ Extremely unlikely

Appendix F

Guilt Rating Scale

Given the information so far, please rate the likelihood that the prime suspect, [name], is guilty of the [crime] on a scale from -5 (Most Probably Not Guilty) to 5 (Most Probably Guilty).

-5 Most probably not guilty
-4
-3
-2
-1
0 Not sure
1
2
3
4
5 Most probably guilty

Appendix G

Confidence Rating Scale

Given the information so far, please rate your level of confidence in the rating you provided for prime suspect, [name], being guilty of the [crime] on a scale from 0% (No Confidence) to 100% (Complete Confidence).

0% No Confidence
10%
20%
30%
40%
50% Neutral
60%
70%
80%
90%
100% Complete Confidence

Appendix H

Social Desirability Scale

Listed below are a number of statements concerning personal attitudes and traits. Read each item and decide whether the statement is true or false as it pertains to you

It is sometimes hard for me to go on with my work if I am not encouraged

- True
- False

I sometimes feel resentful when I don't get my way.

- True
- False

On a few occasions, I have given up doing something because I thought too little of my ability.

- True
- False

There have been times when I felt like rebelling against people in authority even though I knew they were right.

- True
- False

No matter who I'm talking to, I'm always a good listener.

- True
- False

There have been occasions when I took advantage of someone.

- True
- False

I'm always willing to admit it when I make a mistake.

- True
- False

I sometimes try to get even rather than forgive and forget.

- True
- False

I am always courteous, even to people who are disagreeable.

- True
- False

I have never been irked when people expressed ideas very different from my own.

- True
- False

There have been times when I was quite jealous of the good fortune of others.

- True
- False

I am sometimes irritated by people who ask favours of me.

- True
- False

I have never deliberately said something that hurt someone’s feelings.

- True
- False

Appendix I of this thesis has been removed as it may contain sensitive/confidential content

Appendix J

Demographics Questions

What is your Gender?

- Male
- Female
- Other
- Not Specified

What is your age in whole years?

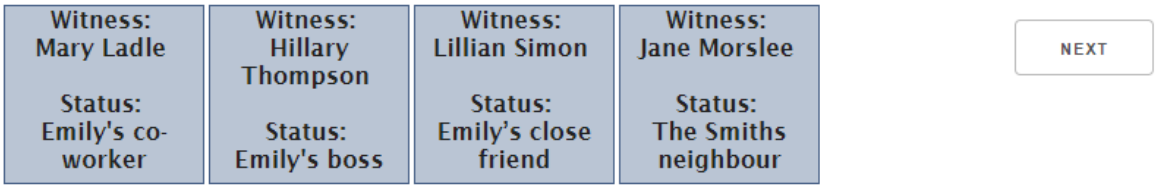
I am currently enrolled in:

- PSYC104
- PSYC105
- PSY246

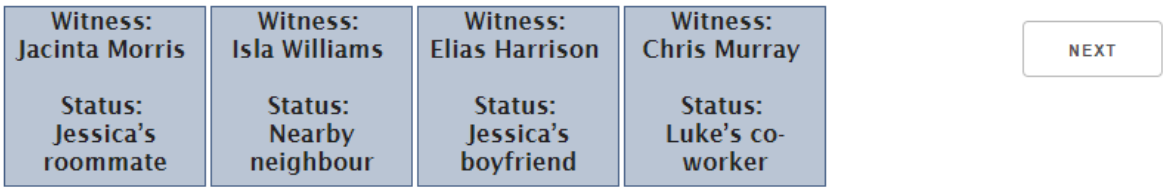
Appendix K

Evidence Summary

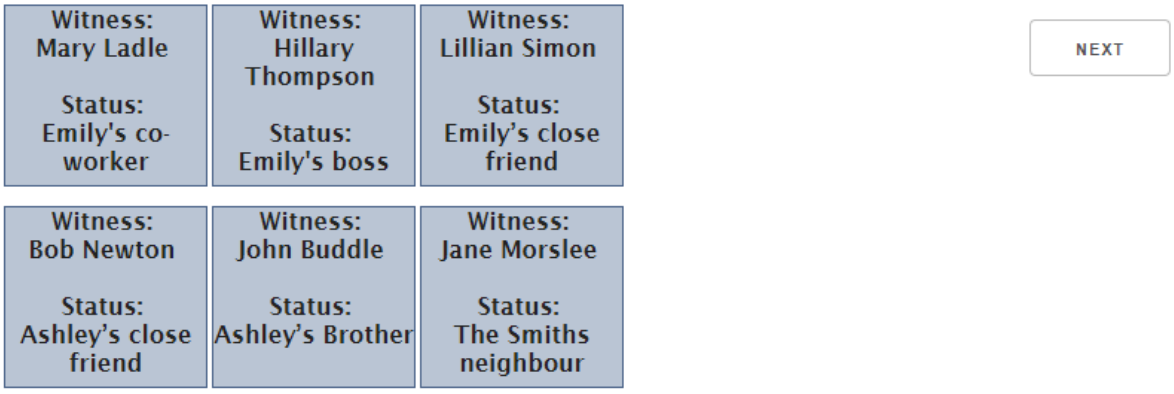
K1- Experiment 1 Murder Scenario Example



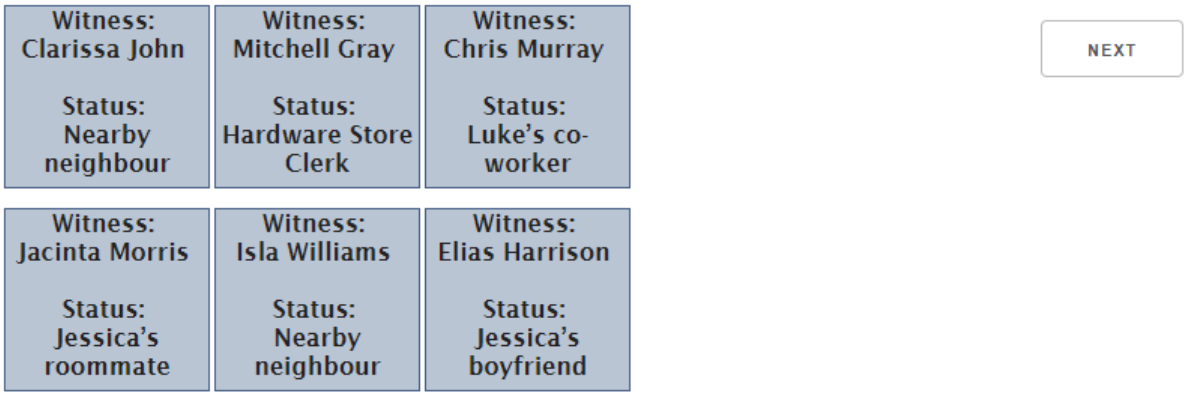
K2- Experiment 1 Robbery Scenario Example



K3- Experiment 2 Murder Scenario Example



K4- Experiment 2 Robbery Scenario Example



Appendix L

Outlier Tables

L1 -- Experiment 1

Table 2. Experiment 1 Excluded Participants Frequencies

Experiments allocated to	Reason Excluded			Total
	Technical errors	Response greater than 2 standard deviation from the group mean	Socially desirable response style	
1 and 6	1	7	2	10
2 and 5	2	3	6	11
3 and 8	0	1	4	5
4 and 7	1	5	3	9
Total	4	16	15	35

Note. Experiment 1 = murder, high plausibility, non-incriminating first; 2 = murder, high plausibility, non-incriminating last; 3 = murder, low plausibility, non-incriminating first; 4 = murder, low plausibility, non-incriminating last; 5 = robbery, high plausibility, non-incriminating first; 6 = robbery, high plausibility, non-incriminating last; 7 = robbery, low plausibility, non-incriminating first; 8 = robbery, low plausibility, non-incriminating last.

L2 -- Experiment 2

Table 6. Experiment 2 Excluded Participants Frequencies

Experiments allocated to	Reason Excluded			Total
	Technical errors	Response greater than 2 standard deviation from the group mean	Socially desirable response style	
1 and 6	1	6	2	9
2 and 5	4	8	1	13
3 and 8	3	5	4	12
4 and 7	2	4	6	12
Total	10	23	13	46

Note. Experiment 1 = murder, high plausibility, non-incriminating first; 2 = murder, high plausibility, non-incriminating last; 3 = murder, low plausibility, non-incriminating first; 4 = murder, low plausibility, non-incriminating last; 5 = robbery, high plausibility, non-incriminating first; 6 = robbery, high plausibility, non-incriminating last; 7 = robbery, low plausibility, non-incriminating first; 8 = robbery, low plausibility, non-incriminating last.

Appendix M

Descriptive Statistics for Experiment 1

M1 – Murder

Table 3. Experiment 1 Descriptive Statistics for Murder

Plausibility	Order		Guilt Rating M(SD)	Confidence Rating M(SD)
High	Non-incriminating first	1	1.78(1.851)	57.14(21.115)
		2	1.38(1.989)	55.60(21.491)
		3	1.72(1.785)	59.20(21.174)
		4	1.94(1.743)	60.21(20.052)
		5	2.37(1.577)	62.24(21.435)
		6	2.55(1.659)	64.79(22.121)
High	Non-incriminating last	1	2.07(1.528)	57.22(22.354)
		2	2.19(1.545)	60.77(18.455)
		3	2.46(1.397)	61.51(19.748)
		4	2.70(1.409)	65.58(16.735)
		5	2.51(1.678)	63.14(21.212)
		6	2.58(1.679)	64.80(21.308)
Low	Non-incriminating first	1	-0.45(2.292)	48.49(26.414)
		2	-0.51(2.172)	49.81(23.817)
		3	0.96(1.969)	53.20(22.985)
		4	1.57(2.022)	57.65(23.374)
		5	2.42(1.736)	62.71(22.193)
		6	2.26(1.648)	63.19(23.601)
Low	Non-incriminating last	1	-0.51(1.977)	47.17(27.060)
		2	0.68(2.208)	54.26(25.298)
		3	0.76(2.119)	58.33(21.697)
		4	2.16(1.346)	63.20(20.745)
		5	0.72(2.382)	58.15(25.334)
		6	1.44(2.218)	65.19(20.910)

Note: 1 = Initial, 2 = After evidence 1, 3 = After evidence 2, 4 = After evidence 3, 5 = After evidence 4, 6 = Final. Guilt ratings and Confidence ratings scales

M2 – Robbery

Table 4. Experiment 1 Descriptive Statistics for Robbery

Plausibility	Order		Guilt Rating M(SD)	Confidence Rating M(SD)
High	Non-incriminating first	1	1.96(1.636)	57.59(21.626)
		2	-.38(1.821)	50.75(21.109)
		3	1.11(1.577)	59.06(20.406)
		4	2.05(1.508)	63.09(21.071)
		5	2.57(1.475)	66.79(20.636)
		6	2.65(1.494)	66.18(22.235)
High	Non-incriminating last	1	1.44(1.939)	56.23(25.209)
		2	2.37(1.341)	58.43(21.668)
		3	2.86(1.327)	63.14(23.366)
		4	3.21(1.348)	66.35(25.051)
		5	1.10(2.277)	51.35(23.766)
		6	1.53(2.176)	56.67(22.949)
Low	Non-incriminating first	1	-1.51(2.054)	52.00(26.345)
		2	-1.70(2.129)	49.12(26.409)
		3	0.31(2.017)	56.00(21.222)
		4	1.47(1.864)	64.00(19.110)
		5	2.30(1.537)	68.70(18.938)
		6	2.33(1.812)	70.19(19.952)
Low	Non-incriminating last	1	-1.44(1.708)	49.45(25.706)
		2	0.92(1.591)	55.09(21.179)
		3	2.09(1.290)	63.40(20.091)
		4	2.81(1.442)	65.56(22.120)
		5	1.40(2.015)	60.18(21.300)
		6	2.07(1.747)	64.63(21.431)

Note: 1 = Initial, 2 = After evidence 1, 3 = After evidence 2, 4 = After evidence 3, 5 = After evidence 4, 6 = Final.

Appendix N

Main Effects and Interactions

N1 – Experiment 1

Table 5. Main Effects and Interactions for Experiment 1

Main effects	F(df)	Sig.	η_p^2
1. Guilt ratings	123.007(2.767)	< .00005	.246
2. Plausibility	24.400(1)	< .00005	.061
3. Order	16.735(1)	< .00005	.043
4. Crime	1.788(1)	.182	.005
Interactions	F(df)	Sig.	η_p^2
1*2	14.666(2.767)	< .00005	.037
1*3	87.056(2.767)	< .00005	.188
1*4	7.923(2.767)	< .00005	.021
2*3	0.762(1)	.383	.002
2*4	1.286(1)	.257	.003
3*4	3.581(1)	.059	.009
1*2*3	1.197(2.767)	.309	.003
1*2*4	1.046(2.767)	.368	.003
1*3*4	16.062(2.767)	< .00005	.041
2*3*4	1.878(1)	.171	.005
1*2*3*4	4.527(2.767)	.005	.012

Note. Effects including 1. Guilt Rating were Greenhouse-Geisser corrected.

N2 -- Experiment 2

Table 9. Main Effects and Interactions for Experiment 2

Main effects	F(df)	Sig.	η_p^2
1. Guilt ratings	19.923(3.560)	< .00005	.054
2. Plausibility	86.489(1)	< .00005	.200
3. Order	0.116(1)	.734	.000
4. Crime	21.886(1)	< .00005	.059
Interactions	F(df)	Sig.	η_p^2
1*2	26.000(3.560)	< .00005	.070
1*3	115.999(3.560)	< .00005	.251
1*4	3.024(3.560)	.021	.009
2*3	43.134(1)	< .00005	.111
2*4	.004(1)	.951	.000
3*4	20.194(1)	< .00005	.055
1*2*3	3.077(3.560)	.020	.009
1*2*4	5.155(3.560)	.001	.015
1*3*4	26.354(3.560)	< .00005	.071
2*3*4	40.689(1)	< .00005	.105
1*2*3*4	2.861(3.560)	.028	.008

Note. Effects including 1. Guilt Rating were Greenhouse-Geisser corrected.

Appendix O

Descriptive Statistics for Experiment 2

O1 – Murder

Table 7. Experiment 2 Descriptive Statistics for Murder

Plausibility	Order		Guilt Rating M(SD)	Confidence Rating M(SD)
High	Non-incriminating first	1	1.80(1.789)	52.55(26.251)
		2	1.58(1.819)	53.40(24.042)
		3	1.39(1.845)	54.00(22.857)
		4	1.14(2.315)	55.40(21.966)
		5	1.71(1.979)	57.29(21.212)
		6	1.25(2.317)	56.92(20.247)
		7	1.90(1.865)	60.20(20.946)
		8	1.86(1.990)	61.84(19.437)
High	Non-incriminating last	1	2.14(1.443)	58.16(17.281)
		2	2.13(1.204)	62.17(20.211)
		3	2.47(1.309)	63.88(18.576)
		4	2.74(1.273)	66.74(17.519)
		5	2.70(1.517)	65.53(17.546)
		6	2.36(1.882)	63.40(20.566)
		7	1.88(2.247)	63.54(21.781)
		8	1.83(2.596)	66.46(21.086)
Low	Non-incriminating first	1	-0.06(2.082)	50.59(22.575)
		2	-0.73(2.268)	51.57(25.090)
		3	-1.00(2.458)	53.47(25.703)
		4	-1.08(2.480)	54.31(25.865)
		5	-0.18(2.543)	55.29(22.747)
		6	0.82(2.247)	57.80(21.313)
		7	1.22(2.540)	62.94(20.129)
		8	1.12(2.582)	60.59(21.299)
Low	Non-incriminating last	1	-0.38(2.294)	50.60(22.261)
		2	0.46(1.935)	55.21(18.449)
		3	1.17(1.948)	60.64(17.119)
		4	1.79(1.829)	61.49(17.443)
		5	1.26(1.799)	57.23(16.511)
		6	0.94(1.983)	56.67(18.141)
		7	0.51(2.083)	53.47(20.971)
		8	0.42(2.711)	59.40(22.715)

Note: 1 = Initial, 2 = After evidence 1, 3 = After evidence 2, 4 = After evidence 3, 5 = After evidence 4, 6 = After evidence 5, 7 = After evidence 6, 8 = Final. Guilt ratings and Confidence ratings scales

O2 – Robbery

Table 8. Experiment 2 Descriptive Statistics for Robbery

Plausibility	Order		Guilt Rating M(SD)	Confidence Rating M(SD)
High	Non-incriminating first	1	1.71(1.272)	54.22(21.584)
		2	0.47(1.660)	50.22(20.393)
		3	-0.89(2.305)	53.40(21.597)
		4	-1.96(2.654)	51.49(27.739)
		5	-0.02(1.994)	52.34(18.441)
		6	1.56(1.436)	57.44(20.011)
		7	1.96(1.673)	60.22(20.601)
		8	1.58(1.803)	60.70(21.201)
High	Non-incriminating last	1	1.96(1.508)	53.08(22.970)
		2	2.36(1.382)	56.20(25.225)
		3	2.71(1.404)	60.00(25.846)
		4	3.12(1.596)	66.27(24.328)
		5	1.94(1.953)	57.80(23.412)
		6	1.25(2.325)	51.76(23.126)
		7	0.13(2.896)	51.13(25.770)
		8	0.69(2.631)	53.08(24.458)
Low	Non-incriminating first	1	-1.88(2.048)	53.27(24.791)
		2	-2.51(2.073)	56.00(19.272)
		3	-3.19(1.907)	59.60(20.400)
		4	-3.64(1.721)	67.92(19.347)
		5	-1.31(2.763)	61.25(19.310)
		6	-0.10(2.978)	59.80(20.553)
		7	0.94(2.470)	57.45(18.851)
		8	0.90(2.435)	60.38(24.250)
Low	Non-incriminating last	1	-1.19(2.068)	43.47(24.200)
		2	0.54(1.798)	49.18(28.638)
		3	1.64(1.675)	52.13(32.231)
		4	2.81(1.123)	52.22(33.971)
		5	1.71(1.543)	55.92(26.530)
		6	0.72(2.157)	55.80(23.219)
		7	0.39(2.384)	62.04(21.211)
		8	0.92(2.488)	60.41(21.791)

Note: 1 = Initial, 2 = After evidence 1, 3 = After evidence 2, 4 = After evidence 3, 5 = After evidence 4, 6 = After evidence 5, 7 = After evidence 6, 8 = Final.