Suggestibility in Adult Witnesses:

Exploring The Impact of Different Factors That Can Impact Eyewitness Memory Accuracy

KARA DEERING, FORENSIC PSYCHOLOGY PRACTICE DOCTORATE, THE CENTRE FOR APPLIED PSYCHOLOGY. UNIVERSITY OF BIRMINGHAM

SEPTEMBER 2021

Abstract

Mistaken eyewitness memory is thought to be the most common cause of wrongful convictions. This thesis explores the impact of factors that can contribute to eyewitness memory inaccuracy, including the misinformation effect. The misinformation effect refers to the tendency for post-event information to interfere with a person's memory for an original event. As discussed in Chapter One, the misinformation effect is commonly studied using a misinformation paradigm: participants experience a to-be-remembered event and then are presented with information that is either consistent or inconsistent with the information that was presented during the witnessed event. Afterwards, participants are tested to examine the extent to which they can distinguish between their memory of the event and the post-event information they encountered. Chapter Two presents a systematic literature review of experimental studies that used the misinformation paradigm. The key aim was to explore whether the congruence between the modality (e.g., visual, auditory, text) in which information presented across the different stages of the misinformation paradigm (encoding, misinformation, test) plays a role in misinformation susceptibility. The findings suggest that congruence between the modality in which information is presented at encoding and at test reduces misinformation acceptance. However, although there was some evidence that modality manipulations across the different stages of the testing paradigm influence misinformation susceptibility, the evidence was limited and results across studies were conflicting. Chapter Three presents empirical research that was undertaken to explore the impact of misinformation on a lineup identification accuracy. Participants watched a mock crime wherein the perpetrator's face was shown from the front or profile facial angle, and then they were presented with misinformation about the perpetrator's appearance. Misinformation presentation was controlled by exposing participants to a video of a news report that featured an innocent suspect, who was presented either in the same pose as the perpetrator or shown from a different angle. Memory for the perpetrator (i.e., the guilty

suspect) was tested using a simultaneous lineup procedure, wherein the test faces matched the same facial angle that participants saw at encoding. This study found that participants were more ***likely to be misinformed when profile faces were presented at encoding compared to frontal faces***. Chapter Four presents a psychometric critique of the only validated measure of adult interrogative suggestibility, the Gudjonsson Suggestibility Scale. The tool was found to be both a reliable and valid measure of interrogative suggestibility. The theses findings are discussed within the context of the misinformation literature and practical implications are considered in Chapter Five.

Acknowledgements

The completion of this thesis would not have been possible without the unwavering support of Dr Melissa Colloff and Professor Heather Flowe. I feel so grateful to have learnt from you. Thank you for all of your reassurance, encouragement, and expertise. I could not have wished for nicer or more supportive supervisors. I would also like to extend thanks to the rest of the ForenPsyD faculty. In particular, Dr Caroline Oliver, for inspiring my enrolment on the course and your warmth. Thanks also to Christina Moreton, for your support during my placements and for your kindness throughout.

To all of the service users I have had the privilege of speaking with over the course of the last three years. I have valued every story, experience and emotion you chose to share with me. I will be forever grateful for your time.

To my friends, for your understanding, patience and compassion. Although I certainly could have done with less invites to social events, I could not have gone without your laughter. Your words of support have been instrumental in helping me to reach the finish line. This of course, includes my course mates, who I have had the privilege of sharing this journey with.

Finally, to my parents, Michael, Lucy and Matt. Without you, none of this would have been possible. No sacrifice was too much, no problem could not be solved, and no IT disaster could not be fixed. Your love and belief have been felt so profoundly throughout this journey and you have been at the heart of everything I do. There will never be a "thank you" big enough. This is for you.

Table of Contents

List of Appendices

List of Tables

List of Figures

Chapter One: Introduction

Memory can be fallible. Whilst memories are meaningful in our daily lives and hold emotional significance, they are not recorded and played back veridically (truthfully), as we might expect. Rather, our memory can be vulnerable to contamination and distortion. These distortions can occur in different kinds of memories (e.g., working memory and veridical memory), and no single type of person is either vulnerable or resistant to this contamination (Nichols & Loftus, 2019). One significant way our memory can be distorted is through misinformation. The misinformation (MI) effect refers to the impairment in memory that occurs after exposure to misleading information. That is, misinformation can cause people to falsely believe they saw or heard details that were only suggested to them. These false memories can be durable and, once they are embraced, people can 'recall' and express the false memories with confidence and with detail (Loftus, 2005). For example, imagine an eyewitness saw a crime and then overheard an incorrect description of that crime on the news that night. In a subsequent police interview, the witness might include the incorrect information they heard on the news in their account of what they witnessed during the crime. This provides one example of how misinformation can contaminate eyewitness memory reports.

A prominent context explored in the misinformation literature is eyewitness testimony and identification in the legal system. Eyewitness testimony refers to the verbal statements made by witnesses regarding what and who they observed during a crime. Eyewitness identification is a specific type of eyewitness testimony, where a witness is asked to recognise an individual as the person who committed the crime (Wells & Olsen, 2003). Eyewitness testimony is frequently used to guide police investigations (MacLin et al., 2008). The testimony is most often collected through police interviews and police identification (ID) procedures and is later reiterated by the witness in court. Eyewitnesses are typically unaware of the influence post-event information may have had on their memory. Post-event information refers to information that is received and processed following an event that has the potential to interfere with the original memory of the actual event. Therefore, eyewitnesses may confidently report what they remember and remain confident in their suspect identification, incorporating the misleading post-event information (Sporer et al., 1995). Therefore, it is important for the police and legal decision makers to minimise the impact of memory contamination and to support correct identifications of guilty suspects.

One of the ways police support witnesses to correctly (or incorrectly) identify a guilty suspect is by presenting them with a lineup. Lineups typically consist of the police suspect and a number of other individuals who are known to be innocent, but who physically resemble the suspect. These individuals are known as fillers. Although in some countries the default lineup procedure is conducted live, other criminal justice departments use videos and photographs (Fitzgerald et al., 2018). The images can be presented one at a time (a sequential lineup) or all at the same time (a simultaneous lineup). Once presented with the lineup, witnesses are asked to identify who they believe is the perpetrator. Eyewitness identification evidence is very influential within legal proceedings (Koriat et al., 2000; Wells & Olson, 2003). However, eyewitness identification has been found to be one of the most common errors leading to incorrect convictions (Findley, 2001). In particular, post-conviction DNA profiling has found that misidentification by at least one eyewitness played a significant role in 70% of incorrect convictions in the USA which were later overturned using DNA evidence (Innocence Project, 2017). Based on these figures, researchers have estimated that up to one hundred innocent people could be wrongly convicted of serious violent offences in the UK each year due to eyewitness misidentification (Mojtahedi, 2017). This estimate needs to be interpreted with caution as the USA and UK use different lineup procedures to aid eyewitness identification. The USA typically use a simultaneous lineup procedure, whereas a sequential lineup procedure is used in the UK. This can be problematic when making estimates as research has argued that sequential lineups are superior to simultaneous (Stebley et al., 2011), and that simultaneous lineups are superior to sequential (Meisters et al., 2017) in the context of eyewitness identification. However, this highlights that memory errors could result in a “false positive”, where an innocent suspect is incorrectly identified as the perpetrator. In the case of Lydell Grant, an innocent man who was sentenced to life in prison in 2012 for murder, six eyewitnesses incorrectly identified him as the perpetrator in a lineup. It was only due to later DNA evidence that Grant was cleared of the crime and another person was implicated. Furthermore, memory errors could lead to a “false negative”, where a perpetrator is not identified as guilty, thus walking free.

Theoretical Explanations of the MI Effect

Several theoretical accounts have been proposed for why the MI effect occurs. Loftus (1975, 1979a) proposed an overwrite/trace alteration account. This argues that the original memory for an event is altered or overwritten by post-event misinformation. Therefore, the original memory no longer exists in its original form. Instead, when subjects are asked to recall the original information, they access the updated memory trace containing the misinformation in its place. Support for the overwrite/trace alteration account has been found in literature (see Loftus, 1975; Morton et al., 1985). However, research by Lindsay and Johnson (1989) found that misinformation can still have an effect without it overwriting the original memory. Bekerian and Bower’s (1983) study investigating this account also found evidence that the original memory for the to-be-remembered information remains intact despite post-event misinformation. These findings are not consistent with Loftus’ (1975) overwrite/trace alteration account because in these studies there was evidence that the original memory was not overwritten. Therefore, other explanations for the MI effect are explored.

In contrast to Loftus' theory, McCloskey and Zaragoza (1985) argued the strategic effects account. This postulates that the MI effect does not occur due to memory impairment (as Loftus suggests). Instead, they argue that subjects simply forget the to-be-remembered information, due to the demand of the task, independent of the misinformation. Specifically, the original information is no longer remembered, but alternative information is recognised as having been presented at the time of the original information. This theory was supported by Zaragoza et al. (1987), who conducted two studies investigating the effects of misleading post-event information on subjects' ability to recall details of a to-be-remembered event. Results from both studies showed that misinformation had no effect on subjects' ability to recognise the original to-be-remembered information. This supports the strategic effects theory, as memory for the original event was not impaired by the misinformation. In contrast, a more recent review by Ayers and Reder (1998) highlighted that research designed to reduce or eliminate strategic effects has found evidence of memory impairment. For example, Ceci et al. (1987) presented children with stories and found that there was memory impairment following misinformation. Likewise, Chandler (1991) presented adults with visual scenes followed by similar visual scenes as misinformation. The findings supported that there had been memory impairment as subjects were unable to discriminate between the original and misinformation scenes. Therefore, Ayers and Reder (1998) argued that none of the aforementioned theoretical explanations offer a full account of why the MI effect occurs.

To explain the MI effect, Ayers and Reder (1998) put forward the activation-based framework. This assumes that once information is stored in memory, representation of that information in memory gets stronger or weaker depending on how frequently and how recently it has been encountered and activated. They postulated that the amount of activation dictates the strength of the memory relative to the strength of competing misinformation. The MI effect, therefore, is more likely to occur when an original memory is not repeatedly activated and post-event misinformation is introduced. Whilst Ayers and Reder's (1998) activation-based framework has been extensively referenced throughout misinformation and memory literature, less has been done to empirically prove or disprove the framework. Therefore, it could be argued that this theory of misinformation has not yet been fully assessed for its ability to account for the MI effect.

Another theory used to explain the MI effect is the Source Monitoring Framework (SMF). The SMF is concerned with how people differentiate between memories from different sources. One of the core concepts underpinning the SMF is that thoughts and feelings that are experienced as memories are attributed to particular sources of past experience (Lindsay & Johnson, 2000). Simply put, subjects remember the misleading information, but misattribute it to the witnessed event, rather than the source of the misleading information (Brubacher et al., 2020). Reyna and Lloyd (1997) famously criticised the SMF, arguing that the complexity of the framework makes it unfalsifiable and of limited value. In response, Lindsay and Johnson (2000) supported the SMF, saying it was a comprehensive framework used to understand false memories. Despite debate regarding the validity of the SMF, it has been widely used throughout false memory research (e.g., Hekkanen et al., 2002; Henkel & Coffman, 2004; Lindsay & Johnson, 2000).

Whilst people may be influenced by misinformation in reporting their memories, this can also be due to informational or normative influence. Informational influence refers to when someone accepts information obtained from another as evidence about reality, even if this is not the case (Deutsch & Gerard, 1955). In contrast, normative influences are based on when a person tries to gain social approval to avoid alienation from others (Deutsch & Gerard, 1955; Goodwin et al., 2013). For example, when someone is faced with peer pressure. Although both informational and normative influences can lead an individual to report misinformation, neither influence result in false memories. Rather, they explain how a person may be motivated towards conforming behaviours. However, both informational and normative influences can be affected by post-event information (Goodwin et al., 2013).

Together, there appears to be a consensus amongst researchers that no single process is responsible for the different ways we can be misinformed. While researchers favour varying theoretical explanations for why the MI effect occurs, none has been proven to fully account for the MI effect. Whichever explanation is correct, it is clear that the vast body of misinformation research has supported the view that memory is reconstructive, and that the reconstruction process leaves memory vulnerable to errors (Pickrell et al., 2016). The majority of empirical research investigating the MI effect has focused on measuring misinformation acceptance and understanding the circumstances under which these errors occur.

Experimentally Testing MI

Researchers testing the MI effect typically use the misinformation paradigm (MI paradigm). The MI paradigm is comprised of three stages; first, in the encoding stage, original information is presented to the subject, often in the form of a video or slideshow; second, misinformation is introduced to the subject. This can be achieved in several ways, including the introduction of misleading questioning. This is where an interviewer will express opinions or make inferences in the wording of their questions (Roebers & Schneider, 2000), lastly, subjects are tested on their memory for the original event. The MI effect is said to have occurred if subjects who have been exposed to misinformation are more likely to report erroneous information or false memories at test (Ayers & Reder, 1998; Loftus, 1975).

The MI paradigm has been used to investigate factors that influence the size of the MI effect. These factors include individual differences such as age, wherein general, younger children are more susceptible to misinformation compared with older children and adults (Ceci & Bruck, 1993). The individual presenting the misinformation can have influence over the size of the MI effect. For example, Skagerberg and Wright (2009) conducted a study where subjects watched an event and were then given testimony from a co-witness, either a child or police officer. The MI effect was only found when the misinformation was given by the police officer. Skagerberg and Wright (2008) also witnessed this effect in a lineup identification task, whereby subjects were more likely to feel confident in their incorrect suspect identification if they were told a number of police officers also made the same choice. Research seems to suggest that the MI effect is larger when the person presenting the misinformation is deemed to be more credible. The impact of the timing of misinformation has also been considered, with findings suggesting that subjects can be more suggestible after longer time delays between encoding and misinformation (Roberts & Powell, 2007). However, the opposite has been found, where a stronger MI effect was found when misinformation was presented closer in time to the original event (Melnyk & Bruck, 2004). Collectively, these studies show that there are various factors that may impact the MI effect and that findings are equivocal. Furthermore, there are some factors, such as modality, that have been less researched. Modality is considered, next.

Modality and MI

The MI effect has been investigated by myriad researchers for at least forty-five years (Loftus, 2005). Therefore, it is not surprising that there has been much progress in misinformation research. Although researchers have addressed a number of different issues related to misinformation, little is known about how modality effects may impact misinformation susceptibility. Modality can be defined as the mode in which information is experienced or expressed (e.g., visually or auditorily). In literature the modality effect specifies that, under some circumstances, memory will be enhanced if textual information (i.e., written text) is presented in an auditory format compared to a visual format (Ginns, 2005). The misinformation literature has, however, given little consideration to the modality manipulations used in studies and the impact these may have on the size of the effect observed in the MI paradigm. This is important for the misinformation literature as the majority of misinformation studies change modalities at each stage of the MI paradigm (e.g., encoding video, followed by a written narrative introducing misinformation, and then an interview at test). However, few have considered the impact of these manipulations on their research findings.

Thesis Aims

The aim of this thesis is to explore the suggestibility of adult witnesses. More specifically, to examine the impact of modality manipulations within the MI paradigm and explore and critique how suggestibility can be measured in practice. To achieve these aims, a systematic literature review was undertaken (Chapter Two) with two main objectives; first, to explore whether similarity in presentation modality (e.g., same modality used throughout stages of the MI paradigm) will increase or decrease susceptibility to the MI effect; second, to explore whether specific modalities (i.e., auditory or visual) at each of the three stages of the MI paradigm appear to be more effective at reducing susceptibility to the MI effect. Overall, results were mixed, with varying findings for the impact of modality congruence and the use of different modalities in the MI paradigm.

Next, an eyewitness identification experiment was conducted (Chapter Three) to explore the impact of viewing angle on misinformation susceptibility using an MI paradigm. Participants watched a mock crime depicting the perpetrator's face from a front or profile facial angle. Next, an innocent suspect was either shown from a front or profile facial angle during a news report video that contained misleading information. Memory for the perpetrator (i.e., the guilty suspect) was tested during a simultaneous lineup procedure, which always matched the facial angle shown at encoding. Participants were randomly assigned to one of eight conditions: 2 encoding view (front, profile) x 2 misinformation view (front, profile) x 2 lineup (target-absent, target-present). The position of the faces shown during the lineup always matched the facial position shown at encoding, whereas the misinformation face could either be presented in the same or a different facial position to the encoding and lineup. Two research questions were posed. First, we asked if discrimination accuracy would be higher when the innocent suspect's face is presented in a different pose at the misinformation stage compared to the encoded event and lineup test (modality incongruent), compared to when the innocent suspect's face is presented in the same pose at the misinformation stage as the encoded event and lineup test (modality congruent). There was no significant difference between discrimination accuracy between the congruent and incongruent facial angle conditions. Second, we asked if discrimination accuracy would be higher when participants viewed the innocent suspect's face from a profile viewing angle at the misinformation stage when the encoding and test faces were frontal, compared to when participants viewed the innocent suspect's face from a frontal viewing angle at the misinformation stage when the encoding and test faces were profile. There was significantly better discrimination accuracy when participants saw the perpetrator's face from a frontal viewing angle at encoding followed by a profile facing innocent suspect at the misinformation stage compared to a when they viewed the perpetrator's face from a profile viewing angle at encoding followed by frontal facing innocent suspect at the misinformation stage. Further analysis showed that the confidence-accuracy relationship was weaker for profile compared to frontal encoding conditions.

Having explored the impact of experimental, external, and situational factors to the person that may influence the MI effect, the internal factors to the person, such as interrogative suggestibility, were also considered. Specifically, it was considered how interrogative suggestibility is measured and its relevance in forensic settings. A critique of the Gudjonsson Suggestibility Scale (GSS) of interrogative suggestibility was conducted and the reliability and validity of the measure was examined (Chapter Four). The GSS is, at present, the only existing validated measure of suggestibility for adult populations. Strong evidence for both the reliability and validity of the measure was found, supporting its practical use in forensic settings.

Finally, the general discussion (Chapter Five) considers the thesis as a whole, and the collective findings across chapters. Together, this thesis provides a unique contribution to the empirical understanding of how adult witnesses can be susceptible to misinformation and how interrogative suggestibility is measured.

Chapter Two

A Literature Review Following a Systematic Approach: An Assessment of The Impact of Modality on Susceptibility to Misinformation

Abstract

Studies that have used the misinformation (MI) paradigm have traditionally presented the to-be-remembered event in a visual modality (e.g., a video), the misinformation in an auditory modality (e.g., misleading questions), and later tested participants in a different visual or auditory modality (e.g., written recognition test or interview questions). Despite this, little is known regarding the impact that modality may have on susceptibility to memory errors across the different stages of the MI paradigm. The widely accepted encoding specificity hypothesis predicts that congruence between encoding and test stages should result in fewer memory errors. The current review aimed to explore the role that manipulations in modality at different stages of the MI paradigm have on the MI effect. The review included eleven papers, published between 1996 and 2021, assessed as either moderate or strong in quality. There were mixed findings regarding the impact of different modalities at different stages of the MI paradigm, including several papers that found non-significant effects of modality on the MI effect. Conflicting findings regarding the impact of specific modality changes at different stages of the MI paradigm are discussed. Findings supported the encoding specificity hypothesis, where modality congruence at encoding and test appear to reduce misinformation acceptance. However, limited evidence also supported that modality congruence between the misinformation and test stages may increase misinformation acceptance. The strengths and limitations of the review are discussed, and recommendations are made for future research and practical application in relation to eyewitness identification.

Introduction

Human memories can be susceptible to misinformation (Loftus, 2005; Straube, 2012), known as the MI effect. The MI effect refers to memory distortions of past events that are the result of post-event exposure to misleading information (Loftus, 2005). Traditionally, psychological studies testing the MI effect have used the MI paradigm, where participants are exposed to an initial encoding event, inconspicuously introduced to misinformation about that event, and then have their memory of the event tested. The format in which each stage of the MI paradigm is presented, whether that be auditory (e.g., voice recording) or visual (e.g., mock crime video), can be referred to as its modality. Commonly in studies using the MI paradigm, participants watch the initial event via a video, the misleading information is introduced via misleading questions or by asking participants to freely recall an event, and then participants' memories are tested verbally (such as a memory test) (Loftus et al., 1978). However, other studies have used different modalities. For example, participants are sometimes exposed to the original event via a live enactment (Gobbo et al., 2002; Roebers et al., 2004). In practice, eyewitnesses are frequently exposed to different presentation modalities, such as witnessing the crime (visual-auditory), police questioning (auditory), and written statements (visual).

Currently, it is not clear how modality changes across the stages of the MI paradigm can impact on participants' misinformation susceptibility (Dijkstra & Moerman, 2012). The current review aims to systematically examine the impact of modality on the MI effect. Specifically, the primary objectives of the review are to explore whether similarity in presentation modality (e.g., same modality used throughout stages of the paradigm) will increase or decrease susceptibility to the MI effect; and whether specific modalities at each of the three stages of the MI paradigm appear to be more effective at reducing susceptibility to the MI effect.

Modality and the MI Paradigm

Changes in modality across the different stages of the MI paradigm are often implemented to reflect real-life scenarios, such as watching a crime video (mimicking witnessing a crime), receiving misleading information via a written narrative (mimicking reading a news report containing inaccuracies about the crime you witnessed), and having our memory for the original event tested (such as recognition tests). Although this method is perhaps ecologically valid, the result is that misinformation studies vary in modality changes across different stages of the MI paradigm, with little consideration for how this may influence findings (Ulatowska et al., 2016). Evidence suggests that false memories can occur when an individual is presented with information via a range of different modalities, such as verbal and visual information (Ulatowska et al., 2016). Although many studies have manipulated modality within and across different stages of the MI paradigm, it is important to systematically consider how representations of original and misleading information impact on memory accuracy.

Within experimental psychology, the impact of information presentation on memory is referred to as the modality effect, whereby someone's learning and memory performance is dependent on the mode in which presentation items are studied (Sweller et al., 2011). For example, a meta-analysis showed that subjects who were presented with information in a combination of modalities (e.g., auditory spoken text and visual graphical information) outperformed on a memory test than those who were only presented with information in a visual format (Ginns, 2005). As well as the importance of the modality effect for encoding of information, research has also focused on how modality can impact memory retrieval, evaluation and endorsement. For example, research has found evidence that modality effects on memory accuracy can be caused by differences in retrieval orientation (Pierce & Gallo, 2011; Smith & Hunt, 1998).

One explanation for the link between modality and memory is the Source Monitoring Framework (SMF). The SMF proposes that memories are comprised of perceptual details, semantic information, traces of cognitive operations, emotions and temporal details (Mitchell & Johnson, 2009). When a memory is retrieved, a subset of the combined details present at the time of encoding are activated (Henkel & Coffman, 2004). This implies that the origin of a memory can be inferred from sensory attributes, such as the colour or pitch of the original information (Garrison et al., 2017). Other situational elements are also activated, such as expectations, assumptions, emotional state and task orientation (Johnson et al., 1996). These elements are particularly important within the MI paradigm as they are introduced during the misinformation stage, forcing participants to process external information. When considering real-world examples, the MI effect closely mimics situations where a witness sees a crime, they then read misinformation about that crime in a newspaper, and are then interviewed by the police about the crime they witnessed following misinformation exposure via a newspaper article. Therefore, it is necessary to consider the propensity for false memories explicitly within the MI paradigm to ascertain the level at which external and internal factors increase the MI effect (Schacter et al., 2012).

The Modality Effect Across Encoding and Test

When considering how modality may influence misinformation susceptibility within the MI paradigm, the majority of existing literature has focused on encoding and test stages. The existing encoding specificity hypothesis and research suggest that modality congruence across encoding and test benefits memory performance. In memory literature, the encoding specificity hypothesis predicts that modality congruence between the encoding and test stages should produce fewer memory errors (Ulatowska et al., 2016). In part, this is due to the similarity in processing required between similar modalities (i.e., visual encoding stage, visual test stage or verbal encoding stage and verbal test stage). Other research suggests that when modalities differ from each other at different stages, the task of making sense of coding information becomes more strenuous, leading to errors (Penney, 1989). For example, verbal information presented in a visual modality (such as a written narrative) requires explicit translation to a phonological format before it can access the phonological buffer (Shallice & Vallar, 1990). Specifically, verbal information presented in a visual (e.g., written) format must undergo visual analysis which is then recoded from a visual to a phonological format. This means that verbal information (i.e., information that contains either written or spoken words) is stored with a certain level of mental representation, helping individuals to distinguish between different linguistic structures. Therefore, it can be argued that when information is presented in the same format at encoding and test, there will be fewer memory errors. This is because the similarity in processing between the pieces of information support memory retrieval of the to-be-remembered information. In contrast, where information is presented in different formats, there are likely to be more memory errors. This is because the different processing requirements needed to process dissimilar types of information are more strenuous, making it more difficult to correctly retrieve the to-be-remembered information. In short, these theories predict a modality congruency hypothesis, which states that there will be a reduction in false memories when there is modality congruence at encoding and test (Gallo et al., 2001; Smith & Hunt, 1998).

The MI paradigm used in research traditionally uses different modalities (modality incongruence) at encoding, misinformation, and test stages. Modality incongruence can also produce differences in false memories, in different sensory modalities, due to the different regions of the brain being used to process each modality (Hendrich, 2019). Therefore, modality congruency across the three different stages of the paradigm may have different, and conflicting effects. I explain this, next.

The Modality Effect Across Encoding and The Misinformation Stage

While current literature largely supports the idea that modality congruence at encoding and test reduces false memories, little is known about how modality congruency at encoding and the misinformation stage of the MI paradigm may impact misinformation susceptibility. However, the limited research that does exist suggests that modality is an important determinant of misinformation acceptance. Okado and Stark (2005) used fMRI scans to explore brain activity during the original encoding event and the misinformation stage. Their findings suggest that the processes that occur during original event encoding and the misinformation stage play a critical role in determining the true and false memory outcomes in the MI paradigm. Specifically, when similar processing occurs at encoding and the misinformation stage (such as the processing of two similar modalities), this may result in more false memories. Stark et al. (2010) further explored this fMRI research, presenting the encoding stage as visual, and the misinformation stage as auditory. Reactivation of both the auditory and visual cortex was observed during the retrieval (test) stage. False recollections of auditory based misinformation were found to relate to greater auditory activity compared to when the misinformation was correctly rejected. This suggests that there are biological processes that impact on misinformation susceptibility at the encoding and misinformation stages of the MI paradigm.

Other researchers have also suggested that modality congruency across the encoding and the misinformation stages might increase susceptibility to the MI effect (Greene, 1992). Loftus (1979b) conducted a study where participants either received misinformation that either subtly contradicted or that blatantly contradicted the original information. The findings suggested that obvious differences between the encoding event and misinformation can reduce susceptibility to misinformation and result in participants being more resistant to suggestion. This is because when an individual is presented with a piece of information that clearly contradicts the original information, they are less likely to use this new information to update their memory. Therefore, it could be hypothesised that an obvious contradiction in modality may produce a similar effect, meaning different modalities may decrease susceptibility to misinformation. Whilst modality congruence at encoding and test may be beneficial for reducing false memories, it could be hypothesised that modality congruence at encoding and the misinformation stage may have the reverse effect in the MI paradigm and increase susceptibility to false information. However, Loftus' (1979b) study primarily focused on obvious discrepancies in verbal information, not obvious discrepancies in modality at the different stages of the MI paradigm. Further research is needed to determine whether the concept of blatant contradiction applies across stages of the MI paradigm.

The Modality Effect Across the Misinformation Stage and Test

Little is known about the relationship between modality congruency across misinformation and test stages. To make sense of this, Okado and Stark (2005) described both the original encoding event and the introduction of post-event misinformation as encoding events. Namely, both stages can be considered to be encoding events. Considering the encoding specificity principle, modality congruence between encoding and test has been found to reduce false memories. If the misinformation stage can also be considered as an encoding event, this same modality congruence between misinformation and test could result in *increased* recall of misinformation at test. Put another way, encoding specificity is usually associated with a reduction in false memories in situations where the encoding and test stages match. However, when the principle is applied at misinformation (another encoding event) and test, the opposite pattern may occur because more false information (from the misinformation stage) may be recalled.

Overall, there has been an increasing number of studies exploring the relationships between modality manipulations and the MI effect. However, this area of research is still in its infancy compared to the myriad studies that have utilised the MI paradigm. Reviews investigating the modality effect have largely focused on its impact within educational environments (Ginns, 2005). Moreover, few studies have considered the impact of modality on the MI effect observed in the MI paradigm. Most misinformation studies have utilised similar modality changes in their procedure. For example, the encoding stage is commonly presented as a video (visual), misinformation is presented as a misleading narrative/questioning (verbal/auditory), and memory tests are presented visually (i.e., written questions regarding the event). However, for the purpose of this review all *between* modality changes (i.e., auditory or visual) and *within* modality changes (i.e., video, written narrative and picture – difference modalities with the visual modality) will be considered. Therefore, the current review aimed to explore the role that manipulations in modality at different stages of the MI paradigm have on the MI effect.

Method

Systematic Search

An initial scoping exercise was conducted to explore the current body of literature surrounding modality and its impact on misinformation acceptance. This informed the potential scope of the synthesis and helped to refine the aims of the review. Initial scoping highlighted that there was sufficient literature on modality and misinformation to conduct the present review. To date, no other reviews exist that explore the use of modality across the different stages of the MI paradigm.

Initial searches were conducted on 23rd October 2019 to establish the originality of the proposed review using the following databases:

- The Campbell Library of Systematic Reviews
- The Centre of reviews and Dissemination (DARE)
- Cochrane Database of Systematic Reviews
- Medline
- Ovid
- PsycArticles
- PsycInfo.

Preliminary searches yielded no relevant systematic reviews or meta-analyses regarding modality and the MI effect. Searches were also conducted using Google Search and Google Scholar, where no evidence of existing reviews in this topic area were uncovered.

A scoping search was conducted to identify whether there was enough literature to conduct a full review. The following processes were conducted: 1) a Google search was conducted to identify existing literature, and articles of interest that were identified through their abstracts were also reviewed via their full article to ensure relevance and 2) reference lists of identified papers were scoped for further relevant articles. Search terms were explored by identifying prominent terms within the literature and relevant synonyms. All identified terms were subsequently combined and input into databases to ascertain the relevance of searches. From this process search terms were selected, and Boolean operators were refined. Multiple combinations of search terms were attempted, and a more specific, smaller number of search terms yielded the most relevant journal papers, as well as excluding a larger number of unrelated papers. The final search terms are displayed in Figure 2.1.

Misinformation effect*, by itself or with any letters following this, for example effect/effects ***OR***

Misinformation paradigm*, by itself of with any letters following this, for example paradigms

AND

Modalit*, by itself or with any letters following this, for example modality/modalities ***OR***

Presentation Modalit*, by itself or with any letters following this, for example modality/modalities ***OR***

Presentation of information ***OR***

Format* of stimul*, by itself or with any letters following this, for example formatting and stimulus/stimuli

AND

Differ*, by itself or with any letters following this, for example differing/difference ***OR***

Alternat*, by itself or with any letters following this, for example alternate/alternative ***OR***

Contrast*, by itself or with any letters following this, for example contrasting/contrasted ***OR***

Similar*, by itself or with any letters following this, for example similarity

(misinformation effect* OR misinformation paradigm*) AND (modalit* OR presentation modalit* OR presentation of information OR format* of stimul*) AND (differ* OR alternat* OR contrast* OR similar*)

Figure 2.1. Search Terms

Potential studies for inclusion in this review were initially sourced on 27th December 2019 (and again on 12th July 2021) using the following electronic databases:

- Cochrane Library (1990-2021)
- Web of Science (1990-2021)
- OVID PsycArticles (1990-2021)
- OVID PsycInfo (1990-2021)

- PUBMED (1990-2021)
- EBSCO Host (1990-2021)
- SCOPUS (1990-2021)
- Science Direct (1990-2021)
- ProQuest (1990-2021)
- OVID (journals) (1990-2021).

Search terms were applied to the ten databases yielding a total of 533 results. Once duplicates were removed a total of 459 articles remained. An additional sixteen were identified through alternative means: ten were identified via existing reference lists and six via Google searching. An initial sifting process excluded 435 papers from the review based on their titles, abstracts, and full articles where relevance was unclear. Reasons articles were removed during initial sifting included: no reference made to the MI effect, no comparisons between modality manipulations, no use of a MI paradigm, and because they were reviews, theoretical discussions, or opinion papers. A total of forty papers from systematic searching of databases, Google searching, and scoping of relevant reference lists were considered for application of PICO criteria. For the remaining forty papers identified as relevant to the review and the sixteen articles identified through alternative means, full copies of the texts, where available, were obtained through the University of Birmingham e-library or access online. When papers could not be accessed in full, where possible, attempts were made to contact researchers in the field. PICO criteria were applied to the remaining thirty-five out of a total forty papers as five papers could not be accessed in full (see Appendix One).

Following application of PICO inclusion criteria, a total of eighteen studies across eleven papers were relevant to the review. Appendix Two provides information regarding the studies that were found to be unsuitable for the review after inclusion and exclusion criteria were applied. The majority of the remaining papers included multiple experiments where all were found to be relevant. Where specific experiments were not relevant to answering the review questions they were ignored, and only pertinent aspects of the paper were considered (e.g., Yamashita, 1996). Appendix Three provides information regarding the studies found to meet inclusion criteria. A diagrammatical flow chart detailing the data selection process is provided in Figure 2.2.

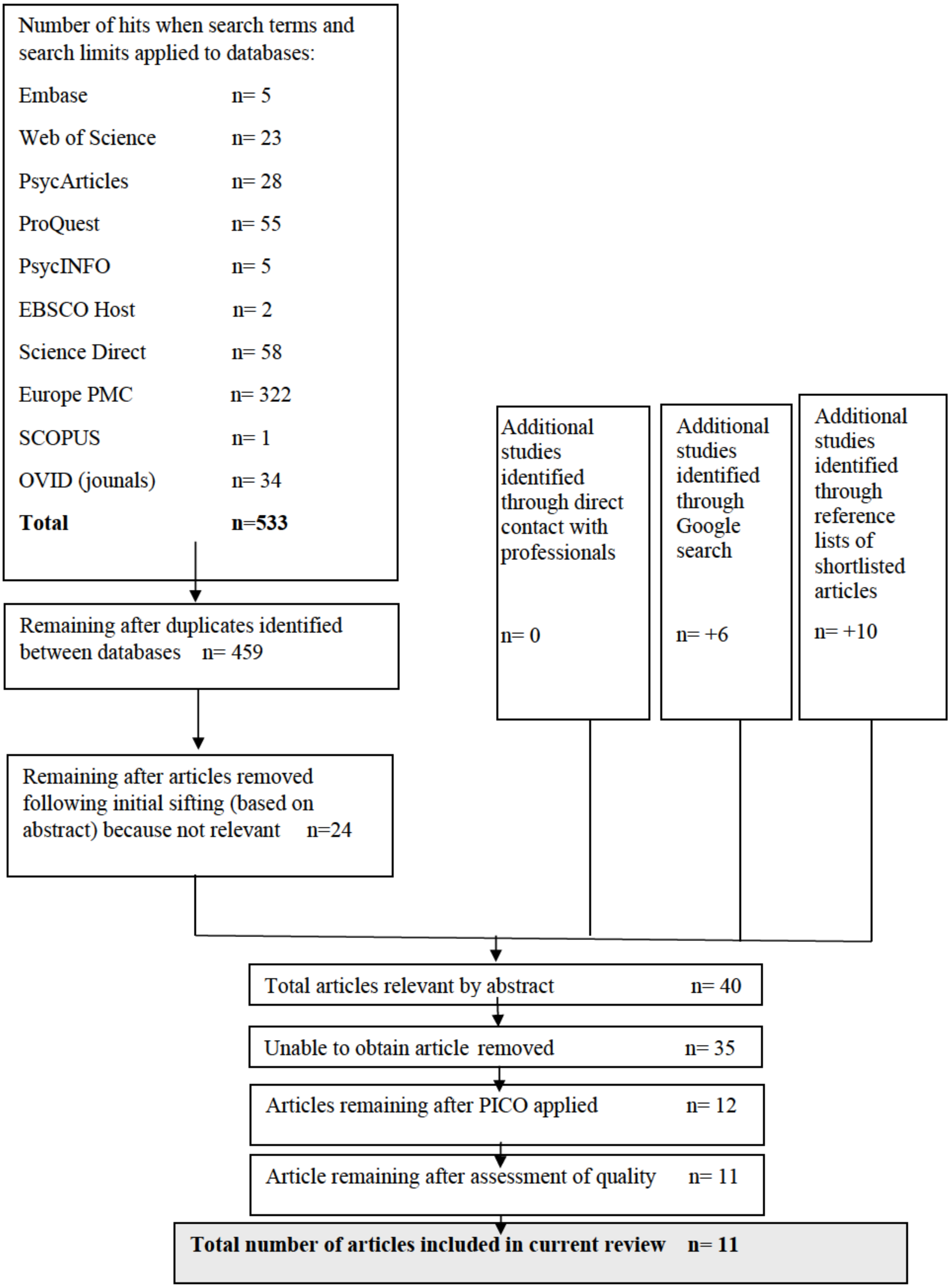

Figure 2.2. Data selection process.

Inclusion/ Exclusion Criteria

To aid the identification of relevant articles, inclusion/exclusion criteria were developed (Table 2.1).

Table 2.1 Inclusion/Exclusion Criteria

PICO	Inclusion Criteria	Exclusion Criteria
Population	Males and/or Females. Any Nationality. Any Ethnicity. Any Age.	Any participant groups who are unable to adequately access presentation modality/formats used.
Intervention/ Exposure	Different modalities (e.g., auditory or visual). Differences within modalities (e.g., video tape, written narrative, and pictures are different visual modalities). Modality manipulation at encoding, misinformation stage or test, including studies with modality manipulations at multiple stages. Two or more modality manipulations at one stage of paradigm (e.g., original event presented as video or live action) or two or more modality manipulations across paradigm stages (e.g., original event in video, misinformation presented auditorily).	Articles that do not explicitly measure the impact of modality on the MI effect. Specifically, studies where modality changes may be present, however the impact of this has not been considered within the results or discussion sections.
Comparator	Comparisons between congruent (modalities are the same across MI Paradigm stages, e.g., auditory encoding and auditory test) and incongruent (modalities are different across MI Paradigm stages, e.g., auditory encoding and visual test) modalities.	Comparison across studies within a single research paper where one or more studies being compared were excluded from the review.

PICO	Inclusion Criteria	Exclusion Criteria
Outcome	Interaction effects between modality and the MI effect. In a single paper, cross-study comparisons of the impact of modality manipulations on MI effect where explicit links are made in results/discussion section between studies.	Studies that do not report results about the influence of modality on the MI effect.
Study Design	Use of traditional MI Paradigm. Use of non-traditional MI Paradigm (e.g., reversed MI Paradigm).	Studies explicitly exploring false memory through the use of alternative paradigms e.g., DRM Paradigm. Qualitative studies, narratives, commentaries, editorials; other varieties of opinion papers.
Other Factors	Year of publication: 1990-2021. Language of publication: English.	

A PICO framework was favoured over other protocols as it has been found to identify a greater number of relevant articles in comparison to other tools, such as SPIDER (Methley et al., 2014). Given that the current review only encompassed quantitative studies, the PICO framework was deemed suitable due to its recognised use in locating quantitative research. The exposure and comparator elements of the PICO framework enabled the researcher to explicitly consider what modalities the population were exposed to at each stage of the MI paradigm, and whether modality exposures were congruent or incongruent. The following rationales were identified for inclusion/exclusion criteria for Population, Intervention or Exposure, Comparator and Outcome (PICO) framework:

Population

Given the limited number of studies identified during the initial scoping exercise, combined with the fact that anyone can witness an event and receive subsequent post-event misinformation, the criteria for population were deliberately broad. Specifically, the MI effect has been witnessed across different ethnicities, ages and genders (Lee, 2004; Loftus, 2005; Sutherland & Hayne, 2001). Therefore, it was important that all were included in the current review. Only participants who could not access different modalities, such as those with sight issues or hearing difficulties, were excluded from this review.

Intervention/Exposure

In the context of the review, a manipulation in modality is defined as a change in modality at one or more stages of the MI paradigm. Both between-modality manipulations (e.g., visual and auditory) and within-modality manipulations (e.g., visual- written narrative or video) were included. The current review did not include studies examining differences in susceptibility to misinformation for central versus peripheral scene information, unless central and peripheral information were presented in different modalities. This was decided as the effect of central and peripheral information is distinct from traditional sensory modalities (Yeari et al., 2015). Similarly, central and peripheral details are features of the initial event and do not necessarily represent modality changes across stages of the MI paradigm (Dalton & Daneman, 2006).

Although most studies utilising a MI paradigm typically manipulate modality between the encoding and misinformation stages, the current review includes studies that have manipulated modality at different stages. Specifically, differences in modality between encoding, misinformation, and test (for example, visual encoding, auditory misinformation and visual test). Given that modality can be manipulated at any stage of the MI paradigm (Ulatowska et al., 2016), the inclusion of different stage manipulations was deemed important to fully understand the existing knowledge of how modality impacts the MI effect.

Comparator

Studies where a comparison was made between congruent and incongruent modalities at any combination of the three stages of the MI paradigm were included. This enabled the current review to explore the impact of modality similarity on the MI effect, as well as assess the impact of this at different stages of the MI paradigm. Papers that conducted more than one study but did not comment on the links between modality and the MI effect within and across experiments were excluded from the current review. Papers of this kind were excluded as they did not fully assess the impact of modality on the MI effect, as no conclusions were drawn from the findings across the studies. For example, most misinformation research uses changes of modality across the MI paradigm, and sometimes across studies within a single paper, despite it not being an aim of the study to examine modality effects.

Outcome

The initial scoping exercise revealed that some research papers contained multiple studies and made cross-study comparisons regarding the modality effect and misinformation findings. These papers were deemed appropriate for inclusion as they made overall conclusions in relation to the MI effect and modality within the MI paradigm. Single study papers were included if they made explicit links between the impact of modality on the MI effect. Papers were excluded if they did not explicitly analyse the impact of modality in the MI paradigm on misinformation susceptibility. This was decided as many misinformation studies do manipulate modality at different stages of the MI paradigm, however, as noted above, the majority do not specifically analyse or consider the impact this has on misinformation susceptibility.

Study Design

To assess the specific impact of modality on the MI effect, only studies that had used a traditional or adapted MI paradigm were included. The exclusion of studies that employed the

Deese-Roediger-McDermott (DRM; Roediger & McDermott, 1995) paradigm is argued for three main reasons; first, false memories in the DRM and MI paradigm are largely linked to differing psychological mechanisms. Whilst false memories are internally generated in the DRM paradigm (i.e., there are no external influences), the MI paradigm introduces external misinformation that results in the false memories (Zhu et al., 2013). Similarly, the DRM effect can be categorised as a gist-based/associative memory error (i.e., errors that occur when an individual is required to learn and remember the relationship between unrelated items), whereas the MI effect has been considered a distinct category of memory distortion, which is due to the external influence of post-event misinformation (Schacter et al., 2012). Namely, the mechanisms that underpin false memories in the MI effect differ from those in the DRM (Zhu et al., 2013). This distinction was further supported by Ost et al. (2013), who argued that none of the misinformation measures used in the MI paradigm were related to DRM measures, indicating that performance using the MI paradigm differed to performance in the DRM. Findings supported that each effect is likely the result of different cognitive processes and the two phenomena are not significantly related across subjects (Calvillo & Parong, 2016); secondly, the current review is interested in assessing encoding included in the MI paradigm rather than reconsolidation associated with the DRM effect (Hardt et al., 2010). The DRM focuses on the process of replacing or disrupting an existing memory with a new version of that memory, whereas the MI paradigm focuses specifically on how memory can be influenced, not replaced (Ecker et al., 2011). Previous reviews have considered the MI paradigm separately from the DRM for this reason (Frenda et al., 2011).

Finally, the DRM effect is primarily driven by memory traces, whereas the MI effect may arise from a combination of memory traces and social influence (Otgaar et al., 2012). Evidence suggests that the DRM paradigm may not be as readily applicable to naturalistic contexts compared to the MI paradigm. Research has highlighted that the use of the MI paradigm is more applicable to real-life memory distortions (Frenda et al., 2014; Lo et al., 2016; Stark et al., 2010). The differences between the two effects mean that there is evidence to suggest that the two forms of false memories are largely unrelated. As evidence suggests there are different processes underpinning the MI effect, the MI effect found in the MI paradigm is considered to be independent of the DRM in this review.

Studies that used an MI paradigm in their research, but who did not explicitly state this was the case, were considered suitable for inclusion in the review. Similarly, variations on the phrase “MI effect”, such as “susceptibility to misinformation”, “impact of misinformation”, and “false memory acceptance” were considered relevant to answering the review questions.

Other Factors

Only papers written in English were considered, due to the time and a financial restraint associated with the translation of articles. Given the specific nature of the topic being reviewed, a date perimeter was set for database searches (1990-2021). This allowed the search to capture earlier developments in MI literature, as well as most recent studies. Given the dearth of research regarding the impact of modality manipulations on the MI effect and MI paradigm, a wide-ranging date parameter was set for the current review. The year 1990 was chosen as the scoping exercise revealed that exploration into the modality impact on the MI effect had largely taken place in the recent decades, since 1990. Although the term MI effect first appeared in literature in 1974 (Loftus & Palmer, 1974), the scoping exercise indicated that a more recent date parameter would reflect developments in the field and increase the relevance of included studies. Note that only one study was excluded based on the date parameter and this paper was published before 1974 (i.e., before Loftus & Palmer’s seminal study using the misinformation paradigm to examine the MI effect).

Papers were excluded from the review if they encompassed reviews, commentaries, narratives, purely theoretical or opinion papers. Peer reviewed, non-peer reviewed, dissertation and thesis papers were acceptable for inclusion in an attempt to mitigate sampling bias of relevant research. For one of the included papers that conducted multiple studies, not all studies met the inclusion criteria of the review (Yamashita, 1996). Only selecting relevant studies from the paper helped to focus the review and separated relevant findings from the context of other findings included in the paper. For the majority of papers, all conducted studies were included in the review as a result of their relevance to the review questions.

Quality Assessment

A quality assessment was applied to each study because the application of quality assessment to systematic reviews has been found to be essential for the interpretation of primary research (Sanderson et al., 2007). Following guidance set out by Sanderson et al. (2007), this review set out to use a quality assessment that distinguished between quality of the report and the quality of research procedure. Quality assessment checklists and scales were evaluated for their relevance to the review topic (Egger et al., 2003). Relevant quality assessment tools were identified via previous systematic literature reviews regarding memory, broader Google searching, and the Joanna Briggs Institute Critical Appraisal Tools list.

The Quality Assessment Tool for Quantitative Studies (Thomas et al., 2004) was selected for this review based on ease of use, existing commentary on reliability and validity of the tools, and relevance to the review. The Quality Assessment Tool for Quantitative Studies published by the Effective Public Health Practice Project was identified as having been used in multiple systematic reviews and appeared to be adaptable to different areas of research (Thomas et al., 2004). This tool was also chosen due to its fair inter-rater agreement for individual domains and excellent agreement for final global scoring, which has been shown to have higher validity than other identified tools (Armijo-Olivio et al., 2012). It was deemed that the individual domain scores and final global rating would provide clear guidance regarding the inclusion and exclusion of papers.

One of the quality assessment domains in the original tool, titled “Intervention Integrity”, was not relevant to the review. Given that the current review did not focus on a health intervention, this section was removed to ensure the assessed domains were related to the topic and methodology of research assessed. An additional domain, titled “Outcomes”, was added to the original tool. This domain was included to support the quality assessment regarding how outcomes were presented and if this was accessible to the reader. The inclusion of “Outcomes” was particularly useful within the context of the review. Specifically, some of the included papers comprised multiple studies. Therefore, quality assessing how outcomes were reported across studies and whether this supported the overall understanding of findings was useful. Therefore, the adapted tool (see Appendix Four) incorporated eight quality assessment domains: 1) Selection Bias, 2) Study Design, 3) Confounding Variables, 4) Blinding, 5) Data Collection Methods and Reporting, 6) Withdrawals, Drop-Outs, and Missing Data, 7), Analysis, and 8) Outcomes. Each domain was assessed sequentially to determine a rating of “strong”, “moderate” or “weak” and the process for the scoring for each independent domain is shown in Appendix Four. Overall ratings were determined in relation to how many sections were determined to be “weak”, where no weak ratings equated to “strong”, one weak rating equated to “moderate”, and more than one weak rating equated to “weak”. The selection bias for the majority of the studies was scored as weak as samples were largely taken from specific populations (educational institutions) and therefore could not be claimed to be representative of the general population. This was taken into consideration during quality assessment and studies were therefore considered strong if the only area of weakness identified was within the selection bias domain. This enabled a more accurate rating of strength in relation to other studies in the field.

To provide examples of how studies were scored an overall quality rating of “strong”, “moderate”, or “weak”, I will give one example of each. Ulatowska et al. (2016) was “strong” as the only area that was rated “weak” was Selection Bias. As mentioned earlier, the fact that the majority of the papers had recruited student samples was taken into consideration. Because this was the only “weak” domain, the paper was rated “strong”. Yamashita (1996) received an overall rating of “moderate” as two domains received a “weak” rating, one being the Selection Bias domain, which was then taken into consideration. The Blinding domain was also rated “weak” as the researcher had not described whether the participants were randomly allocated to conditions or whether they were aware of the research question. Only one study, Itsukushima et al. (2006), was given an overall rating of “weak”. This paper had three “weak” domains, Study Design and Data Collection Methods and Reporting (including the Selection Bias domain). Study Design was rated “weak” as the paper did not adequately answer the research question as cross study comparisons were not made to conclude regarding the impact of modality on the MI effect. The Data Collection Methods and Reporting domain was rated “weak” because explicit links to the MI effect were not be made and were not reported.

The twenty studies (twelve papers) meeting the PICO criteria were quality assessed (see Appendix Five for quality assessment outcomes). Upon application of the quality assessment tool, six studies had a global rating of “strong”, twelve were rated “moderate” and one was rated as “weak” and therefore excluded. Eleven papers were included and examined for the purpose of the review. The articles were quality assessed by two independent coders. The papers were assessed separately without knowledge of the other coder’s ratings. Inter-rater reliability was considered and the percentage of agreement between coders was 100% for overall quality rating. The percentage of agreement for each of the eight domains across the eleven articles was calculated, giving an agreement percentage of 93.2%. Areas of disagreement were then discussed between coders and considered when giving the overall quality rating.

Data Extraction

A standardised form was used to extract data from each individual study included in the review (see Appendix Six). The form was designed to extract relevant information regarding each study to identify its strengths and limitations, and to identify the modalities manipulated and at which stage this manipulation occurred. Information was extracted on the following key areas:

- Sample characteristics, including sample size, age, ethnicity, gender, and the country the sample was collected from
- Modalities manipulated
- Stage of paradigm at which modalities were manipulated
- Outcomes, including results, significance of findings, statistical test used, impact of modality manipulations
- Study strengths and limitations
- Suggestions made for future research.

In cases where information was unclear, “unclear” was listed next to the relevant extraction item. Authors were not contacted for further clarification on these points as those identified as missing were not considered detrimental to the review process.

Results

Data were extracted from a sum of eighteen studies conducted across the eleven included papers. Table 2.2 provides a summary of the population, modality manipulations and findings from the studies. Next, data were synthesised across studies.

Data synthesis

The aims of the included studies (see Appendix Seven) were not always explicitly consistent with the objectives of the current review. For example, 75% of study aims directly related to the review question, compared to 25% which did not explicitly mention the study's manipulation of modality. Similarly, some papers included in the review conducted multiple studies with different aims. There was heterogeneity between studies in relation to population, modality manipulations, stage at which modality was manipulated, and the impact of this on the MI effect. Given the range of modalities manipulated and the different stages of the MI paradigm at which this was accomplished, studies were often not directly comparable. Therefore, data could not be quantitatively synthesised because there was not enough critical mass for a meaningful meta-analysis to be conducted. To encompass all findings expressed in this review the data has been examined qualitatively, with relevant numerical information provided where possible. Some studies viewed modalities, such as written narratives, as purely visual. Alternatively, other studies considered written narratives to be both visual and verbal. Therefore, Table 2.2 has listed any modality that contained written or spoken words as verbal for consistency.

Table 2.2 Summary of data from included studies

Authors and year of study **(*Quality assessment score)***	**Population** **Sample size, gender, age, population type, country**	**Summary of study conditions**	**Modality presentation used at each stage and modality manipulations assessed** **(*Type of modality/modalities covered in the condition)***	**Impact of modality manipulation on misinformation (MI) effect**
Abeles & Morton, 1999 ***(Moderate)***	***N* = 48** Mean age 23.3 years Range= 17-58 Undergraduate students UK	Unlike a traditional MI paradigm, the misinformation was presented in the form of a written narrative prior to participants viewing the encoding event photograph. The narrative either contained information regarding objects that were not present in the photograph (misinformed condition), objects that appeared in both the narrative and the photograph (picture-and-narrative condition), objects only appearing in the picture and not mentioned in the narrative (picture-only condition), and objects that neither appeared in the narrative or the photograph (distractor condition). There was also a control	**Encoding:** Written text (visual/verbal) *AND* Picture (visual) **MI:** Written narrative (visual/verbal) **Test:** Pictures (visual) *OR* written test (visual/verbal) **Test manipulated** **Manipulations were modality congruent**	The MI effect was found when participants were given a written misleading narrative prior to viewing a stimulus picture and then tested with verbal probes. However, those misinformed in the pictorial group were no different from the pictorial control group. Therefore, there were increased responses to the misinformed objects by the verbally tested group.

Authors and year of study **(*Quality assessment score)***	**Population** **Sample size, gender, age, population type, country**	**Summary of study conditions**	**Modality presentation used at each stage and modality manipulations assessed** **(*Type of modality/modalities covered in the condition)***	**Impact of modality manipulation on misinformation (MI) effect**
		group, who received no misinformation. The subsequent photograph shown to participants contained multiple different objects depicted in a scene. They were informed they would be tested on their memory for the objects in the picture. For the test, participants were either tested pictorially or using a written verbal test response sheet. In the pictorial test condition had each mini photograph (from the encoding event) placed in front of them. They were then asked to respond "yes" if they recognised the object from the picture or "no" if they did not. Participants in the written test condition were given a response sheet (with no pictures) and were asked to put a tick in the "yes" box by an object if they recognised it and "no" if they did not.		

Authors and year of study **(*Quality assessment score)***	**Population** **Sample size, gender, age, population type, country**	**Summary of study conditions**	**Modality presentation used at each stage and modality manipulations assessed** **(*Type of modality/modalities covered in the condition)***	**Impact of modality manipulation on misinformation (MI) effect**
Campbell et al., 2007 **<u>Study 1</u>** ***(Moderate)***	***N* = 60** Mean age 20.61 years 41 female (f)/19 male (m) Undergraduate students UK	For the encoding event, participants were shown a video of a simulated theft without sound. For the misinformation stage, participants were randomly assigned to auditory or text conditions. Text was provided in the form of a written narrative, whereas the auditory condition was presented as a recorded voice. Each condition contained restated (items were presented in the narrative exactly as they occurred in the video), misled (items gave misleading information), and neutral (gave no descriptive information about the event) items. For the test, all participants completed the same forced-choice standard (choosing between misled and original items) and modified (choice between original and novel items) testing procedures to assess	**Encoding:** Video (visual) **MI:** Narrative (auditory) *OR* written narrative (visual/verbal) **Test:** Written recognition task (visual/verbal) **MI manipulated** **Manipulations were both modality congruent and incongruent**	When misinformation was presented via an auditory modality and a written recognition test was administered, memory for the source information was unaffected by post-event modality. When misinformation was presented via a written narrative, participants made fewer errors in the restated condition compared to the neutral and misled conditions, with the misled condition producing the most errors. In the modified test, there was no evidence that post-event information affected memory for the encoding event in the auditory condition. However, in the written narrative

Authors and year of study **(*Quality assessment score)***	**Population** **Sample size, gender, age, population type, country**	**Summary of study conditions**	**Modality presentation used at each stage and modality manipulations assessed** **(*Type of modality/modalities covered in the condition)***	**Impact of modality manipulation on misinformation (MI) effect**
		contextual cues on recognition memory.		misinformation condition, memory for the encoding information was significantly better in the restated condition compared to the neutral and misled conditions. They concluded that encoding specificity may negatively impact memory when modalities match (modality congruence) at misinformation and test stages. This would suggest that the modality congruence of the misinformation can interfere with the traditional encoding specificity hypothesis, perhaps even when encoding and test modalities are also congruent, leading to

Authors and year of study **(*Quality assessment score)***	**Population** **Sample size, gender, age, population type, country**	**Summary of study conditions**	**Modality presentation used at each stage and modality manipulations assessed** **(*Type of modality/modalities covered in the condition)***	**Impact of modality manipulation on misinformation (MI) effect**
				higher levels of misinformation acceptance.
Campbell et al., 2007 **<u>Study 2</u>**	***N* = 36** Mean age 20.81 23f /13 m Undergraduate students UK	Procedure was identical to Study 1 apart from participants were instructed at test to answer questions according to what they remembered seeing in the original video. Standard and modified forced-choice recognition procedures at test as in Study 1.	**Encoding:** Video (visual) **MI:** Narrative (auditory) *OR* written narrative (visual/verbal) **Test:** Written recognition task (visual/verbal) **MI manipulated** **Manipulations were both modality congruent and incongruent**	No significant effects were found for information modality. A similar pattern of results was observed across auditory and visual modalities, suggesting they exerted a similar influence on recall performance. . Whilst the data from Experiment 1 suggested that the MI effect is increased when the misinformation stage is presented in an auditory modality, the data from Experiment 2 suggested that this effect is limited to recognition tests.

Authors and year of study (*Quality assessment score)*	Population Sample size, gender, age, population type, country	Summary of study conditions	Modality presentation used at each stage and modality manipulations assessed (*Type of modality/modalities covered in the condition)*	Impact of modality manipulation on misinformation (MI) effect
Dijkstra & Moerman, 2012 ***(Strong)***	***N* = 115** Mean age 20.4 years Range= 17-42 94f/ 21m Students Netherlands	For the encoding event, participants were exposed to an action sequence of a man visiting a friend at home. The sequence was either presented auditorily, (where participants listened to a story of the event), as slides (participants observed a series of slides with no sound or verbal cues), or as an enactment (participants performed actions presented on slides, no auditory enactment). For the misinformation stage, participants were randomly assigned to a misinformation or control condition. All participants were presented with a written narrative describing the encoding event. In the misinformation condition, this narrative contained misleading information. For the control condition, there was no misinformation.	**Encoding:** Narrative (auditory/verbal) *OR* slides (visual) *OR* enacted visual slides (enactment/visual) **MI:** Written narrative (visual/verbal) **Test:** Written multiple choice test (visual/verbal) **Encoding manipulated** **Manipulations were both modality congruent and incongruent**	For immediate recognition, the results indicated a lower rate of misinformation acceptance in the auditory instead of the enactment conditions. However, there was a significant increase in misinformation acceptance over time in the auditory condition, whereas visual and enactment remained the same. They found support for the encoding specificity hypothesis. That is, matching encoding and retrieval modalities (i.e., visual) could facilitate performance and reduce the MI effect.

Authors and year of study **(*Quality assessment score)***	**Population** **Sample size, gender, age, population type, country**	**Summary of study conditions**	**Modality presentation used at each stage and modality manipulations assessed** **(*Type of modality/modalities covered in the condition)***	**Impact of modality manipulation on misinformation (MI) effect**
		For the test, participants completed a multiple-choice recognition test containing correct and misleading answers. Participants were administered a multiple-choice test immediately after encoding and again two weeks after the study (delayed recognition).		
Gobbo et al., 2002 **<u>Study 1</u>** ***(Strong)***	***N* = 60** Group 1 (*n* = 30) mean age 3 years 6 months 13f/ 17m Group 2 (n = 30). mean age 5.4 years 15f/ 15m Range= 3.2-5.9 School children Italy	For the encoding event, participants were randomly assigned to either participation, observation, or narration condition. The event activity was the same for each condition and related to making an animal figure out of dough. The event included narration of each step, but no demonstration in the participation (enactment) condition. For the misinformation stage, participants were immediately interviewed and asked to freely recall what they had experienced	**Encoding:** Enacted event (enactment/visual) *OR* observed event (visual) *OR* narrative of event (auditory/verbal) **MI:** Interview questions (auditory/verbal) **Test:** Interview questions (auditory/verbal) **Encoding manipulated**	The modality presented at encoding did not affect the reporting of misinformation and results were not significant. However, overall, there was a low level of suggestibility in sample.

Authors and year of study **(*Quality assessment score)***	**Population** **Sample size, gender, age, population type, country**	**Summary of study conditions**	**Modality presentation used at each stage and modality manipulations assessed** **(*Type of modality/modalities covered in the condition)***	**Impact of modality manipulation on misinformation (MI) effect**
		during the encoding stage. After the free recall, the interviewer presented misinformation via two misleading questions. Participants were tested one week after the original event and misinformation in another free recall interview.	**Manipulations were both modality congruent and incongruent**	
Gobbo et al., 2002 <u>**Study 2**</u>	***N* = 118** Mean age 4.8 years Range= 3.7-5.3 55f/ 63m School children Italy	For the encoding event, participants were randomly assigned to either enact an event or receive a narrative of the event. The event focused on "visiting a pirate", which involved becoming a pirate, making a map, find the key, and finding treasure. Half of the children in each group received a single exposure to event followed by a learning test. The remaining children received additional, consecutive presentations and learning trials. For the misinformation stage, participants were asked immediate	**Encoding:** Enacted event (enactment) *OR* narrative (auditory/verbal) **MI:** Interview questions (auditory/verbal) **Test:** Interview questions (auditory/verbal) **Encoding manipulated** **Manipulations were both modality congruent and incongruent**	This study replicated findings from study 1 regarding encoding modality and misinformation acceptance. However, the misinformation had a stronger impact when children had experienced an event vicariously, compared to when they directly took part in the event. This was shown through uncertainty in

Authors and year of study (*Quality assessment score)*	Population Sample size, gender, age, population type, country	Summary of study conditions	Modality presentation used at each stage and modality manipulations assessed (*Type of modality/modalities covered in the condition)*	Impact of modality manipulation on misinformation (MI) effect
		cued recall questions to assess learning. Then an immediate memory test was completed. During this memory test, participants in the immediate misinformation condition were exposed to misleading questions. Those in the leading information condition were exposed to leading questions. Those in the delayed misinformation condition were only asked the questions in the immediate test. For the test, the participants were all interviewed one week after the encoding event. Participants in the immediate misinformation and leading information conditions were asked a set of test recall questions, with no repetition of misleading or leading information. Those in the delayed misinformation condition were asked a set of misleading questions (the same that were presented in the immediate		their responses (i.e., "I do not know") and misled responses.

Authors and year of study **(*Quality assessment score)***	**Population** **Sample size, gender, age, population type, country**	**Summary of study conditions**	**Modality presentation used at each stage and modality manipulations assessed** **(*Type of modality/modalities covered in the condition)***	**Impact of modality manipulation on misinformation (MI) effect**
		misinformation condition), followed by test recall questions.		
Hendrich, 2019 ***(Strong)***	***N* = 92** Mean age 32.14 years Range- 18-67 58f/ 33m/ 1 transgender Random Sample- Mechanical Turk *And* Undergraduate students US	For the encoding event, participants were exposed to pictures depicting a man breaking into a car. Participants were randomly assigned to experience the event visually or auditorily. In the visual condition, participants were presented with 50 pictures showing the event. In the auditory condition, participants were exposed to 50 voice recorder sentences. For the misinformation stage, participants were also randomly assigned to a visual or auditory modality. In the visual condition, participants were shown 50 pictures, with 12 containing misinformation. For the auditory misinformation condition, participants were exposed to 50 recorded sentences, where 12 contained misinformation.	**Encoding:** Pictures (visual) *OR* recorded sentences (auditory/verbal) **MI:** Pictures (visual) *OR* recorded sentences (auditory/verbal) **Test:** Written questions (visual) **Encoding and MI manipulated Manipulations were both modality congruent and incongruent**	The researcher did not find a significant interaction between encoding modality or misinformation modality. The hypothesis that there would be a higher misinformation acceptance rate in the auditory encoding conditions compared to visual conditions was not supported. The hypothesis that there would be a higher misinformation acceptance rate in the modality incongruent conditions was not supported.

Authors and year of study **(*Quality assessment score)***	**Population** **Sample size, gender, age, population type, country**	**Summary of study conditions**	**Modality presentation used at each stage and modality manipulations assessed** **(*Type of modality/modalities covered in the condition)***	**Impact of modality manipulation on misinformation (MI) effect**
		For the memory test, participants first completed an open-ended response memory test, and then a multiple-choice memory test. These were accompanied by source memory questions.		However, subjects in auditory misinformation conditions had higher true memory rates than those in visual condition, which supports previous modality effect research.
Kiat, 2018 **<u>Study 1</u>** ***(Moderate)***	***N* = 87** Mean age 19.33 Range= 18-25 59f/ 28m Students US	Two paradigms, the MI paradigm and the source monitoring paradigm, were used to conduct the study. For the encoding event, participants viewed two events sequentially. The first depicted a man breaking into a car, and the second a woman's wallet being stolen. Both events were depicted in a series of digital colour slides. For the misinformation stage, participants were given two written narratives to read that contained misleading information. One	**Encoding:** Slides (visual) *FOLLOWED BY* picture-word study (visual/verbal) **MI:** Two written narratives (visual/verbal) *OR* **Test:** Picture-word test (visual/verbal) **Encoding and Test manipulated** **Manipulations were modality congruent**	Word-as-picture errors on the source memory test had marginal negative relationship with misinformation susceptibility. Higher levels of Picture-As-Word errors were associated with both increased levels of perceptual misinformation endorsement and reduced levels of accurate perceptual control item endorsement levels. Participants were more likely to correctly attribute items shown as words as

Authors and year of study **(*Quality assessment score)***	**Population** **Sample size, gender, age, population type, country**	**Summary of study conditions**	**Modality presentation used at each stage and modality manipulations assessed** **(*Type of modality/modalities covered in the condition)***	**Impact of modality manipulation on misinformation (MI) effect**
		narrative related to the car break-in, and the second to the stolen wallet. For the test stage, a true/false test format with a source monitoring component was administered. In the test, participants were presented with a statement that either accurately described the original event (assessment of original memory strength), or inaccurately described the event but accurately described the misinformation. The source monitoring task was a picture-word source monitoring task, where participants were presented with a set of pictures and words under specific task instruction. After the study phase, participants were tested for their memory for items using an old-presented- as-picture, old-presented-as-word or new source monitoring test.		being presented as words relative to misattributing them as having been presented as pictures. Likewise, they were also more likely to correctly attribute items shown as pictures as being presented as pictures relative to misattributing them as having been presented as words.

Authors and year of study **(*Quality assessment score)***	**Population** **Sample size, gender, age, population type, country**	**Summary of study conditions**	**Modality presentation used at each stage and modality manipulations assessed** **(*Type of modality/modalities covered in the condition)***	**Impact of modality manipulation on misinformation (MI) effect**
Kiat, 2018 **<u>Study 2</u>**	***N* = 179** Mean age 19.52 Range= 18-28 135f/ 44m Students US	Participants were tested as they were in Study 1. However, the test format was changed from a true/false format to a two alterative forced-choice structure.	**Encoding:** Slides (visual) *FOLLOWED BY* picture-word study (visual/verbal) **MI:** Two written narratives (visual/verbal) **Test:** Picture-word test (visual/verbal) **Encoding and Test manipulated** **Manipulations were modality congruent**	Replicated findings of study 1. The researcher concluded that verbal processes appear to play a stronger role in misinformation susceptibility compared to visual processes.
Kiat, 2018 **<u>Study 3</u>**	***N* = 19** Mean age 19.84 Range= 18-23 12f/ 7m Students US	Participants were tested similarly as to how they were in Study 1 and 2. The encoding slides were increased to include four scenarios. This study aimed to examine differences in neural activity in MI effect and picture-word source	**Encoding:** Slides (visual) *FOLLOWED BY* picture-word study (visual/verbal) **MI:** Written narrative (visual/verbal)	Subjects more likely to correctly attribute pictures compared to misattributing them as being words. EEG data suggested there was a greater number of picture-as-word errors related to higher levels of

Authors and year of study **(*Quality assessment score)***	**Population** **Sample size, gender, age, population type, country**	**Summary of study conditions**	**Modality presentation used at each stage and modality manipulations assessed** **(*Type of modality/modalities covered in the condition)***	**Impact of modality manipulation on misinformation (MI) effect**
		monitoring. This was achieved by collecting electroencephalogram (EEG) data at test only. Participants were tested on the following day.	**Test:** Picture or word test whilst EEG data was taken (visual/verbal) **Encoding and Test manipulated** **Manipulations were modality congruent**	misinformation susceptibility.
Kiat, 2018 **<u>Study 4</u>**	***N* = 30** Mean age 19.70 Range= 17-28 17f/ 13m Students US	Participants were tested as they were in Study 1. The encoding event was shown for longer than in previous studies. Aimed to examine the relationship between individual differences in misinformation susceptibility and neural activity between stages when exposed to different modalities. EEG data was once again used to explore this, this time at each stage of the experiment.	**Encoding:** Event slides (visual) *AND* narrative slides (visual/verbal) **MI:** Misinformation narrative (visual/verbal) **Test:** Picture or word test (visual/verbal) **Encoding and Test manipulated** **Manipulations were modality congruent**	Results indicated that verbal processes may play a stronger role in misinformation susceptibility compared to visualisation processes.

Authors and year of study **(*Quality assessment score)***	**Population** **Sample size, gender, age, population type, country**	**Summary of study conditions**	**Modality presentation used at each stage and modality manipulations assessed** **(*Type of modality/modalities covered in the condition)***	**Impact of modality manipulation on misinformation (MI) effect**
Mitchell & Zaragoza, 1996 **<u>Study 1</u>** ***(Moderate)***	***N* = 180** Undergraduate students US	Participants were randomly allocated between mixed modality (received misinformation in all three modalities one at a time) and single modality group (received MI in one of the three modalities). For the encoding event, participants viewed a video depicting two youths burgling a home followed by a police chase. The video included dialogue. For the misinformation stage, participants were instructed to answer questions focused on the events in the video. This included misleading questions related to the encoding event. Participants were tested using a source memory test and given both written and verbal instructions. The source memory test was presented	**Encoding:** Video (visual/auditory/verbal) **MI:** Printed questions (visual/verbal) *OR/AND* videotaped auditory questions (visual/auditory/verbal) *OR/AND* audiotaped questions (auditory/verbal) **Test:** Recorded statements (auditory/verbal) *AND* written questions (visual/verbal- (written *AND* verbal instructions were given on how to conduct the test) **MI manipulated** **Manipulations were both modality congruent and incongruent**	Subjects who were exposed to repeated suggestions in different modalities were more likely to misattribute suggestions to an incorrect source than when they were all presented in one modality, which met the authors’ hypothesis. As such, the researchers concluded that contextual variability (modality incongruence at the misinformation stage) can weaken a person’s ability to make accurate source attributions. This may be impactful when considering the encoding specificity hypothesis, as even when encoding and test modalities matched, and the misinformation

Authors and year of study **(*Quality assessment score)***	**Population** **Sample size, gender, age, population type, country**	**Summary of study conditions**	**Modality presentation used at each stage and modality manipulations assessed** **(*Type of modality/modalities covered in the condition)***	**Impact of modality manipulation on misinformation (MI) effect**
		via recorded statements and written questions. They were asked to circle "yes" under video or question depending on which they recognised the statement from. They were then asked to circle no under video or question depending on which they did not believe the statement came from.		modality did not match, participants were more likely to make source misattributions compared to when all stages were presented in one modality.
Mitchell & Zaragoza, 1996 **<u>Study 2:</u>**	***N* = 396** No further sample information provided Undergraduate students US	The materials and procedure for Study 2 was identical to that used in study apart from a further two added filler tasks.	**Encoding:** Video (visual) **MI:** Printed questions (visual/verbal) *OR/AND* videotaped questions (visual/auditory/verbal) *OR/AND* audiotaped questions (auditory/verbal) **Test:** Recorded statements (auditory/verbal) *AND* written questions (visual/verbal- written *AND* verbal instructions	This study closely replicated findings from study 1. The introduction of spacing manipulation (by adding two extra filler tasks) had no impact on any of the reported measures, with no reliable main effects or interactions.

Authors and year of study **(*Quality assessment score)***	**Population** **Sample size, gender, age, population type, country**	**Summary of study conditions**	**Modality presentation used at each stage and modality manipulations assessed** **(*Type of modality/modalities covered in the condition)***	**Impact of modality manipulation on misinformation (MI) effect**
			were given on how to conduct the test) **MI manipulated** **Manipulations were both modality congruent and incongruent**	
Roebers et al., 2004 ***(Moderate)***	***N* = 270** 90**-** 5- and 6-year-olds Mean age 5 years 7 months 51f/ 39m 92- 7- and 8-year-olds Mean age 7 years 8 months 49f/ 43m 88- 9- and 10-year-olds Mean age 9 years 11 months 43f/ 45m	For the encoding event, participants were randomly assigned to watch a magic show in a live format, in a videotaped format or as a slideshow with narration. For the misinformation stage, participants were questioned by a researcher individually a week after the encoding event. The interview included misleading questions. After this round of questioning, followed by a short break, participants were tested using a	**Encoding:** Live enacted event (visual/enactment- acted out by someone who was not a participant. Does not specify if verbal communication was used during acting) *OR* video (visual) *OR* narrated slideshow (auditory/visual/verbal) **MI:** Interview questions (auditory/verbal)	The MI effect was equally strong across the visual video and narrated slides encoding modalities. However, in the live condition there was consistently less misinformation acceptance compared to other modalities. The enacted event encoding condition consistently produced a higher percentage of

Authors and year of study (*Quality assessment score)*	Population Sample size, gender, age, population type, country	Summary of study conditions	Modality presentation used at each stage and modality manipulations assessed (*Type of modality/modalities covered in the condition)*	Impact of modality manipulation on misinformation (MI) effect
	School children Germany	forced choice interview. The forced choice questions included a correct answer and a plausible, but incorrect answer (related to the misinformation).	**Test:** Interview questions (auditory/verbal) **Encoding manipulated** **Manipulations were both modality congruent and incongruent**	correct answers compared to all other conditions. However, participants in this condition were not found to have smaller memory deficits due to misinformation compared to the other modality conditions.
Stoll, 2021 ***(Moderate)***	***N* = 103** 73f/ 27m/ 2 no response Mean age 21 Age range 18-52 Undergraduate students US	For the encoding event, participants were shown a video of a repairman entering a home to complete work. However, the video depicted him stealing items from inside the house. The video was comprised of photos that were displayed on screen. For the misinformation stage, participants were randomly assigned to receive misinformation via indirect interviews, direct interview, or written narrative. Indirect interviews included misleading statements followed by questions	**Encoding:** Video (visual) **MI:** Written narrative (visual/verbal) *OR* indirect interviews (visual/verbal) *OR* direct interviews (visual/verbal) **Test:** Recognition task (visual/verbal) **MI manipulated** **Manipulations were congruent**	Subjects in the narrative condition were less accurate than the interview conditions for true memories. The MI effect for all three misinformation conditions were found to be non-significant.

Authors and year of study **(*Quality assessment score)***	**Population** **Sample size, gender, age, population type, country**	**Summary of study conditions**	**Modality presentation used at each stage and modality manipulations assessed** **(*Type of modality/modalities covered in the condition)***	**Impact of modality manipulation on misinformation (MI) effect**
		(i.e., “The man did several things in the kitchen including eating an apple. Did he repair any items in the kitchen?”). Direct interviews included misleading questions (i.e., “Did the man eat an apple?”). The written narrative contained a description of the video events with added misleading details. For the test, participants were given either one or two written recognition tests. No details were given regarding the explicit differences between the two tests.		
Ulatowska et al., 2016 **Study 1** ***(Strong)***	***N* = 161** 114f/ 47m Mean age 20.28 years. Undergraduate students US	Two groups were tested. The first was exposed to misleading information. The second was a control group, who were provided with correct information only. For the encoding event, participants were randomly assigned to watch a video or slides, read a narrative, or listen to a recorded auditory	**Encoding:** Written narrative (visual/verbal) *OR* Auditory narrative (auditory/verbal) *OR* slides (visual/auditory/verbal- each followed by a spoken comment) *OR* video (visual)	A lower misinformation acceptance rate was found following both narrative encoding conditions (auditory and written) compared to the slide condition. The interaction of encoding modality was

Authors and year of study **(*Quality assessment score)***	**Population** **Sample size, gender, age, population type, country**	**Summary of study conditions**	**Modality presentation used at each stage and modality manipulations assessed** **(*Type of modality/modalities covered in the condition)***	**Impact of modality manipulation on misinformation (MI) effect**
		narrative of a street mugging. In the slide condition, an accompanying spoken sentence describing what the slide depicted was included. For the misinformation stage, participants were asked 25 questions about the event. For those in the misinformation group, some of the questions contained changed or new information. In the control group the questions only contained correct information. For the test, all participants tested verbally using a forced choice recognition test.	**MI:** Written questions (visual/verbal) **Test:** Written recognition test (visual/verbal) **Encoding manipulated** **Manipulations were both modality congruent and incongruent**	significant. There was increased misinformation acceptance where modalities differed significantly from one another across stages of the paradigm. They could not draw a general conclusion regarding the influence of encoding and test modality congruence (encoding specificity hypothesis). As such, they conducted Study 2.
Ulatowska et al., 2016 **<u>Study 2</u>**	***N* = 207** 152 f/ 55m Mean age 19.91 years Undergraduate students	For the encoding event, participants either watched a sequence of slides (without sound) or heard an auditory version of a street mugging. However, this time slides were not accompanied by a spoken descriptive sentence.	**Encoding:** Slides (visual) OR auditory narrative (auditory/verbal) **MI:** Slides (visual) *OR* auditory narrative (auditory/verbal)	Results indicated that the format of encoded information and the incongruence between encoding and test conditions might make participants less likely to retrieve the original

Authors and year of study **(*Quality assessment score)***	**Population** **Sample size, gender, age, population type, country**	**Summary of study conditions**	**Modality presentation used at each stage and modality manipulations assessed** **(*Type of modality/modalities covered in the condition)***	**Impact of modality manipulation on misinformation (MI) effect**
	US	For the misinformation stage, participants were either exposed to either a sequence of slides or an audio narrative containing four changed details from the original event. For the recognition test, participants were once again presented with slides or auditory test formats. All participants were given ten written questions, either accompanied by picture cues or verbal cues by a male voice. This meant that participants were either presented with a visual or auditory modality at each stage of the MI paradigm	**Test:** Pictures (visual) *OR* auditory questions (auditory/verbal) **Encoding, MI and Test manipulated** **Manipulations were both modality congruent and incongruent**	information. Thus, partially supporting the encoding specificity principle (across event and recognition test). However, this was modified by the format in which event or misinformation was presented. They concluded that auditory modality resulted in stronger memory trace, meaning auditory/verbal information is responsible for creating rich and strong memory representation. Therefore, when misinformation was auditory it can overwrite original information to a larger degree. A such, the MI effect was stronger after misinformation was presented auditorily,

Authors and year of study **(*Quality assessment score)***	**Population** **Sample size, gender, age, population type, country**	**Summary of study conditions**	**Modality presentation used at each stage and modality manipulations assessed** **(*Type of modality/modalities covered in the condition)***	**Impact of modality manipulation on misinformation (MI) effect**
				compared to when it was presented in slides. The interaction effect between encoding modality and misinformation acceptance was not significant.
Yamashita, 1996 **<u>Study 2</u>** ***(Moderate)*** **<u>Study 1:</u>** ***Not Relevant to the review***	***N* = 113** Undergraduate students Japan	For the encoding event, participants were shown a series of colour pictorial slides showing the stages of an auto-pedestrian accident. For the misinformation stage, participants were randomly assigned to receive consistent information or misleading information in the form of written questionnaire. Finally, participants were randomly assigned to one of two recognition tests. One test was presented in the form of slides, where participants were then asked to choose the slide	**Encoding:** Slideshow (visual) **MI:** Written questionnaire (visual/verbal) **Test:** Slides (visual) *OR* verbal questions (auditory) **Test Manipulated** **Manipulations were both modality congruent and incongruent**	The MI effect was only significant in the verbal recognition test condition. There was some evidence of an MI effect in the visual test condition, however this was non-significant. The MI effect was higher when the modality of the recognition test and post-event information were the same (modality congruence) compared to when they were different

Authors and year of study **(*Quality assessment score)***	**Population** **Sample size, gender, age, population type, country**	**Summary of study conditions**	**Modality presentation used at each stage and modality manipulations assessed** **(*Type of modality/modalities covered in the condition)***	**Impact of modality manipulation on misinformation (MI) effect**
		they had seen earlier. The other was presented as a written sentence containing a question.		(modality incongruent). This would suggest that the modality congruence of the misinformation can interfere with the traditional encoding specificity hypothesis, even when encoding and test modalities are also congruent, leading to higher levels of MI acceptance.

Population

Sample sizes ranged from 19 (Kiat, 2018, study 3) to 396 (Mitchell & Zaragoza, 1996, study 2). The total number of participants across the included studies was 2,274 (Mean= 126.3, SD= 92.6). Participants' ages ranged from 3 to 67; however, explicit age ranges were not provided by seven of the studies, who presumably sampled adults (Campbell et al., 2007, study 1 and 2; Mitchell & Zaragoza, 1996, study 1 and 2; Ulatowska et al., 2016, study 1 and 2; Yamashita, 1996). Three of the studies specifically focused on the susceptibility of children to misinformation (Gobbo et al., 2002, study 1 and 2; Roebers et al., 2004). This was useful for the review as the importance of considering the MI effect on child eyewitness testimony has been expressed in literature (Cossins, 2006; Davies et al., 2004; Powell, 2005). Although children are viewed as poor eyewitnesses and perhaps more susceptible to misinformation (Sutherland & Hayne, 2001), recent research has found that children are, in some instances, less vulnerable to misinformation than adults. Otgaar et al. (2018) argued that a child's lack of knowledge can protect them from misinformation. They explained that throughout the course of life, people acquire more knowledge, resulting in faster and more automatic associative activation of memories. Due to increased knowledge, under some circumstances (e.g., when an individual is surrounded by associatively related cues), adults are more susceptible to false memories compared to children because they are more likely to generate faulty associations. Therefore, there has been a movement away from viewing children as having poor memory performance. On this basis, studies that included children in their sample have been included in the review.

Four studies did not provide a gender breakdown for their sample (Abeles & Morton, 1999; Mitchell & Zaragoza, 1996, study 1 and 2; Yamashita, 1996). The reported sample was predominantly female, with only two studies reporting a higher number of male participants (Gobbo et al., 2002, study 1 and 2). Of the total sample, at least 42% (*n*= 958) were female, 21% (*n*= 486) were male, and 0.04% were transgender (*n*= 1). The gender information was not provided for approximately 37% of participants included in this review.

Studies were conducted in a variety of different countries. Ten studies were undertaken in the US (Hendrich, 2019; Kiat, 2018, studies 1, 2, 3 and 4; Mitchell & Zaragoza, 1996, study 1 and 2; Stoll, 2021; Ulatowska et al., 2016, study 1 and 2), three studies in the UK (Abeles & Morton, 1999; Campbell et al., 2007, study 1 and 2), one study in the Netherlands (Dijkstra & Moerman, 2012), two studies in Italy (Gobbo et al., 2002, study 1 and 2), one study in Germany (Roebers et al., 2004), and one study in Japan (Yamashita, 1996), indicating that American researchers have been most active in this area of research. It is of note that only five of the included studies reported the ethnic composition of their sample (Hendrich, 2019; Kiat, 2018, studies 1 ,2 3 and 4). Therefore, it has not been possible to quantify data regarding participant ethnicity. Given that there is no theoretical reason why ethnicity would be an important factor in moderating the influence of modality and the MI effect, this was not deemed to be detrimental to the review.

All of the participants were recruited from educational facilities, such as schools (Gobbo et al., 2002, study 1 and 2; Roebers et al., 2004) and universities (Abeles & Morton, 1999; Campbell et al., 2007, study 1 and 2; Dijkstra & Moerman, 2012; Hendrich, 2019; Kiat, 2018, studies 1, 2, 3 and 4; Mitchell & Zaragoza, 1996, study 1 and 2; Stoll, 2021; Ulatowska et al., 2016, study 1 and 2; Yamashita, 1996). One of the studies used Amazon Mechanical Turk to recruit participants outside of an educational opportunity sample (Hendrich, 2019).

Stage of Modality Manipulation

The current review included studies that manipulated modality at different stages of the MI paradigm. From the included studies, eleven manipulated modality at one stage of the MI paradigm (Campbell et al., 2007, study 1 and 2; Dijkstra & Moerman, 2012; Gobbo et al., 2002, study 1 and 2; Mitchell & Zaragoza, 1996, study 1 and 2; Roebers et al., 2004; Stoll, 2021; Ulatowska et al., 2016, study 1; Yamashita, 1996), and six studies manipulated modality at two stages (Abeles & Morton, 1999; Hendrich, 2019; Kiat, 2018, studies 1, 2, 3 and 4). One study manipulated modality at all three stages of the MI paradigm (Ulatowska et al., 2016, study 2). Eleven studies manipulated modality at encoding (Dijkstra & Moerman, 2012; Gobbo et al., 2002, study 1 and 2; Hendrich, 2019; Kiat, 2018, studies 1, 2, 3 and 4; Roebers et al., 2004; Ulatowska et al., 2016, study 1 and 2), eight at the misinformation stage (Abeles & Morton, 1999; Campbell et al., 2007, study 1 and 2; Hendrich, 2019; Mitchell & Zaragoza, 1996, study 1 and 2; Stoll, 2021; Ulatowska et al., 2016, study 2), and seven at test (Abeles & Morton, 1999; Kiat, 2018, studies 1, 2, 3, and 4; Ulatowska et al., 2016, study 2; Yamashita, 1996).

Modality Manipulations

Modalities manipulated across the papers reflected visual and auditory manipulations. In studies that manipulated the visual modality, information was presented as pictures (Abeles & Morton, 1999; Hendrich, 2019; Ulatowska et al., 2016, study 2), slide shows (Dijkstra & Moerman, 2012; Kiat, 2018, study 4; Ulatowska et al., 2016, study 1; Yamashita, 1996), videos (Roebers et al., 2004), written narratives (Campbell, Edwards, Horswill & Helman, 2007, study 1 and 2; Stoll, 2021; Ulatowska et al., 2016, study 1), printed questions (Mitchell & Zaragoza, 1996, study 1 and 2; Stoll, 2021), and watching live action enacted events (Gobbo et al., 2002, study 1 and 2; Roebers et al., 2004). In papers that manipulated the auditory modality, information was presented as auditory narratives (Campbell et al., 2007, study 1 and 2; Dijkstra & Moerman, 2012; Gobbo et al., 2002, study 1 and 2; Hendrich, 2019; Ulatowska et al., 2016, study 1), audiotaped questions (Mitchell & Zaragoza, 1996, study 1 and 2), and live verbal questioning (Yamashita, 1996). Studies also utilised manipulations that combine auditory and visual modalities, such as verbally narrated slide shows (Roebers et al., 2004), videotaped auditory questions (Mitchell & Zaragoza, 1996, study 1 and 2), and watching live action enacted events (Gobbo et al., 2002, study 1 and 2; Roebers et al., 2004). One of the included studies explored a manipulation where participants were instructed to act out an encoding event themselves without verbalisation (Dijkstra & Moerman, 2012). Another asked participants to actively take part in making a figure (at the encoding stage) with verbal instructions on how to do so (Gobbo et al., 2002, study 1). In total, fourteen different modality presentation formats are represented in the composition of the manipulations explored in this review.

The included studies have also examined *within* modality changes (e.g., *within* the visual modality: video, slideshow, live action) or *between* modality changes (e.g., *between* the auditory and visual modalities: voice recording, written narrative). Seven studies manipulated within modality (Gobbo et al., 2002, study 1; Kiat, 2018, studies 1, 2, 3 and 4; Roebers et al., 2004; Stoll, 2021), and nine manipulated between modalities (Abeles & Morton, 1999; Campbell et al., 2007, study 1 and 2; Gobbo et al., 2002, study 1 and 2; Hendrich, 2019; Ulatowska et al., 2016, study 1 and 2; Yamashita, 1996). Four studies included at least one study which included manipulation both within and between modalities (Dijkstra & Moerman, 2012; Mitchell & Zaragoza, 1996, study 1 and 2; Ulatowska et al., 2016, study 1).

Comparisons between Auditory and Visual Modality Manipulations

Eleven of the included studies compared auditory to visual modalities. Results regarding enactment modalities are considered independently, given that they can contain both auditory and visual information. Six of these studies manipulated between auditory and visual conditions at encoding (Dijkstra & Moerman, 2012; Gobbo et al., 2002, study 1 and 2; Hendrich, 2019; Ulatowska et al., 2016, study 1 and 2). Of the six studies, two found that there was a reduced MI effect when an auditory modality was presented at encoding, compared to a visual modality (Dijkstra & Moerman, 2012; Ulatowska et al., 2016, study 1).

Dijkstra & Moerman (2012) found that when an auditory modality (auditory narrative) was used at encoding, and when an immediate recognition test was administered, participants were less susceptible to the MI effect compared to both visual and enactment modalities. Similarly, Ulatowska et al. (2016, study 1) found that there was a lower misinformation acceptance rate following both auditory narrative and written narrative condition compared to when participants were shown slides (note, the visual slides were accompanied by spoken comments). This was explained using the encoding specificity hypothesis, arguing that the written narrative and auditory narrative encoding modalities matched the written test format, making it easier for participants to retrieve the original information. They also argued that information presented in an auditory modality resulted in a stronger memory trace, and that verbal information creates rich and strong representation, reducing misinformation susceptibility.

In contrast, four studies did not find the same support for auditory encoding. For example, Hendrich (2019) hypothesised that there would be a higher misinformation acceptance rate in the auditory encoding conditions compared to visual conditions. Despite this prediction, the findings did not support this. That is, a significant main effect was not found for the encoding modality ($F(3, 85) = .70$, $p= .554$, partial $\eta 2 = .02$). Results from Gobbo et al. (2002, study 1) also showed that the modality of the event presentation (i.e., the modality shown at encoding) did not affect the reporting of misinformation in response to test recall questions, that is, mislead responses. This result was replicated in Gobbo's et al. (2002, study 2) second study, where the main effect of event presentation on misinformation did not reach significance. Interestingly, in contrast to their findings in their first study, Ulatowska et al. (2016, study 2) found that the main effects for the encoding modality $F(1, 199)= 1.85$, $p= .17$, was not significant.

Six studies manipulated between auditory and visual conditions at the misinformation stage (Campbell et al., 2007, study 1 and 2; Hendrich, 2019; Mitchell & Zaragoza 1996, study 1 and 2; Ulatowska et al., 2016, study 2). Of the six studies, two concluded that auditory misinformation led to reduced misinformation acceptance (Campbell et al., 2007, study 1 and 2). Campbell et al. (2007, study 1) found a marginally significant main effect for information modality ($F(1, 58)= 3.99$, MSE=0.02, $p=.051$, partial $\eta2= .06$), with the researchers concluding that their findings suggest that the MI effect is attenuated when the misinformation stage is presented in an auditory modality, compared to a visual modality (Campbell et al., 2007, study 1 and 2). It is of note, however, that Campbell et al. (2007, study 2) found this effect is limited to recognition tests. In contrast, in one study, Hendrich (2019) found that the misinformation modality did not have a significant effect on the misinformation-consistent response rate, $F(1, 87)= .08$, $p= .781$, partial $\eta2= .10$. However, one study found evidence that auditory misinformation increased misinformation acceptance (Ulatowska et al., 2016, study 2). Analysis by Ulatowska et al. (2016, study 2) revealed a main effect of post-event misinformation ($F(1, 199)= 9.34$, $p< .01$, $\eta2= .05$). They also found that the MI effect was stronger after misinformation was presented auditorily ($M= 0.29$, $SD= 0.25$) than in the form of visual slides ($M= 0.19$, $SD= 0.21$). This suggests that individuals who are exposed to auditory misinformation may be more likely to accept misinformation compared to those who receive visual misinformation. Although the remaining two studies by Mitchell and Zaragoza (1996, study 1 and 2) manipulated between auditory and visual modalities at the misinformation stage, their analysis primarily focused on the use of single or mixed modality presentations. They do not comment on the performance of specific modalities in these conditions. Therefore, these results will be discussed separately in this results section.

Four studies manipulated between auditory and visual conditions at test (Mitchell & Zaragoza, 1996, study 1 and 2; Ulatowska et al., 2016, study 2; Yamashita, 1996). One study found evidence to suggest that visual tests, in the form of picture slides, reduced the MI effect or supported recognition (Yamashita, 1996). Yamashita, (1996) found a main effect of information (consistent vs misleading) was significant ($F(1, 109)= 24.606$, $p< .001$) and that there was an interaction effect with test modality, which was either presented auditorily as verbal questions, or through visual picture slides ($F(1, 109)= 5.749$, $p< .05$). The MI effect was only significant in the verbal (i.e., auditory) test condition ($F(1, 109)= 15.078$, $MSE = 1.169$, $p< .001$), but not in the visual condition. Although there was some evidence of an MI effect in the visual test condition, this result was non-significant. This means that participants were more likely to be susceptible to the MI effect when tested auditorily compared to visually, after exposure to visual encoding misinformation stages. One study commented on the modality congruence between encoding and test, but not the independent effects of the test modality (Ulatowska et al., 2016, study 2). Therefore, this will be discussed with the modality congruence findings. For the remaining two studies by Mitchell and Zaragoza (1996, study 1 and 2), you may recall they did not comment on individual modality performance. Therefore, it is not possible to ascertain if auditory or visual modalities improved memory performance at test.

The conflicting results regarding the impact of the auditory modality on the MI effect could be due to difference in timing across studies. When exploring the impact of visual and auditory modalities over time, Dijkstra and Moerman (2012) initially found that misinformation acceptance was lowest in the auditory encoding condition compared to the visual encoding condition ($t(30)= 2.89$, $p< .01$) and enactment conditions ($t(40)= 2.07$, $p< .05$) after immediate recognition. However, misinformation acceptance increased significantly from immediate to delayed retrieval in the auditory condition ($t(16)= 2.50$, $p< .05$) compared to the visual and enactment conditions, which remained the same after delay. Therefore, misinformation acceptance was higher in the auditory condition compared to the visual and enactment conditions in the delayed retrieval conditions. This suggests that there is an auditory encoding benefit (a reduced MI effect) when retrieval is immediate, but not when retrieval is delayed. This also suggests that encoding visual and enactment modalities may produce less of an MI effect over time compared to the auditory encoding modality. However, this evidence is limited.

Despite the majority of included studies utilising both auditory and visual modalities, one study solely considered within-modality comparisons (Stoll, 2021). Stoll (2021) manipulated the misinformation stage between three visual modalities; written narrative, indirect interviews, and direct interviews (where questions were presented visually on screen). There was no main effect of misinformation modality on misinformation acceptance ($F(2, 100) = 2.58$, $p = .08$, $\eta 2p = .05$). Therefore, the MI effect for all three misinformation conditions were found to be non-significant.

The impact of multiple modality exposures at one stage of the MI paradigm was explored. Mitchell and Zaragoza (1996, study 1) randomly assigned participants to receive misinformation in one modality (single modality condition; print, audiotape, or videotape) or the three different modalities (mixed modality condition, where suggestions were given to participants in three different modalities; print, audiotape, and videotape) that reflected visual and auditory presentation formats. Findings from study one suggested there was an increase in source attribution errors in the mixed modality condition compared to the single modality condition ($F(1, 178) = 39.98$, $MSE = - .05$, $p < .0001$). Specifically, participants in the mixed modality condition were more likely to accept post-event suggestion in the form of suggested items on the source memory test. Likewise, the mixed modality group were more susceptible to misinformation ($F(1,178) = 4.00$, $MSE = .25$, $p < .05$) and were more likely to be impacted by repeated exposures to suggestibility via the three modalities, ($F(1, 178) = 5.82$, $MSE = - .05$, $p < .05$). These findings were replicated in study two, ($F(1, 394) = 3.27$, $MSE = .15$, $p = .07$).

Therefore, participants who were exposed to repeated misinformation in different modalities were more likely to misattribute the suggestions to an incorrect source.

Findings regarding the strength of specific modalities were mixed. Regarding modality at encoding, some evidence suggested that there may be a reduction in misinformation acceptance when the encoding event is presented auditorily. However, over time, evidence also suggested that auditory modalities at encoding will increase the MI effect compared to visual modalities (Dijkstra & Moerman, 2012). This is because auditory encoding initially led to reduced misinformation acceptance, but when this increased over time, there was increased misinformation acceptance in the auditory encoding condition compared to both visual encoding conditions (enactment and slides). However, the majority of evidence reviewed found no effect of encoding presentation modality on misinformation acceptance. This tells us that, at encoding, the modality shown may not be an influential factor impacting the MI effect. The results for the misinformation stage were also mixed. Contrasting studies found that auditory modalities at the misinformation stage were associated with reduced misinformation acceptance, that auditory modalities at the stage were associated with increased misinformation acceptance, as well as analysis that found no significant effect of misinformation modality on the MI effect. At the test stage, some evidence suggested visual modalities are better at reducing the MI effect than auditory modalities. However, this evidence was limited. Given that the findings regarding different modalities at different stages of the MI paradigm are so mixed, the strongest evidence appears to be that the modality shown at each stage is not hugely impactful on the MI effect. Although it cannot be ignored that some studies did find an effect, more research is needed to clarify these results.

Combined Visual-Auditory Modality Manipulations

Only three of the studies included a combined visual-auditory modality presentation in the form of videotaped auditory questions (Mitchell & Zaragoza, 1996, study 1 and 2) and visual slides accompanied by spoken sentences (Ulatowska et al., 2016, study 1). However, specific findings regarding the combined modality from Mitchell & Zaragoza (1996, study 1 and 2) were not reported in the studies as only exposure to a single or mixed range of modalities was assessed. However, Ulatowska et al. (2016, study 1) did provide some details regarding their combined modality encoding condition. Specifically, there was a lower misinformation acceptance rate following both narrative encoding conditions (one auditory, the other visual) compared to the slide condition. This suggests that there was higher misinformation acceptance in the combined modality encoding condition (slides). However, the researchers did not explicitly comment on the significance of this finding within the context of it being a combined modality. Therefore, it has not been possible to fully account for the significance of combined modality conditions. It is of note that a further study by Roebers et al. (2003) used a live magic show enactment at encoding. Whilst it was made clear in the paper that this was a visual modality, there was no information regarding whether the visual live show included auditory information (such as talking) as this was not explicitly stated by the researchers. Therefore, it is unclear whether or not this could be considered a combined modality condition.

Whilst some studies included in this review considered modalities, such as written narratives and written text, to be purely visual, five studies considered the addition of written words to account for both a visual and auditory modality (Abeles & Morton, 1999; Kiat, 2018, studies 1, 2, 3 and 4). This is partially supported in literature, as words require similar processing to auditory information (Nishiyama et al., 2017). Firstly, Abeles and Morton (1999) considered their written text conditions to contain verbal, and therefore auditory, information. They found a significant difference between misinformed and control subjects who were tested with a written test (verbal) ($F(1, 44) = 16.58$, $p < .001$), but not when they were tested pictorially

($F(1, 44)= 1.54$, $p> .05$). This indicated an increased susceptibility to the MI effect when participants were exposed to a verbal conditions.

Research conducted by Kiat (2018, studies 1, 2, 3 and 4) compared a visual modality to a verbal modality (in the form of a visual written narrative). Given that both modalities were viewed, as opposed to heard, they could be considered as modality congruent. Kiat (2018, study 1) found that word-as-picture errors, where a set of objects were presented as words instead of pictures (verbal modality), had a marginally negative relationship with misinformation susceptibility ($B= -.12$, SE= .06, $F(1,84)= 3.68$, $p= .06$). However, picture-as-word errors, where a set of objects were presented as pictures instead of words (visual modality) errors had a significantly positive relationship ($B= .32$, SE= .13, $F(1,84)= 5.81$, $p= .02$). This means that increased word-as-picture errors were marginally associated with fewer perceptual misinformation endorsements. Moreover, increased picture-as-word errors were associated with increased levels of perceptual misinformation endorsement (Kiat, 2018, study 1 and 3). That is, endorsement of misinformation was more strongly associated with verbalisation (auditory) source monitoring errors relative to visualisation errors. This suggests that verbal processes may play a role in driving misinformation susceptibility, more so than visual. Kiat (2018, study 2) supported this, finding that there was a larger MI effect when participants were presented with a verbal modality compared to a visual modality at encoding. Therefore, it was concluded that verbal processes may play a more significant role in misinformation acceptance compared to visual process (Kiat, 2018, study 4).

Enactment Modality Manipulations

Three studies used an enactment modality, and the results were mixed across studies (Dijkstra & Moerman, 2012; Gobbo et al., 2002, study 1 and 2). Dijkstra and Moerman (2012) hypothesised that participants would be less susceptible to misinformation if parts of the original encoding event were acted out by participants (i.e., motor encoding). This was hypothesised because previous research has found enhanced recall of stimuli that is acted out or performed, compared to verbally encoded items (Engelkamp & Cohen, 1991; Nyberg et al., 1991). However, this hypothesis was only partially supported. The researchers summarised that an advantage of enactment at encoding could not be supported in the context of the misinformation paradigm. Specifically, motor encoding was not deemed to be totally protective against the MI effect in comparison to encoding modalities that match at original event and retrieval. Therefore, the MI effect was still present, even when children had enacted the encoding event. However, analysis exploring misinformation errors found that participants in the enactment condition made fewer errors compared to the visual condition at immediate and delayed performance ($t(39) = 2.17$, $p < .05$, and $t(39) = 2.27$, $p < .05$). Together, this suggests that enactment modalities at encoding may not be as effective at mitigating against the MI effect as previously believed.

Furthermore, Gobbo et al. (2002, study 1) found that the modality of the encoding event did not affect reporting of misinformation. Although Gobbo et al. (2002, study 1 and 2) found a main effect for event presentation on recall questions following the presentation of misleading questions ($F(2, 54) = 4.72$, $p < .05$), this did not specifically relate to misinformation acceptance levels. This suggests that the enactment modality at encoding can support later recall at test, but less is known about how the enactment modality may protect against misinformation acceptance. This links with the previous point that enactment encoding may not be as effective at mitigating against the MI effect as previously hypothesised. However, there were lower than expected levels of suggestibility in the sample.

Alternative to participants engaging in enactment, one study exposed participants to a live action event presented by a non-participant. Those live events would have contained both auditory and visual information. Roebers et al. (2004) found a significant difference between the correctly answered memory questions in the three visual modalities (slide show, video, and live enaction) tested at encoding. The MI effect in children was equally strong across three presentation modalities; slideshow, video, and watching a live enaction. Despite participants having better overall memory for the live action event compared to the other modalities (live event 81.5%, video 79.1%, slide show 77.0%), this did not reduce the MI effect. Similarly, Gobbo et al. (2002, study 1) found no significant difference in true memories between participants who observed the encoding event compared to the enactment or written narrative conditions. However, the authors commented that there was a trend for children who observed a live event to be less susceptible to misinformation (i.e., increased accurate responses) than children who listened to a narrative.

This review has found conflicting findings regarding enactment at encoding. On the one hand, enactment at encoding (where participants directly participated in an enactment) appears to have little to no attenuating effects of misinformation acceptance (Dijkstra & Moerman, 2012; Gobbo et al., 2002, study 1 and 2). However, when participants were instructed to watch a live enactment (rather than participate), there was evidence that this was associated with reduced misinformation acceptance, compared to participants who listened to a narrative (Roebers et al., 2004). Enactment has also been found to have no encoding benefit over other auditory and visual presentations at immediate recall (Dijkstra & Moerman, 2012). However, a significant increase in misinformation acceptance was found over time (two weeks) in the auditory condition, which was not the case in the enactment and visual modalities, with the enactment condition performing best over time. Therefore, using an enactment modality at encoding may have more benefit over time as opposed to immediately after a to-be-remembered event. It is of note that only three of the included papers applied an enactment modality. Therefore, these findings should be interpreted with caution. This suggests that further research is needed in order to fully assess the impact of the enactment modality.

Findings from the combined audio-visual modality is limited to two studies. As mentioned, these findings were not adequately reported, meaning there is scope for future research to explore the effect of combined audio-visual modality manipulations on the MI effect.

Modality Congruence

The impact of modality congruence across the stages of the MI paradigm was considered. Three studies tested encoding and test modality congruence (Dijkstra & Moerman, 2012; Ulatowska et al., 2016, study1 and 2). All three studies found evidence that congruence between encoding and test made participants more likely to retrieve original information. In Ulatowska's et al. (2016, study 1) first study, there was increased misinformation acceptance where modalities differed significantly from one another across stages of the paradigm. Although it was not possible to draw a general conclusion regarding the influence of encoding and test modality congruence, a second study was conducted to make sense of this finding. After further analysis in study 2, an interaction effect was found between encoding format (presented pictorially or auditorily) and test format (presented pictorially or auditorily) on misinformation acceptance ($F(1, 199) = 4.12$, $p < .05$, $\eta^2 = .02$). Therefore, participants who were exposed to the same modality at encoding and test were less susceptible to the MI effect compared to participants who were exposed to a different modality at encoding and test (Ulatowska et al., 2016, study 2).

Dijkstra and Moerman (2012) also suggested that modality congruency between encoding and test facilitated performance and may be more important than the encoding modality. That is, enactment encoding was predicted to result in stronger memories than auditory encoding, thus reducing misinformation acceptance. This was predicted due to the motor encoding hypothesis, which postulates that the motor system is activated during encoding, enhancing encoding of item specific information and gives a free recall advantage because movements are distinct and separate actions. However, Dijkstra found an auditory modality ($t(30) = 2.89$, $p < .01$) benefit over the predicted benefit of an enactment ($t(40) = 2.07$, $p < .05$). The researchers suggested that this was because encoding and test conditions were more similar in the auditory condition. Specifically, they hypothesised that the verbal narrative encoding and the written test in the auditory condition were more similar than the visual and enactment encoding conditions and written test in the visual conditions (Dijkstra & Moerman, 2012). Therefore, this may be considered to be evidence for the benefit of encoding and test congruency on reducing susceptibility to MI. However, it is of note that were no exact modality matches in this study.

Two studies have considered modality congruence of the misinformation stage with test stage (Campbell et al., 2007, study 1; Yamashita, 1996). Yamashita (1996) found a main effect of modality at the misinformation stage ($F(1,109) = 24.606$, $p < .001$), which interacted with test modality ($F(1,109) = 5.749$, $p < .05$). The findings indicated that there were higher levels of misinformation susceptibility when the modality of the misinformation stage and test matched (Yamashita, 1996). Similarly, Campbell et al., (2007, study 1) postulated that, based on their findings, misinformation acceptance would increase when the misinformation and test stages were both presented in a written modality (i.e., written narrative followed by written recognition test) compared to when the modalities did not match. Therefore, the researchers concluded that the MI effect reported in the study may be explained by encoding specificity biases. Such that, participants match the written cues stored in memory at the time of misinformation with the cues available at test.

Finally, one study considered modality congruence of the encoding stage and misinformation stage by examining the effects of visual and auditory modalities, and encoding-misinformation incongruence, on misinformation acceptance using the MI paradigm (Hendrich, 2019). Significant effects on misinformation acceptance were not found for encoding modality ($F(3, 85) = .70$, $p = .55$, partial $\eta^2 = .02$) and no interactions were found between encoding modality and misinformation modality ($F(3, 85) = 2.07$, $p = .11$, partial $\eta^2 = .02$). Therefore, no evidence was found for a reduced MI effect when there was modality congruence between the encoding and MI stages of the MI paradigm (Hendrich, 2019).

Discussion

Main findings

The current review aimed to use a systematic approach to examine the association between changes in modality and the impact this has on the MI effect. This is believed to be the first attempt to systematically review the impact of different modality manipulations on the occurrence of the MI effect. Two primary objectives for the review were proposed: To explore whether specific modalities at each of the three stages of the MI paradigm are more effective at reducing susceptibility to the MI effect, and to explore whether modality congruence (e.g., same modality used throughout stages of the MI paradigm) increases the presence of the MI effect. For the first objective, the main findings show that the type of modality (i.e., auditory or visual) within the MI paradigm has a limited impact in influencing misinformation susceptibility. Although findings did suggest that auditory modalities at the encoding/misinformation stages can reduce the MI effect, a larger number of studies found no significant main effect of modality on misinformation acceptance. For the second objective, the strongest evidence supported the encoding specificity hypothesis, when encoding and test modalities were congruent. Evidence, albeit limited, also found there were higher levels of misinformation acceptance when the modality of the misinformation and test stages were congruent. No such evidence was found for congruency between encoding and misinformation stages. Therefore, key findings suggest that modality congruence between encoding and test stages can mitigate the MI effect, though research is limited.

These findings are important because the majority of studies within misinformation research have utilised the MI paradigm whereby the original information (encoding stage) and misleading information (misinformation stage) are presented in different modalities (Titcomb & Reyna, 1995). Although there is a wealth of literature using modality changes within the MI paradigm, many studies do not comment on the impact this has on findings. This is problematic, as the current review has highlighted that modality manipulations at encoding, misinformation and test may impact susceptibility to misinformation. Given that different modalities may require different processing, such as sequential processing for reading or auditory, and parallel processing for videos or slides, current misinformation literature has not accounted for these differences (Ulatowska et al., 2016). Therefore, researchers give little consideration to what modality they are using at each stage of the MI paradigm and the different processes required to interact with them. Likewise, the majority of misinformation literature has not considered the impact of modality congruence at different stages of the MI paradigm. This means that a large proportion of misinformation research, specifically those that utilise a MI paradigm, have not adequately accounted for the modality effect and the impact this may have on findings. Although we cannot be sure of the influence the modality effect has in the broader misinformation literature, it is a strength that the current review has considered the scope of this impact. It is likely that modality congruency/incongruency effects are present within literature, but they have heretofore not been explicitly explored.

The Impact of Different Modalities on the MI Effect

The first objective of the current systematic review was to explore whether specific modalities at each of the three stages of the MI paradigm are more effective at reducing susceptibility to the MI effect. The findings reviewed to investigate this objective were mixed. Two of the included papers that explored between modality manipulations indicated that individuals were less likely to be susceptible to the MI effect when they are exposed to auditory information at the encoding stage of the MI paradigm compared to auditory modalities (Dijkstra and Moerman, 2012; Ulatowska et al., 2016, study 1). However, this was contested by several studies, who found no significant effect of encoding modality on misinformation acceptance (Hendrich, 2019; Gobbo et al., 2002, study 1 and 2; Ulatowska et al., 2016, study 2). There is also a further complication when considering the quality rating of the studies. That is, both of the papers that supported the use of auditory modalities at encoding were rated as "strong" (Dijkstra and Moerman, 2012; Ulatowska et al., 2016, study 1). In comparison, all three of the papers who found no effect were rated as "strong" (Hendrich, 2019; Gobbo et al., 2002, study 1 and 2; Ulatowska et al., 2016, study 2). Within the context of this, however, it is also important to consider that although Ulatowska's (et al., 2016, study 1) found support for using an auditory modality at encoding, a follow-up study by the same authors did not replicate this finding (Ulatowska et al., 2016, study 2). What this tells us, however, is that there are discrepancies in the findings regarding modality in the MI paradigm amongst high quality papers. Therefore, this further highlights the need for replication of studies and clarification regarding the impact of modality changes on the MI effect.

Mixed conclusions were also found regarding misinformation modality. Campbell et al. (2007, study 1 and 2) found evidence that the MI effect is attenuated when the misinformation stage is presented in an auditory modality (compared to a visual/text modality), but that this effect is only limited to recognition tests. This conclusion should be interpreted with caution, however, because the design in both experiments did not utilise an auditory testing procedure. However, alternative findings by Ulatowska et al. (2016, study 2), suggested that misinformation acceptance rates were shown to be significantly higher when the misinformation stage was presented in an auditory modality compared to a visual modality. In practice, this might be particularly concerning as the majority of witnesses are interviewed using verbal questioning by police. Further analysis by Ulatowska et al. (2016, study 2) revealed that the misinformation rate differed between auditory and visual misinformation only when the encoding event was presented auditorily. This additional finding provides evidence (albeit limited) for the strength of auditorily encoded information. However, Ulatowska's et al. (2016) paper was given a quality rating of "strong", emphasising the quality of the studies that produced this finding. Kiat (2018, studies 1, 2, 3 and 4) also concluded that auditory processes, specifically verbal, may be more significant in driving the MI effect relative to visualisation. These finding appear to be consistent across different types of within auditory (e.g., voice recordings and verbal questions) and visual modality (e.g., pictures and videos) manipulations.

The impact of auditory misinformation can be partially understood within the context of verbal acoustic coding. Penney (1989) argued that the acoustic code created as a result of exposure to auditory presentation is distinct, whereas the phonological code created when exposed to visual presentations is common to both auditorily and visually presented modalities. This suggests that auditory presentations are rich and durable in comparison to visual presentations. For example, subjects who received visual presentations during questioning compared to auditory were found to be less susceptible to leading questions (Cardone & Dent, 1996). Likewise, subjects have been found to give more enriched accounts when the information is presented verbally (i.e., auditorily) as opposed to a written narrative (i.e., visually; Sauerland & Sporer, 2011). Therefore, the auditory strength may encourage increased recall of auditory items, even if such items are presented at the misinformation stage. However, it is important to note that other research has not found better memory performance for auditory encoding information compared to visual modalities, which we might expect to occur, if auditory presentation results in stronger memories (Roebers et al., 2003). Therefore, future research should further explore the verbal acoustic code as an explanation for the observed effect that auditory misinformation increases the MI effect.

The reviewed studies also considered auditory and visual modalities at test. However, only one study commented on the independent effects of the modalities (Yamahsita, 1996). Although this study found that there was a reduced MI effect when participants were tested visually compared to auditorily, this finding cannot be compared to any of the present studies. Likewise, Yamashita's (1996) paper was given a quality rating of "moderate", so there is the possibility for improved research in this area. Therefore, more research is needed to explore test modality and the impact this may have on the MI effect.

Discrepancies between the findings for visual and auditory modalities are also found in wider eyewitness literature. Further research also highlights how visual and auditory information at encoding interfere with one another. Specifically, there is strong evidence that visual information interferes with auditory processing, and mixed support for auditory interference with visual processing (McAllister et al., 1993). This might suggest that using both visual and auditory modalities across stages of the MI paradigm could interfere with processing of information entirely (Mitchell & Zaragoza, 1996 study 1 and 2). Specifically, modality incongruence at different stages of the MI paradigm could result in one type of modality (i.e., visual) interfering with another (i.e., auditory). On the one hand, interfering processing may be beneficial for memory in instances where the misinformation processing is the stage that is interfered with. On the other hand, it could be detrimental to memory, particularly in instances where there are interferences with processing at encoding. This has a potential impact on current misinformation literature, where the majority of studies utilise different modality presentations at different stages of the MI paradigm.

The current review has highlighted different results for the use of the enactment format at encoding modality on memory accuracy and misinformation acceptance reporting. Dijkstra and Moerman (2012) found that enactment at encoding led to better recognition compared to visual and auditory encoding conditions. When this effect was explored over time (two weeks later), recognition performance in the enactment modality also appeared to deteriorate less than the auditory and visual modalities. This suggests that there is an enactment encoding benefit. However, after a delay of a year, Dijkstra and Moerman (2012) found that the recognition performance for subjects in the enactment condition was somewhat contaminated by the misinformation. Out of the original sample number (115) only 27 participants responded and were included in the follow-up study after a year. Despite this, Dijkstra and Moerman's (2012) study was still rated as "strong" in quality, due to the studies thorough research procedure and open reporting. However, this open reporting also included the researchers' own acknowledgement that because of the limited participant numbers in the follow-up study, these results need to be interpreted with caution. Despite Dijkstra and Moerman's (2012) support for the use of enactment encoding, they also found evidence to support the use of auditory encoding above that of enactment encoding. Specifically, there was a lower rate of misinformation acceptance at immediate recognition in the auditory encoding condition compared to the enactment and visual encoding conditions. However, there was a significant increase in misinformation acceptance over time (two weeks) in the auditory condition, which was not the case in the enactment and visual modalities, with the enactment condition performing best over time. Therefore, using an enactment modality at encoding may have more benefit over time as opposed to immediately after a to-be-remembered event.

On the other hand, Roebers et al. (2003) found lower levels of the MI effect in children who watched a live enactment show, although this did not reach statistical significance. Although findings did not reach statistical significance, there are good theoretical reasons to believe that enactment might improve memory accuracy and prevent MI acceptance. For example, previous research has found superior recall of stimuli that is acted out, compared to verbally encoded items (Engelkamp & Cohen, 1991; Nyberg et al., 1991). Whilst the actor in the study did not directly engage with the children, Roebers et al. (2003) believed that the laughing and shouting out by the children indicated higher levels of emotion compared to if they had been less expressive. One idea is that emotional engagement may contribute to better memory performance; which has been found in multiple studies (Forgas et al., 2005; Hess et al., 2012; Porter et al., 2003). A second idea, by Henson and Gagnepain (2010), is that information encoded during a live enactment is perceptually rich, thus aiding the process of accurate remembering. This idea is supported by the finding that the video condition (presumably more perceptually rich) yielded significantly better memory performance than the slideshow (presumably less perceptually rich; Roebers et al., 2003). Despite theoretical support for enactment strength, the reviewed literature has provided inconsistent findings regarding whether enactment can support memory and reduce the MI effect. Roebers et al. (2003) were given a quality score of "moderate", which is lower than Dijkstra and Moerman's (2012) rating. This is significant as Dijkstra and Moerman (2012) found conflicting results compared to Roebers et al. (2003) regarding the enactment modality. In weighing up this evidence, it is important to consider that Dijkstra and Moerman (2012) asked participants to directly partake in the enactment, whereas Roebers et al. (2203) asked participants to watch a live enactment. Therefore, findings are not directly comparable. Future research is needed to ascertain the impact of enactment on susceptibility to misinformation at the encoding and misinformation stage of the MI paradigm.

Age dependent effects of presentation modality were also explored by three studies. Roebers et al. (2003) found no interaction effect between presentation modality at encoding and age in regard to misleading questions or recall measures. This suggests that the effect of modality on misinformation acceptance is not dependent on age. However, findings did suggest that older children (aged seven) were less susceptible to the MI effect than younger children (aged five). It is of note, however, that the five-year-olds in the control group produced more incorrect answers (77.9%) than those who had been misinformed (73.1%). This may suggest that younger children are more likely to answer incorrectly compared to older children, regardless of the interference of misinformation. Gobbo et al. (2002, study 1 and 2) also explored the interaction between modality presentation, age, and the MI effect, finding that younger children recalled fewer original details compared to older children. This suggests that younger children may have poorer veridical memory (i.e., decreased true memories) after exposure to misinformation compared to older children. No other interaction effects were significant, again indicating that the effect of modality on the misinformation acceptance is not dependent on age.

The second objective of the current systematic review was to explore whether similarity in presentation modality (e.g., modality congruence, the same modality used throughout stages of the paradigm) impacts upon the presence of the MI effect. Therefore, findings relevant to this objective are discussed in relation to modality congruence across the different stages of the MI paradigm.

Modality Congruency at Encoding and Test

Two studies tested and found that modality congruence between encoding and test reduced susceptibility to misinformation. For example, Ulatowska et al. (2016, study 1), found that there was increased misinformation acceptance when modalities differed significantly from one another at encoding and test. Ulatowska et al. (2016, study 2) also found that when encoding and test conditions were presented in incongruent modalities participants were less likely to retrieve the original information. Findings from both studies were from a single paper that was quality assessed as “strong”, giving further weight to the results. The authors concluded that this provided evidence of the encoding specificity hypothesis, as recall was better in conditions where encoding and test were modality congruent. However, there are three stages in the MI paradigm, so congruency across different stages might be important. I consider this, next.

Modality Congruency at The Misinformation Stage and Test

Two studies tested congruency between the misinformation and test stages of the MI paradigm and found that congruency between misinformation and test may be significant in increasing misinformation acceptance (Campbell et al., 2007, study 1; Yamashita, 1996). Yamashita (1996) found evidence that participants were more susceptible to misinformation when the modality presented at the misinformation stage was also the same at test (e.g., visual misinformation and visual test) opposed to when the modalities were incongruent (e.g., visual misinformation and auditory test). Campbell et al. (2007, study 1) found that when the misinformation stage and test were presented in similar modalities, participants’ memory for the encoding event was negatively affected. They proposed that the contextual match in misinformation and test modalities will improve memory for the misinformation, impairing retrieval of the to-be-remembered event. Campbell et al. (2007) explained this using the encoding specificity effect, whereby modality congruence at misinformation (an encoding event) and test could be detrimental to memory retrieval. Abeles and Morton’s (1999) findings also supported that modality congruence at the misinformation stage and test can impair retrieval with fewer correct details of the encoding event, despite identical source information and test modalities. However, the three papers that considered modality congruence between the misinformation and test stages were rated as having “moderate” quality. As such, there is scope for improved research to explore this area and add to our understanding of modality congruence in the MI paradigm.

Also of note is a study that solely considered modality congruence at only the misinformation stage. Mitchell and Zaragoza (1996, study 1 and 2) considered the impact of modality congruence when there were repeated (up to three times) misinformation exposures. Each misinformation exposure was presented either in the same or different modalities. Exposure to repeated misinformation, presented each time in a similar modality, was found to strengthen subjects' ability to make accurate source attributions compared to those who were exposed to a single modality. They concluded that when the goal was to simply remember the source of an encoding event, modality congruence at the misinformation stage, in instances where there are repeated exposures to the misinformation, may decrease misinformation acceptance. Put another way, modality incongruence between multiple misinformation exposures can result in an increased tendency to misattribute post-event misinformation to the originally witnessed encoding event. The researchers theorised that modality congruence makes it more challenging for the participant to accurately discriminate between original and post-event exposures. Of note is that the paper was given a quality score of "moderate". This was due to limited discussion around the implication of source attributions and how this links to the MI effect. Although some attempt to do this was made, further clarity is needed to draw firm conclusions from these findings.

Modality Congruency at Encoding and The Misinformation Stage

Limited evidence for the impact of modality congruence at encoding and the misinformation stage was found in the reviewed studies. One study did not find a significant interaction effect between encoding modality and misinformation modality (Hendrich, 2019), and that this paper had a "strong" quality rating. Therefore, the hypothesis that participants would have decreased rates of misinformation acceptance when modality was congruent across the three stages of the MI paradigm compared to when there was modality incongruence was rejected (Hendrich, 2019). However, the researcher notes that their study produced lower misinformation acceptance rates overall compared to other misinformation studies, and so the researcher's lack of significant findings may have been due to floor effects. Given that only one of the studies included in the review explored modality congruence at encoding and the misinformation stage, a firm conclusion about the impact this has on misinformation acceptance cannot be made. Likewise, further research has suggested that individuals make familiarity-based memory judgements (Nishiyama et al., 2017). Therefore, it is likely discrepancies between objects and presentation modalities at encoding and the misinformation stage will help to reduce familiarity and mitigate the MI effect (Loftus, 1979b). This suggests that modality congruency at encoding and the misinformation stage will improve an individuals' ability to make accurate source attributions, compared to when modality is incongruent. Therefore, the importance of understanding how subjects can be misled through modality changes or congruence is imperative to avoiding eyewitness errors (Ulatowska et al., 2016).

Considering the modality congruency findings across the stages of the MI paradigm, what are the implications for theory? The encoding specificity principle proposes that modality congruence between information at encoding and test assists memory retrieval. That is, when modality is congruent at encoding and test, there is increased recall of correct information (Lee & Hirota, 1980; Tulving & Thomson, 1973). The principle has been used to describe two different patterns of modality congruence findings. First, this hypothesis has been used to explain why modality congruence between the encoding event and test reduced susceptibility to misinformation (e.g., Ulatowska et al., 2016). The traditional principle of encoding specificity states that matching encoding and recall (test) contexts will assist the retrieval of true memories (Tulving & Thomson, 1973). The original encoding event and the misinformation stages can both be considered to be distinct encoding events (Okado & Stark, 2005), and therefore the encoding specificity principle can be used flexibly to account for different findings. When considering what this means for the encoding specificity hypothesis, this suggest that the principles underpinning encoding specificity can be applied to misinformation research. However, its application to MI research is not simple, given the multiple stages in the MI paradigm. Equally, it could be argued that the encoding specificity principle's ability to explain different patterns of findings means that it is of limited value (Greenwald et al., 1986). This could mean the encoding specificity principle does not adequately explain the MI findings and more specific theories could be developed.

Limitations of the review

At present, the studies that have been conducted in this area have primarily used small sample sizes recruited from educational institutions. Although statistical analysis within each of the included papers was assessed to be appropriate, the generalisability of these findings cannot be confirmed. Positively, the issues associated with this have been acknowledged by one study, noting that results need to be interpreted with caution due to the small sample size (Dijkstra & Moerman, 2012). Samples were almost exclusively recruited via educational institutions, and some may be concerned that this limits the generalisability of these findings to other participant pools, such as members of the public who are not currently in education (Hultsch et al., 2002). However, Patihis et al. (2013) found that even people with superior memories are susceptible to misinformation. Therefore, it is unlikely education or ability make an individual less susceptible to misinformation. Future research may benefit from recruiting larger samples from a wider range of participant pools. This will increase the generalisability of results to help better understand how modality manipulation can impact the MI effect.

Comparison between studies manipulating different modalities (i.e., visual, auditory, enacted) at different stages of the paradigm (i.e., encoding, misinformation stage, test) has been challenging. Whilst the heterogeneity between studies may be problematic for data synthesis, an attempt to summarise and collate the findings was made. The current review is the first of its kind and should be interpreted considering the limitations of the included studies. For example, different methodologies were used across the studies, such as variations in encoding events (i.e., differences in the story or event being encoded), delivery of misinformation (i.e., different levels of similarity between the misinformation and the encoding event), and test procedures (i.e., the different timing test procedures were administered). All of these experimental decisions could potentially influence the size of the MI effect (Titcomb & Reyna, 1995). Similarly, results should be interpreted with consideration of the particular experimental designs used. It is recommended that higher levels of comparable study features, such as similar modality manipulation at the same stages of the MI paradigm are likely to produce more reliable, comparable and generalisable conclusions. Whilst the presence of the MI effect is mostly consistent when the MI paradigm is used, some of the included studies used non-traditional forms of the paradigm (e.g., Abeles & Morton, 1999). Differences between experimental designs may impact the production of the MI effect, thus making the specific impact of modality more challenging to analyse (Hendrich, 2019). Therefore, the current body of research would likely benefit from more studies that directly compare modality effects within a single experiment.

Whilst some studies considered written narratives to be a visual modality, some also considered them to be a verbal modality. Given that the view on this was not consistent, studies that have used two different modality labels have been considered separately in the results. This made it difficult to collate and make sense of the findings for the verbal modality, if indeed it is significantly different to auditory modalities. However, given that verbal and auditory processing are more similar than visual processing (Fletcher et al., 1995; Grasby et al., 1993), this may have further implications for findings. Therefore, future reviews should consider the specific impact of verbal modalities within the context of auditory and visual modality manipulations.

Dissertations and theses were included in the review, because this is deemed appropriate when there are fewer relevant studies (Hartling et al., 2017). This was particularly true of the current review topic given the dearth of research in the subject area. Including dissertations and theses in this review may also have aided the reduction of publication bias (Vickers & Smith, 2000) and provided a more complete reflection of the research that has been conducted in this area. Nevertheless, unpublished research is less likely to have been peer reviewed and critically analysed by experts in the field. This may mean methodological issues and the quality of the research have not been accounted for (Paez, 2017). However, these issues were considered during the adaptation of the quality assessment tool to ensure papers were only included when their methodology was moderate or strong in rating.

It is important to note that the quality assessment tool was edited to encompass components of quality most relevant to the current review. Although the original tool has been found to have validity across assessment domains (Armijo-Olivo et al., 2012), it cannot be guaranteed that this validity applies to the adapted version. Steps were taken to ensure adapted sections followed identical scoring procedures to the original tool, including making sure sections were adapted in relation to the relevant headings of each domain. The adaptations made were aimed at maximising the relevance and usefulness of the quality assessment to the field of misinformation research. Previous reviews have also adapted tools to quality assess papers within their topic area (e.g., Lumbreras et al., 2008)

Conclusions

This review firstly aimed to investigate whether specific modalities at each of the three stages of the MI paradigm are more effective at reducing misinformation susceptibility. Findings highlighted that although specific modality manipulations have been found to impact on the MI effect across the different stages of the MI paradigm, this evidence is limited. At encoding, there was strong evidence to suggest that the impact of the modality presented has a minimal impact on misinformation acceptance. It is of note that in a smaller number of studies a significant effect of modality at encoding on the MI effect was found. Specifically, that an auditory modality at encoding may reduce the MI effect. However, given that this was only shown in two studies these findings are not generalisable. There were also conflicting findings regarding the misinformation stage, whereby there was limited evidence to both support and counter the view that auditory misinformation reduced the MI effect. Findings were also explored at test, where only one study found that using a visual modality at test reduced the MI effect, making it difficult to conclude whether this finding is universal. Therefore, further research is needed to expand these findings so firm conclusions can be made.

Second, this review aimed to explore whether modality congruence (e.g., same modality used throughout stages of the MI paradigm) increases the MI effect. There was support for the encoding specificity hypothesis, where modality congruence at encoding at test was found to reduce misinformation acceptance. One study found that modality congruence between the misinformation stage and test, may make subjects more susceptible to the MI effect. Future research should aim to expand these findings to ascertain the full extent of the impact modality congruence presented at the misinformation stage may have on misinformation acceptance. At present this is an undeveloped area of the misinformation research, and therefore the current impact of modality congruence may be underestimated, with researchers making experimental decisions without considering the effect that modality may have on misinformation acceptance.

Chapter Three: Research Report

Does Presenting Perpetrator and Innocent Suspect Faces from Different Facial Angles Influence the Susceptibility of Eyewitness Memory? An Investigation into The Misinformation Effect and Eyewitness Misidentification

Abstract

This research investigates the misinformation effect in the context of a simultaneous lineup test, wherein the lineup faces are presented simultaneously in an array to an eyewitness. We used a misinformation paradigm to test whether the misinformation effect is differentially likely depending on the angle at which the perpetrator is encoded relative to the misinformation and lineup test stages. Participants viewed a mock crime that showed the perpetrator in frontal or profile view. Next, participants viewed a news report that showed the face of a person (presented in front or profile view) who was suspected of committing the crime. However, in actuality, he was a different person (i.e., he was an innocent suspect). Thus, the angle of the faces shown across the stages of the MI paradigm were either congruent (i.e., matched across all stages) or incongruent (did not match across all stages). There were no significant differences in discrimination accuracy depending upon congruency. These results suggest that the misinformation effect is not differentially likely depending on whether the misleading face is presented in the same or different angle relative to encoding and test. Further, participants who encoded and were tested with faces shown in frontal rather than profile pose had significantly lower discrimination accuracy. These results are in line with the encoding strength hypothesis, whereby the encoding of frontal positioned face is superior to encoding of a profile face. This has implications for best practice in eyewitness procedures and provides further insight into the impact of facial angle congruency and the MI effect.

Introduction

In June of 1984, infamous serial killer Ted Bundy appeared before the Court of Appeal to argue that the police had used an impermissibly suggestive procedure by asking one of the eyewitnesses to select his picture from a photographic lineup (see Ted Bundy v. State, 1984). In particular, he argued that the witness was influenced by having seen his picture in the newspaper in the days before being shown the lineup. The prosecution countered this, arguing the witness had not been influenced by the picture of Bundy in the newspaper because it showed his face in frontal pose (i.e., head on) whereas she saw the perpetrator from the profile angle during the offence. The witness herself also argued that because the lineup picture was of Bundy's profile, she was not misled by the newspaper photo because she identified him based on her original memory of seeing him from a profile position. Bundy's appeal was rejected, suggesting the Court found the prosecution's argument more persuasive. While there is much research on the misinformation effect (MI effect), there has been little research on whether the effect is differentially likely depending on the facial angle congruence across the encoding, misinformation, and test phases. In this chapter, we test the prosecution's argument, and ask whether discrimination accuracy is higher among misled witnesses if the facial angle of the misleading face differs from the angle at which witnesses saw the perpetrator commit the crime.

The misinformation (MI) effect refers to memory impairment that arises after exposure to misleading information (Volz et al., 2017). That is, when an individual is exposed to misleading information after witnessing (i.e., encoding) an event, they are likely to include the misleading information when reporting their memory of what they witnessed (Ayers & Reder, 1998). Several theories as to why the MI effect occurs have been proposed (see Chapter One). The overwrite explanation postulates that the memory trace for the original information is degraded, impaired, or altered due to misinformation (Loftus, Miller & Burns, 1978). With the Ted Bundy example, this explanation would hold that the newspaper picture had overwritten the witness' memory of the perpetrator, and therefore, the witness mistakenly identified Bundy. Others argue that the original and misleading information coexist in memory, without any alteration to the original memory trace. Specifically, the retrieval interference explanation maintains that whilst the original memory and misinformation coexist, difficulty accessing the original information is a direct cause of exposure to misinformation (Bekerian & Bowers, 1983; Chandler, 1991). Alternative coexistence theories argue that the original information and misinformation are both remembered and retained; however, the misinformation is selected due to source confusion (Johnson et al., 1993). Referring back to the Ted Bundy case, the coexistence account would hold that while the witness held in memory both the perpetrator and the newspaper image of Bundy, the witness based her identification on the image of Bundy, and thereby committed a source confusion error. Despite ongoing debate about the underlying mechanisms, it is well-accepted that individuals can report post-event misleading information when their memory of the original event is tested.

When considering the real-world impact of the MI effect, the most prominent area of research concerns eyewitness testimony. The consequences of erroneous eyewitness testimony are enormous, such as an innocent person being wrongly convicted, or a guilty person being free to commit further crimes. Research has found that both single and multiple eyewitnesses can be misled regarding a single event when they are asked to recall details (Frenda et al., 2011). As well as recalling details from an event, witnesses are also required to describe what people look like and to identify guilty suspects from police lineups (Frenda et al., 2011; See Chapter One). Therefore, the impact of post-event misinformation on memory for faces could influence a witness's ability to make a correct identification from a lineup.

What is known about factors that affect susceptibility to post-event misinformation in face recognition? Researchers have postulated that internal systems used to process faces are unique, arguing that faces are processed differently compared to other objects (Farah et al., 1998; Leder & Bruce, 2000; Werner et al., 2013). As a result, early research claimed that faces are not susceptible to interference from other faces (Davies et al., 1979). However, developments in this field have found that memory for faces, much like other objects, is susceptible to interference. For example, faces have been found to be susceptible to interference after the introduction of misleading written information.

Loftus and Greene (1980) found that when participants viewed a face and later heard a written description of the face that contained misinformation, they were more likely to incorporate false details in their description and reconstruction of the target face. Other research has found that inaccurate facial composites, where a portrait of a perpetrator is completed based on an eyewitnesses' memory (often using a sketch artist), may also affect susceptibility to misinformation for faces. A study conducted by Topp-Manriquez et al. (2016) found that participants who created or simply viewed a composite of a face after viewing a to-be-remembered target face had lower levels of identification accuracy compared to those who did not. This suggests, along with earlier studies, that viewing post-event faces that differ to that of the target face can hinder identification accuracy (Kempen & Tredoux, 2012; Sporer et al., 2020; Wells et al., 2005).

Research has largely studied facial recognition under conditions in which the to-be-remembered face is presented in the frontal viewpoint (e.g., Jenkins & Burton, 2008). However, facial pose variation across study and test may be a key variable with respect to accurate face recognition. A face presented in profile pose, for example, has one less eye, eyebrow, and cheek compared to a face in frontal pose. Therefore, different facial angles can result in the obstruction of facial features and surfaces (Favelle et al., 2017; Swystun & Logan, 2019). Studies have concluded that changes in the viewpoint of a face is associated with a reduction in the accuracy of face identification (Hill et al., 1997; O'toole et al., 1998). Likewise, multiple studies have concluded that facial viewpoint changes can impair the ability to discriminate between unfamiliar faces (Favelle et al., 2017; Guy et al., 2017; Meinhardt-Injac et al., 2009). Bruce (1982) found that when individuals studied a frontal face and were subsequently tested with a frontal face, they were able to recognise faces more accurately and quickly compared to when they were tested with faces at a 45° angle. This is also supported by recent research where larger, more significant viewpoint changes, such as 45° (three-quarters) and 90° (profile), are associated with poorer face discrimination performance (Lee et al., 2006). Swystun and Logan (2019) explored more minor degree changes in the viewpoint of unfamiliar faces. They found that sensitivity to face identification was greatest for faces presented from a frontal facial angle. However, sensitivity to face identification declined linearly as faces were rotated away from the frontal facial angle, even when rotations were completed in smaller increments (5°). The researchers also concluded that changes in facial angle further impaired facial discrimination ability. Sensitivity declined linearly as the magnitude of the facial angle change increased. This suggests that even smaller changes in the facial angle shown can impair face identification and discrimination, with this impairment increasing with more extreme facial viewpoint changes (such as a 90° profile facial angle). This suggests that there would be significant benefits to encoding a frontal posed face, compared to profile face, for later identification and discriminability.

It is also practically important to consider facial angle, given that an eyewitness may see a perpetrator from the side or with parts of their face concealed. Therefore, the existing research may not be adequate in explaining how witnesses process and recognise faces when they are not encoding from a frontal position. Misinformation acceptance has important implications for facial recognition during criminal proceedings. For example, if a witness is unable to remember or encode the original face, it is probable that they will be more prone to misinformation in the form of an interfering face (Loftus & Greene, 1980). Likewise, different types of faces or facial angles may make witnesses more susceptible to poor encoding or misinformation acceptance. This is theoretically important because it can help us learn more about how faces are stored in memory and retrieved when interfering faces, at different facial angles, are presented. One way of testing this is to manipulate the angle of a face at encoding (e.g., profile or front) and the misinformation stage (e.g., profile or front) in the MI paradigm, with angle of encoding always matching encoding at test. At test, participants either see a target-present lineup, containing the original perpetrator, or a target-absent lineup, containing the misinformation suspect. This will enable the measurement of discrimination accuracy where the viewing angle of the faces shown are either the same (congruent) across encoding, misinformation and test, or different (incongruent) at encoding and test compared to the innocent misinformation suspect. It also enables for the measurement of the encoding strength within each condition. That is, whether encoding a frontal positioned face will produce a stronger encoded memory compared to encoding a profile view face. Or, put another way, if discrimination accuracy is better when the perpetrator at the encoding stage is presented from the front or profile view.

On the one hand, we might predict that susceptibility to misinformation depends on the strength of the memory that is encoded (at either initial encoding or the misinformation stage). Research suggests that viewing faces from the front provides people with more information to aid the process of encoding, compared to viewing faces from the profile (McKelvie, 1976). One idea is that the front of the face contains more information about facial features. Researchers concerned with the contributions of internal and external facial features have identified the importance of certain facial features for face recognition. The “feature hierarchy” for unfamiliar faces indicates that the outline of the head, the eyes, mouth, and nose are the most important for facial recognition (Fraser, 1990). That is, seeing more of these facial features, such as both eyes simultaneously, may improve recognition of faces (Goldstein & Mackenberg, 1966; McKelvie, 1976). Another idea is that faces are thought to be processed holistically. The holistic encoding hypothesis maintains that instead of processing faces as a collection of separate facial features, we instead process the face as a perceptual whole (Taubert et al., 2011). Therefore, holistic processing of a face may be impacted by facial angles that provide less information (e.g., the profile view), as inversion and misalignment have both been associated with poor facial recognition performance (Meltzer & Bartlett, 2019). If the front of the face provides a stronger memory trace for later recognition, then this leads to the prediction that when people encode a face in frontal view, but later receive misinformation in profile view, they will be more likely to have a stronger memory trace for the frontal to-be remembered face compared to the profile view misinformation. But by the same logic, when people encode a face from profile view, but receive misinformation in frontal view, individuals may have a stronger memory trace for the frontal misinformation compared to the to-be-remembered profile face. Therefore, we might predict that discrimination accuracy will be higher when participants view a profile misinformation face when the encoding and test faces are frontal, compared to when participants view a frontal misinformation face when the encoding and test faces are profile. This would mean that the false alarm rate to an innocent suspect in the target-absent lineup will be lower, and the hit rate to the perpetrator in the target-present lineup will be higher when participants view a profile misinformation face when the encoding and test faces are frontal, compared to when participants view a frontal misinformation face when the encoding and test faces are profile.

On the other hand, it could be predicted that susceptibility to misinformation depends on the congruency of the facial angle shown across the stages of the MI paradigm. Specifically, facial angle congruence between encoding, misinformation and test phases may make an intervening face more difficult to distinguish from the to-be-remembered face, and therefore result in lower discrimination accuracy, compared to when the misinformation face is presented in a different position (incongruent). This prediction is supported by facial recognition research, where sensitivity to face identification accuracy has been found to be greatest for faces presented from a front facial angle, with sensitivity to face identification declining as faces are rotated away from a frontal angle (Swystun and Logan, 2019). Therefore, it could be predicted that when a frontal facial angle is shown at encoding and test, but a profile facial angle is shown at the misinformation stage, participants will be more able to distinguish between the faces shown and correctly identify the perpetrator. This is because their sensitivity to the frontal facial angle should produce richer memories that aid identification compared to those presented in a profile facial angle (Goldstein & Mackenberg, 1966; McKelvie, 1976). What is less understood is what might happen when all three stages of the MI paradigm are presented with a congruent facial angle (i.e., front, front, front or profile, profile, profile). Better recognition performance has been found when study and test format of faces is the same (Liu & Chaudhuri, 1997; Shapiro & Penrod, 1986; Wells & Hryciw, 1984; Wogalter & Laughery, 1987). However, recent research by Carpenter et al. (2022) found that when there was shared contextual information between targets and lures, this led to increased false memories. Therefore, it may be more challenging for individuals to discriminate between more similar information (i.e., similar facial angles). As such, viewing more of the face, such as a frontal face position, may not be the most important factor in boosting recognition. Instead, facial angle congruency may be important.

To make predictions regarding whether congruence between facial positioning at the event and misinformation stage of the MI paradigm increases misinformation acceptance, it is important to consider the encoding specificity hypothesis. Traditionally, the encoding specificity hypothesis argues that matched encoding context at encoding and test will enhance true memory retrieval. Whilst this hypothesis has been widely supported in literature, the misinformation stage can also be viewed as an encoding event. Indeed, Campbell et al. (2010) and Yamashita (1996) tested event recall and both argued that the MI effect was greater when the recognition test was presented in a more similar format as the misinformation. This suggests that congruence between misinformation and test, may be detrimental to encoding memory retrieval. These findings support the prediction that the false alarm rate to an innocent suspect in the target-absent lineup condition will be lower in the incongruent facial angle conditions compared to congruent facial angle misinformation condition, and the hit rate to the perpetrator in target-present lineups will be higher in the incongruent facial angle condition compared to congruent facial angle misinformation condition. Put another way, discrimination accuracy will be better when the facial angle shown at the misinformation stage is incongruent to the facial angle shown and the encoding and test stage, compared to when it is congruent.

To summarise, this project investigates: (1) If discrimination accuracy is higher when participants view a profile misinformation face when the encoding and test faces are frontal, compared to when participants view a frontal misinformation face when the encoding and test faces are profile (encoding strength hypothesis). This is because evidence suggests that seeing a frontal face may facilitate encoding and recognition processes; therefore, participants may be more likely to accept the frontal misinformation, as the encoding of this will be superior to the original event (profile information). Likewise, participants may be less likely to accept the profile misinformation, as the encoding of this will be inferior to the original event (front information). This project also investigates (2) If discrimination accuracy is higher when misinformation is presented in a different pose to the encoded event and lineup test, compared to when misinformation is presented in the same pose as the encoded event and lineup test. It is hypothesised that discrimination accuracy will be lower when participants see the misinformation suspect and the lineup test from the same facial position (facial angle congruency prediction). This is because it will be easier to mislead an eyewitness if the cues stored in their memory match at misinformation and test (see Chapter Two; Campbell et al., 2007).

Methodology

Design

The current hypotheses and analysis plan were pre-registered on the Open Science Framework before data were collected. A factorial between subjects design was used, where participants were randomly assigned to one of eight conditions: 2 encoding view (front, profile) x 2 innocent suspect view (front, profile) x 2 lineup (target-absent, target-present). Target-absent lineups contained the innocent suspect presented among five fillers. The target-present lineups contained the guilty culprit among five fillers. The innocent suspect and guilty culprit were never presented in the same lineup. The facial position of the faces shown during the lineup always matched the facial position shown at encoding. Therefore, it was possible to collapse across conditions such that participants either received congruent facial angles (front encoding, front innocent suspect, front lineup; profile encoding, profile innocent suspect, profile lineup) or incongruent facial angles (front encoding, profile innocent suspect, front lineup; profile encoding, front innocent suspect, profile lineup) information. Table 3.1 shows the summary of each condition with the relevant abbreviations.

Table 3.1: Table to show Front (F), Profile (P), Target-Present (TP) and Target-Absent (TA) experimental conditions.

Encoding Facial Position	***Facial Position of Innocent Suspect***	***Test: Lineup Condition and Facial Position***	***Condition Summary***	***Facial Angle Congruency***	***Total Per Condition***
Front	Front	Front; Target-Present	FFF-TP	Congruent	258
Front	Front	Front; Target-Absent	FFF-TA	Congruent	269
Front	Profile	Front; Target-Present	FPF-TP	Incongruent	268

Encoding Facial Position	***Facial Position of Innocent Suspect***	***Test: Lineup Condition and Facial Position***	***Condition Summary***	***Facial Angle Congruency***	***Total Per Condition***
Front	Profile	Front; Target-Absent	FPF-TA	Incongruent	252
Profile	Profile	Profile; Target-Present	PPP-TP	Congruent	251
Profile	Profile	Profile; Target-Absent	PPP-TA	Congruent	251
Profile	Front	Profile; Target-Present	PFP-TP	Incongruent	251
Profile	Front	Profile; Target-Absent	PFP-TA	Incongruent	251

Participants

Our pre-registered data collection stopping rule was 2,000 participants. The sample size was agreed so that data could be collapsed across conditions to answer the research questions. Using mean differences and standard deviations observed in Mickes et al. (2012) as a guide, a power analysis indicated that, with a minimum of 250 participants per between-subjects condition, power would exceed 80%. We determined the sample size needed for >80% power to detect significant MI effect within each lineup condition. A bespoke power calculation tool developed for eyewitness lineup procedures was used (https://github.com/E-Y-M/poweROC). The MI effect size was based on effect sizes from the literature (Bülthoff et al., 2019; Colloff et al., 2021; Longmore et al., 2008), and it was reframed in terms of possible condition pAUC ratios, and used a Bonferroni-corrected alpha level based on the number of comparisons to be made (i.e., alpha = .05/2). An initial 2,947 participants were recruited using Amazon Mechanical Turk; all of whom were located in the UK or America and aged 18 years or older.

Individuals who had previously taken part in studies using the same crime video or lineup photographs were prevented from taking part in this study. Participants were paid 35 cents for taking part in the study, which took approximately five minutes. Participants were excluded from final analysis if they incorrectly answered the attention check question or stated they had experienced significant technological issues that prevented them from witnessing either video (total *N* excluded = 896).

The final sample was 2,051 participants (55% female, 44% male, 1% preferred not to say or stated “other”; 18-89 years old, *M* age= 38; 71% White Caucasian, 9% Black or African American, 6% Hispanic or Latino or Spanish, 5% East Asian, 2% South Asian, <1% American Indian or Alaska Native, <1% Native Hawaiian or Other Pacific islander, 3% said other and 3% preferred not to say).

The study was advertised online via Amazon Mechanical Turk. Research has shown that collecting data online, through websites such as Amazon Mechanical Turk, can produce large amounts of high-quality and valid data (Buhrmester et al., 2011; Mason & Suri, 2012). Participants were initially provided with an on-screen participant information sheet. This included information about the study and the participant’s right to withdraw. Participants were required to select “continue” on-screen to consent before they could take part. When they began the study, participants were asked several demographic questions (i.e., age, sex, and ethnicity/race).

Materials

A traditional MI paradigm was used in this study as it allowed the researcher to replicate witnesses’ experiences under experimental conditions. That is, it allowed the researcher to introduce misinformation that eyewitnesses may be exposed to in real-life scenarios (Takarangi et al., 2006). The MI paradigm also allowed for misinformation errors that involve the construction of "incorrect" memories based on suggestions and misleading information. Therefore, the MI paradigm is able to test how an individual takes an external suggestion and misattributes this to their own personal memory of an event (Zhu et al., 2013). Whilst other tools used to measure memory, such as the Deese-Roediger-McDermott paradigm, can test internally generated false memories, the MI paradigm explicitly accounts for the external suggestion witnesses may encounter before identifying a suspect. Likewise, even if a witness cannot recall or does not fully encode the original event, they are still likely to be aware, in emotional or subconscious memory, that an event has occurred (Hämmerer et al., 2017; Macleod, 2002). This supports the use of a traditional MI paradigm as witnesses are still misled even when they have not encoded the original event, as they do not reject the misinformation.

Videos

The video stimuli presented at the encoding stage was a mock crime video from Colloff et al. (2021), lasting approximately seventeen seconds, depicting a Caucasian male perpetrator, aged approximately thirty years, stealing a handbag from a female victim. There were two videos: The perpetrator was shown either from a front facing or profile position. The video stimuli presented at the misinformation stage was a news report video containing a photograph of an innocent suspect. The video lasted approximately thirty-six seconds and contained an auditory narrative and subtitles explaining that a suspect had been arrested in connection with a recent handbag theft in the area (see Appendix Eight). Specifically, the suspect was apprehended after police reviewed CCTV footage of the crime, and believe that the culprit looked like a local resident. A picture of an innocent suspect's face was then shown on screen, either from a front facing or profile view. The innocent suspect was male, aged approximately thirty years, and was similar in appearance to the perpetrator in the encoding video. The innocent suspect was chosen based on pre-existing data from Colloff et al. (2021). These data showed that amongst the six filler faces used in the target-absent condition in the study, the innocent suspect chosen was considered the most similar in appearance to the perpetrator. Faces shown in the encoding stage and the innocent suspect stage were both displayed for a duration of seven seconds.

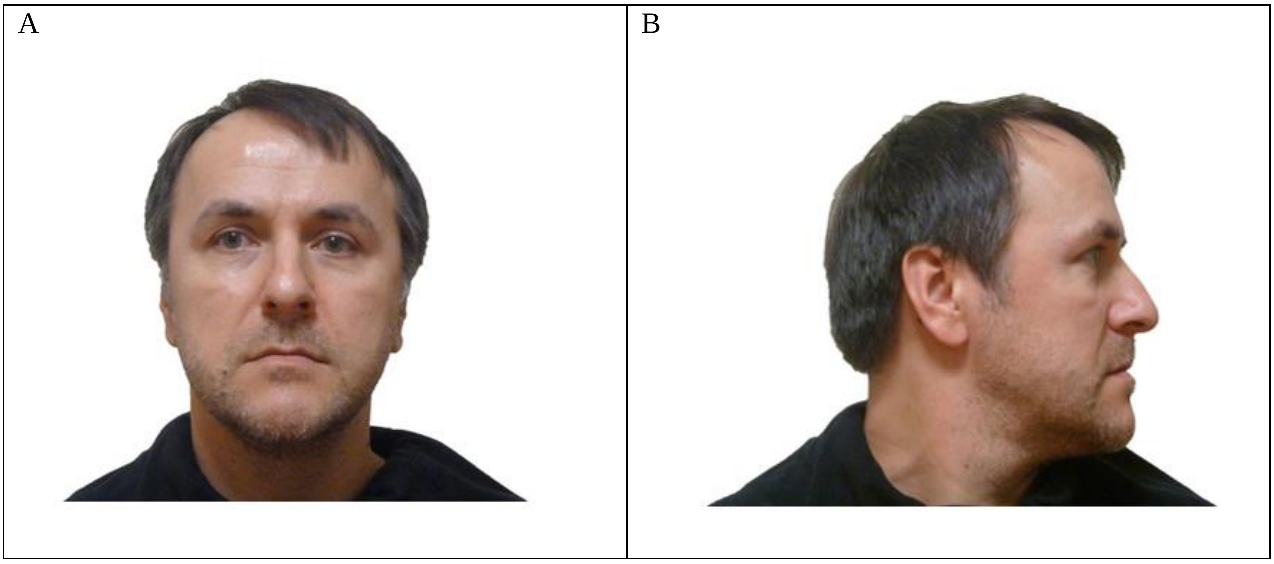

Figure 3.1. Guilty suspect lineup faces from the front (A) and right-profile (B)

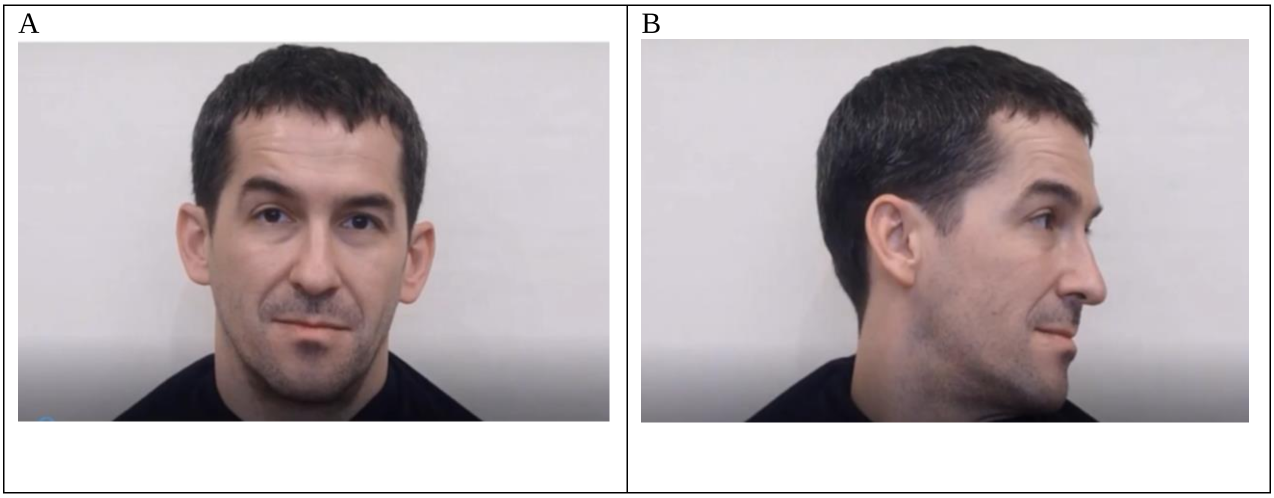

Figure 3.2. Innocent suspect lineup faces from the front (A) and right-profile (B)

Lineups

For the final stage of the MI paradigm, participant memories were tested using a six-person simultaneous lineup procedure, which is used in many countries worldwide (Fitzgerald et al., 2021). The lineup consisted of photographs of six people from the shoulder upwards that have been successfully used in prior research (Colloff et al., 2021). In the target-present lineup conditions, the guilty suspect was shown amongst five fillers. In the target-absent lineup conditions, the misinformation suspect was shown amongst five fillers. Lineups were presented with either profile view or frontal facing suspects that always matched the facial position presented to the participant at encoding. Colloff et al. (2021) selected fillers who had similar facial attributes to the perpetrator in the mock crime video, as dictated by police guidance where it states that police officers should select fillers such that the suspect does not stand out (Police and Criminal Evidence Act 1984, Code D, 2017; Technical Working Group for Eyewitness Evidence, 1999). For example, police officers select fillers based on the similarity of appearance to the suspect or based on the witness's description of the perpetrator, including similarity between age, build, gender, skin tone, hair colour, eye colour, facial hair, and hairstyle. Colloff et al. (2021) established through mock witness-testing that the lineups were fair.

At the lineup, the faces were either shown from the front or from the right profile view. At present, there is a dearth of literature examining the effects of the different sides of the face on facial recognition performance. For example, some research has suggested that the right side of the human face has greater saliency as it bears more resemblance to the face as a whole (Gilbert & Bakan, 1973). On the other hand, Butler et al. (2005) found that when chimeric faces are used (where the left and right side of the face are combined from two different people), participants were more likely to bias their responses towards information on the left-hand side of the face. However, the current research did not use chimeric faces, instead using photographs and videos of sole individuals. Therefore, the right profile faces were used in the "profile" conditions.

Procedure

All participants completed the three primary stages of the MI procedure: the encoding stage, the misinformation stage and lineup test stage. First, in the encoding stage, participants were randomly assigned to watch one of two versions of the video: 1) the perpetrator's face was shown in right profile view for the duration of the video, or 2) the perpetrator's face was shown from the front, head on, for the duration of the video. After watching the video, participants completed a one-minute filler task, consisting of anagram puzzles.

Next, the misinformation stage began. Participants watched the news report video and were randomly assigned to view the innocent suspect either in the same (front encoding, front MI; profile encoding, profile MI) or different pose (front encoding, profile MI; profile encoding, front MI) compared to the facial positioning shown in the initial video. After viewing the news report video, participants then completed a further one-minute anagram filler task.

Finally, participants were presented with a simultaneous lineup test. Participants were randomly assigned to view either a target-present or target-absent lineup. Before the lineup, participants were told that they needed to identify the person who they saw in the mock crime video. They were informed that the guilty suspect may or may not be present in the lineup. Participants were asked to identify whether the guilty suspect was present, or to indicate 'not present' if they believed the perpetrator was not present in the lineup. If a suspect was selected, participants were asked to indicate how confident they were in their identification response on a scale raining from "guessing that he is the culprit" (50%) to "completely certain this is the culprit" (100%). If "not present" was selected, participants were presented with a forced choice lineup, comprising of the same lineup members in the same facial position as they had seen before, and asked to guess which suspect was the one they had seen in the crime video. They were then asked to indicate how confident they were that the person selected was *not* the person seen in the original crime video, on scale from “completely certain he is not the culprit” (-100%) to “guessing this is not the culprit” (-50%). This was done as there have been instances where police have tested individuals in lineups more than once. For example, in the case of Foster v. California (see Foster v. California, 1969), a suspect was presented to the same witness in two separate lineups. In this case, the court held that the way the lineups had been conducted may have been unfair to the suspect, providing a compelling example of unfair lineup procedures. Therefore, it was important to ensure we did not screen out the real-world possibility this could occur. On completion of the confidence scale, participants were asked an attention check question (“How many people were in the first video you watched?”) and a technical check question (“Did you experience any technical issues when watching the mock-crime video (the first video) or the news report video (the second video). If “yes” was selected for the technical check question, participants were asked to briefly explain the technical issue they had experienced. Participants who answered the attention check incorrectly, or who described experiencing significant technical issues (that prevented them from watching the videos), had their data excluded from final analysis. Upon completing checks, participants were shown an on-screen debrief form which reiterated the details of the study, withdrawal procedures, and provided contact details for the researchers. Participants completed the study by closing the study tab on their computer.

Ethics

Full ethical approval for the current research was granted by The University of Birmingham Ethics Committee on the 14th October 2019 (See Appendix Nine). Consent was gathered from each participant after providing an on-screen participant information sheet. The information sheet informed participants that taking part in the study would involve watching a non-violent crime video. Therefore, participants could make an informed decision as to whether they would find this distressing. Materials such as videos and photographs used in this study have been used in previous research. Therefore, there were no anticipated risks to participant wellbeing. A non-violent crime was chosen to minimise the risk of harm to participants. Participants indicated consent by pressing the “continue” option on-screen after the participant information sheet.

All participants signed up to the study using their unique Amazon Mechanical Turk ID number. Participants were only identifiable by their Amazon Mechanical Turk ID numbers and no personally identifying details were accessible to the researcher. Participants were informed of this in the information sheet and were made aware that their ID numbers would not be included in write-up. The right to withdraw from the study was outlined in the initial information and debrief sheets. There were no consequences for withdrawing from the study and participants were informed their data would be destroyed once consent was withdrawn by using their unique Amazon Mechanical Turk ID. Participants were informed they would still receive payment for their participation even if they later withdrew from the study.

Only authorised researchers in the research team had access to participants’ unique Amazon Mechanical Turk ID numbers and data. The data will be stored until completion of the research study and thesis, for at least ten years thereafter, in accordance with the University of Birmingham’s Code of Practice for Research. The anonymised data will be made available to other researchers, in line with ethical approval, on the open science framework.

Results

The number of subjects in each of the eight conditions is displayed in Table 3.1. Recall that when presented with the lineup at test, participants either selected a suspect from the six photographs presented (first lineup selection), or selected "Not Present", which subsequently lead to a second forced choice lineup. Response frequencies for the perpetrator, misinformation suspect, filler and rejection (i.e., not present) decisions at each level of confidence for each condition are shown in Tables 3.2 and 3.3, for first lineup selection and second forced choice lineup selection, respectively. The overall incorrect ID rate of the innocent suspect (displayed in the proportion row in Table 3.2) is equal to the total number of innocent suspect IDs from the target-absent lineups divided by the total number of target-absent lineups for each facial angle condition. Similarly, the overall correct ID rate of the guilty suspect (also displayed in the proportion row in Table 3.2) is equal to the total number of perpetrator IDs from target-present lineups divided by the total number of target-present lineups for each facial angle condition.

The overall false ID rates of the innocent suspect (shown at the misinformation stage) when a selection was made during the first lineup (Table 3.2) were .12, .50, .14, and .49 for the FFF, PPP, FPF and PFP target-absent conditions. The corresponding overall correct ID rate of the guilty suspect (shown at the encoding stage) for the target-present conditions were .77, .61, .79. 65. For the second forced choice lineup (Table 3.3), the overall false ID rates of the innocent suspect were .34, .69, .46, and .76 for the FFF, PPP, FPF and PFP target-absent conditions. The corresponding overall correct ID rate of the guilty suspect for the target-present conditions were .66, .71, .75, .70. Further analyses were conducted to explore these results, considering discrimination accuracy.

Table 3.2: Frequencies of perpetrator, innocent suspect and filler identification decisions by pose condition for first lineup respondents.

	FFF						PPP						FPF						PFP					
	Target-Present			Target-absent			Target-Present			Target-absent			Target-Present			Target-absent			Target-Present			Target-absent		
Confidence rating	Perp	Filler	Reject	MI	Filler	Reject	Perp	Filler	Reject	MI	Filler	Reject	Perp	Filler	Reject	MI	Filler	Reject	Perp	Filler	Reject	MI	Filler	Reject
100	82	2	-	5	2	-	36	2	-	45	0	-	94	0	-	3	1	-	46	0	-	31	0	-
90	63	1	-	8	5	-	51	3	-	27	1	-	58	1	-	6	5	-	46	1	-	32	3	-
80	28	2	-	12	4	-	32	1	-	18	3	-	27	1	-	8	5	-	33	6	-	24	1	-
70	12	4	-	2	1	-	17	1	-	17	0	-	19	2	-	8	4	-	17	0	-	21	1	-
60	10	0	-	3	5	-	13	3	-	14	3	-	10	3	-	9	1	-	19	2	-	11	1	-
50	3	1	-	2	0	-	5	1	-	4	0	-	3	2	-	1	1	-	3	1	-	5	2	-
Total	198	10	50	32	17	220	154	11	86	125	7	119	211	9	48	35	17	200	164	10	77	124	8	119
Proportion	0.77	0.04	0.19	0.12	0.06	0.82	0.61	0.04	0.34	0.50	0.03	0.47	0.79	0.03	0.18	0.14	0.07	0.79	0.65	0.04	0.31	0.49	0.03	0.47

Table 3.3: Frequencies of perpetrator, innocent suspect and filler identification decisions by pose condition for second forced choice lineup respondents.

	FFF				PPP				FPF				PFP			
	Target-Present		Target-absent		Target-Present		Target-absent		Target-Present		Target-absent		Target-Present		Target-absent	
Confidence rating	Perp	Filler	MI	Filler	Perp	Filler	MI	Filler	Perp	Filler	MI	Filler	Perp	Filler	MI	Filler
-50	10	3	4	8	18	2	3	4	12	2	7	11	12	5	16	4
-60	5	2	3	10	8	2	2	3	10	2	9	10	14	2	12	3
-70	4	2	10	11	12	3	14	3	1	4	13	10	9	1	10	2
-80	6	3	9	18	5	3	11	3	2	0	18	15	8	6	9	4
-90	3	3	13	28	8	4	22	9	7	1	20	17	7	3	13	8
-100	5	4	35	71	10	11	30	15	4	3	24	46	4	6	30	8
Total	33	17	74	146	61	25	82	37	36	12	91	109	54	23	90	29
Proportion	0.66	0.34	0.34	0.66	0.71	0.29	0.69	0.31	0.75	0.25	0.46	0.55	0.70	0.30	0.76	0 24

Note: Each proportion is the proportion within each of the second forced choice lineups only.

ROC Analysis

Receiver Operating Characteristic (ROC) analysis was used to assess whether discrimination accuracy was higher when participants were exposed to different facial angles at encoding and the misinformation stage (facial angle congruency hypothesis) and whether discrimination accuracy was higher when participants viewed an innocent suspect from a profile facial angle at the misinformation stage when the encoding and test faces were frontal, compared to those who viewed the innocent suspect's face from a frontal angle when the guilty suspect's face at the encoding stage and test faces were profile (encoding strength hypothesis). To conduct ROC analysis, an ROC curve is produced. This curve represents the diagnostic ability of binary classifiers, where classifiers have four possible outcomes: true negative, false negative, true positive, false positive. Put more simply, ROC curves are useful for organising classifiers and visualising their performance. An example of a binary classification issue would include medical testing to ascertain whether patient has a specific disease or not. The true positive rate is then calculated and plotted against the false positive rate for a single classifier at different thresholds. Whilst historically ROC analysis has been used in medical decision-making research, this type of analysis has more recently been used in lineup identification research.

Previous lineup research relied on the diagnosticity ratio, which is the ratio of correct and incorrect identification probabilities (Rotello & Chen, 2016). This can be misleading because it is not possible to characterise the performance of a lineup procedure using a single diagnosticity ratio (Mickes et al., 2012). Rather, each lineup procedure yields a family of hit-rates (HR; the rate of correct guilty suspect identifications) and false alarm rates (FAR; the rate of incorrect innocent suspect identifications) and the analytical method used should take that into account. Moreover, evidence shows that the diagnosticity ratio increases as responding becomes more conservative. Therefore, this further supports the view that the diagnosticity ratio cannot determine which lineup procedure is diagnostically superior (i.e., which lineup procedure enhances the ability of witnesses to discriminate between innocent and guilty suspects; Clark et al., 2011). Finally, ROC analysis is better able to separately measure discrimination accuracy and response bias compared to alternative analysis techniques, such as logistic regression (Gronlund & Neuschatz, 2014). Discrimination accuracy can be defined as an eyewitnesses' ability to distinguish (or discriminate) the presence, or absence, of a guilty perpetrator. Moreover, response bias refers to the tendency to provide inaccurate or false answers, or in this case, inaccurate or false identifications. ROC analysis is a useful measure of lineup performance, because a lineup is a memory test to identify the presence or absence of a perpetrator (Mickes et al., 2012). ROC analysis is also useful for applied purposes because it provides an atheoretical measure of empirical discriminability (Wixted et al., 2017).

However, the benefits of using ROC analysis in eyewitness identification research are contended. Wells et al. (2015) argued that ROC analysis of lineups does not measure underlying discriminability or control for response bias. Although it is true that findings from empirical and underlying (theoretical) discriminability can diverge (e.g., see Rotello & Chen, 2016), the findings usually converge, and it is empirical discriminability (i.e., as measured by an ROC analysis) that is relevant for practice. This is because policy-makers are not concerned with theoretical debates, but with which lineup procedures result in the most accurate identification outcomes. ROC analysis has become a prominent analysis method in the field and studies have used ROC analysis to test theories of eyewitness memory (Humphries & Flowe, 2015) and differentiate between opposing theories of recognition memory (Hautus et al., 2008; Heathcote, 2003; Mickes et al., 2010). Moreover, the view that ROC does measure discriminability separately from response bias is misguided, and *p*AUC from ROC analysis has been used for decades in medical research to measure empirical discriminability separately from response bias (Alemayehu & Zou, 2012). In 2014 the National Academy of Sciences endorsed the use of ROC analysis for eyewitness research.

In the current study, the ROC curves were created by plotting the hit rate (HR) against the false alarm rate (FAR). That is, when witnesses correctly identified the perpetrator in target-present lineups (HR) or incorrectly identified an innocent suspect in target-absent lineups (FAR). Much previous lineup literature has plotted only positive IDs in ROC curves. Here, because participants in the study were forced to make an identification decision in the second forced choice lineup task, it was possible to extend the curves to contain negative IDs (second forced choice lineup selections). In order to plot the extended ROC curves, we took the six-point confidence scale from the first lineup selections (50%: guessing he is the culprit to 100%: certain he is the culprit) and the six-point confidence scale from the second, forced-choice lineup selections (-50%: guessing he is not the culprit to -100%: certain he is not the culprit) and combined them to create a single twelve-point scale (-100% to 100%). This followed a similar analysis procedure used by Colloff and Wixted (2020), where both partial and full ROCs were plotted. In both partial and full ROC analysis, the procedure with the ROC curve that falls furthest from the dashed line in the best at enhancing empirical discriminability (Colloff & Wixted, 2020).

To statistically compare ROC curves, pairwise comparisons between two conditions were made. To complete this pairwise comparison, the partial area under the curve (*p*AUC) was computed using the statistical package *p*ROC (Robin et al., 2011). The difference between the two *p*AUCs was then calculated and divided by the standard deviation of the difference estimated by bootstrapping, and therefore *D* is the measure of effect size. D is defined as (AUC1 – AUC2) / s, where s is the standard error of the difference between the two AUCs estimated by the bootstrap method, with the number of bootstraps set to 10,000 (Mickes et al., 2012). In a *p*AUC analysis, the specificity cut-off must be set in the analysis. In each set of analyses, a cut-off that was applied at the most liberal ROC point on the most conservative procedure.

As noted above, to increase the power of our analysis, “extended” ROCs were constructed that included both first lineup decisions (positive IDs where a face was selected) and second forced choice decisions (made after a negative “not present” decision), and the plan was to calculate the *p*AUC for the extended ROCs. However, when the extended ROCs were plotted, it was evident that the portion of the ROCs for the second forced choice lineup decisions were noisy. Previous research has found different results for positive and negative portions of ROCs (see Colloff et al., 2018; Colloff & Wixted, 2020). Therefore, for each research question, we plotted the extended ROCs (as we had initially planned) and also plotted the ROCs for the first lineup decisions only (i.e., the positive IDs, in the way that has typically been done in the lineup literature). For each research question, we present the *p*AUC results for extended ROCs that contain the positive and negative IDs (following our preregistered plan) and then the *p*AUC results for the positive IDs in the first lineups.

Testing the Facial Angle Congruency Hypothesis

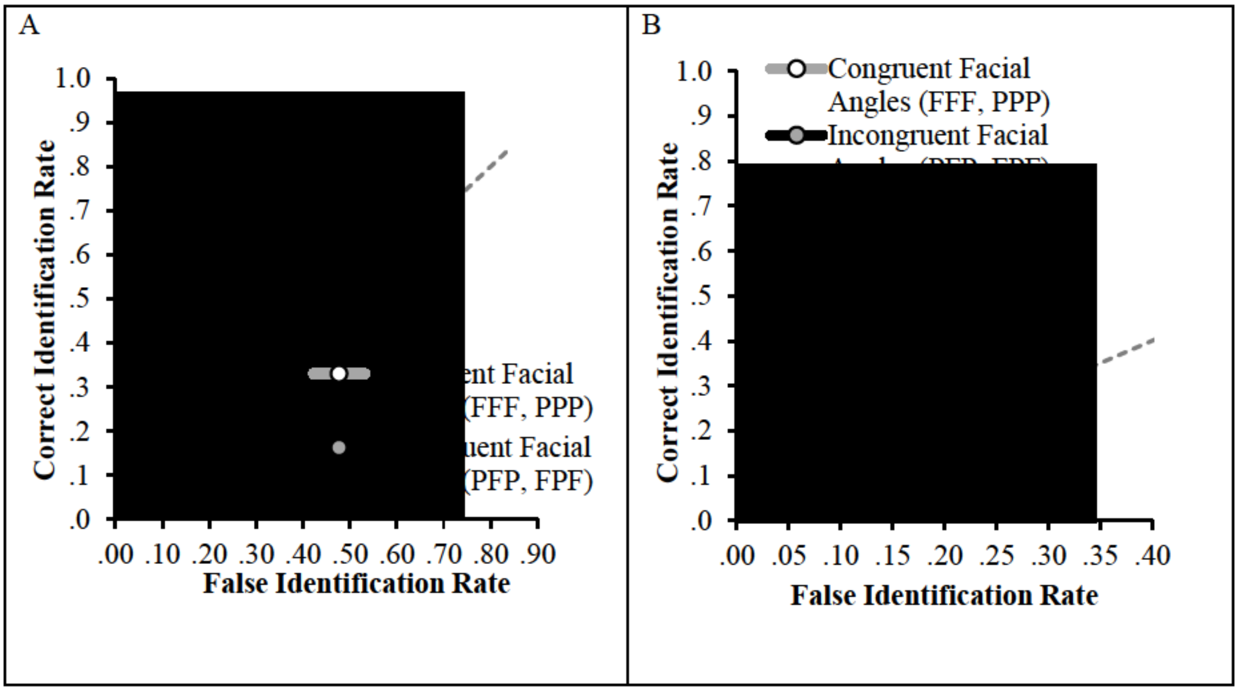

Figure 3.3. ROC data in the congruent facial angle (FFF, PPP) and incongruent facial angle (FPF, PFP) conditions for (A) positive IDs and negative ID decisions (extended ROCs) and (B) positive ID decisions only. The circular icons represent the empirical data. The dashed line indicates chance-level performance.

First, we asked if discrimination accuracy would be higher when an innocent suspect's face is presented in a different pose at the misinformation stage compared to the same facial angle presented during the to be remembered event. To answer that question, data were collapsed across the congruent facial angle (i.e., FFF and PPP, n= 1029) and incongruent facial angle (i.e., FPF and PFP, n= 1022) conditions. The extended ROCs are displayed in Figure 3.3A. Although the incongruent facial angle condition yielded a slightly higher pAUC (.377, 95% CI [0.358-0.402]) compared to the congruent condition (.362, 95% CI [0.336-0.387]), this difference was not statistically significant, D= 0.78, p= .44 (specificity cut-off of 0.60).

The same was true for the positive portion of the ROC (Figure 3.3B; specificity cut-off of 0.30), where once again the incongruent facial angle condition yielded a slightly higher pAUC (.131, 95% CI [0.114-0.148]) compared to the congruent condition (.114, 95% CI

[0.096-0.131]), but this difference was not statistically significant, *D*= 1.33, *p*= .19. Together, this suggests that, contrary to the facial angle congruency hypothesis, discrimination accuracy was similar, regardless of facial angle congruency.

Testing the Encoding Strength Hypothesis

Second, we asked if the MI effect was larger in the congruent condition in the profile compared to the frontal encoding condition. Put another way, we investigated whether participants were more likely to mistake the innocent person as the perpetrator when the perpetrator's face was seen from a frontal angle, relative to profile angle. To answer that question, we compared the ROC curves for the FPF and PFP conditions (see Figure 3.4).

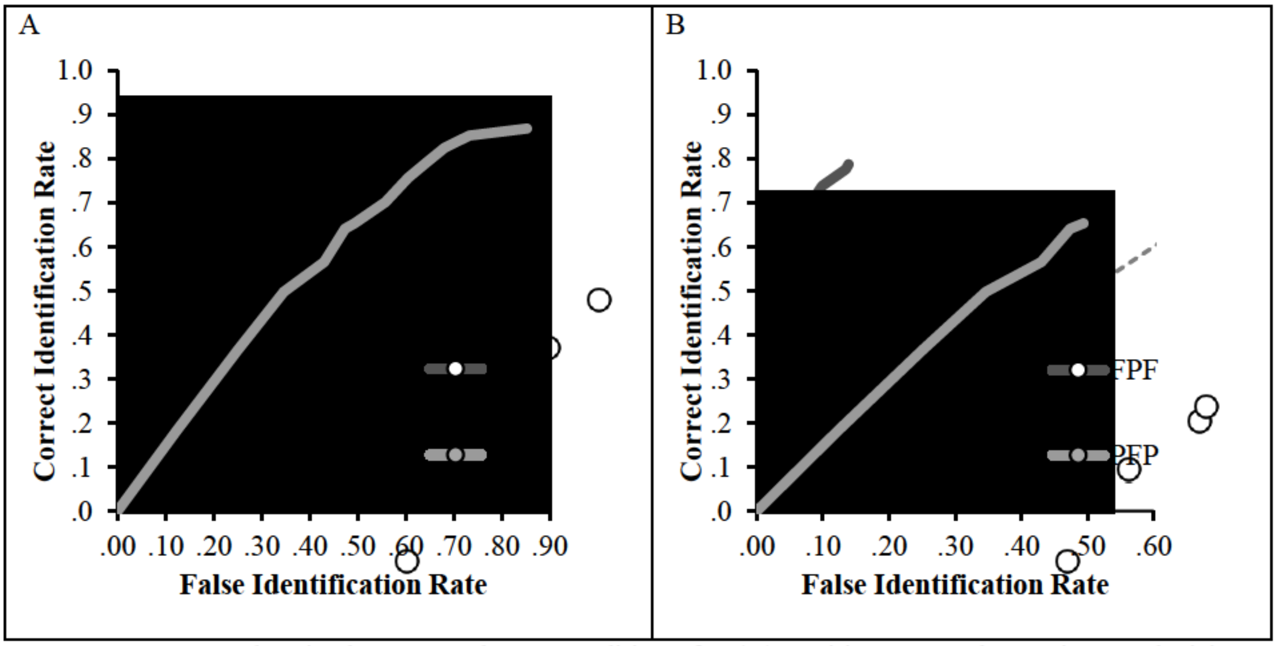

Figure 3.4. ROC data in the FPF and PFP conditions for (A) positive IDs and negative ID decisions (extended ROCs) and (B) positive ID decisions only. The circular icons represent the empirical data. The dashed line indicates chance-level performance.

Considering the extended ROCs plotted in Figure 3.4A. Those in the FPF condition discriminated the guilty suspect from innocent suspects better than those in the PFP condition. The *p*AUC for the FPF condition (.404, 95% CI [0.381-0.426]) was significantly greater than the *p*AUC for the PFP condition (.176, 95% CI [0.143-0.209]), D= 10.97, p< .001 (specificity cut-off of .50).

The same was true for the positive portion of the ROC (Figure 3.4B). The *p*AUC for the FPF condition (.101, 95% CI [0.088-0.112]) was significantly greater than the *p*AUC for the PFP condition (.015, 95% CI [0.010-0.021]; D= 11.97, p< .001 (specificity cut-off of 0.14). Together, this suggests that discrimination accuracy was significantly higher in the FPF condition compared to the PFP condition. This is consistent with the encoding strength hypothesis, as performance was better when the face was encoded and tested in frontal view compared to profile view.

However, it is possible that better performance in the FPF condition compared to the PFP condition result may be partially explained by the facial angle shown at the misinformation stage. That is, seeing a frontal face during the misinformation stage may result in poorer performance when the perpetrator's face at encoding is shown from a profile facial angle (i.e., in the PFP condition), compared to when the innocent suspect was shown from a profile facial angle during the misinformation stage when the perpetrator's face was presented from a frontal facial angle at encoding (i.e., in the FPF condition). To further explore this, ROC curves for each of the facial angle conditions (i.e., FFF, PPP, FPF, PFP) were plotted on a single plot. Figure 3.5 shows the ROC curves for the FFF, PPP, FPF and PFP conditions.

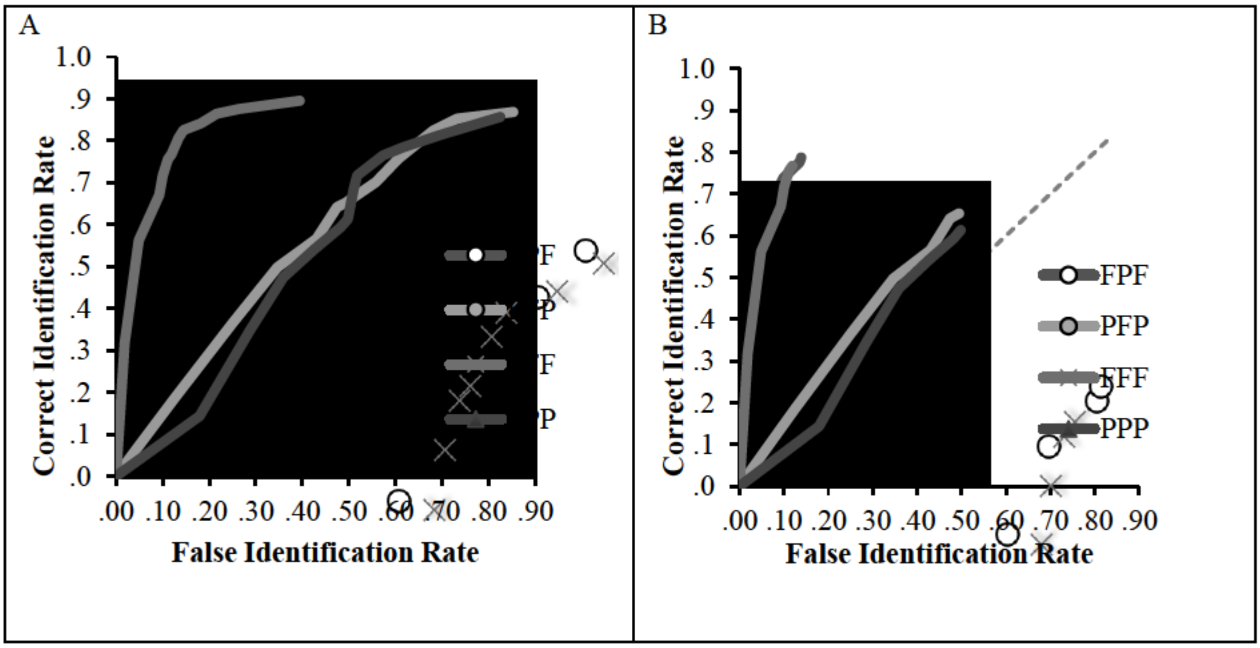

Figure 3.5. ROC data in the FFF, PPP, FPF and PFP conditions for (A) positive IDs and negative ID decisions (extended ROCs) and (B) positive ID decisions only. The circular icons represent the empirical data. The dashed line indicates chance-level performance.

Considering the extended ROCs (3.5A, specificity .39), as predicted by the encoding strength hypothesis, the *p*AUC for the FFF condition (.297, 95% CI [0.275-0.317]) was significantly higher than that for the PPP condition (.084, 95% CI [0.06-0.109]), $D= 12.73$, $p< .001$, and the PFP condition (.110, 95% CI [0.086-0.138]), $D= 11.39$, $p< .001$. The *p*AUC for the FPF (.303, 95% CI [0.282-0.323]) condition was significantly higher than the *p*AUC for the PPP condition (.084), *D*= 13.48, *p*<.001. No significant differences were found between the *p*AUC for PPP (.084) and PFP (.110) conditions, *D*= 1.45, *p*=.15, or between the FFF (.297) and FPF (.303) conditions, *D*=.43, *p*=.67. For completeness, the *p*AUC for the FPF and PFP conditions were calculated again during this analysis using the new specificity cut-off. Again, the *p*AUC for the FPF condition (.303) was significantly higher than that for the PFP condition (.110), *D*= 11.97, *p*< .001. This indicates that discrimination accuracy was significantly higher when participants were exposed to a frontal face at encoding and test compared to if they were exposed to a profile face at encoding and test. This suggests that the difference between the FPF and PFP in the previous analysis was due to a benefit effect of viewing frontal faces and encoding and test, rather than a detrimental effect of viewing frontal faces at the misinformation stage.

However, it should be noted that the results differed slightly for the positive portion of the ROC (Figure 3.5B; specificity .12). Unlike in the extended ROC analysis, the *p*AUC for the FPF condition (.083, 95% CI [0.071-0.094]) was significantly higher than the FFF condition (.063, 95% CI [0.051-0.076), *D*= 2.20, *p*= .01. Similar to the extended ROC analysis, the *p*AUC for the PFP condition (.169, 95% CI [0.136-0.200]) was not significantly higher than *p*AUC for the PPP condition (.140, 95% CI [0.110-0.173]), *D*= 1.29, *p*= .20. This suggests that discrimination accuracy was significantly higher when participants were exposed to the incongruent frontal encoding conditions (FPF) compared to the congruent frontal encoding conditions (FFF), but only for those who made IDs in the first lineup.

Confidence-Accuracy Characteristic (CAC) Analysis

The relationship between confidence and accuracy was also explored in the current study. Mickes (2015) recommended using CAC analysis to assess the reliability of IDs.

Reliability refers to the probability than an identification made with a certain level of confidence is correct (Mickes, 2015). This differs to discriminability, which can be defined as people's collective ability to tell the difference between innocent and guilty suspects (Colloff et al., 2020). The link between high confidence ratings taken at the time of the identification and accurate lineup IDs has been well documented in recent research (Kebbell et al., 1996; Seale-Carlisle et al., 2019; Wixted et al., 2015; Wixted & Wells, 2017). Yet, there is a dearth of research looking at CACs for misinformation studies.

CAC analysis consists of plotting identification accuracy of suspect IDs (ignoring fillers IDs) for each level of confidence. For a six-person lineup procedure, CAC is given by;

$$CAC = \frac{CID_{conf}}{CID_{conf} + FID_{conf}}$$

CID*conf* is the number of correct guilty suspect IDs made with each level of confidence from target-present lineups. Alternatively, FID*conf* is the number of false IDs of innocent suspects made with that same level of confidence from the target-absent lineups (Mickes, 2015; Seale-Carlisle et al., 2019). In this study, confidence ratings were binned into four levels of confidence: -100 to -80 and -70 to -50 (for the forced-choice lineup decisions, or negative IDs), and 50-70 and 80-100 (for the first lineup decisions, or positive IDs). Unlike ROC analysis, the goal of CAC is to measure the relationship between confidence and accuracy (Mickes, 2015). Therefore, accuracy is plotted on the *y*-axis and confidence is plotted on the *x*-axis. This is useful from a practical standpoint, whereby the legal system is most interested in knowing the probability that a suspect who has been identified is actually guilty (Wilson et al., 2018).

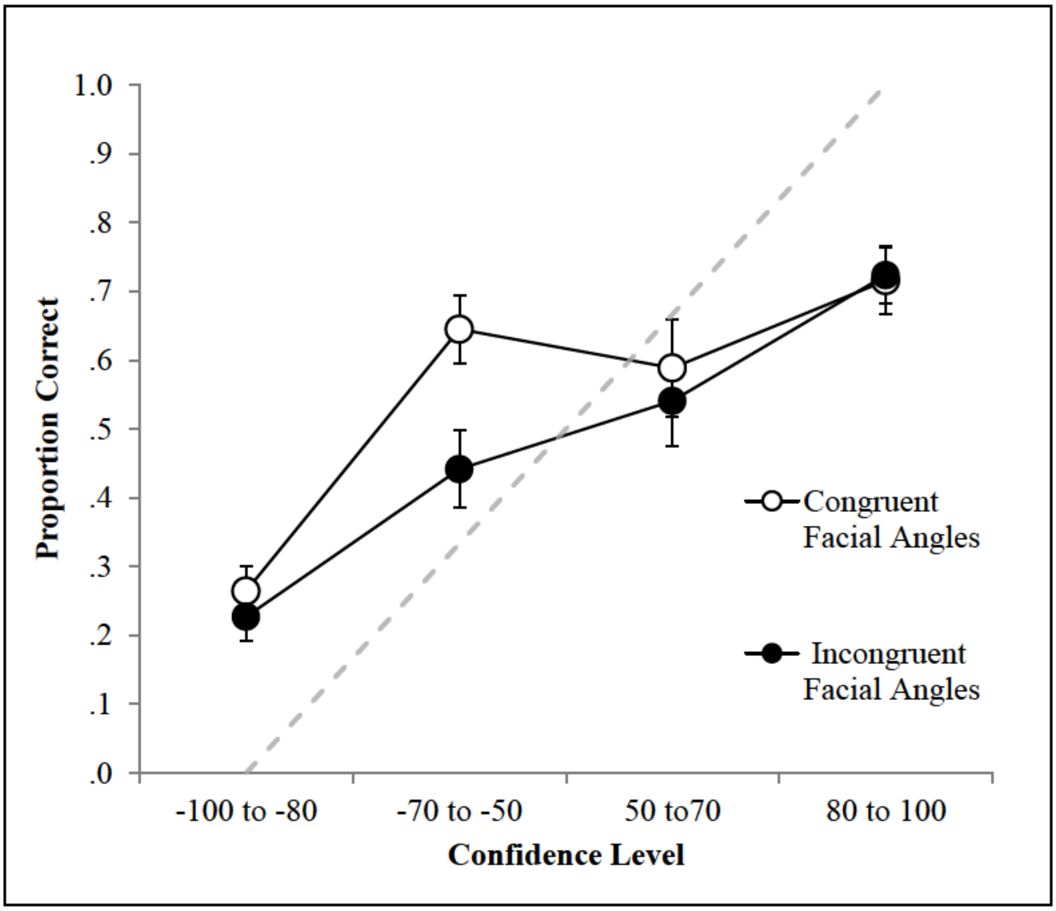

Figure 3.6. CAC data for the facial angle congruent (FFF, PPP) and incongruent (PFP, FPF) conditions for first- and second-line up decisions. The circular icons represent the empirical data. The dashed line indicates chance-level performance. The error bars also represent the standard error.

First, CAC curves were plotted for the congruent facial angle (FFF, PPP) and incongruent facial angle (FPF, PFP) conditions. Figure 3.6 shows that there appeared to be a relationship between confidence and ID accuracy in both conditions, because, generally speaking, as accuracy increased, so did confidence. However, the relationship was stronger in the incongruent facial angle conditions. In the congruent facial angle conditions, there was a relationship within the negative IDs (i.e., -70 to -50 yielded a higher proportion correct than -100 to -80) and within the positive IDs (i.e., 80 to 100 yielded a higher proportion correct than 50 to 70) but, for some reason, IDs made with a confidence rating of 50 to 70 were less accurate than those made with -70 to -50. For both conditions, it is important to note that high confidence did not indicate high accuracy, as participant were overconfident at high confidence. Participants who made 80-100% confidence judgements where only approximately 70% accurate in their suspect IDs. This is likely due to the deleterious effect of MI.

To further explore the relationship between confidence and accuracy in the frontal and profile facial angle encoding conditions, all four conditions were plotted for the CAC analysis. Figure 3.7 shows the CAC analysis for the FFF, FPF, PPP and PFP conditions.

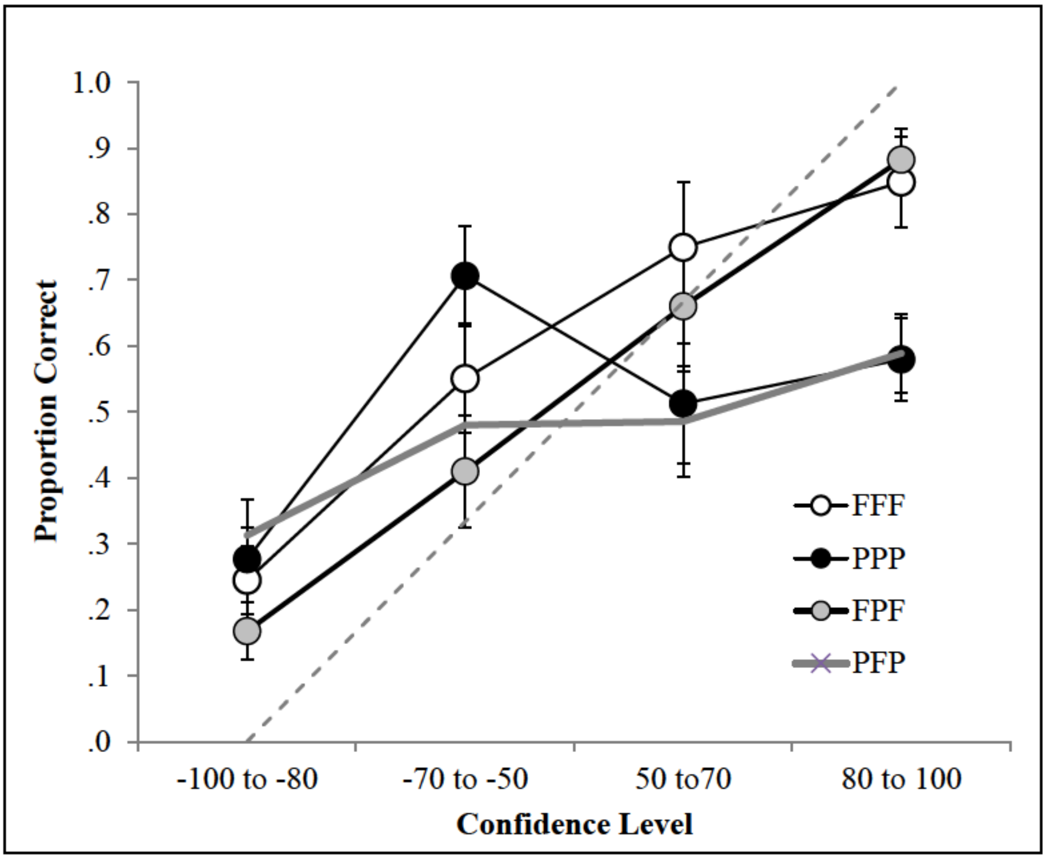

Figure 3.7. CAC data for the all four facial angle (FFF, PPP, FPF, PFP) conditions for first- and second-line up decisions. The circular icons represent the empirical data. The dashed line indicates chance-level performance.

For the frontal encoding and test conditions (FFF and FPF) there was a relationship between confidence and accuracy, because as confidence increased, so did accuracy. IDs made with high confidence (i.e., 80%-100% confidence rating) were also higher in accuracy (around 80% accurate) in the frontal facial angle encoding conditions compared to the profile facial angle encoding conditions. For the profile encoding conditions (PPP and PFP), there appeared to be a weaker relationship between confidence and accuracy. Moreover, participants were overconfident that they had identified the perpetrator when they provided high confidence ratings; they were only approximately 55% accurate when they were 80 to 100 confident.

Discussion

The current research explored the impact of facial positioning on misinformation susceptibility. It was hypothesised based on previous research regarding the strength of frontal face encoding, that frontal encoding would enable better discrimination accuracy compared to profile encoding. This encoding strength hypothesis was supported because performance was generally better when the encoded face was front facing compared to profile. This suggests that frontal face encoding and test is superior in memory to profile face encoding and test. It was also hypothesised that facial angle congruence between post-event misinformation and test would decrease discrimination accuracy compared to facial angle incongruence. The facial angle congruence hypothesis was not supported, as there was no significant difference between congruent and incongruent facial angle conditions. This suggests that participants were no more likely to be misinformed if the misinformation was more similar to encoding and test, compared to when the misinformation was more different to encoding and test.

The prediction that discrimination accuracy would be better when participants are exposed to different-pose misinformation (incongruent facial angle) compared to same-pose misinformation (congruent facial angle) was not met. This indicated that the viewing angle of a misinformation suspect's face is not likely to impact on the encoding specificity between encoding and test. One explanation could be that because the facial angle at encoding and test were always congruent, this may have had a stronger impact compared to congruent facial angles at misinformation and test stages. That is, matching the context at misinformation and test is less problematic for discrimination accuracy, so long as the test context remains the same as that experienced at encoding. This supports previous research by Bruce (1982), who found that when individuals learned a frontal face and were subsequently tested with a frontal face, they were able to recognise faces more accurately and quickly compared to when they were tested with faces posed a 45° angle (profile). Although the prediction was not met, a dearth of previous research has fully explored congruent and incongruent facial angles at different stages of the MI paradigm. Therefore, this finding has contributed to the growing understanding of facial manipulations in the MI paradigm.

An additional encoding strength hypothesis proposed that discrimination accuracy would be higher when participants were presented with a profile facing misinformation suspect when the encoding and test faces are frontal (FPF), compared to when participants view a frontal misinformation face when the encoding and test faces are profile (PFP). This hypothesis was supported, as discrimination accuracy was better in the FPF condition than the PFP condition. To explain this result, we initially proposed that frontal posed misinformation would result in a stronger memory trace when encoding and test were profile, compared to profile posed misinformation when encoding and test were frontal. That is, it is possible that the finding discrimination accuracy was better in the FPF condition than the PFP condition may be partially explained by the strength of the facial angle at the misinformation stage. Put another way, discrimination accuracy in the PFP condition may have been lower due to the stronger encoding of the front facing misinformation, opposed to the profile facing encoding and test stages. Likewise, higher discrimination accuracy found in the FPF condition may be because profile misinformation would not have had the same encoding strength as the original front facing perpetrator, making it easier for participants to discriminate between faces. This would support previous research, whereby frontal faces have been considered to provide more information than a profile face (McKelvie, 1976), thus leaving a stronger memory trace (Fraser, 1990; Meltzer & Bartlett, 2019).

However, our further analyses suggest this is not the case. Further support for a front face encoding benefit was evident when comparing the four facial angle combinations, whereby a frontal encoding (FFF and FPF) benefit over profile encoding (PPP and PFP) was observed. This difference cannot be explained by differences of facial angle at the misinformation stage, and instead must be explained by difference of facial angle at encoding (and test). Together, the findings support the encoding strength hypothesis and previous face memory literature, where frontal face encoding is argued to be superior to other poses (Colloff et al., 2021). This supports the holistic encoding hypothesis, which suggests that instead of processing faces as a collection of separate, distinct, facial features, we instead process the face as a perceptual whole (Taubert et al., 2011). Therefore, seeing a criminal's face from a frontal view at encoding and test means participants can engage in holistic facial encoding and recognition. We also know that a frontal face provides more perceptual information than a profile face (Meltzer & Bartlett, 2019) and that this perceptual information can be beneficial for facial recognition.

For the majority of the findings, the ROC analysis of the positive lineup IDs replicated the findings from the extended ROC IDs. However, when results for the partial positive portion of the curve were calculated for the FFF and FPF condition, discrimination accuracy was significantly higher in the FPF condition compared to the FFF condition (p=.03). This significant difference was not observed in the extended ROC analysis. A possible explanation for the significant finding is that the congruence between encoding, misinformation and test in the FFF condition may have made it more difficult for participants to discriminate between the guilty suspect and the innocent misinformation suspect than the FPF. This would, in part, support the proposed facial angle congruence hypothesis. But if that were true, it is not clear why the same pattern of results was not observed in the profile encoding conditions (i.e., no significant difference between PPP and PFP), or on the extended ROC. What we do know is that the analysis found significantly better discrimination accuracy in the frontal encoding conditions compared to the profile encoding conditions. One reason this finding may not have been observed in the profile encoding condition is due to the overall poor discrimination accuracy in the PPP and PFP conditions, where discrimination accuracy was only marginally better than chance. Moreover, other research has found the predicted pattern of results only in the positive IDs and not the negative IDs (see Colloff et al., 2018; Colloff & Wixted, 2020), but it is not yet clear why that is the case. Nevertheless, because this result was only found in front encoding conditions (i.e., FFF, FPF), but not profile encoding conditions (i.e., PPP, PFP), and was only observed in the positive ID ROC and not the extended ROC, the significant result should be interpreted with caution and further research is needed.

The results of this study also have practical implications. Appeals have been made in practice due to the angle of a face that may have been seen between encoding and the lineup test. The case of Ted Bundy is a prominent example. Recall that Ted Bundy's prosecution team argued that the victim had been misinformed by a newspaper photograph of Bundy prior to the lineup selection. The current findings do not appear to mirror the witness's claims regarding her own misinformation susceptibility. That is, she argued she could not have been misinformed by seeing Ted Bundy's front facing photograph in the newspaper, because during the offence she only saw him from a profile view. However, we found that the angle of the misinformation (congruent or incongruent with study at test) was not an important determinant of identification accuracy. Instead, we found that when the encoding face was presented from a profile view discrimination accuracy was significantly poorer than when the encoding face was presented from the front. The witness in the Ted Bundy case did encode Ted Bundy from the profile view. Whilst it is highly probable that she has correctly identified Bundy (due to the evidence against him and subsequent conviction), the lower discrimination accuracy results for profile encoding in the current study cannot be ignored. This has important implications for future legal arguments regarding facial misinformation and the awareness of investigators to the apparent limitations of witnessing a perpetrator from a profile view.

Moreover, the results suggest that witnesses who have encoded perpetrators from profile view may be less reliable because they were found to have lower accuracy at high-confidence and have a poorer confidence-accuracy relationship than witnesses who have encoded perpetrators from the front. One explanation for this is that because the discrimination performance was so low in the PPP and PFP conditions (only marginally higher than chance), this impacted participant's ability to assign appropriate confidence ratings. The poor confidence-accuracy relationship in the PPP and PFP conditions are consistent with previous research that found a poor confidence-accuracy relationship when memory accuracy is below chance (see Nguyen et al., 2017; Weber & Brewer, 2003). Theoretically, participants who are guessing should not be more confident in their guess that resulted in a correct identification than a guess that resulted in an incorrect identification (Nguyen et al., 2017). Furthermore, participants who are guessing (i.e., whose memory signal is weak) would have more relaxed criterion for identifying faces. Therefore, they are predicted to be less confident in their responses than participants who make recognition judgements based on more information in memory (i.e., stronger feelings of familiarity with a face). This suggests that accuracy is more likely to fluctuate around chance levels at lower levels of confidence.

The differences between front and profile encoding suggests that legal decision makers should be wary of IDs made when witnesses have encoded the perpetrator from the profile. This is because participants were making high confidence IDs, with considerably low accuracy rates. Court systems may not always consider confidence when evaluating eyewitness IDs (Juslin et al., 1996). It can be argued that the reason for this is because confidence ratings are susceptible to interference. For example, other research has found that poor confidence rating reliability has also been associated with conformity to misinformation, whereby participants are misled but still provide high confidence ratings (Foster et al., 2012; Mudd & Govern, 2004; Spearing & Wade, 2021). This suggests that practitioners should be wary of high-confidence IDs or memory reports when it is possible that witnesses have been exposed to post-event misleading information.

In considering these findings, it is important to note a methodological limitation of the current research. Given that participants were always exposed to the same facial position at encoding and test, this research has not considered the potential influence that incongruent facial angles between encoding and test in the MI paradigm may have on misinformation susceptibility. Previous research suggests that people will be slower to recognise a face and less accurate in their recognition if the viewing angle of a face is changed (for example, between front facing and ¾ facing) between initial presentation and test compared to when it remains unchanged (Bruce, 1982). However, it is noted this finding has not been explicitly explored in the MI paradigm. Likewise, the full impact of facial viewing angle manipulations across the three stages of the MI paradigm have not been explored in this single study. It will be important for future research to explore how further facial manipulations at test could impact misinformation susceptibility.

Moreover, it might be useful for future research to consider whether a combined lineup procedure would have implications for these findings. That is, the lineup procedure at test could contain both the guilty suspect and misinformation suspect amongst fillers in a single lineup. A similar procedure has been used by some police departments, whereby everyone in the lineup in suspected of being the person (all-suspect design) who committed the offence (Wells & Luus, 1990). Whilst this lineup design has been used in forensic contexts, it is certainly not the norm and it would be unusual to have multiple suspects (i.e., one guilty and one innocent) in a single lineup. However, Loftus' (1975, 1979a) overwriting memory theory offers insight into why this may be theoretically interesting for future research to consider. Putting the guilty suspect and the misinformation suspect in the same lineup allows the researcher to explicitly see if the misinformation suspect is chosen instead of the guilty suspect. If Loftus' (1975, 1979a) overwriting theory is supported, then the original memory for the guilty suspect would no longer be available, instead updated with the misinformation (Loftus et al., 1992). Therefore, presenting the guilty and misinformation suspects simultaneously in a lineup would test this theory. Although it may be interesting for future research to explore this different method, the current research has reflected a more ecologically valid approach by presenting the guilty suspect and misinformation suspect in different lineups.

Conclusion

The impact of facial angle congruence across encoding, misinformation presentation, and test on the MI effect was explored. Participants were not differentially likely to be misled by an innocent suspect depending on congruency across the angles in which the faces were shown across encoding, the misinformation, and lineup phases. This suggests that participants are no less likely to be misled if the innocent suspect's face is presented in the same as opposed to different angle across encoding, misinformation, and test. Discrimination accuracy was significantly higher when the participants encoded the perpetrator from the front compared to the profile angle, suggesting memory is stronger for faces that are originally encoded in frontal view. ROC analysis for all four conditions (FFF, PPP, FPF, PFP) also supported the encoding benefit of encoding a face from the front compared to the profile. Moreover, CAC analysis revealed a weak relationship between confidence and accuracy in the profile encoding (PPP and PFP) conditions compared to a stronger relationship in the frontal encoding (FFF and FPF) conditions. Given that legal decision makers rely heavily on eyewitness confidence in court (Garrett et al., 2020), they should be aware that the reliability of eyewitness identifications could be impaired when a witness has encoded a perpetrator from a profile posed face compared to when the face is encoded from the front.

It would be interesting for future research to examine the impact of the viewing angle of faces at test to further isolate the impact of congruent facial angles at encoding and the misinformation stage. This may provide further insight into how the superior encoding strength of information, such as a frontal angled face, may lead individuals to accept misinformation. Findings from this research therefore illustrate the significance of considering the viewing angle of faces and the importance of accounting for effects in future misinformation research.

Chapter Four: Psychometric Critique

The Gudjonsson Suggestibility Scale:

A Psychometric Test Critique

Abstract

Interrogative suggestibility can be defined as the extent to which subjects are susceptible to altering their memory reports during questioning (Thorley, 2011). Therefore, interrogative suggestibility is considered a unique and distinct form of suggestibility (Lee, 2004). The majority of suggestibility studies have focused on the effects of misleading questions (Eisen & Carlson, 1998) and interrogative suggestibility (Gudjonsson, 1988; Roebers & Schneider, 2001). However, Lee (2004) argued that while there is the potential that findings from these studies could be generalised to the MI effect, differences between the MI effect and interrogative suggestibility are noted throughout literature. This chapter will review and critique how interrogative suggestibility is measured. At present, there is only one validated measure of interrogative suggestibility for adults. Therefore, this critique will focus on the reliability, validity, and practical application of the Gudjonsson Suggestibility Scale (GSS). This review finds overall support for the reliability and validity of the GSS, as well as for its use in practical settings.

Introduction

The importance of testimony from suspects, victims and eyewitnesses in criminal justice proceedings is profound. Accurate memory reports can help to ensure justice is served, whereas inaccurate or false memories can result in the innocent being found guilty, and the guilty being found innocent. However, experimental research has shown that people can be susceptible to post-event misleading information. Loftus (1979a) used an experimental MI paradigm to demonstrate that subjects could be easily led to remember incorrect information regarding an event if they are exposed to post-event misinformation (Loftus et al., 1978; Loftus & Palmer, 1974). Research has shown that when witnesses are exposed to leading questions during interview, they can come to report inaccurate information (Loftus & Zanni, 1975; Schooler & Loftus, 1986). An important scenario where these errors can occur is during police interviews, where a suspect, victim, or witness may be questioned. Interrogation techniques specifically aimed at manipulating confidence and self-esteem (such as the REID method in the US) may result in some subjects becoming especially vulnerable to suggestive influences during interrogative processes (Gudjonsson & Lister, 1984; Hooper et al., 2016). For example, the REID technique involves offering two choices for what happened, one being more socially acceptable than the other. This technique is specifically designed to manipulate the suspect into selecting the more socially desirable scenario. Although the REID technique is not used in the UK, subjects can still be vulnerable to suggestion in the form of leading questions during an interview.

Loftus (1979a) developed experimental procedures to measure individual responses to leading questions (which Gudjonsson termed “yield”), however Gudjonsson (1997) argued the experimental methods were unsatisfactory for use in practical forensic contexts. Specifically, he argued that experimental methods do not consider individual vulnerability to interrogative processes (which Gudjonsson termed “shift”). Furthermore, Gudjonsson noted that a significant body of suggestibility research focused solely on hypnotic suggestibility, which is a subject's tendency to respond to hypnosis and hypnotic suggestion (Dienes et al., 2009). He argued measures of hypnotic suggestibility were insufficient, due to being poorly correlated, unreliable, and unable to further the understanding of interrogative suggestibility (Gudjonsson, 1984). Gudjonsson and Clark (1986) highlighted the importance of ascertaining individual levels of suggestibility from a quantitative, objective viewpoint. This is so that witnesses who are more susceptible to giving false accounts can be identified, and the negative impact of their account mitigated. Therefore, Gudjonsson argued that a tool was needed to measure individual vulnerability to interrogative pressure. Accordingly, the Gudjonsson Suggestibility Scale (GSS) was developed (Gudjonsson 1984, 1987a).

The GSS comprises two forms, the GSS-1 and the GSS-2 (Gudjonsson, 1984, 1987a). The GSS-1 was developed as a psychological test to measure the interrogative suggestibility of a subject and, specifically, to assess the reliability of retracted confessions (Gudjonsson, 1984). During administration of the GSS subjects are read a short story followed by free recall, where subjects are asked to report everything they remember from the story. Subjects are then asked a series of questions regarding the story, including leading questions. Subjects are then told they have made errors in their answers, and must answer the questions a second time. Subjects are scored on their memory recall, which refers to the number of facts the subject correctly remembered during free recall. They are also scored on yield, which refers to the number of suggestive questions they answer correctly; shift, which refers to any notable changes in the subject's answers after they were told they had made errors; and total suggestibility, which refers to the sum of both yield and shift scores. The GSS-2 was developed by Gudjonsson (1987a) after it was proving difficult to examine the test-retest reliability of the GSS-1 because of the subjects' residual memory of the first administration of the tool at the second testing. Therefore, the GSS-2 is almost identical to the GSS-1, apart from the story and interrogative questions used.

Gudjonsson and Clark (1986) define interrogative suggestibility as the extent to which, within a closed social interaction, subjects come to accept messages communicated during formal questioning, as a result of which their subsequent behavioural response is affected. Gudjonsson (1987b) postulated that interrogative suggestibility comprises of three distinctive features: (1) it involves a questioning procedure within a closed social interaction; (2) the questioning is primarily concerned with past experiences and events; and (3) it has a strong element of "uncertainty" which relates to the cognitive processing capacity and functioning of a subject. Researchers have argued that interrogative suggestibility bears little resemblance to other types of suggestibility (Raymond, 2020). Unlike previous theories of hypnotic suggestibility, interrogative suggestibility focuses on recollections of past events rather than sensory experiences (Raymond, 2020).

Raymond's (2020) emphasis on the link between interrogative suggestibility and recollections of past events has some resemblance to Loftus' conceptualisation of suggestibility. Both propose that in real-life contexts, multiple environmental and internal stimuli (e.g., heightened arousal) compete for attentional resources (Nichols & Loftus, 2019; Zaragoza et al., 2007). However, Loftus' conceptualisation of suggestibility does differ to that of interrogative suggestibility. In Loftus' MI paradigm, suggestions are content specific, relating to observed items or details. In an interrogative suggestibility paradigm, two types of suggestions are generally used: misleading questions and critical feedback. Although evidence shows that the MI effect is sensitive to social demands (Ceci et al., 1987), the introduction of feedback in interrogative paradigms is likely to result in subject responses being more sensitive to interpersonal pressure (Gudjonsson, 1987a; Lee, 2004).

The GSS is widely used, namely during criminal justice proceedings and the preparation of court reports, to obtain information regarding memory, suggestibility and confabulation in suspects, victims and witnesses (Clare et al., 1994). The GSS scale can be of use in situations where individuals are being questioned by police to provide insight into their vulnerability to suggestion (Willner, 2011). The GSS is also widely used in research to investigate the influence of interrogative pressure and suggestibility. Because of its widespread use, it is important to consider if the scale is an appropriate, reliable and valid measure of interrogative suggestibility by reviewing the relevant literature. It is important to consider normative data that supports the use of the GSS in forensic settings and research.

Overview

The Gudjonsson and Clark Model

The theoretical approach underpinning the GSS differs from that used in traditional MI paradigm studies introduced by Loftus and colleagues. That is, Gudjonsson and Clark (1986) take an individual differences approach within their integrated model of interrogative suggestibility, whereby Loftus and colleagues take an experimental approach (Loftus et al., 1978). A key way in which the experimental approach differs from the individual differences approach is that the experimental approach emphasises gaining an understanding of the conditions that increase a subject's vulnerability to leading questions, regardless of an eyewitness' individual characteristics. In contrast, the individual differences approach focuses more on the independent coping strategies that subjects generate and implement when dealing with interrogative suggestion. Despite this difference, several studies have examined the impact of individual differences in susceptibility to misinformation using the MI paradigm (Lee, 2004;

Zhu et al., 2010). This suggests that both the experimental and individual approaches can be considered in unison.

Based on Gudjonsson and Clark's (1986) theoretical model of interrogative suggestibility, two forms of suggestibility were identified. The first was yield, referring to the propensity of subjects to concede to leading questions (Binet, 1900; Stern, 1938, 1939). The second was shift, which refers to any significant change in a subject's answer after they have been exposed to negative feedback from the administrator (Gudjonsson, 1983). For example, a subject may initially report there were three perpetrators involved in a crime. However, after receiving feedback from the administrator that their initial report contained errors, the subject may then change their report to state there were only two perpetrators, even if this was not the case. By measuring yield and shift scores, the administrator is able to determine the degree to which a subject yields to leading or misleading questioning and gives in to the interrogative pressure and negative feedback during interview.

The Gudjonsson Suggestibility Scale 1 (GSS-1, Gudjonsson, 1984)

The GSS comprises two parallel forms, the GSS-1 and the GSS-2. Gudjonsson (1987a) described the forms as parallel because the GSS-1 and GSS-2 are largely identical, apart from differences in the content of the narrative and interrogative questions. By creating a parallel form, Gudjonsson (1987a) was able to research the test-retest reliability of suggestibility. The original version of the GSS, known as the GSS-1 (Gudjonsson, 1984), was the first measure of interrogative suggestibility that was developed to assess a subject's vulnerability to suggestion. Gudjonsson developed the GSS-1 with the purpose of objectively measuring interrogative suggestibility from a consolidative socio-cognitive approach (Gudjonsson & Clark, 1984). That is, the proneness of subjects to accept suggestive influences that result in erroneous accounts as a result of both social and cognitive factors.

In developing the GSS-1, Gudjonsson decided to construct the scale where the content of the interrogation context was based on a story that could be presented verbally prior to interrogation. When administering the GSS, the subject is first introduced to a narrative describing a fictitious robbery, which is either read or played from a recording to the subject. The subject is then required to recall everything they can about the fictious robbery immediately after hearing the narrative and once again after a 50-minute time period, before initial questioning. The scoring aimed to ensure the subjects' memory was accurate prior to suggestive questioning, helping GSS administrators to identify when errors in memory are the result of suggestion (Gudjonsson, 1983). While the story contains believable details relevant to the time of its construction, certain elements of the story contents, such as "traveller's cheques", could be outdated for modern use. The story also contains British place names, something that may hold little relevance to subjects from other countries. This highlighted that an updated story may improve the measure for modern administration and for its use across other nations.

During initial questioning, the subject is asked 20 questions, with 15 of the questions considered as "suggestive", and 5 "true" questions where answers are affirmative. Gudjonsson distinguishes the suggestive questions into three categories: 1) leading questions, which are worded in such a way that they seem plausible ("Did the woman's glasses break during the struggle?"); 2) affirmative questions, that present facts that did not appear in the story, but did, however, prompt an affirmative response bias ("Were the assailants convicted six weeks after their arrest?"); and 3) false alternative questions, which imply the presence of persons, objects and events that were not present in the story ("Did the woman have one or two children?") (Gudjonsson, 1983).

Once answered, the interviewer informs the subject that they have made mistakes, regardless of whether this is true. The subject is then told that it is necessary to repeat the questioning and that they should try to be more accurate. Therefore, the 20 questions are repeated. The intent of this is to measure shift, by assessing the subject's ability to manage interrogative pressure in the form of negative feedback. Shift, or changes in the subject's answers, are then documented. The instructions given by the administrator follow Kelman's (1950) argument that feedback of failure will increase suggestibility (i.e., "You have made a number of errors, try to be more accurate this time"). Despite the development of suggestibility research since Kelman's (1950) conclusions, research still supports the idea that feedback about failure will increase suggestibility. For example, Tata and Gudjonsson (1990) found that negative feedback produced dramatic changes and erroneous accounts in a population of 40 male hospital employees. Further research has expanded these findings, arguing that negative feedback increases the subject's uncertainty to a greater extent for anxious persons than for control groups (Wolfradt & Meyer, 1998). The impact of negative feedback on suggestibility is supported in more recent literature (e.g., Biondi et al., 2020; Szpitalak & Polczyk, 2020). Therefore, the current research supports the instructions used in the GSS, as a way to increase susceptibility to suggestion.

The GSS produces four scores: the extent of suggestibility to misleading questions is scored as yield 1 (with a possible score range of 0 to 15, with higher scores indicating higher levels of suggestibility). Yield 2 is the measure of vulnerability to misleading questions following negative feedback (with a score range of 0 to 15, with higher scores indicating increased levels of vulnerability). Changes in response to the 20 questions, after they are administered for the second time, produce a shift score (with a score range of 0 to 20, with a point given every time a subject shifts from their initial response (i.e., from "yes" to "no"). Yield 1 and shift scores are combined to create a total suggestibility score, which produces a possible maximum combined score of 35 (Drake et al., 2008).

The Gudjonsson Suggestibility Scale 2 (GSS-2, Gudjonsson, 1987a)

Gudjonsson (1987a) developed an adapted version of the GSS-1, the GSS-2, following a review of the GSS-1 by Grisso (1986), who suggested that a non-forensic narrative be used with less specific content in relation to the use of UK place names. Grisso (1986) argued that the forensic narrative used in the GSS-1 may have little relevance to actual suggestibility scores. Therefore, the GSS-2 is identical to the GSS-1, except for the content of the story, which contains no reference to a crime. Instead, the GSS-2 narrative comprises a fictional story, where a couple prevent a boy from having a bicycle accident. Furthermore, the subsequent 20 questions differ from those in the GSS-1. This is because the questions for the GSS-2 directly relate to items in the non-forensic story.

The story was designed, much like the GSS-1, to be subjectively scored on 40 distinct items. A strength of the new story is that it has been specifically designed to be appropriate for use in other countries outside of the UK (Gudjonsson, 1987). Unlike the GSS-1, the GSS-2 story contains no outdated language or references to specific UK locations. This could be considered an improvement from the original story, as it is more applicable to different populations. The development of the GSS-2 also enabled the assessment of test-retest reliability.

In recent years, the GSS-2 has been adapted for online use (Wachi et al., 2019a, 2019b). This differs from standard in-person testing, as participants are required to type their answers to each question and negative feedback is presented in text on-screen. During free recall, participants must type as much of the narrative as they can remember into an on-screen box. This is problematic as Gudjonsson defines interrogative suggestibility as a phenomenon that occurs within a closed social interaction, the ecological validity of which is challenging to replicate online. However, Wachi et al. (2019a) found that there was a small difference in yield

1 scores, but no difference in yield 2, sand total suggestibility scores between the standard and online versions. The online version also removes the interference of individual differences between interviewers. That is, interviewer style can impact on the responses given by the subject given during the GSS and therefore impact the GSS result (Bain & Baxter, 2000).

Characteristics of the Gudjonsson Suggestibility Scale

Level of Measurement

The GSS uses an interval level of measurement where participants are scored on each subscale based on whether or not they have been influenced by each misleading question. Interval scales are often considered to be desirable as interval data supports multiple statistical tests compared to nominal and ordinal levels of measurement (Harwell & Gatti, 2001). It is also important to consider the appropriateness of norms, reliability and validity of the GSS in order to fully understand the scale's quality (Kline, 1986), which will be considered next.

Psychometric properties of the Gudjonsson Suggestibility Scale

The GSS-1 and GSS-2 have been used in myriad published articles to measure individual interrogative suggestibility. The GSS is also a validated tool, meaning the psychometric items have been assessed for validity (how accurately a tool measures what it is intended to measure), tested for reliability (the extent to which a tool yields the same results over time and repeated trials), and their practical application evaluated (Boateng et al., 2018). To date, the GSS remains the only validated psychometric tool to measure interrogative suggestibility.

Despite its validated status, the GSS is not without criticism. The GSS aims to be subtle so subjects are unable to identify the true nature of the test. Therefore, the GSS is presented to subjects as a memory assessment. Although deliberately misleading a subject is ethically problematic, evidence suggests that informing subjects that the GSS measures vulnerability to interrogative interviewing significantly reduces the effect of suggestive questioning (Warren et al., 1991). One argument against the validity of the GSS is that it may not successfully measure the internalisation of suggestion, but instead measures subjects' compliance with the interrogator (Mastroberardino & Marruci, 2013). The main difference between suggestibility and compliance is that suggestibility infers personal acceptance of the suggested information, whereas compliance does not. The reliability of the GSS has also been questioned, with researchers arguing that the GSS does not consider direct and indirect suggestibility. Direct suggestions can be defined as "outspokenly expressed", where the intention to influence is overt. Indirect suggestions are described as being masked, where the intention to influence is concealed from the subject (Gheorghiu et al., 1966; Polczyk & Pasek, 2006). This potentially leads to methodological issues with the GSS (Polczyk, 2005). Therefore, the validity and reliability of the GSS will be examined.

Reliability

Reliability refers to the consistency of a specific measure. A measure is considered reliable if results are replicated repeatedly. There are different types of reliability, such as internal consistency, inter-rater, and test-retest reliability. Each type of reliability is discussed below in relation to the GSS.

Internal Consistency

Internal consistency refers to the extent to which items in a tool measure aspects of the same construct, with a Cronbach's alpha of 0.6-0.7 considered to indicate an acceptable level of internal reliability and 0.8 considered as very good internal reliability (Ursachi et al., 2015). Analysis of the internal consistency of the GSS-1 found a Cronbach's alpha coefficient of 0.77

for the yield scale and 0.67 for the shift scale (Polczyk, 2005). According to Ursachi et al. (2015), these values indicate an "acceptable" level of internal consistency and homogeneity for both the yield and shift scales. Gudjonsson (1992b) repeated this analysis for the GSS-2 using factor analysis. Factor analysis is a statistical analysis used to investigate the underlying facets of a measure. Gudjonsson found alpha coefficients of 0.87 for yield 1 and 0.90 for yield 2. These are considered "very good", and an improvement from the GSS-1 factor analysis. A score of 0.79 for shift was indicative of "acceptable" levels, similar to the GSS-1 findings.

Further support for acceptable internal consistency in the GSS has been found in various studies. An acceptable level of internal consistency was found using a sample of 40 undergraduate students, with Cronbach's alphas being 0.79 for yield, 0.75 for shift, and 0.82 for total suggestibility (Merckelbach et al., 1998). Likewise, during development of the GSS-1, Gudjonsson (1984) did not include five questions pertaining to the shift score. In order to then develop the GSS-2, Singh and Gudjonsson (1987) recruited a sample of 285 completed GSS forms from clinical practice and previous research. The five questions were then included, raising the alpha coefficient of shift from 0.67 to 0.70. A final amendment was made to include instances where the subject's response changed, once again raising the alpha coefficient to 0.71. These results are considered to be satisfactory and support the GSS being internally reliable.

In reviewing the internal reliability of the shift subscale scores, Gignac and Powell (2009) argued that shift scoring was only a simple composite, in turn suggesting an alternative method for scoring the subscale. Instead of the simple composite shift score, they proposed a separate score for negative indicators of susceptivity to interrogative pressure (shift-negative), and a second for positive indications of susceptivity to such pressure (shift-positive). Using previous research data from a sample of 220 children, both shift-positive and shift-negative were found to be associated with unacceptably low internal consistency reliability of $w2 < .60$

(*w* representing the auto-regression effects between the corresponding yield 1 and shift items). Shift-positive and negative scoring is not included in Gudjonsson's (1997) recommended scoring guidelines. However, Gignac and Powell (2009) argued that issues with the shift score limit the use of the GSS to the yield subscale. That is, the shift score is too basic a score to capture the true nature of vulnerability to interrogative pressure. Given that Gignac and Powell (2009) were unable to offer a suitable alternative to shift scoring, Gudjonsson's (1997) original scoring guidance remains the most reliable method for scoring shift.

Inter-Rater Reliability

Inter-rater reliability is the measure of consistency in scoring between two or more scorers for the same measurement tool. Richardson and Smith (1993) studied a sample of 57 young people and found high inter-rater reliability for the GSS-1 (correlations > .94 for yield 1, yield 2, shift and total suggestibility), with the highest being for yield 2 (r= .994). Clare et al. (1994) examined the inter-rater reliability of the GSS-2 with specific focus on memory, suggestibility and confabulation. The inter-rater reliability was assessed across these domains by three independent raters, using data from 101 subjects who had been recruited from job centres and community day centres in the UK providing adult education classes or learning difficulty support. The GSS-2 was administered to all participants. Intra-class correlation coefficients to assess consistency across raters were calculated from one point being allocated for each item recalled from the story, and half a point for slightly inaccurate details. The researchers also produced confabulation scores, where one point was given in instances where an obvious inaccuracy in the recall of the narrative was present. Intra-class correlation coefficients for the measures of memory and total suggestibility were both high and statistically significant (p< .001): .951 and .996 respectively. Correlations for confabulation on delayed and immediate recall were statistically significant (p< .001), however the correlations were lower than the other domains at .724 and .803 respectively. In response, the researchers felt it necessary to clarify the definition of confabulation. Therefore, confabulation was defined as an obvious inaccuracy in the recall of the narrative, arising from the distortion or incorrect introduction of ideas Clare et al. (1994). Together, this suggests that the inter-rater reliability of the GSS-1 and 2 is high, meaning the tool is reliable between scorers.

Test-Retest Reliability

The use of test-retest methods enables a direct assessment of the degree to which test scores are consistent over time across test administrations (Murphy & Davidshofer, 2005). Given the nature of the GSS, it is not possible to assess test-retest reliability within each scale, because prior memory of the narrative or questions will impact subsequent testing. However, comparisons can be made between the GSS-1 and the GSS-2, indicative of temporal consistency.

Gudjonsson (1987a) examined the test-retest reliability (for the story at immediate and delayed recall after 50 minutes) of the GSS using three groups of adult participants. Group 1 comprised “normal” men and women, group 2 comprised forensic patients referred to the author for court reports, and group3 comprised forensic cases similar to those in group 2. Both groups 2 and 3 included both males and females. Gudjonsson (1987b) found a highly significant correlation in memory (group1 = .77, group2 = .93, group3 = .87) and total suggestibility scores (group1 = .90, group2 = .92, group3 = .81) across the GSS-1 and 2. Correlations for shift (group1 = 0.79, group2 = 0.80, group3 = 0.73) were found to be consistently lower than those for yield 1 (group1 = 0.84, group2 = 0.86, group3 = 0.78) and yield 2 (group1= 0.86, group2= 0.90, group3= 0.84). This suggests that there is temporal consistency of interrogative suggestibility between the GSS-1 and 2. Similar findings were reported by Gudjonsson (1992a) when they explored the test-retest reliability (between immediate and delayed test of 4 weeks) of the GSS-1 yield scores using an opportunity sample from the general population. A yield test-retest correlation of 0.81 was reported, meaning there was a strong correlation between yield scores.

Merckelback et al. (1998) also explored the internal consistency and test-retest stability of the GSS-1. A total of 40 undergraduate students were tested immediately after receiving the GSS-1 narrative, and again four weeks later. A moderate, however significant test-retest reliability correlation of 0.55 was observed for the yield scale (Merckelbach et al., 1998). This finding was considerably lower than the yield test-retest correlation of 0.81 reported by Gudjonsson (1992a). However, note that Merckelback et al. (1998) did not specify which explicit yield scale (yield 1 or 2) they were investigating. Merckelbach et al. (1998) offer the possible explanation that their use of an undergraduate sample may have produced less variation than Gudjonsson's (1992a) sample taken from the general population. This is because the general population sample was more likely to be characterised by a broader range of suggestibility. Whilst it is possible that there was less variation in suggestibility amongst a homogeneous student sample, the difference in correlations brings into question the accuracy of test-retest reliability of the GSS.

Other factors that can impact test-retest reliability of the GSS are also noted in literature. Boon and Baxter (2000) argued that the interviewer may be responsible for up to two thirds of variance in interrogative suggestibility scores. In part, this can be due to the demeanour of the interviewer, with an abrupt approach producing higher GSS-1 scores than an interviewer who is more amicable (Bain & Baxter, 2000). Likewise, research has found that shift scores are dependent on whether the instructions are presented in a convincing manner (Merckelbach et al., 1998). Therefore, the reliability of the GSS scores can be dependent on the way in which the individual(s) administering the measure act during administration. It is difficult to control for subtle differences between administrators and their approach to administering the GSS.

Therefore, it is likely test-retest reliability of GSS scoring could be impacted, in some instances, by the tool's administrator.

Other issues associated with interrogative interviewing, such as malingering (where a subject fabricates, feigns or exaggerates psychological or physiological symptoms to achieve a desired outcome), do not appear to reduce the reliability of GSS scores. Hansen et al. (2010) found that heightened suggestibility was difficult to malinger, even when deliberate malingering instructions were given to participants. Further research has explored whether the GSS can distinguish malingering from compliance. 66 participants were randomly assigned to one of three conditions: Misled, compliant, or standard instruction. Results indicated that malingerers scored differently on almost all suggestibility measures compared to truly vulnerable subjects. This supports the hypothesis that the GSS can identify some patterns of malingering. However, the GSS was less successful at distinguishing compliant responding from genuine vulnerability. The authors argued that "fakers" may recognise that they should accept leading questions in an attempt to appear vulnerable, but not that they should also change their shift responses following the negative feedback, as a genuinely vulnerable witness would be likely to do (Woolston et al., 2006). Therefore, this suggests that the GSS scoring still holds acceptable levels of reliability, even in instances of deliberate malingering.

Validity

It is necessary to understand the validity of the GSS-1 and GSS-2 scales to ascertain whether the tool is successful in measuring what it aims to measure.

Construct Validity

Construct validity refers to the degree to which a test successfully measures what is purports to be measuring (Cronbach & Meehl, 1955). Gudjonsson (1984, 1992b) used factor analysis to explore the individual factors that make up the scale. Factor analysis is a statistical analysis used to investigate the underlying facets of a measure. Specifically, Gudjonsson (1984) used factor analysis to investigate the individual factors the GSS-1 scale comprised of. Using a Varimax Rotation, Gudjonsson (1984) aimed to maximise the dispersion of loading within each factor, resulting in the loading of a lesser number of variables into factors. Analysis is then interpreted into clusters (Field, 2000). Gudjonsson's sample consisted of 195 participants (58 females, 56 males, 40 forensic patients, 41 primarily delinquent male children), with analysis revealing two main factors: yield 1 items and shift items. Alpha coefficients for scoring on the 15 yield and 15 shift items were 0.77 and 0.67 respectively. Therefore, the independent yield and shift factors were identified. Whilst Varimax Rotation is a widely used statistical technique that can clarify the relationship among different factors, it has limitations. Varimax Rotation can be limited in its appropriateness in testing hypotheses, or revealing meaningful patterns (Schmitt & Sass, 2011). However, because Gudjonsson was attempting to maximise the loading of variables onto factors, his use of Varimax Rotation was appropriate. However, the term "Varimax Rotation" does not refer to a unique procedure as several different types of rotation are possible (Forina et al., 1989). Gudjonsson (1984) simply referred to the Varimax Rotation used as a "Varimax procedure". This is problematic, as it does not provide details for replication, or indeed, details about the specifics of the rotation used.

A review of the GSS-1 conducted by Grisso (1986) determined that the concepts underpinning the development of the GSS-1 are adequately conceptualised, supporting its use in forensic settings. When considering these concepts, Gudjonsson (1992b) found that yield 2 is the most sensitive measure of vulnerability to interrogative suggestibility. This finding was also supported by Baxter & Boon (2000). This suggests that there is an overreliance on total suggestibility scores as this does not provide enough specific information regarding the amount a subject shifts or yields. This may be especially true if yield 2 is an important measure of interrogative suggestibility in its own right (Baxter & Boon, 2000). Furthermore, this focus on total suggestibility could draw attention away from those subjects who may be more vulnerable to interrogative processes, as evidenced by individual yield and shift scores. The authors proposed that more focus on the individual yield (yield 1 and 2) and shift scores was needed.

Further factor analysis was then conducted for the GSS-2, where Gudjonsson (1992b) explored whether the shift and yield parts of the GSS-2, like the GSS-1, load onto separate factors. Likewise, the internal consistency of shift and yield were analysed. Gudjonsson (1992b) also conducted a Varimax Rotation on data from the GSS-2 by recruiting 129 participants (100 forensic males, 29 males from general population). Yield and shift items were again loaded on two distinct factors. Alpha coefficients for yield 1, yield 2 and shift ranged from acceptable to good, with values of 0.87, 0.90 and 0.79 respectively. Therefore, the findings from both Gudjonsson's (1984; 1992b) studies support the view that shift and yield are measuring independent and distinct constructs.

Mastroberardino and Marucci (2013) explored the prevalence of compliance and internalised suggestion in the GSS. Two experiments were conducted to test whether compliance with the interrogator was more prevalent than internalisation of the suggested materials. This is important for the GSS, as suggested materials during interrogation can be internalised by the subject being questioned. Despite the GSS purporting to measure interrogative suggestibility, Mastroberardino and Marucci (2013) argued that it is unclear whether the GSS instead measures a subject's compliance with the interrogator. In experiment one, the GSS-2 was administered to participants and immediately followed by a source identification task, which comprised of a questionnaire containing 20 items related to the items encountered during the scale (e.g., the boy's bicycle got damaged when it fell on the ground). Participants were then asked to indicate when they first encountered each item (e.g., during the original story, during questioning, or do not know). This was designed to measure the degree to which subjects internalised the suggested information in the GSS or complied with the interrogator. In the second experiment, half of the participants were given the source identification task immediately after GSS-2 administration, whereas the other half completed it 24 hours after GSS-2 administration. The results from both studies found increased compliant responses during the shift part of the assessment, with participants internalising more suggested information after yield 1. Mastroberardino and Marucci (2013) concluded that it is likely that different processes underpin the yield 1 and shift portions of the GSS-2, with internalisation underpinning yield 1, and compliance underpinning shift. This is problematic for the validity of the total score on the GSS, as the tool struggles to differentiate between suggestibility and compliance, given that the outcomes of these behaviours appear to be the same. In contrast, it could be argued that this provides evidence that the yield and shift scales measure concepts that are meaningfully different from each other.

Researchers have found evidence that higher total suggestibility scores on the GSS were also related to higher levels of subsequent false memory (Gudjonsson et al., 2016). This would suggest that the GSS is measuring what it sets out to measure. However, note that this finding may also map onto convergent validity, which refers to how closely a scale is related to other variables and measures of the same construct. Contrastingly, other studies have found this not to be the case (Lee, 2004; Vagni et al., 2015). To explain this discrepancy in findings, Gudjonsson et al. (2016) theorised that variance in sample size may account for the differences in the significance of the relationship across studies. Specifically, unlike the findings of Lee (2004) and Vagni et al. (2015), Gudjonsson et al. (2016) found a significant relationship between immediate and delayed suggestibility (i.e., false memories) with a sample size of 1,183 participants. However, the effect sizes were small (Cohen's d= .25 and .28), which may account for the absence of a significant relationship in other studies where a much smaller sample size was used. For example, in Lee's (2004) study with a sample of 35 participants, and a study conducted by Vagni et al. (2015) with a sample of 180 participants. However, a study conducted by Raymond (2020) found a relationship between interrogative suggestibility and false memory despite recruiting a smaller sample size (*N*= 53). It should be noted that it is possible that studies with small sample sizes can be underpowered and produce imprecise effect sizes; however, these issues were not explicitly considered in the discussion sections by the aforementioned researchers. Therefore, these findings should be interpreted with caution. Together, the literature seems to indicate that there is support for the relationship between total suggestibility scores and false memories. Which, in turn, supports the construct validity of the GSS.

Research conducted by Miles et al. (2007) administered the GSS-2 and an independent suggestibility paradigm to 69 children with an intellectual disability and 50 children without an intellectual disability. The independent suggestibility paradigm involved all of the children participating in a magic show. Each child subsequently attended a "biasing interview" 3 days after the magic show, where the interviewer suggested details to the child that did not occur in the show. The next day the children completed a second "memory interview", which aimed to assess the impact of the biasing interview on the children's recall of the show. No significant association between performance on the GSS-2 and the suggestibility paradigm were found for children with an intellectual disability. Whilst findings suggest that the GSS-2 did not provide a reliable means for distinguishing the relative suggestibility of children with an intellectual disability, the GSS-2 was found to be useful in distinguishing the relative suggestibility of older children without an intellectual disability. The researchers concluded that the GSS-2 appears to be more useful when administered to children without intellectual disabilities. A limitation of this study is that the independent suggestibility paradigm is not a validated tool. Still, given a distinct lack of validated suggestibility tools, it can be argued that this research goes some way to highlighting the concurrent validity of the GSS-2.

Concurrent Validity

Concurrent validity is demonstrated when a test correlates well with an existing validated measure. Given that the GSS remains the only validated measure of interrogative suggestibility, the true extent of concurrent validity for the measure is challenging to assess (Grabner, 2018). The majority of alternative suggestibility scales are specifically concerned with hypnotic phenomena (Barber & Wilson, 1978; London, 1962; Shor & Orne, 1962; Spiegel, 1974) which do not map directly onto interrogative suggestibility. However, the GSS-2 acts as a parallel form to the GSS-1, providing some insight into concurrent validity of the GSS as a whole (for results see test-retest reliability section).

An attempt has been made to compare to GSS-2 with other measures. Roma et al. (2011) compared the ability of the GSS-2 with the ability of the Bonn Test of Statement Suggestibility (BTSS) to measure interrogative suggestibility of 84 children. To achieve this, the researchers explored the correlation between the yield, shift and total suggestibility scores for the GSS-2 and BTSS. Analyses between the corresponding scale found strong correlation between yield variables (r= .71; p= .001) and total suggestibility (r= .72; p= .001). However, there was a weak correlation between shift variable (r= .33; p= .05). Given that the GSS-2 and BTSS assess shift differently to one another, it is not surprising that a weak correlation was found. Specifically, the BTSS repeats eight questions immediately after the child responds to the first question, whereas the GSS-2 repeats 20 questions after the subject has responded to all questions. The authors concluded that the BTSS was more suitable for measuring child suggestibility during police and forensic interviews. This is because the BTSS procedure for assessing shift is more pressing than that of the GSS-2, enabling better discrimination of the better child's failure to handle leading questions. This is understandable, as the BTSS was developed from the GSS (which was originally developed with an adult focus) to test children specifically.

Recent research has explored the concurrent validity of the existing GSS when compared to adapted online versions of GSS. Wachi et al. (2019a) found that the Japanese online version of the GSS (developed from the Japanese standard GSS) did not produce significantly different scores for yield 2, shift, or total suggestibility compared to the paper-based version of the GSS. Online scores for yield 1 (M= 3.22, SD= 3.38) were significantly lower than the standard GSS (M= 3.84, SD= 2.95), however, in line with Cohen's (2013) classified effect sizes, the effect size was small (d= 0.19). As such, Wachi and colleagues concluded that the Japanese online version of the GSS was promising in terms of its ability to measure interrogative suggestibility in line with the standard version of the GSS. This supports the adaptation of the GSS for online use, further highlighting the concurrent validity of the measure, even when it is presented in a non-traditional format.

Predictive Validity

Predictive validity refers to the extent the results of a test can predict future behaviour. Researchers appear to have come to conflicting conclusions regarding the predictive validity of the GSS. Grisso (1986) argued the predictive validity of the GSS has not been examined during a police interrogation, and therefore has not been studied sufficiently. Grisso (1986) suggested, therefore, that predictive validity of the GSS could not be assumed within the context of interrogative interviewing. In Merckelbach et al. (1998) second study examining the predicative validity of the GSS, 53 students completed the measure. A Pearson correlation between subjects' GSS yield score and the extent subjects gave in to leading questions from the interviewer was found to be small, but statistically significant: $r(53)= 0.22$, $p< .05$. The researchers concluded that they had only found "some" indications of predicative validity for the GSS yield scale. However, the researchers noted that there was no free recall of the story in their procedure, which is a deviation from the standard GSS procedure outlined by Gudjonsson (1992). Although they argued the free recall was not needed as it was not related to measuring yield and shift, it could also be argued that this negatively impacted the validity of the findings as the procedure did not match that of the GSS procedure used in the original research by Gudjonsson (1992).

In contrast, Gudjonsson (1997) claimed that results obtained by Tully and Cahill (1984), who explored the police interviewing of subjects with learning difficulties, suggest that the GSS-1 had some power to predict the accuracy of witness accounts during police interviews. Tully and Cahill (1984) compared the total memory and total suggestibility scores on the GSS-1 of two groups with learning disabilities with a control group of subjects with "normal" intellectual functioning. The result showed that the most intellectually disadvantaged groups performed worse on both measures compared to the control group. However, Tully and Cahill's (1984) sample selection was idiosyncratic, as they only selected participants based on their attendance at specialised facilities without formally testing their intellectual functioning. Similarly, their research only focused on a total suggestibility score, which has previously been found to be unsuitable (see test-retest section; Baxter & Boon, 2000). Therefore, it is not possible to conclude whether the higher scores found for subjects with learning difficulties reflect tendencies to shift or yield. Together, these findings suggest that the GSS has only a low level of predicative ability. This is problematic given that the GSS is designed to help predict whether someone will be susceptible to interrogative suggestibility.

On the other hand, however, Sigurdsson and Gudjonsson (1996) also countered Grisso's (1986) argument by investigating false confessions during police interviews. A sample of 62 prisoners were asked if they had ever made a false confession to police and were subsequently administered the GSS-1. The yield 1 (Z= -.0.68, p< .05), total suggestibility (Z= -2.21, p< .05) and confabulation (Z= -2.01, p< .05) scales were found to have significant predictive power in predicting false confessions. However, it was not possible for the researchers to verify the truthfulness of the prisoners' accounts regarding whether they had previously given a false confession. This may have negatively impacted the validity of the findings in instances where a participant either falsely claimed they had or had not given a false confession. Despite this, the current finding supports the predictive validity of the GSS-1.

Normative Samples

Normative data refers to data that summarise what is usual or typical in a defined population, culture, or institution at a specific point of period of time (O'Connor, 1990). Without understood "normal" levels, the scores obtained on a test may be less meaningful. The GSS was developed using normative data from Great Britain and Iceland. However, substantial research has been conducted with the GSS-1 and GSS-2 to provide further normative data. It is of note that norms are not presented individually for males and females or across age, as research suggests there is no significant difference between vulnerability to suggestibility between the sexes across different adult age groups (Gudjonsson et al., 2016).

First, considering adult samples, normative data for the GSS-1, including yield 1, yield 2, shift, total suggestibility, means/standard deviations for memory at immediate recall, and delayed recall are available for a variety of populations through Gudjonsson's research. Primarily, Gudjonsson's normative data is representative of adults, including Icelandic prisoners (Gudjonsson & Sigurdson, 1995), members of the forensic population with intellectual vulnerabilities (Gudjonsson, 1997), and court referrals, primarily composed of defendants in criminal trials, including victims and witnesses (Gudjonsson, 1997). Similar findings have been replicated for the GSS-2, providing normative data for forensic persons with intellectual vulnerabilities (Gudjonsson, 1997), suspects detained for questioning in London (Gudjonsson et al., 1993), and court referrals, similar to those used in the GSS-1 assessment (Gudjonsson, 1997). A strength of this normative data is the breadth of forensic contexts from which data have been collected. This provides insight into the "normal" suggestibility levels of subjects at different stages of the criminal justice system, including those with vulnerabilities. Therefore, the availability of the normative data makes it easier to meaningfully use the GSS in these variety of forensic contexts.

Gudjonsson (1997) also explored normative data for the GSS-1 and 2 for adults from the general population. Whilst the general population are not representative of forensic populations, as mentioned earlier, the GSS can be used with witnesses. Specific research for the GSS-2 also explored normative data for persons with intellectual vulnerabilities not related to forensic contexts (Gudjonsson, 1997). However, the GSS was specifically designed to be used in forensic contexts. Therefore, these norms may be less applicable for practical application of the GSS in forensic settings. Whilst this is an apparent limitation of the findings, the results could provide useful comparison of norms between subjects in general and forensic settings, and be useful for research and theory development on suggestibility.

Next, considering young people, normative data for the GSS-1 have also been presented for young offenders in Iceland (Gudjonsson & Sigurdson, 1995), and in Britain (Gudjonsson & Singh, 1984; Singh & Gudjonsson, 1992). Likewise, normative data has also been collected for young people and the GSS-2. This includes findings from Icelandic male and female children (Danielsdottir et al., 1993). Whilst this normative data is helpful for understanding what the scores for the different age groups mean, the GSS-1 and 2 were primarily designed for use with adults. Therefore, it could be argued that normative data for children is less informative than that for adults. However, tools such as the BTSS are specifically designed for and more effectively used with children (Roma et al., 2011).

A key strength of the GSS-1 and 2 in relation to its underpinning normative data is that findings are based on populations around the world. That is, there is support for Gudjonsson's normative data and its consistency with norms from other countries. Pollard et al. (2004) collected normative data on the GSS-2 in the US and compared this to normative data collected in the UK. Data were collected from 72 US participants, where it was found that the UK sample (as reported in the GSS manual) scored significantly higher on yield, whereas the US sample scored significantly higher on shift. The authors argued this raised the possibility there may be cross-national differences in interrogative suggestibility. Later research by Frumkin et al. (2012) administered the GSS to a forensic sample of 332 adult and juvenile Americans. Scores were found to be consistent with that of Gudjonsson's British and Icelandic samples, where yield 1 scores were most impacted by intellectual and cognitive vulnerabilities, and yield 2 and shift scores were most related to emotional and personality characteristics, as outlined by Gudjonsson and Sigurdsson (1999).

Normative data have also been explored in other nations. Polczyk (2005) administered the Polish adaptation of the GSS-1 and 2 to 148 and 207 Polish students respectively. Polczyk (2005) found that the Polish versions of the GSS-1 and 2 were comparable to the original scales, with means and standard deviations for the suggestibility indices being consistent with those found in British samples (Gudjonsson, 1997). Pires et al. (2014) administered the Portuguese adaptation of the GSS-1 and 2 to 40 Portuguese prisoners. The Portuguese versions of the GSS were found to have validity through the comparison of means of the original and the translated scales. These results were consistent with those in Gudjonsson's (1997) original study. However, the study conducted by Pires et al. (2014) recruited a small sample. Therefore, a limitation of the study is the heterogeneity of the Portuguese sample. It is argued that these findings are still helpful as they align with those first obtained by Gudjonsson. Pires et al. (2014) conducted a second study where they administered the Portuguese GSS-1 to explore the relationship between memory and vulnerability to suggestion in the sample of prisoners, but also in the general population. As with Gudjonsson's (1997) study using Icelandic samples, the comparison of Portuguese inmates to a general population had also shown that the forensic sample produced significantly higher scores in the interrogative suggestibility measure compared to the general population. This further confirmed the results based on data collected by Gudjonsson (1997). Bianco and Curci (2015) conducted research to evaluate the validity of the Italian adaptation of the GSS-1 and 2 in Italian samples. The first study showed that when coefficients were calculated for yield 1, 2, shift and total suggestibility, they were lower than those reported by Gudjonsson (1984). However, the Italian version of the GSS has been used in subsequent research with Gudjonsson (Curci et al., 2017), and has provided helpful normative data for Italian samples. Therefore, the current evidence supports that the GSS scales have been found to be reliable and valid in a large number of countries and different populations.

Although there is little evidence to suggest populations become more or less susceptible to suggestibility over time, police interviewing processes and questioning continue to develop and change (Milne et al., 2011). Future research should consider the relevance of the GSS and its components (such as the story) to police interrogative interviewing processes. Furthermore, a large proportion of the existing literature reviewing the reliability and validity of the GSS has been conducted by Gudjonsson himself, inviting the potential for confirmation and publication bias (Hergovich et al., 2010). Therefore, the current review should be interpreted with caution.

.

Conclusions

The aim of the current psychometric critique was to provide an updated assessment of the tool's validity, and reliability in forensic contexts. Regarding reliability, a disproportionate amount of research exploring the reliability and validity of the GSS has focused solely on yield scoring. This is problematic as yield 1 and 2 have been found to be more homogenous than shift (Gudjonsson, 1992b). However, the research examined suggests that the GSS-1 and 2 are highly reliable measures of interrogative suggestibility. Considering validity, the GSS experiences some issues when distinguishing between suggestibility and compliance, making the true extent of the construct validity difficult to measure. This suggests that further research is needed to distinguish these concepts and to examine whether the GSS-1 and 2 accurately measures them distinctly. Further, there is a need for validated tools that directly assess the concurrent validity of the GSS. However, the GSS-1 and 2 are considered to have at least satisfactory validity. Therefore, the GSS remains the most appropriate measure of interrogative suggestibility in forensic settings with a strong evidence base for both its reliability and validity. The GSS should continue to be used in forensic settings, as long as its limitations are considered by its administrators.

Chapter Five: General Discussion

Thesis Aims

This thesis aimed to contribute to the growing body of research exploring the issue of eyewitnesses giving honest, but inaccurate, testimony (Ceci & Bruck, 1995; Gudjonsson, 1992; Loftus, 1986). Despite the vast number of empirical studies that have explored the MI effect, little is known about how modality may impact on eyewitness susceptibility to misinformation. Academic researchers who have begun to explore the impact of modality on misinformation acceptance have highlighted the need for further research into these effects (Ulatowska et al., 2016). Likewise, there have been similar gaps in research relating to the impact of different facial angles on the MI effect within a MI paradigm. With a limited number of validated measures of suggestibility, it is important to consider the reliability and validity of psychometrics that aim to measure suggestibility. Therefore, this thesis aimed to provide increased understanding into the suggestibility of adult eyewitnesses. To achieve this, the thesis firstly aimed to explore how individuals can be externally misinformed by modality in the MI paradigm. Second, this thesis aimed to investigate within a MI paradigm whether presenting a perpetrator's face, followed by an innocent suspect's face shown in a different facial angle, influences eyewitness susceptibility to misinformation in the context of a lineup task. Thirdly, this thesis aimed to assess how internal factors to the person, such as interrogative suggestibility, are measured. Thus, the work provides a deeper understanding of how both external and internal factors contribute to adult suggestibility.

To achieve the outlined aims, three interrelated pieces of work were undertaken, which are discussed in detail below. First, a systematic review explored the existing literature regarding modality manipulations in the MI paradigm. Second, an empirical study was undertaken to explore how different facial angles may impact the MI effect within the MI paradigm. Finally, given that there was strong evidence throughout the reviewed memory literature that eyewitnesses can be misinformed through suggestion, a critique of the

Gudjonsson Suggestibility Scale (GSS; Gudjonsson, 1984) was undertaken. The reliability and validity of the scale was assessed. In this chapter, I will attempt to consolidate the main findings in broad terms, consider the consistent themes across previous chapters, and discuss the implications for future research and practice.

Main Findings

Chapter Two: A Literature Review Following a Systematic Approach: An Assessment of the impact of Modality on Susceptibility to Misinformation.

The majority of MI paradigm studies manipulate modality at each stage, but rarely consider the impact of this, which is problematic for misinformation research. In Chapter Two a systematic literature review was conducted to explore the impact of modality on the MI effect, within the MI paradigm. This is the first systematic review of its kind, where an extensive search of ten electronic databases identified eleven papers exploring the impact of modality in the MI paradigm. Results were discussed in relation to auditory and visual modality comparisons, combined auditory-visual modalities, and modality congruence.

One key finding from Chapter Two is that findings are mixed across the studies reviewed regarding the impact of specific modalities on the MI effect. That is, in some papers, modality manipulations within the MI paradigm appeared to impact misinformation acceptance; however, the results of other papers reviewed indicated that modality had no such effect on misinformation susceptibility. Even when studies found that modality affected the MI effect, there were discrepancies regarding which specific modalities (i.e., auditory or visual) were more impactful on the MI effect. Due to these discrepancies, it has not been possible to draw conclusions about modality from the reviewed findings.

Chapter Two also reviewed findings regarding modality congruence. Specifically, modality congruence at the encoding and test stages appeared to reduce the MI effect. This supports the encoding specificity hypothesis within the MI paradigm, contributing to our understanding of how modality congruence can not only support recall (Ulatowska et al., 2016), but also reduce the MI effect. In contrast, modality congruence at the misinformation and test stages was found to increase the MI effect. However, this was only found in one of the reviewed studies. This finding can be explained by the encoding specificity hypothesis, whereby it is theorised that the modality congruence at misinformation presentation and at test lead to similar memory performance as when there is modality congruence at encoding and test (Campbell et al., 2007).

Chapter Three: Presenting Perpetrator and Innocent Suspect Faces from Different Facial Angles Influence the Susceptibility of Eyewitness Memory? An Investigation into The Misinformation Effect and Eyewitness Misidentification

Academic researchers and legal decision makers have both expressed concerns regarding the impact of misinformation on accurate eyewitness testimony. It is without doubt that the MI effect can be catastrophic in instances where an eyewitness accepts misinformation about a crime they witnessed and makes erroneous testimony or an incorrect identification (Loftus, 2005). We also know that incorrect eyewitness identifications, often made during a police lineup, are the leading cause of innocent persons being incorrectly convicted of crimes (Innocence Project, 2017). Chapter Three investigated the impact of the viewing angle of a perpetrator's face on the MI effect using the MI paradigm. Using a quantitative design, over 2000 participants completed an online MI paradigm procedure where the viewing angle of the perpetrator's face and intervening face was manipulated at encoding/test and the misinformation stage to be either posed from profile or the front. The procedure involved watching a mock crime video, followed by a misinformation news report. After which, participants were randomly allocated to be tested using a target-present or target-absent lineup.

Analysis was conducted using ROC analysis and CAC analysis. No statistically significant differences were found between conditions where all three stages were shown in the same facial position (congruent facial angle) or when the encoding and test face position matched, but the misinformation face differed (incongruent facial angle). However, discrimination accuracy was significantly better when the encoding face was viewed from a frontal angle compared to a profile angle. This supported the proposed encoding strength hypothesis and also previous research, whereby frontal faces have been found to provide more information than a profile viewed face (McKelvie, 1976). Therefore, it is argued that encoding strength appeared to be more important than the congruence of the facial angles across the misinformation and encoding/test stages.

Chapter Four: The Gudjonsson Suggestibility Scale: A Psychometric Critique

Chapters Two and Three considered the experimental, external and situational factors to a person that may influence the MI effect. The interrogative suggestibility model theorises that suggestibility is the result of interaction between the individual and the environment (Bruck et al., 1997). Therefore, individual factors that may influence adult eyewitnesses are important to consider. As much as it is important to consider the ways in which an individual can be suggestible, it is necessary to consider ways individual suggestibility can be measured and subsequently mitigated. Therefore, Chapter Four reviewed the reliability and validity of the GSS (GSS-1 and 2) looking at evidence of its reliability and validity, which are both vital. The objective of this chapter was to provide an up-to-date critique of the scale.

This critique found clear evidence for reliability and validity for both the GSS-1 (Gudjonsson, 1984) and GSS-2 (Gudjonsson, 1987a) in the general population and in forensic samples. In addition, there was support for adaptations of the measure, such as online administration (Wachi et al., 2019a, 2019b). Researchers administering the measure in other countries supported its use, indicating that the GSS is applicable to other populations and cultures. In sum, the scale appears to have a strong a varied evidence base, supporting its use in forensic settings.

Implications for Theory

The thesis has considered both external (Chapter Two and Chapter Three) and internal (Chapter Four) influences on adult memory accuracy. Findings in Chapter Two highlighted key considerations regarding the importance of considering modality effects at encoding and test. First, the systematic literature review has provided a unique contribution to our understanding of the encoding specificity hypothesis within the MI paradigm. Specifically, when encoding and test modalities are congruent as opposed to incongruent, the MI effect was reduced. Second, the literature review adds to our understanding of the literature with respect to memory strength in relation to encoding modality. Previous research and theory indicate that under some circumstances, memory will be enhanced if textual information (i.e., written text) is presented in an auditory format compared to a visual format (Ginns, 2005). Two studies in the review found that misinformation acceptance was lower when the encoding event was presented in an auditory modality, compared to a visual modality (Dijkstra & Moerman, 2012; Ulatowska et al., 2016, study 1). These findings need to be considered alongside the fact that the majority of studies reviewed that manipulated modality at encoding found no significant effects of modality on the MI effect. In fact, there was very little support for the impact of modality on the MI effect overall when manipulations were made at encoding, the misinformation stage, and test. Findings were largely mixed across the reviewed studies. As such, the current review did not find strong evidence for modality effects on misinformation acceptance, and further research is needed in this area, particularly with respect auditorily encoded information.

In Chapter Three, there were no significant findings relating to facial angle congruence at the three stages of the MI paradigm. One explanation could be the encoding specificity hypothesis. In the study procedure, encoding and test always matched. Therefore, the modality congruence at encoding and test may have improved performance, despite instances where there was also modality congruence at the misinformation stage and test. Nevertheless, it is important to remember that the empirical chapter findings supported an encoding strength hypothesis, whereby the strength of the encoded information was more important than modality congruence. Therefore, the research in Chapter Three has proposed an alternative explanation for why false memory may occur in the MI paradigm. This has implications for future research, where further investigation into the encoding strength hypothesis is needed to fully assess the extent of its impact on the MI effect across different experimental tasks (e.g., ID tasks, MI paradigm).

There are also theoretical implications relating to the proposed encoding strength hypothesis from Chapter Three, whereby participants who encoded a frontal posed face had better discrimination accuracy compared to those who encoded a face shown at a profile facial angle. This finding supports previous facial recognition research, where sensitivity to face identification has been found to be greatest for faces presented from a front facial angle compared to any deviation from this angle (Swystun and Logan, 2019). Furthermore, this supports the theory that faces shown from a frontal angle should positively aid identification compared to those presented at a profile facial angle (Goldstein & Mackenberg, 1966; McKelvie, 1976). The current finding aligns with the holistic encoding hypothesis, whereby faces are thought to be processed as a perceptual whole (Taubert et al., 2011). Given that a frontal face arguably provides more perceptual details than a profile facial angle (i.e., two eyes, two eyebrows, etc.), it is possible participants in this study processed frontal pose faces more holistically, which in turn boosted their memory performance compared to when the perpetrator was encoded in profile view. Overall, the findings from Chapter Three are in keeping with the extant face processing and identification literatures, further contributing to these bodies of evidence.

Implications for Practice and Future Research

Chapter Two supported that modality congruence between encoding and test can reduce the MI effect. This could have positive implications for criminal justice proceedings, as eyewitnesses often witness the perpetrator of a crime visually, and are tested visually on their memory for the perpetrator using lineups. Future research is needed to examine the effects of modality congruence, particularly to test whether misinformation acceptance is affected depending on whether information is presented in the same modality across the stages of the MI paradigm. This research will provide those working in the criminal justice system with further information about how existing procedures may mitigate or exacerbate the MI effect. Chapter Two highlighted that, although specific modality manipulations may impact on the MI effect across the different stages of the MI paradigm, findings across studies are mixed. Therefore, it was not possible to draw firm conclusions about the significance of specific modality manipulations (i.e., visual or auditory) at different stages of the MI paradigm on the MI effect due to a paucity of data on the issue. Further research is needed to advance our understanding of how modality may be impactful, and inform legal decision makers.

Findings from Chapter Three suggest that individuals who witness a crime and the perpetrator is only viewed from a profile facial position are likely to have poorer discrimination accuracy compared to if they viewed the perpetrator from a frontal face position (Swystun & Logan, 2019). CAC results also found that confidence-accuracy was higher in frontal face encoding conditions compared to profile. As a result of these findings, legal decision makers should consider the facial position a perpetrator was seen in and how this may impact later identification. For example, it could be recommended that when an eyewitness has seen a perpetrator from a profile facial position that investigators need to give more weight to corroborating evidence. However, investigators need to be cautious when purporting the idea that an eyewitness, having witnessed a profile facing perpetrator, is less credible. Research indicates that it can be damaging for witnesses if they do not feel believed when discussing aspects of the witnessed crime (Patterson, 2011; Randell et al., 2018). Therefore, investigators should be sensitive to the lived experiences of the eyewitness and not entirely rule out their testimony on the basis of a profile encoded face. Whilst is recommended investigators are cautious of profile encoded perpetrators, further research is needed to consolidate these findings.

To consider how we may mitigate adult vulnerability to suggestibility we can consider the findings from Chapter Four. As highlighted, the GSS was found to be both a reliable and valid measure of interrogative suggestibility. Likewise, it was concluded that the GSS is appropriate for use in forensic settings, including for use with eyewitnesses. Therefore, it is suggested that the GSS could be used more widely and frequently to assess the internal suggestibility of eyewitnesses. This would help identify eyewitnesses at higher risk of interrogative suggestibility. This would be particularly useful for interviewers, as they can use the results of the GSS to consider how to proceed with an eyewitness. It is recommended, however, that the GSS psychometric data is used with some caution by the prosecution and defence in UK trials. From a legal perspective, submission of GSS psychometric data in a trial could be challenged as trespassing on the role of the jury in deciding the credibility of a witness based on primary evidence heard in court. Therefore, it is recommended that consideration is given to the legal context relating to the individual prior to administration of the GSS.

The conclusions drawn throughout this thesis provide some suggestions for the direction of future research. Chapter Two highlighted the discrepancy in findings regarding the impact of auditory and visual modalities on the MI effect. Therefore, it would be beneficial to investigate the impact of these modalities further, with a particular view to explore the impact of specific modality presentations (i.e., written narratives, videos, audio recordings etc.). This will help researchers to identify if specific auditory or visual modality presentations are particularly likely to misinform subjects and whether this needs to be taken into account by legal decision makers. Of particular interest is the enactment modality, as it could be argued enactment most closely resembles an eyewitness being present for and experiencing a crime. Future research exploring enactment would also be relevant in investigating the enactment encoding benefit hypothesis (Dijkstra & Moerman, 2012).

Chapter Three concluded that participants had improved discrimination accuracy when they were presented with a frontal face at encoding compared to a profile posed face. This is the first study of its kind to consider the implications of the viewing angle of faces on eyewitness susceptibility to misinformation. Future research is needed to further our understanding of face angle manipulations and the impact of this on eyewitness' susceptibility to the MI effect. This is of particular importance, as many eyewitnesses will see a perpetrator's face and later be required to identify them. In the USA, it is common for eyewitnesses to be given a photo-lineup prior to a suspect being charged, with a live parade conducted later. Therefore, it is possible that the photo-lineup could be a misinformation stage, whereby eyewitnesses could be introduced to misleading faces. Therefore, future misinformation research should consider the effects of timing of facial presentations and the authority of the individual delivering facial misinformation in the MI paradigm, as these factors have both been influential in misinformation research (Roberts & Powell, 2007; Skagerberg & Wright, 2009).

Conclusion

This thesis explored the impact of factors that can contribute to eyewitness memory inaccuracy, including the MI effect. A systematic review explored the effect of modality congruence across the stages of the MI paradigm (i.e., at encoding, during a subsequent interpolated event, and at test) on misinformation acceptance. Support was found for the encoding specificity hypothesis. Findings regarding specific modalities were mixed and firm conclusions could not be drawn regarding their impact on the MI effect in the MI paradigm. An empirical study was undertaken that investigated whether mistaken lineup identification is influenced by the congruence between the facial angle in which the perpetrator was encoded versus the viewing angle in which an innocent suspect was subsequently encountered prior to the lineup. No statistically significant differences were found between conditions where all three stages were shown in the same facial position or when the encoding and test face position matched, but the innocent face differed. However, findings supported that encoding and being tested on a frontal pose face results in better identification performance than encoding and being tested on a profile posed face. Finally, the reliability and validity of an established psychometric measure of interrogative suggestibility was examined and supported.

In conclusion, it is hoped that this thesis, including the recommendations it makes for practice and future research, will support the ever-growing field of misinformation and eyewitness memory. It is also hoped that the current findings can support legal decision makers to consider the viewing angle from which the perpetrator’s face was encoded, and how this may subsequently impact later identification accuracy.

References

Abeles, P., & Morton, J. (1999). Avoiding misinformation: Reinstating target modality. *The Quarterly Journal of Experimental Psychology: Applied*, *52*(3), 581-592. https://doi/abs/10.1080/713755830

Alemayehu, D., & Zou, K. H. (2012). Applications of ROC analysis in medical research: recent developments and future directions. *Academic radiology*, *19*(12), 1457-1464. https://doi.org/10.1016/j.acra.2012.09.006

Armijo-Olivo, S., Stiles, C. R., Hagen, N. A., Biondo, P. D., & Cummings, G. G. (2012). Assessment of study quality for systematic reviews: A comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: Methodological research. *Journal of evaluation in clinical practice*, *18*(1), 12-18. https://doi/abs/10.1111/j.1365-2753.2010.01516.x

Ayers, M. S., & Reder, L. M. (1998). A theoretical review of the misinformation effect: Predictions from an activation-based memory model. *Psychonomic Bulletin & Review*, *5*(1), 1-21. https://doi.org/10.3758/BF03209454

Bain, S. A., & Baxter, J. S. (2000). Interrogative suggestibility: The role of interviewer behaviour. *Legal and Criminological Psychology*, *5*(1), 123-133. https://doi.org/10.1348/135532500168029

Barber, T. X., & Wilson, S. C. (1978). The Barber Suggestibility Scale and the Creative Imagination Scale: Experimental and clinical applications. *American Journal of Clinical Hypnosis*, *21*, 84-108. https://doi.org/10.1080/00029157.1978.10403966

Baxter, J. S., & Boon, J. C. (2000). Interrogative suggestibility: The importance of being earnest. *Personality and Individual Differences*, *28*(4), 753-762. https://doi.org/10.1016/S0191-8869(99)00136-1

Bekerian, D.A., & Bowers, J.M. (1983). Eyewitness testimony: Were we misled? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 139-145. https://doi.org/10.1037/0278-7393.9.1.139

Bianco, A., & Curci, A. (2015). Measuring interrogative suggestibility with the Italian version of the Gudjonsson Suggestibility Scales (GSS): Factor structure and discriminant validity. *Personality and Individual Differences*, *82*, 258-265. https://doi.org/10.1080/1068316X.2016.1269901

Binet, A. (1900). *La suggestibilite*. Paris: Doin.

Biondi, S., Mazza, C., Orrù G., Monaro, M., Ferracuti, S., Ricci, E., Di Domenico, A., Roma, P. (2020) Interrogative suggestibility in the elderly. PLoS ONE 15(11): e0241353. https://doi.org/10.1371/journal.pone.0241353

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Frontiers in Public Health*, *6*, 149. https://doi.org/10.3389/fpubh.2018.00149

Boon, J. C., & Baxter, J. S. (2000). Minimizing interrogative suggestibility. *Legal and Criminological Psychology*, *5*(2), 273-284. https://doi.org/10.1348/135532500168137

Brubacher, S. P., Sharman, S. J., Scoboria, A., & Powell, M. B. (2020). The effect of question type on resistance to misinformation about present and absent details. *Applied Cognitive Psychology*, *34*(6), 1323-1334. https://doi.org/10.1002/acp.3742

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British journal of psychology*, *73*(1), 105-116. https://doi.org/10.1111/j.2044-8295.1982.tb01795.x

Bruck, M., Ceci, S. J., & Melnyk, L. (1997). External and internal sources of variation in the creation of false reports in children. *Learning and Individual Differences*, *9*(4), 289-316. https://doi.org/10.1016/S1041-6080(97)90011-X

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3-5. https://doi.org/10.1177/1745691610393980

Bülthoff, I., Mohler, B. J., & Thornton, I. M. (2019). Face recognition of full-bodied avatars by active observers in a virtual environment. *Vision research*, *157*, 242-251. https://doi.org/10.1016/j.visres.2017.12.001

Butler, S., Gilchrist, I. D., Burt, D. M., Perrett, D. I., Jones, E., & Harvey, M. (2005). Are the perceptual biases found in chimeric face processing reflected in eye-movement patterns? *Neuropsychologia*, *43*(1), 52-59. https://doi.org/10.1016/j.neuropsychologia.2004.06.005

Calvillo, D. P., & Parong, J. A. (2016). The misinformation effect is unrelated to the DRM effect with and without a DRM warning. *Memory*, *24*(3), 324-333. https://doi.org/10.1080/09658211.2015.1005633

Campbell, J. M., Edwards, M. S., Horswill, M. S., & Helman, S. (2007). Effects of contextual cues in recall and recognition memory: The misinformation effect reconsidered. *British Journal of Psychology*, *98*(3), 485-498. https://doi.org/10.1348/000712606X160768

Cardone, D., & Dent, H. (1996). Memory and interrogative suggestibility: The effects of modality of information presentation and retrieval conditions upon the suggestibility scores of people with learning disabilities. *Legal and Criminological Psychology*, *1*(2), 165-177. https://doi.org/10.1111/j.2044-8333.1996.tb00316.x

Carpenter, C., Webb, C., Overman, A., & Dennis, N. (2022). Contextual similarity negatively affects associative memory performance in both younger and older adults. Department of Psychology, The Pennsylvania State University. https://psyarxiv.com/ytbw4/

Ceci, S. J., & Bruck, M. (1995). *Jeopardy in the courtroom: A scientific analysis of children's testimony*. American Psychological Association. https://doi.org/10.1037/10180-000

Ceci, S. J., & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. *Psychological bulletin*, *113*(3), 403. https://doi.org/10.1037/0033-2909.113.3.403

Ceci, S. J., Ross, D. F., & Toglia, M. P. (1987). Suggestibility of children's memory: Psycholegal implications. *Journal of Experimental Psychology: General*, *116*(1), 38. https://doi.org/10.1037/0096-3445.116.1.38

Chandler, C.C. (1991). How memory for an event is influenced by related events: Interference in modified recognition tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(1), 115-125. https://doi.org/10.1037/0278-7393.17.1.115

Clare, I. C., Gudjonsson, G. H., Rutter, S. C., & Cross, P. (1994). The inter-rater reliability of the Gudjonsson Suggestibility Scale (Form 2). *British Journal of Clinical Psychology, 33*(3), 357-365. https://doi.org/10.1111/j.2044-8260.1994.tb01132.x

Clark, S. E., Erickson, M. A., & Breneman, J. (2011). Probative value of absolute and relative judgments in eyewitness identification. *Law and Human Behavior*, *35*(5), 364-380.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press. Hillsdale, N.J.: Lawrence Erlbaum.

Colloff, M. F., Flowe, H. D., Smith, H. M. J., Seale-Carlisle, T. M., Meissner, C. A., Rockey, J. C., Pande, B., Kujur, P., Parveen, N., Chandel, P., Singh, M. M., Pradhan, S., & Parganiha, A. (2020). *Active exploration of faces in police lineups increases discrimination accuracy for own- and other-race faces* [Manuscript submitted for publication]. School of Psychology, University of Birmingham. https://doi.org/10.31234/osf.io/tvga4

Colloff, M., Seale-Carlisle, T., Karoğlu, N., Rockey, J., Smith, H. M., Smith, L., Maltby, J & Flowe, H. D. (2021). *Enabling witnesses to reinstate perpetrator pose during a lineup test increases accuracy* [Manuscript submitted for publications]. School of Psychology, University of Birmingham.

Colloff, M. F., Wade, K. A., & Strange, D. (2016). Unfair lineups make witnesses more likely to confuse innocent and guilty suspects. *Psychological Science*, *27*(9), 1227-1239. https://doi.org/10.1177/0956797616655789

Colloff, M. F., Wade, K. A., Strange, D., & Wixted, J. T. (2018). Filler-Siphoning theory does not predict the effect of lineup fairness on the ability to discriminate innocent from guilty suspects: Reply to Smith, Wells, Smalarz, and Lampinen (2018). *Psychological Science*, *29*(9), 1552-1557. https://doi.org/10.1177/0956797618786459

Colloff, M. F., & Wixted, J. T. (2020). Why are lineups better than showups? A test of the filler siphoning and enhanced discriminability accounts. *Journal of Experimental Psychology: Applied*, *26*(1), 124-143. https://doi.org/10.1037/xap0000218

Cossins, A. (2006). Prosecuting child sexual assault cases: are vulnerable witness protections enough? *Current Issues in Criminal Justice, 18*, 299–317. https://doi.org/10.1080/10345329.2006.12036391

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281-284. https://doi.org/10.1037/h0040957

Curci, A., Bianco, A., & Gudjonsson, G. H. (2017). Verbal ability, depression, and anxiety as correlates of interrogative suggestibility in children exposed to life adversities. *Psychology, Crime & Law*, *23*(5), 445-458. https://doi.org/10.1080/1068316X.2016.1269901

Dalton, A. L., & Daneman, M. (2006). Social suggestibility to central and peripheral misinformation. *Memory*, *14*(4), 486-501. https://doi.org/10.1080/09658210500495073

Danielsdottir, G., Sigurgeirsdottir, S., Einarsdottir, H. R., & Haraldsson, E. (1993). Interrogative suggestibility in children and its relationship with memory and vocabulary. *Personality and Individual Differences, 14*, 499-502. https://doi.org/10.1016/0191-8869(93)90323-U

Davies, E., Devere, H., & Verbitsky, J. (2004). Court education for young witnesses: evaluation of the pilot service in Aotearoa, New Zealand. *Psychiatry, Psychology and Law*, *11*, 226–235. https://doi.org/10.1375/pplt.2004.11.2.226

Davies, G., Shepherd, J., Ellis, H. (1979). Effects of interpolated mugshot exposure on accuracy of eyewitness identification. *Journal of Applied Psychology*, *64*(2), 232-237. https://doi.org/10.1037/0021-9010.64.2.232

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The journal of abnormal and social psychology*, *51*(3), 629. https://doi.org/10.1037/h0046408

Dienes, Z., Brown, E., Hutton, S., Kirsch, I., Mazzoni, G., & Wright, D. B. (2009). Hypnotic suggestibility, cognitive inhibition, and dissociation. *Consciousness and cognition*, *18*(4), 837-847. https://doi.org/10.1016/j.concog.2009.07.009

Dijkstra, K., & Moerman, E. M. (2012). Effects of modality on memory for original and misleading information. *Acta psychologica*, *140*(1), 58-63. https://doi.org/10.1016/j.actpsy.2012.02.003

Drake, K. E., Bull, R., & Boon, J. C. (2008). Interrogative suggestibility, self-esteem, and the influence of negative life-events. *Legal and Criminological Psychology*, *13*(2), 299-307. https://doi.org/10.1348/135532507X209981

Ecker, U. K., Lewandowsky, S., Swire, B., & Chang, D. (2011). Correcting false information in memory: Manipulating the strength of misinformation encoding and its retraction. *Psychonomic Bulletin & Review*, *18*(3), 570-578. https://doi.org/10.3758/s13423-011-0065-1

Egger, M., Juni, P., Bartlett, C., Holenstein, F., & Sterne, J. (2003). How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technology Assessment*, *7*(1), 1-76.

Eisen, M. L., & Carlson, E. B. (1998). Individual differences in suggestibility: Examining the influence of dissociation, absorption, and a history of childhood abuse. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *12*(7), S47-S61. https://doi.org/10.1002/(SICI)1099-0720(199812)12:7<S47::AID-ACP598>3.0.CO;2-P

Engelkamp, J., & Cohen, R. L. (1991). Current issues in memory of action events. *Psychological Research*, *53*(3), 175-182.

Farah, M.J., Wilson, K.D., Drain, H.M., Tanaka, J.W. (1998). What is “special” about face perception? *Psychological Review*, *105*(3), 482-498. https://doi.org/10.1037/0033-295X.105.3.482

Favelle, S., Hill, H., & Claes, P. (2017). About face: matching unfamiliar faces across rotations of view and lighting. *i-Perception*, *8*(6), 2041669517744221. https://doi.org/10.1177/2041669517744221

Field, A. (2000). *Discovering statistics using SPSS for Windows*. London: Sage.

Findley, K. A. (2001). Learning from our mistakes: A criminal justice commission to study wrongful convictions. *Cal. WL Rev*, *38*, 333.

Fitzgerald, R. J., Price, H. L., & Valentine, T. (2018). Eyewitness identification: Live, photo, and video lineups. *Psychology, Public Policy, and Law*, *24*(3), 307-325. http://dx.doi.org/10.1037/law0000164

Fitzgerald, R. J., Rubínová, E., & Juncu, S. (2021). Eyewitness identification around the world. In *Methods, measures, and theories in eyewitness identification tasks* (pp. 294-322). Routledge.

Fletcher, P. C., Frith, C. D., Grasby, P. M., Shallice, T., Frackowiak, R. S. J., & Dolan, R. J. (1995). Brain systems for encoding and retrieval of auditory—verbal memory. An in vivo study in humans. *Brain*, *118*(2), 401-416. https://doi.org/10.1093/brain/118.2.401

Foos, P., & Goolkasian, P. (2005). Presentation format effects in working memory: The role of attention. *Memory & Cognition*, *3*, 499–513. https://doi.org/10.3758/BF03193067

Forgas, J. P., Laham, S. M., & Vargas, P. T. (2005). Mood effects on eyewitness memory: Affective influences on susceptibility to misinformation. *Journal of Experimental Social Psychology*, *41*(6), 574-588. https://doi.org/10.1016/j.jesp.2004.11.005

Forina, M., Armanino, C., Lanteri, S., & Leardi, R. (1989). Methods of varimax rotation in factor analysis with applications in clinical and food chemistry. *Journal of Chemometrics*, *3*(S1), 115-125. https://doi.org/10.1002/cem.1180030504

Foster v. California, 394 U.S. 440 (1969). https://scholar.google.co.uk/scholar_case?case=9759772205335000085&q=foster+v. +california&hl=en&as_sdt=2006&as_vis=1

Foster, J. L., Huthwaite, T., Yesberg, J. A., Garry, M., & Loftus, E. F. (2012). Repetition, not number of sources, increases both susceptibility to misinformation and confidence in the accuracy of eyewitnesses. *Acta psychologica*, *139*(2), 320-326. https://doi.org/10.1016/j.actpsy.2011.12.004

Fraser, I. H., Craig, G. L. & Parker, D. M. (1990) Reaction time measures of features saliency in schematic faces. *Perception, 19*(5), 661-673. https://doi.org/10.1068/p190661

Frenda, S. J., Nichols, R. M., & Loftus, E. F. (2011). Current issues and advances in misinformation research. *Current Directions in Psychological Science*, *20*(1), 20-23.

Frenda, S. J., Patihis, L., Loftus, E. F., Lewis, H. C., & Fenn, K. M. (2014). Sleep deprivation and false memories. *Psychological Science, 25*(9), 1674-1681. https://doi.org/10.1177/0956797614534694

Frumkin, I. B., Lally, S. J., & Sexton, J. E. (2012). A United States forensic sample for the Gudjonsson suggestibility scales. *Behavioral Sciences & The Law*, *30*(6), 749-763. https://doi.org/10.1002/bsl.2032

Gallo, D. A., McDermott, K. B., Percer, J. M., & Roediger,H. L. III. (2001). Modality effects in false recall and false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 339–353. https://doi.org/10.1037/0278-7393.27.2.339

Garrett, B. L., Liu, A., Kafadar, K., Yaffe, J., & Dodson, C. S. (2020). Factoring the Role of Eyewitness Evidence in the Courtroom. *Journal of Empirical Legal Studies*, *17*(3), 556-579. https://doi.org/10.1111/jels.12259

Garrison, J. R., Bond, R., Gibbard, E., Johnson, M. K., & Simons, J. S. (2017). Monitoring what is real: The effects of modality and action on accuracy and type of reality monitoring error. *Cortex*, *87*, 108-117. https://doi.org/10.1016/j.cortex.2016.06.018

Gheorghiu, V., Meiu, G., Negrea, G., Onofrei, A., Pătraşcu, F., & Pungă, C. (1966). Experimental research on suggestibility of patients with asthenic neurosis. *Neurologia, psihiatria, neurochirurgia*, *11*(4), 289-294.

Gignac, G. E., & Powell, M. B. (2009). A psychometric evaluation of the Gudjonsson Suggestibility Scales: Problems associated with measuring suggestibility as a difference score composite. *Personality and Individual Differences*, *46*(2), 88-93. https://doi.org/10.1016/j.paid.2008.09.007

Gilbert, C., & Bakan, P. (1973). Visual asymmetry in perception of faces. *Neuropsychologia*, *11*(3), 355-362. https://doi.org/10.1016/0028-3932(73)90049-3

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and instruction*, *15*(4), 313-331. https://doi.org/10.1016/j.learninstruc.2005.07.001

Gobbo, C., Mega, C., & Pipe, M. E. (2002). Does the nature of the experience influence suggestibility? A study of children's event memory. *Journal of Experimental Child Psychology*, *81*(4), 502-530. https://doi.org/10.1006/jecp.2002.2662

Goldstein, A. G., & Mackenberg, E. J. (1966). Recognition of human faces from isolated facial features: A developmental study. *Psychonomic Science*, *6*(4), 149-150. https://doi.org/10.3758/BF03328001

Goodwin, K. A., Kukucka, J. P., & Hawks, I. M. (2013). Co-witness confidence, conformity, and eyewitness memory: An examination of normative and informational social influences. *Applied Cognitive Psychology*, *27*(1), 91-100. https://doi.org/10.1002/acp.2877

Goolkasian, P., & Foos, P. W. (2002). Presentation format and its effect on working memory. *Memory & Cognition*, *30*, 1096–1105. https://doi.org/10.3758/BF03194327

Grabner, S. S. (2018). *Assessing Instructed Feigner's Response to the Gudjonsson Suggestibility Scale* [Doctoral dissertation, Nova Southeastern University].

Grasby, P. M., Frith, C. D., Friston, K. J., Bench, C. R. S. F., Frackowiak, R. S. J., & Dolan, R. J. (1993). Functional mapping of brain areas implicated in auditory—verbal memory function. *Brain*, *116*(1), 1-20. https://doi.org/10.1093/brain/116.1.1

Greene, R. L. (1992). *Human memory: Paradigms and paradoxes.* Hillsdale, NJ: Erlbaum

Greenwald, A. G., Pratkanis, A. R., Leippe, M. R., & Baumgardner, M. H. (1986). Under what conditions does theory obstruct research progress?. *Psychological review*, *93*(2), 216. https://doi.org/10.1037/0033-295X.93.2.216

Grisso, T. (1986). *Evaluating competencies. Forensic assessments and instruments*. New York: Plenum Press.

Gronlund, S. D., & Neuschatz, J. S. (2014). Eyewitness identification discriminability: ROC analysis versus logistic regression. *Journal of Applied Research in Memory and Cognition*, *3*(2), 54-57. https://doi.org/10.1016/j.jarmac.2014.04.008

Gudjonsson, G. H. (1984). A new scale of interrogative suggestibility. *Personality and Individual Differences*, *5*, 303-314. https://doi.org/10.1016/0191-8869(84)90069-2

Gudjonsson, G. H. (1987a). A parallel form of the Gudjonsson Suggestibility Scale. *British Journal of Clinical Psychology*, *26*(3), 215-221. https://doi.org/10.1111/j.2044-8260.1987.tb01348.x

Gudjonsson, G. H. (1987b). Historical background to suggestibility: How interrogative suggestibility differs from other types of suggestibility. *Personality and Individual Differences*, *8*(3), 347-355. https://doi.org/10.1016/0191-8869(87)90035-3

Gudjonsson, G. H. (1992b). Interrogative suggestibility: Factor analysis of the Gudjonsson Suggestibility Scale (GSS 2). *Personality and Individual Differences*, *13*, 479-481. https://doi.org/10.1016/0191-8869(92)90077-3

Gudjonsson, G. H. (2010). *Interrogative suggestibility and false confessions.* In J. M. Brown & E. A. Campbell (Eds.), *The Cambridge handbook of forensic psychology* (p. 202–207). Cambridge University Press. https://doi.org/10.1017/CBO9780511730290.025

Gudjonsson, G. H. (1983). Suggestibility, intelligence, memory recall and personality: An experimental study. *The British Journal of Psychiatry*, *142*, 35-37. https://doi.org/10.1192/bjp.142.1.35

Gudjonsson, G. H. (1997). *The Gudjonsson suggestibility scales manual.* Hove: Psychology Press.

Gudjonsson, G. H. (2003). *The psychology of interrogations and confessions: A handbook*. John Wiley & Sons.

Gudjonsson, G. H. (1992a). *The psychology of interrogations, confessions and testimony*. Chichester: Wiley.

Gudjonsson, G. H. Clare, I. C. H., Rutter, S., & Pearse, J. (1993). *Persons at risk during interviews in police custody: The identification of vulnerabilities.* Royal Commission on Criminal Justice. London: London: H.M.S.O

Gudjonsson, G. H., & Clark, N. K. (1986). Suggestibility in police interrogation: A social psychological model. *Social Behaviour*, *1*(2), 83-104.

Gudjonsson, G. H., & Lister, S. (1984). Interrogative suggestibility and its relationship with self-esteem and control. *Journal of the Forensic Science Society*, *24*(2), 99-110. https://doi.org/10.1016/S0015-7368(84)72302-4

Gudjonsson, G. H., & Sigurdsson, J. F. (1999). The Gudjonsson Confession Questionnaire-Revised (GCQ-R) factor structure and its relationship with personality. *Personality and Individual Differences*, *27*(5), 953-968. https://doi.org/10.1016/S0191-8869(98)00278-5

Gudjonsson, G. H., & Sigurdsson, J. F. (2004). The relationship of suggestibility and compliance with self-deception and other-deception. *Psychology, Crime & Law*, *10*(4), 447-453. https://doi.org/10.1080/10683160310001634278

Gudjonsson, G. H., & Sigurdsson, J. F. (1995). The relationship of confabulation to the memory, intelligence, suggestibility, and personality of juvenile offenders. *Nordic Journal of Psychiatry*, *49*(5), 373-378. https://doi.org/10.3109/08039489509011930

Gudjonsson, G. H., & Singh, K. K. (1984). Interrogative suggestibility and delinquent boys: An empirical validation study. *Personality and Individual Differences*, *5*(4), 425-430. https://doi.org/10.1016/0191-8869(84)90007-2

Gudjonsson, G., Vagni, M., Maiorano, T., & Pajardi, D. (2016.) Age and memory related changes in children's immediate and delayed suggestibility using the Gudjonsson Suggestibility Scale. *Personality and Individual Differences*, *102*, 25-29. https://doi.org/10.1016/j.paid.2016.06.029

Guy, J., Habak, C., Wilson, H. R., Mottron, L., & Bertone, A. (2017). Face perception develops similarly across viewpoint in children and adolescents with and without autism spectrum disorder. *Journal of Vision*, *17*(1), 38-38. https://doi.org/10.1167/17.1.38

Hämmerer, D., Hopkins, A., Betts, M. J., Maaß, A., Dolan, R. J., & Düzel, E. (2017). Emotional arousal and recognition memory are differentially reflected in pupil diameter responses during emotional memory for negative events in younger and older adults. *Neurobiology of aging*, *58*, 129-139. https://doi.org/10.1016/j.neurobiolaging.2017.06.021

Hansen, I., Smeets, T., & Jelicic, M. (2010). Further data on interrogative suggestibility and compliance scores following instructed malingering. *Legal and Criminological Psychology*, *15*(2), 221-228. https://doi.org/10.1348/135532509X447796

Hardt, O., Einarsson, E. Ö., & Nader, K. (2010). A bridge over troubled water: Reconsolidation as a link between cognitive and neuroscientific memory research traditions. *Annual Review of Psychology, 61*,141–167. https://doi/abs/10.1146/annurev.psych.093008.100455

Hartling, L., Featherstone, R., Nuspl, M., Shave, K., Dryden, D. M., & Vandermeer, B. (2017). Grey literature in systematic reviews: A cross-sectional study of the contribution of non-English reports, unpublished studies and dissertations to the results of meta-analyses in child-relevant reviews. *BMC medical research methodology*, *17*(1), 64. https://doi.org/10.1186/s12874-017-0347-z

Harwell, M. R., & Gatti, G. G. (2001). Rescaling ordinal data to interval data in educational research. *Review of Educational Research*, *71*(1), 105-131. https://doi.org/10.3102/00346543071001105

Hautus, M., Macmillan, N. A., & Rotello, C. M. (2008). Toward a complete decision model of item and source recognition. Psychonomic Bulletin & Review, 15, 889-905. https://doi.org/10.3758/PBR.15.5.889

Heathcote, A. (2003). Item recognition memory and the ROC. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1210-1230. https://doi.org/10.1037/0278-7393.29.6.1210

Hekkanen, S. T., & McEvoy, C. (2002). False memories and source-monitoring problems: criterion differences. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *16*(1), 73-85. https://doi.org/10.1002/acp.753

Hendrich, M. A. (2019). *Modality Effects in False Memory Production Using the Misinformation Paradigm* [Doctoral dissertation, Marietta College].

Henkel, L. A., & Coffman, K. J. (2004). Memory distortions in coerced false confessions: A source monitoring framework analysis. *Applied Cognitive Psychology, 18*(5), 567-588. https://doi.org/10.1002/acp.1026

Henson, R. N., & Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus*, *20*(11), 1315-1326. https://doi.org/10.1002/hipo.20857

Hergovich, A., Schott, R., & Burger, C. (2010). Biased evaluation of abstracts depending on topic and conclusion: Further evidence of a confirmation bias within scientific psychology. *Current Psychology*, *29*(3), 188-209.

Hess, T. M., Popham, L. E., Emery, L., & Elliott, T. (2012). Mood, motivation, and misinformation: Aging and affective state influences on memory. *Aging, Neuropsychology, and Cognition*, *19*(1-2), 13-34. https://doi.org/10.1080/13825585.2011.622740

Hill, H., Schyns, P. G., & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition*, *62*(2), 201-222. https://doi.org/10.1016/S0010-0277(96)00785-8

Hooper, V. R., Chou, S., & Browne, K. D. (2016). A systematic review on the relationship between self-esteem and interrogative suggestibility. *The Journal of Forensic Psychiatry & Psychology*, *27*(6), 761-785. https://doi.org/10.1080/14789949.2016.1201844

Horsley, T., Dingwall, O., & Sampson, M. (2011). Checking reference lists to find additional studies for systematic reviews. *Cochrane Database of Systematic Reviews*, (8). https://doi/10.1002/14651858.MR000026.pub2/abstract

Hultsch, D. F., MacDonald, S. W., Hunter, M. A., Maitland, S. B., & Dixon, R. A. (2002). Sampling and generalisability in developmental research: Comparison of random and convenience samples of older adults. *International Journal of Behavioral Development*, *26*(4), 345-359. https://doi/abs/10.1080/01650250143000247

Humphries, J. E., & Flowe, H. D. (2015). Receiver operating characteristic analysis of age-related changes in lineup performance. *Journal of experimental child psychology*, *132*, 189-204. https://doi.org/10.1016/j.jecp.2014.12.009

Innocence Project. (2017) Eyewitness Misidentification. Retrieved from https://www.innocenceproject.org/causes/eyewitness-misidentification. Accessed 27th October, 2019.

Jenkins, R., & Burton, A. M. (2008). Limitations in facial identification: The evidence. *Justice of the Peace*, *172*, 4-6.

Johnson, M.K., Hashtroudi, S., & Lindsay, D.S. (1993). Source monitoring. *Psychological Bulletin, 114*(1), 3-28.

Johnson, M. K., Nolde, S. F., & De Leonardis, D. M. (1996). Emotional focus and source monitoring. *Journal of Memory and Language, 35*(2), 135-156. https://doi.org/10.1006/jmla.1996.0008

Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence–accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1304-1316. https://doi.org/10.1037/0278-7393.22.5.1304

Kebbell, M. R., Wagstaff, G. F., & Covey, J. A. (1996). The influence of item difficulty on the relationship between eyewitness confidence and accuracy. *British Journal of Psychology*, *87*(4), 653-662. https://doi.org/10.1111/j.2044-8295.1996.tb02614.x

Kelman, H. C. (1950). Effects of success and failure on" suggestibility" in the autokinetic situation. *The Journal of Abnormal and Social Psychology*, *45*(2), 267. https://doi.org/10.1037/h0062561

Kempen, K., & Tredoux, C. G. (2012). 'Seeing is believing': the effect of viewing and constructing a composite on identification performance. *South African Journal of Psychology*, *42*(3), 434-444. https://doi.org/10.1177/008124631204200315

Kiat, J. E. (2018). *The Role of Visual and Verbal Processes in False Memory Susceptibility on the Misinformation Effect* [Doctoral dissertation, University of Nebraska]. *

Kline, P. (1986). *A handbook of test construction*. New York: Metheuen.

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual review of psychology*, *51*(1), 481-537. https://doi.org/10.1146/annurev.psych.51.1.481

Leder, H., & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *The Quarterly Journal of Experimental Psychology*, 53A, 513-536. https://doi.org/10.1080/713755889

Lee, K. (2004). Age, neuropsychological, and social cognitive measures as predictors of individual differences in susceptibility to the misinformation effect. *Applied Cognitive Psychology*, *18*(8), 997-1019. https://doi.org/10.1002/acp.1075

Lee, T. D., & Hirota, T. T. (1980). Encoding specificity principle in motor short-term memory for movement extent. *Journal of Motor Behavior*, *12*(1), 63-67. https://doi.org/10.1080/00222895.1980.10735206

Lee, Y., Matsumiya, K., & Wilson, H. R. (2006). Size-invariant but viewpoint-dependent representation of faces. *Vision Research*, *46*(12), 1901-1910. https://doi.org/10.1016/j.visres.2005.12.008

Lindsay, D. S., & Johnson, M. K. (2000). False memories and the source monitoring framework: Reply to Reyna and Lloyd (1997). *Learning and Individual Differences*, *12*(2), 145-161. https://doi.org/10.1016/S1041-6080(01)00035-8

Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition*, *17*(3), 349-358.

Liu, C. H., & Chaudhuri, A. (1997). Face recognition with multi-tone and two-tone photographic negatives. *Perception*, *26*(10), 1289-1296. https://doi.org/10.1068/p261289

Lo, J. C., Chong, P. L. H., Ganesan, S., Leong, R. L. F., & Chee, M. W. L. (2016). Sleep deprivation increases formation of false memory. *Journal of Sleep Research*, *25*(6), 673- 682. https://doi.org/10.1111/jsr.12436

Loftus, E. F. (1979a). *Eyewitness testimony*. London: Harvard University press.

Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive psychology*, *7*(4), 560-572. https://doi.org/10.1016/0010-0285(75)90023-7

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, *12*, 361-366.

Loftus, E. F. (1979b). Reactions to blatantly contradictory information. *Memory and Cognition*, *7*, 368-374. https://doi.org/10.3758/BF03196941

Loftus, E. F. (1986). Ten years in the life of an expert witness. *Law and Human Behavior*, *10*(3), 241-263.

Loftus, E. F., & Greene, E. (1980). Warning: Even memory for faces may be contagious. *Law and Human Behavior*, *4*(4), 323-334. https://doi.org/ 10.1007/BF01040624

Loftus, E. F., Hoffman, H. G., & Wagenaar, W. A. (1992). The misinformation effect: Transformations in memory induced by postevent information. In *Development of long-term retention* (pp. 159-183). Springer, New York. https://doig.org/10.1007/978-1-4612-2868-4_5

Loftus, E.F., Miller, D.G., Burns, H.J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory*,*4*(1), 19-31. https://doi.org/10.1037/0278-7393.4.1.19

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, *13*(5), 585-589. https://doi.org/10.1016/S0022-5371(74)80011-3

Loftus, E. F., & Zanni, G. (1975). Eyewitness testimony: The influence of the wording of a question. *Bulletin of the Psychonomic Society*, *5*, 86-88. https://doi.org/10.1016/S0022-5371(74)80011-3

London, P. (1962). *Children's Hypnotic Susceptibility Scale.* Consulting Psychologists Press, Palo Alto, California, USA.

Longmore, C. A., Liu, C. H., & Young, A. W. (2008). Learning faces from photographs. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(1), 77. https://doi.org/10.1037/0096-1523.34.1.77

Lumbreras, B., Porta, M., Márquez, S., Pollán, M., Parker, L. A., & Hernández-Aguado, I. (2008). QUADOMICS: An adaptation of the Quality Assessment of Diagnostic Accuracy Assessment (QUADAS) for the evaluation of the methodological quality of studies on the diagnostic accuracy of '-omics'-based technologies. *Clinical biochemistry*, *41*(16-17), 1316-1325. https://doi.org/10.1016/j.clinbiochem.2008.06.018

MacLeod, M. (2002). Retrieval-induced forgetting in eyewitness memory: Forgetting as a consequence of remembering. *Applied Cognitive Psychology: The Official Journal of* *the Society for Applied Research in Memory and Cognition*, *16*(2), 135-149. https://doi.org/10.1002/acp.782

MacLin, M. K., Zimmerman, L. A., Meissner, C. A., MacLin, O. H., Tredoux, C. G., & Malpass, R. S. (2008). The science of collecting eyewitness evidence: Recommendations and an argument for collaborative efforts between researchers and law enforcement. *Police Practice & Research*.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. Behavior Research Methods, 44(1), 1-23. https://doi.org/10.3758/s13428-011-0124-6

Mastroberardino, S., & Marucci, F. S. (2013). Interrogative suggestibility: Was it just compliance or a genuine false memory? *Legal and Criminological Psychology*, *18*(2), 274-286. https://doi.org/10.1111/j.2044-8333.2012.02048.x

Maylor, E. A., & Mo, A. (1999). Effects of study-test modality on false recognition. *British Journal of Psychology*, *90*(4), 477-493. https://doi.org/10.1348/000712699161567 *

McAllister, H. A., Dale, R. H., Bregman, N. J., McCabe, A., & Cotton, C. R. (1993). When eyewitnesses are also earwitnesses: Effects on visual and voice identifications. *Basic and Applied Social Psychology*, *14*(2), 161-170. https://doi.org/10.1207/s15324834basp1402_3

McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General*, *114*(1), 1-16.

McKelvie, S. J. (1976). The role of eyes and mouth in the memory of a face. *The American Journal of Psychology*, 311-323.

Meinhardt-Injac, B., Meinhardt, G., & Schwaninger, A. (2009). Does matching of internal and external facial features depend on orientation and viewpoint?. *Acta Psychologica*, *132*(3), 267-278. https://doi.org/10.1016/j.actpsy.2009.07.011

Meisters, J., Diedenhofen, B., & Musch, J. (2018). Eyewitness identification in simultaneous and sequential lineups: An investigation of position effects using receiver operating characteristics. *Memory*, *26*(9), 1297-1309. https://doi.org/10.1080/09658211.2018.1464581

Melnyk, L., & Bruck, M. (2004). Timing moderates the effects of repeated suggestive interviewing on children's eyewitness memory. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *18*(5), 613-631. https://doi.org/10.1002/acp.1013

Meltzer, M. A., & Bartlett, J. C. (2019). Holistic processing and unitization in face recognition memory. *Journal of Experimental Psychology: General*, *148*(8), 1386-1406. https://doi.org/10.1037/xge0000640

Merckelbach, H., Muris, P., Wessel, I., & Van Koppen, P. J. (1998). correlates. *Social Behavior and Personality: an international journal*, *26*(2), 203-209. https://doi.org/10.2224/sbp.1998.26.2.203

Methley, A. M., Campbell, S., Chew-Graham, C., McNally, R., & Cheraghi-Sohi, S. (2014). PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. *BMC health services research*, *14*(1). https://doi.org/1-10. 10.1186/s12913-014-0579-0

Mickes, L. (2015). Receiver operating characteristic analysis and confidence–accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory and Cognition*, *4*(2), 93-102. https://doi.org/10.1016/j.jarmac.2015.01.003

Mickes, L., Flowe, H. D., & Wixted, J. T. (2012). Receiver operating characteristic analysis of eyewitness memory: comparing the diagnostic accuracy of simultaneous versus sequential lineups. *Journal of Experimental Psychology: Applied*, *18*(4), 361-376. https://doi.org/10.1037/a0030609

Mickes, L., Johnson, E. M., & Wixted, J. T. (2010). Continuous recollection vs. unitized familiarity in associative recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 843-863. https://doi.org/10.1037/a0019755

Miles, K. L., Powell, M. B., Gignac, G. E., & Thomson, D. M. (2007). How well does the Gudjonsson Suggestibility Scale for Children, version 2 predict the recall of false details among children with and without intellectual disabilities? *Legal and Criminological Psychology*, *12*(2), 217-232. https://doi.org/10.1348/135532506X116750

Milne, B., Roberts, K. A., Hill, J. A., & Moston, S. (2011). Police perceptions of investigative interviewing: Training needs and operational practices in Australia. *The British journal of forensic practice*. https://doi.org/10.1108/14636641111134314

Mitchell, K. J., & Johnson, M. K. (2009). Source monitoring 15 years later: What have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin*, *135*(4), 638-677. https://doi.org/10.1037/a0015849

Mitchell, K. J., & Zaragoza, M. S. (1996). Repeated exposure to suggestion and false memory: The role of contextual variability. *Journal of Memory and Language*, *35*(2), 246-260. https://doi.org/10.1006/jmla.1996.0014

Mojtahedi, D. (2017). New research reveals how little we can trust eyewitnesses. *The Conversation*.

Morton, J., Hammersley, R. H., & Bekerian, D. A. (1985). Headed records: A model for memory and its failures. *Cognition*, *20*(1), 1-23. https://doi.org/10.1016/0010-0277(85)90002-2

Mudd, K., & Govern, J. M. (2004). Conformity to Misinformation and Time Delay Negatively Affect Eyewitness Confidence and Accuracy. *North American Journal of Psychology*, *6*(2), 227-238.

Murphy, K. R., & Davidshofer, C. O. (2005). *Psychological testing: principles and applications*. Upper Saddle River, N.J.: Pearson/Prentice.

Nguyen, T. B., Pezdek, K., & Wixted, J. T. (2017). Evidence for a confidence–accuracy relationship in memory for same-and cross-race faces. *Quarterly Journal of Experimental Psychology*, *70*(12), 2518-2534. https://doi.org/10.1080/17470218.2016.1246578

Nichols, R. M., & Loftus, E. F. (2019). Who is susceptible in three false memory tasks?. *Memory*, *27*(7), 962-984. https://doi.org/10.1080/09658211.2019.1611862

Nishiyama, R., Hirano, T., & Ukita, J. (2017). Usage of semantic representations in recognition memory. *Memory*, *25*(10), 1412-1424. https://doi.org/10.1080/09658211.2017.1310252

Nyberg, L., Nilsson, L. G., & Bäckman, L. (1991). A component analysis of action events. *Psychological Research*, *53*(3), 219-225.

O'Connor, P. J. (1990). Normative data: their definition, interpretation, and importance for primary care physicians. *Family Medicine*, *22*(4), 307.

Okado, Y., & Stark, C. E. (2005). Neural activity during encoding predicts false memories created by misinformation. *Learning & memory*, *12*(1), 3-11. https://doi.org/10.1101/lm.87605

Olszewska, J. M., Reuter-Lorenz, P. A., Munier, E., & Bendler, S. A. (2015). Misremembering what you see or hear: Dissociable effects of modality on short-and long-term false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(5), 1316. https://doi.org/10.1037/xlm0000115

Ost, J., Blank, H., Davies, J., Jones, G., Lambert, K.,Salmon, K., & Lu, L. (2013). False memory≠ false memory: DRM errors are unrelated to the misinformation effect. *PloS one*, *8*(4).

Otgaar, H., Howe, M. L., Brackmann, N., & Smeets, T. (2016). The malleability of developmental trends in neutral and negative memory illusions. *Journal of Experimental Psychology: General*, *145*(1), 31. http://dx.doi.org/10.1037/xge0000127

Otgaar, H., Howe, M. L., Merckelbach, H., & Muris, P. (2018). Who is the better eyewitness? Sometimes adults but at other times children. *Current Directions in Psychological Science*, *27*(5), 378-385. https://doi.org/10.1177/0963721418770998

Otgaar, H., Verschuere, B., Meijer, E. H., & van Oorsouw, K. (2012). The origin of children’s implanted false memories: Memory traces or compliance? *Acta Psychologica*, *139*, 397–403. https://doi.org/10.1016/j.actpsy.2012.01.002

O'toole, A. J., Edelman, S., & Bülthoff, H. H. (1998). Stimulus-specific effects in face recognition over changes in viewpoint. *Vision research*, *38*(15-16), 2351-2363. https://doi.org/10.1016/S0042-6989(98)00042-X

Paez, A. (2017). Gray literature: An important resource in systematic reviews. *Journal of Evidence-Based Medicine*, *10*(3), 233-240. https://doi.org/10.1111/jebm.12266

Patihis, L., Frenda, S. J., LePort, A. K., Petersen, N., Nichols, R. M., Stark, C. E., McGaugh, J. L., & Loftus, E. F. (2013). False memories in highly superior autobiographical memory individuals. *Proceedings of the National Academy of Sciences*, *110*(52), 20947-20952. https://doi.org/10.1073/pnas.1314373110

Patterson, D. (2011). The impact of detectives' manner of questioning on rape victims' disclosure. *Violence against women*, *17*(11), 1349-1373. https://doi.org/10.1177/1077801211434725

Penney, C. J. (1989). Modality effect and the structure of short- term verbal memory. *Memory & Cognition*, *17*, 398–422. doi.org /10.3758/BF03202613

Pickrell, J. E., McDonald, D. L., Bernstein, D. M., & Loftus, E. F. (2016). Misinformation effect. In *Cognitive illusions* (pp. 416-433). Hove, United Kingdom & New York: Psychology Press.

Pierce, B. H., & Gallo, D. A. (2011). Encoding modality can affect memory accuracy via retrieval orientation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 516. https://doi.org/10.1037/a0022217

Pires, R., Silva, D. R., & Ferreira, A. S. (2013). Portuguese adaptation of the Gudjonsson Suggestibility Scales (GSS1 and GSS2): Empirical findings. *Personality and Individual Differences*, *54*(2), 251-255. https://doi.org/10.1016/j.paid.2012.09.008

Polczyk, R. (2005). Interrogative suggestibility: Cross-cultural stability of psychometric and correlational properties of the Gudjonsson Suggestibility Scales. *Personality and Individual Differences*, *38*(1), 177-186. https://doi.org/10.1016/j.paid.2004.03.018

Polczyk, R., & Pasek, T. (2006). Types of suggestibility: relationships among compliance, indirect, and direct suggestibility. *International Journal of Clinical and Experimental Hypnosis*, *54*(4), 392-415. https://doi.org/10.1080/00207140600856764

Police and Criminal Evidence Act 1984, Code D. (2017) Retrieved from https://www.gov.uk/government/publications/pace-code-d-2017

Pollard, R., Trowbridge, B., Slade, P. D., Streissguth, A. P., Laktonen, A., & Townes, B. D. (2004). Interrogative suggestibility in a US context: Some preliminary data on normal subjects. *Personality and Individual Differences*, *37*(5), 1101-1108. https://doi.org/10.1016/j.paid.2003.12.004

Porter, S., Spencer, L., & Birt, A. R. (2003). Blinded by emotion? Effect of the emotionality of a scene on susceptibility to false memories. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *35*(3), 165. https://doi.org/10.1037/h0087198

Powell, M. (2005). Improving the reliability of child witness testimony in court: the importance of focusing on questioning techniques. *Current Issues in Criminal Justice*, *17*, 137–143. https://doi.org/10.1080/10345329.2005.12036342

Randell, I., Seymour, F., Henderson, E., & Blackwell, S. (2018). The experiences of young complainant witnesses in criminal court trials for sexual offences. *Psychiatry, psychology and law*, *25*(3), 357-373. https://doi.org/10.1080/13218719.2017.1396866

Raymond, A. (2020). *How Individual Differences in Psychosocial Factors Affect Interrogative Suggestibility and False Memory* [Doctoral dissertation, Southern Illinois University at Edwardsville].

Reyna, V. F., & Lloyd, F. (1997). Theories   false memory in children and adults. *Learning and individual differences*, *9*(2), 95-123. https://doi.org/10.1016/S1041-6080(97)90002-9

Richardson, G., & Smith, P. (1993). The inter-rater reliability of the Gudjonsson Suggestibility Scale. *Personality and Individual Differences*, *14*(1), 251-253. https://doi.org/10.1016/0191-8869(93)90197-B

Roberts, K. P., & Powell, M. B. (2007). The roles of prior experience and the timing of misinformation presentation on young children's event memories. *Child Development*, *78*(4), 1137-1152. https://doi.org/10.1111/j.1467-8624.2007.01057.x

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC bioinformatics*, *12*(1), 1-8.

Roebers, C. M., Gelhaar, T., & Schneider, W. (2004). "It's magic!" The effects of presentation modality on children's event memory, suggestibility, and confidence judgments. *Journal of Experimental Child Psychology*, *87*(4), 320-335. https://doi.org/10.1016/j.jecp.2004.01.004

Roebers, C. M., & Schneider, W. (2000). The impact of misleading questions on eyewitness memory in children and adults. *Applied Cognitive Psychology: The Official Journal of* *the Society for Applied Research in Memory and Cognition*, *14*(6), 509-526. https://doi.org/10.1002/1099-0720(200011/12)14:6<509::AID-ACP668>3.0.CO;2-W

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803-814.

Roma, P., Sabatello, U., Verrastro, G., & Ferracuti, S. (2011). Comparison between Gudjonsson Suggestibility Scale 2 (GSS2) and Bonn Test of Statement Suggestibility (BTSS) in measuring children's interrogative suggestibility. *Personality and Individual Differences*, *51*(4), 488-491. https://doi.org/10.1016/j.paid.2011.05.003

Rotello, C. M., & Chen, T. (2016). ROC curve analyses of eyewitness identification decisions: An analysis of the recent debate. *Cognitive Research: Principles and Implications*, *1*(1), 1-12.

Sanderson S., Tatt I. D., Higgins J. P. (2007). Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: A systematic review and annotated bibliography. *International Journal of Epidemiology*, *36* (3), 666-676. https://doi.org/10.1093/ije/dym018

Sauer, J. D., Palmer, M. A., & Brewer, N. (2019). Pitfalls in using eyewitness confidence to diagnose the accuracy of an individual identification decision. *Psychology, Public Policy, and Law*, *25*(3), 147-165. https://doi.org/10.1037/law0000203

Sauerland, M., & Sporer, S. L. (2011). Written vs. spoken eyewitness accounts: Does modality of testing matter?. *Behavioral sciences & the law*, *29*(6), 846-857. https://doi.org/10.1002/bsl.1013

Schacter, D. L., Guerin, S. A., & St. Jacques, P. L. (2012). Memory distortion: An adaptive perspective. *Trends in Cognitive Sciences*, *15*, 467–474. https://doi.org/10.1016/j.tics.2011.08.004

Schmitt, T. A., & Sass, D. A. (2011). Rotation criteria and hypothesis testing for exploratory factor analysis: Implications for factor pattern loadings and interfactor correlations. *Educational and Psychological Measurement*, *71*(1), 95-113. https://doi.org/10.1177/0013164410387348

Schooler, J. W., & Loftus, E. F. (1986). Individual differences and experimentation: Complementary approaches to interrogative suggestibility. *Social Behaviour*, *1*(2), 105-112.

Seale-Carlisle, T. M., Colloff, M. F., Flowe, H. D., Wells, W., Wixted, J. T., & Mickes, L. (2019). Confidence and response time as indicators of eyewitness identification accuracy in the lab and in the real world. *Journal of Applied Research in Memory and Cognition*, *8*(4), 420-428. https://doi.org/10.1016/j.jarmac.2019.09.003

Shallice, T., & Vallar, G. (1990). The impairment of auditory–verbal short- term storage. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short- term memory* (pp. 11–53). Cambridge, England: Cambridge University Press.

Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin*, *100*(2), 139-156. https://doi.org/10.1037/0033-2909.100.2.139

Shor, R. E., & Orne, E. C. (1962). Harvard group scale of hypnotic susceptibility, form A. *International Journal of Clinical and Experimental Hypnosis*. https://doi.org/10.1037/t02246-000

Sigurdsson, J. F., & Gudjonsson, G. H. (1996). The psychological characteristics of "false confessors". A study among Icelandic prison inmates and juvenile offenders. *Personality and Individual Differences*, *20*, 321-329. https://doi.org/10.1016/0191-8869(95)00184-0

Singh, K. K., & Gudjonsson, G. H. (1992). Interrogative suggestibility among adolescent boys and its relationship with memory, suggestibility, compliance, anxiety and self-esteem. *Journal of Adolescence*, *15*, 155-161. https://doi.org/10.1016/0140-1971(92)90044-6

Singh, K. K., & Gudjonsson, G. H. (1987). The internal consistency of the shift factor on the Gudjonsson Suggestibility Scale. *Personality and Individual Differences*, *8*, 265-266. https://doi.org/10.1016/0191-8869(87)90183-8

Skagerberg, E. M., & Wright, D. B. (2008). The prevalence of co-witnesses and co-witness discussions in real eyewitnesses. *Psychology, Crime & Law*, *14*(6), 513-521. https://doi.org/10.1080/10683160801948980

Skagerberg, E. M., & Wright, D. B. (2009). Susceptibility to postidentification feedback is affected by source credibility. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *23*(4), 506-523. https://doi.org/10.1002/acp.1470

Smith, R. E., & Hunt, R. R. (1998). Presentation modality affects false memory. *Psychonomic Bulletin & Review*, *5*,710–715. https://doi.org/10.3758/BF03208850

Smith, R. E., Hunt, R. R., & Gallagher, M. P. (2008). The effect of study modality on false recognition. *Memory & cognition*, *36*(8), 1439-1449. https://doi.org/10.3758/MC.36.8.1439

Spearing, E. R., & Wade, K. A. (2021). *Providing Eyewitness Confidence Judgements During Versus After Eyewitness Interviews Does Not Affect the Confidence-Accuracy Relationship* [Manuscript submitted for publication]. Department of Psychology, University of Warwick. https://doi.org/10.31234/osf.io/bm2ua

Spiegel, H. (1974). *Manual for hypnotic induction profile: eye-roll levitation method*. New York: Soni Medica.

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: a meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin*, *118*(3), 315. https://doi.org/10.1037/0033-2909.118.3.315

Sporer, S. L., Tredoux, C. G., Vredeveldt, A., Kempen, K., & Nortje, A. (2020). Does exposure to facial composites damage eyewitness memory? a comprehensive review. *Applied Cognitive Psychology*, *34*(5), 1166-1179. https://doi.org/10.1002/acp.3705

Stark, C. E., Okado, Y., & Loftus, E. F. (2010). Imaging the reconstruction of true and false memories using sensory reactivation and the misinformation paradigms. *Learning & Memory*, *17*(10), 485-488.

Steblay, N. K., Dysart, J. E., & Wells, G. L. (2011). Seventy-two tests of the sequential lineup superiority effect: A meta-analysis and policy discussion. *Psychology, Public Policy, and Law*, *17*(1), 99. https://doi.org/10.1037/a0021650

Stern, W. (1938). *General psychology: from the personalistic standpoint*. New York: Macmillan.

Stern, W. (1939) The psychology of testimony. *The Journal of Abnormal and Social Psychology*, *34*(1), 3-20. https://doi.org/10.1037/h0054144

Stoll, E. M. (2021). *Misinformation modality and its effects on memory* [Unpublished thesis]. Oklahoma State University.

Straube, B. (2012). An overview of the neuro-cognitive processes involved in the encoding, consolidation, and retrieval of true and false memories. *Behavioral and Brain Functions*, 8(35), 1-10. https://doi.org/10.1186/1744-9081-8-35

Sutherland, R., & Hayne, H. (2001). Age-related changes in the misinformation effect. *Journal of experimental child psychology*, *79*(4), 388-404. https://doi.org/10.1006/jecp.2000.2610

Sweller J., Ayres P., & Kalyuga, S. (2011). *The Modality Effect. In: Cognitive Load Theory. Explorations in the Learning Sciences, Instructional Systems and Performance Technologies*, (1st ed). Springer, New York, NY.

Swystun, A. G., & Logan, A. J. (2019). Quantifying the effect of viewpoint changes on sensitivity to face identity. *Vision research*, *165*, 1-12. https://doi.org/10.1016/j.visres.2019.09.006

Szpitalak, M., & Polczyk, R. (2020). Reducing interrogative suggestibility: The role of self-affirmation and positive feedback. *Plos one*, *15*(7), e0236088. https://doi.org/10.1371/journal.pone.0236088

Takarangi, M. K., Parker, S., & Garry, M. (2006). Modernising the misinformation effect: The development of a new stimulus set. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *20*(5), 583-590. https://doi.org/10.1002/acp.1209

Tata, P. R., & Gudjonsson, G. H. (1990). The effects of mood and verbal feedback on interrogative suggestibility. *Personality and Individual Differences*, *11*(10), 1079-1085. https://doi.org/10.1016/0191-8869(90)90136-F

Taubert, J., Apthorp, D., Aagten-Murphy, D., & Alais, D. (2011). The role of holistic processing in face perception: Evidence from the face inversion effect. *Vision research*, *51*(11), 1273-1278. https://doi.org/10.1016/j.visres.2011.04.002

Technical Working Group for Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement.* Washington, DC: United States Department of justice, Office of justice Programs, National Institute of justice.

Thomas, B. H., Ciliska, D., Dobbins, M., & Micucci, S. (2004). A process for systematically reviewing the literature: providing the research evidence for public health nursing interventions. *Worldviews on Evidence-Based Nursing*, *1*(3), 176-184. https://doi.org/10.1111/j.1524-475X.2004.04006.x

Thorley, C. (2013). Memory conformity and suggestibility. *Psychology, Crime & Law*, *19*(7), 565-575. https://doi.org/10.1080/1068316X.2011.648637

Titcomb, A. L., & Reyna, V. F. (1995). Memory interference and misinformation effects. *Interference and inhibition in cognition*, 263-294. https://doi.org/10.1016/B978-012208930- 5/50009-X

Topp-Manriquez, L. D., McQuiston, D., & Malpass, R. S. (2016). Facial composites and the misinformation effect: How composites distort memory. *Legal and criminological psychology*, *21*(2), 372-389. https://doi.org/10.1111/lcrp.12054

Tully, B., & Cahill, D. (1984). *Police interviewing of mentally handicapped persons: An experimental study.* London: The Police Foundation of Great Britain.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological review*, *80*(5), 352-373. https://doi.org/10.1037/h0020071

Ulatowska, J., Olszewska, J., & Hanson, M. D. (2016). Do format differences in the presentation of information affect susceptibility to memory distortions? The three-stage misinformation procedure reconsidered. *American Journal of Psychology*, *129*(4), 407-417. https://doi.org/10.5406/amerjpsyc.129.4.0407

Ursachi, G., Horodnic, I. A., & Zait, A. (2015). How reliable are measurement scales? External factors with indirect influence on reliability estimators. *Procedia Economics and Finance*, *20*, 679-686. https://doi.org/10.1016/S2212-5671(15)00123-9

Vagni, M., Maiorano, T., Pajardi, D., & Gudjonsson, G. (2015). Immediate and delayed suggestibility among suspected child victims of sexual abuse. *Personality and Individual Differences*, *79*, 129-133. https://doi.org/10.1016/j.paid.2015.02.007

Van der Cruyssen, I., Regnath, F., Ben-Shakhar, G., Pertzov, Y., & Verschuere, B. (2020). The Picture Superiority Effect in Memory Detection: Modality Effects in the Concealed Information Test. https://doi.org/10.31219/osf.io/tg7us

Vickers, A. J., & Smith, C. (2000). Incorporating data from dissertations in systematic reviews. *International journal of technology assessment in health care*, *16*(2), 711-713.

Volz, K., Leonhart, R., Stark, R., Vaitl, D., & Ambach, W. (2017). Psychophysiological correlates of the misinformation effect. *International Journal of Psychophysiology*, *117*, 1-9. https://doi.org/10.1016/j.ijpsycho.2017.04.004

Wachi, T., Watanabe, K., Yokota, K., Otsuka, Y., & Hirama, K. (2019a). Comparison between Japanese online and standard administrations of the Gudjonsson Suggestibility Scale 2 and effects of post-warning. *Legal and Criminological Psychology*, *24*(1), 71-86. https://doi.org/10.1111/lcrp.12147

Wachi, T., Watanabe, K., Yokota, K., Otsuka, Y., & Hirama, K. (2019b). The immediate and delayed suggestibility examined by the online version of the Gudjonsson Suggestibility Scale 2. *Personality and Individual Differences*, *146*, 20-25. https://doi.org/10.1016/j.paid.2019.03.040

Warren, A., Hulse-trotter, K., & Tubbs, E. C. (1991). Inducing resistance to suggestibility in children. *Law and Human Behaviour*, *15*, 273-285.

Weber, N., & Brewer, N. (2003). The effect of judgment type and confidence scale on confidence-accuracy calibration in face recognition. *Journal of Applied Psychology*, *88*(3), 490-499. https://doi.org/10.1037/0021-9010.88.3.490

Wechsler, D. (1945). *Wechsler memory scale.* Psychological Corporation.

Wells, G. L., Charman, S. D., & Olson, E. A. (2005). Building face composites can harm lineup identification performance. *Journal of experimental psychology: Applied*, *11*(3), 147. https://doi.org/10.1037/1076-898X.11.3.147

Wells, G. L., & Hryciw, B. (1984). Memory for faces: Encoding and retrieval operations. *Memory & Cognition*, *12*(4), 338-344. https://doi.org/ 10.3758/BF03198293

Wells, G. L., & Luus, C. A. (1990). Police lineups as experiments: Social methodology as a framework for properly conducted lineups. *Personality and Social Psychology Bulletin*, *16*(1), 106-117. https://doi.org/10.1177/0146167290161008

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology*, *54*(1), 277-295. https://doi.org/10.1146/annurev.psych.54.101601.145028

Wells, G. L., Smith, A. M., & Smalarz, L. (2015). ROC analysis of lineups obscures information that is critical for both theoretical understanding and applied purposes. *Journal of Applied Research in Memory and Cognition*, *4*(4), 324-328. https://doi.org/10.1016/j.jarmac.2015.08.008

Werner, N. S., Kühnel, S., & Markowitsch, H. J. (2013). The neuroscience of face processing and identification in eyewitnesses and offenders. *Frontiers in behavioral neuroscience*, *7*, 189. https://doi.org/ 10.3389/fnbeh.2013.00189

Willner, P. (2011). Assessment of capacity to participate in court proceedings: A selective critique and some recommendations. *Psychology, Crime & Law*, *17*(2), 117-131. https://doi.org/10.1080/10683160903392400

Wilson, B. M., Seale-Carlisle, T. M., & Mickes, L. (2018). The effects of verbal descriptions on performance in lineups and showups. *Journal of Experimental Psychology: General*, *147*(1), 113. https://doi.org/10.1037/xge0000354

Wixted, J. T., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review*, *121*(2), 262. https://doi.org/10.1037/a0035940

Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., & Roediger III, H. L. (2015). Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *American Psychologist*, *70*(6), 515-526. https://doi.org/10.1037/a0039510

Wixted, J. T., Mickes, L., Wetmore, S. A., Gronlund, S. D., & Neuschatz, J. S. (2017). ROC analysis in theory and practice. *Journal of applied research in memory and cognition*, *6*(3), 343-351. https://doi.org/10.1016/j.jarmac.2016.12.002

Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest*, *18*(1), 10-65. https://doi.org/10.1177/1529100616686966

Wogalter, M. S., & Laughery, K. R. (1987). Face recognition: effects of study to test maintenance and change of photographic mode and pose. *Applied Cognitive Psychology*, *1*(4), 241-253. https://doi.org/10.1002/acp.2350010403

Wolfradt, U., & Meyer, T. (1998). Interrogative suggestibility, anxiety and dissociation among anxious patients and normal controls. *Personality and Individual Differences*, *25*(3), 425-432. https://doi.org/10.1016/S0191-8869(98)00023-3

Woolston, R., Bain, S. A., & Baxter, J. S. (2006). Patterns of malingering and compliance in measures of interrogative suggestibility. *Personality and Individual Differences*, *40*(3), 453-461. https://doi.org/10.1016/j.paid.2005.10.010

Yamashita, M. (1996). A re-examination of the misinformation effect by means of visual and verbal recognition tests. *Japanese Psychological Research*, *38*(1), 47-52. https://doi.org/10.1111/j.1468-5884.1996.tb00007.x

Yeari, M., van den Broek, P., & Oudega, M. (2015). Processing and memory of central versus peripheral information as a function of reading goals: Evidence from eye-movements. *Reading and Writing*, *28*(8), 1071-1097. https://doi.org/10.1007/s11145-015-9561-4

Zaragoza, M. S., Belli, R. F., & Payment, K. E. (2007). Misinformation effects and the suggestibility of eyewitness memory. *Do justice and let the sky fall: Elizabeth Loftus and her contributions to science, law, and academic freedom*, 35-63.

Zaragoza, M. S., McCloskey, M., & Jamis, M. (1987). Misleading postevent information and recall of the original event: Further evidence against the memory impairment hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(1), 36. https://doi.org/10.1037/0278-7393.13.1.36

Zhu, B., Chen, C., Loftus, E. F., Lin, C., & Dong, Q. (2013). The relationship between DRM and misinformation false memories. *Memory & cognition*, *41*(6), 832-838. https://doi.org/10.3758/s13421-013-0300-2

Zhu, B., Chen, C., Loftus, E. F., Lin, C., He, Q., Chen, C., Li, H., Moyzis, R. K., Lessard, J & Dong, Q. (2010). Individual differences in false memory from misinformation: Personality characteristics and their interactions with cognitive abilities. *Personality and Individual differences*, *48*(8), 889-894. https://doi.org/10.1016/j.paid.2010.02.016

Appendices

Appendix One: Unable to Obtain Articles

The following table includes articles that could not be accessed even when attempts to contact the authors were made.

Author of publication (date)	Title of publication	How study identified
Bobak (2013)	Refining fuzzy-trace theory by comparing visual and nonverbal auditory witness memory accuracy	Systematic search of databases
Cushman (1994)	The effect of post-event misinformation on event retrieval.	Google search
Dijkstra (2013)	The role of modality on correct recognition and misinformation in younger and older adults	Systematic search of databases
Saakian (2008)	Misinformation effect: vulnerability to misleading information as a function of modality	Reference lists of shortlisted articles

Appendix Two: Articles Accessed in Full and Excluded After Application of Inclusion Criteria

Authors of publication (date)	Title of Publication	How study was identified	Reason for exclusion
Bowman & Zaragoza, 1989	Similarity of encoding context does not influence resistance to memory impairment following misinformation.	Google search	Explicit details regarding the impact of specific modalities were not included
Brandimonte, Schooler & Gabbino, 1997	Attenuating Verbal Overshadowing Through Color Retrieval Cues	Systematic search of databases	No reference to MI effect or use of MI paradigm
Candel, Hayne, Strange & Prevoo, 2009	The effect of suggestion on children's recognition memory for seen and unseen details	Systematic search of databases	Changes in modality were not accounted for in the findings
Chrobak & Zaragoza, 2013	When Forced Fabrications Become Truth: Causal Explanations and False Memory Development	Systematic search of databases	Explicit details regarding the impact of specific modalities were not included
Dijkstra, MacMahon & Misirlisoy, 2008	The effects of golf expertise and presentation modality on memory for golf and everyday items	Reference lists of shortlisted articles	No use of MI paradigm
Dodhia & Metcalfe, 1999	False memories and source monitoring	Systematic search of databases	No use of MI paradigm
Gerlach, Dornblaser & Schacter, 2014	Adaptive constructive processes and memory accuracy: consequences of counterfactual simulations in young and older adults	Systematic search of databases	No reference to MI effect or use of MI paradigm
Ghetti, Qin & Goodman, 2002	False Memories in Children and Adults: Age, Distinctiveness, and Subjective Experience	Systematic search of databases	DRM paradigm used
Gobbo, 2000	Assessing the effects of misinformation on children's recall: how and when makes a difference	Systematic search of databases	Explicit details regarding the impact of specific modalities were not included

Authors of publication (date)	Title of Publication	How study was identified	Reason for exclusion
Heath, 1992	Effects of varying postevent information on memory for central and peripheral actions and props	Systematic search of databases	Focus solely on central and peripheral details
Holliday, Douglas & Hayes, 1999	Children's eyewitness suggestibility: Memory trace strength revisited	Systematic search of databases	No reference to impact of modality
Holliday & Hayes, 2002	Automatic and intentional processes in children's recognition memory: The reversed misinformation effect	Systematic search of databases	No reference to impact of modality
Hunt, 2003	The concept of knowledge and how to measure it	Reference lists of shortlisted articles	No reference to MI effect or use of MI paradigm
Kidorf, 1995	Effects of temporally varied biased and unbiased story summaries on the suggestibility of pre-schoolers	Systematic search of databases	No reference to impact of modality
LaPaglia & Chan, 2019	Telling a good story: The effects of memory retrieval and context processing on eyewitness suggestibility	Systematic search of databases	No reference to impact of modality
Lee & Chen, 2013	Post-event information presented in a question form eliminates the misinformation effect	Reference lists of shortlisted articles	No reference to impact of modality
Maylor & Mo, 1999	Effects of study-test modality on false recognition	Google search	DRM paradigm used
McKone & Murphy, 2000	Implicit false memory: Effects of modality and multiple study presentations on long-lived semantic priming	Google search	DRM paradigm used
Murphy & Greene, 2016	Perceptual Load Affects Eyewitness Accuracy and Susceptibility to Leading Questions	Systematic search of databases	Explicit details regarding the impact of specific modalities were not included

Authors of publication (date)	Title of Publication	How study was identified	Reason for exclusion
Okado & Stark, 2005)	Neural activity during encoding predicts false memories created by misinformation	Systematic search of databases	Explicit details regarding the impact of specific modalities were not included
Olszewska, Reuter-Lorenz, Munier & Bendler, 2015	Misremembering what you see or hear: Dissociable effects of modality on short-and long-term false recognition.	Reference lists of shortlisted articles	DRM paradigm used
Panksy, Tenenboim & Bar, 2011	The misinformation effect revisited: Interactions between spontaneous memory processes and misleading suggestions	Systematic search of databases	No reference to impact of modality
Pierce & Gallo, 2011	Encoding modality can affect memory accuracy via retrieval orientation.	Reference lists of shortlisted articles	No use of MI paradigm
Principe, Haines, Adkins & Guiliano, 2010	False rumors and true belief: memory processes underlying children's errant reports of rumored events	Systematic search of databases	No reference to impact of modality
Smith & Hunt, 1998	Presentation modality affects false memory	Reference lists of shortlisted articles	DRM paradigm used
Stark, Okado & Loftus, 2010	Imaging the reconstruction of true and false memories using sensory reactivation and the misinformation paradigms	Systematic search of databases	Participant “imagination” not considered as a modality under used definition
Tait, 2014	A stimulus control analysis of the misinformation effect	Reference lists of shortlisted articles	No reference to impact of modality
Thierry, Goh, Pipe & Murray, 2005	Source recall enhances children's discrimination of seen and heard events.	Systematic search of databases	DRM paradigm used

Authors of publication (date)	Title of Publication	How study was identified	Reason for exclusion
Zhu et al. 2019	Multiple interactive memory representations underlie the induction of false memory	Systematic search of databases	DRM paradigm used

Appendix Three: Articles Accessed in Full and Meeting Inclusion/Exclusion Criteria

Authors of publication (date)	Title of publication	How study was identified
Abeles & Morton (1999)	Avoiding Misinformation: Reinstating Target Modality	Google search
Campbell, Edwards, Horswill & Helman (2007)	Effects of contextual cues in recall and recognition memory: The misinformation effect reconsidered	Systematic search of databases
Dijkstra & Moerman (2012)	Effects of modality on memory for original and misleading information	Systematic search of databases
Gobbo, Mega & Pipe (2002)	Does the nature of the experience influence suggestibility? A study of children's event memory	Systematic search of databases
Hendrich (2019)	Modality Effects in False Memory Production Using the Misinformation Paradigm	Google search
Itsukushima, Nishi, Maruyama & Takahashi (2006)	The effect of presentation medium of post-event information: impact of co-witness information	Reference lists of shortlisted articles
Kiat (2018)	The Role of Visual and Verbal Processes in False Memory Susceptibility on the Misinformation Effect	Systematic search of databases
Mitchell & Zaragoza (1996)	Repeated exposure to suggestion and false memory: The role of contextual variability	Reference lists of shortlisted articles
Roebers, Gelhaar, and Schneider (2004)	"It's magic!" The effects of presentation modality on children's event memory, suggestibility, and confidence judgments	Reference lists of shortlisted articles
Stoll (2021)	Misinformation Modality and its Effects on Memory	Google Search
Ulatowska, Olszewska and Hanson (2016)	Do format differences in the presentation of information affect susceptibility to memory distortions? The three-stage misinformation procedure reconsidered	Systematic search of databases
Yamashita (1996)	A re-examination of the misinformation effect by means of visual and verbal recognition tests	Systematic search of databases

Appendix Four: Quality Assessment Tool for Quantitative Studies Adapted From Effective Public Health Practice Project (1998) for Studies Used Within the Current Review.

Name of study:

Date of Publication:

Authors:

Quality Assessment (only score items of relevance)	**Scoring Criteria for Each Domain**	**Domain rating guidance**
A) SELECTION BIAS		**STRONG** ☐ **MODERATE** ☐ **WEAK** ☐
(Q1) Are the individuals selected to participate in the study likely to be representative of the target population? a) Very likely b) Somewhat likely c) Not likely d) Cannot tell	***Score very likely-*** Given that any participant groups can be eye witness, this is scored if participants are randomly selected from an accessible population **Score somewhat likely-** considered representative if participants are only obtained via opportunity sampling **Score not likely-** if participants are not able to adequately access changes in modality ***Cannot tell-*** if participants' details are not adequately described.	***Strong*** Both Qu1 <u>&</u> Qu2 =a ***Moderate*** all other combinations not included in 'strong' or 'weak' scoring criteria **Weak** Qu1 <u>or</u> Qu2 = c OR Qu1 = d <u>&</u> Qu2 = d or e
(Q2) What percentage of selected individuals agreed to participate? a) 80-100% b) 60-79% c) Less than 60% d) Not applicable e) Cannot tell	**If Relevant**: Refers to the % of those who agreed to participate at the point of study commission compared to the original number selected. ***Cannot tell-*** Percentage not stated or defined	
B) STUDY DESIGN *(Added to current adapted version for the purpose of the review)*		**STRONG** ☐ **MODERATE** ☐ **WEAK** ☐
Q1) Was the design appropriate for research question(s)? a) Yes b) No c) Cannot Tell	***Yes-*** Design adequately enabled research question(s) to be answered **No-** Did not enable the study aims to be answered research question(s) answered. Another type of research design would have been more appropriate ***Cannot tell-***Not enough information	***Strong*** Qu1, Qu2 <u>&</u> Qu3 =a ***Moderate*** all other combinations not included in 'strong' or 'weak' scoring criteria **Weak** Qu1 =b
Q2) Has the study adequately answered the research question? a) Yes b) Partially c) No d) Cannot Tell	***Yes-*** All research question(s) were answered in line with the in-text explanations **Partially** – Only some of the research question(s) were answered **No-** None of the research question(s) were adequately answered ***Cannot tell-*** Not enough information	

Q3) Were the manipulations within and between modalities clearly defined and useful? a) Yes b) Partially c) No d) Cannot Tell	***Yes-*** The comparisons between different modalities/differences within modalities were well defined, useful to answering the research question(s) ***Partially*** - The manipulations only provided some usefulness when answering the research question(s) or where somewhat defined **No-** The manipulations were not able to answer the research question(s). An inadequate description of the manipulations was provided ***Cannot tell-*** The nature of the modality manipulations were not appropriately described to draw conclusions	OR Qu2 = c OR Qu3 = c OR Qu1 is c & Qu2 is d
C) CONFOUNDING VARIABLES *(Added to current adapted for the purpose of the rveiew)*		**STRONG** ☐ **MODERATE** ☐ **WEAK** ☐
(Q1) Were alternative factors that could have accounted for findings appropriately controlled for? a) Yes b) Partially c) No d) Cannot tell	***Yes-*** Most obvious factors were accounted for and described **Partially** – Some factors were accounted for and described **No-** No controls were implemented ***Cannot tell-*** Not enough information	***Strong*** Qu1 =a ***Moderate*** Qu1 is b OR Qu2 =a or c **Weak** Qu1 =c or d & Qu2 = b
(Q2) Were confounding variables identified? a) Yes b) No c) Partially	***Yes-***An appropriate explanation was included for any confounding variables ***Partial-*** Some confounding variables within study were described, but not to a full extent ***No-*** No information was present that accounted for potential confounding variables, or there are confounding variables that are obviously present in the study	
D) BLINDING		**STRONG** ☐ **MODERATE** ☐ **WEAK** ☐
(Q1) Were the researchers aware of the participants allocation to modality conditions? a) Yes b) No c) Cannot tell	Researchers must state that that participants were randomly assigned to conditions and/or the researchers must be unaware of which condition participants were assigned to.	***Strong*** Qu1 & Qu2=b ***Moderate*** all other combinations not included in 'strong' or 'weak' scoring criteria **Weak** Qu1 & Qu2=a
(Q2) Were the study participants aware of the research question? a) Yes b) No c) Cannot tell	Researchers must state if participants were aware of the research question and aims.	
E) DATA COLLECTION METHODS AND REPORTING		**STRONG** ☐ **MODERATE** ☐ **WEAK** ☐

<table>
<tr><td>(Q1) Were data collection tools shown to be valid?
a) Yes
b) No
c) Cannot tell</td><td rowspan="2">Tools used to assess outcome measures (both regarding changes in modality and the impact on the misinformation effect) must be found to have validity and reliability.

Reliability and validity can be reported by the researchers within the paper or in a separate study. The use of the MI paradigm can be considered valid and reliable.</td><td rowspan="4">Strong
Qu1, Qu2, Qu 3 <u>&</u> Qu4=a

Moderate
all other combinations not included in ‘strong’ or ‘weak’ scoring criteria

Weak
Qu1 = b
OR
Qu1 <u>&</u> 2=c
OR
Qu3 & 4= b</td></tr>
<tr><td>(Q2) Were data collection tools shown to be reliable?
a) Yes
b) No
c) Cannot tell</td></tr>
<tr><td>(Q3) Were explicit links made to the misinformation effect in reporting
a) Yes
b) No
(added to current adapted version)</td><td>Yes- Clear indication that data collection methods considered the misinformation effect explicitly

No- Unclear whether misinformation effect was considered in write-up</td></tr>
<tr><td>(Q4) Was impact of modality on the misinformation effect adequately reported?
a) Yes
b) No
(added to current adapted version)</td><td>Yes – Clear indication provided of the modality manipulations participants were exposed to and at which stage of the MI paradigm manipulations occurred

No- Unclear on the modalities manipulated and at which stage of the MI paradigm these manipulations occurred</td></tr>
<tr><td>F) WITHDRAWALS, DROP-OUTS AND MISSING DATA</td><td></td><td>STRONG ☐
MODERATE ☐
WEAK ☐</td></tr>
<tr><td>(Q1) Were participant withdrawals reported?
a) Yes
b) No
c) Cannot tell
Not Relevant</td><td>Where found to be Relevant:

Yes-if researchers provide clear and complete details of withdrawal information

No-if either the numbers or reasons for withdrawals and drop-outs are not reported.</td><td rowspan="2">This item does not need to be scored if domain is deemed unnecessary

Strong
Qu2 =a

Moderate
Qu2=a or c

Weak
Qu2 =b</td></tr>
<tr><td>(Q2) Was missing data reported and dealt with appropriately?
a) Yes
b) No
c) Cannot tell
Not Relevant
(added to current adapted version)</td><td>If Relevant:

Yes- Missing data accounted for and dealt with appropriately
No- Missing data not accounted for and not dealt with appropriately
Cannot tell- not enough information</td></tr>
<tr><td colspan="3">G) Removed from current adapted version as not relevant, i.e., focus was on interventions</td></tr>
<tr><td>(H) ANALYSIS
(added to current adapted version for the purpose of the review)</td><td></td><td>STRONG ☐
MODERATE ☐
WEAK ☐</td></tr>
</table>

(Q1) Were appropriate statistical tests used to assess the research question (s)? Yes Partially No Unclear	Was the quantitative analysis appropriate to the research question(s) being asked?	**STRONG** –yes **MODERATE** – partial **WEAK** – No or unclear
(I) OUTCOMES *(added to current adapted version for the purpose of the review)*		**STRONG** ☐ **MODERATE**☐ **WEAK**☐
(Q1) Were outcomes clearly described? a) Yes b) Partially c) No	Was it easy to make sense of the findings? Was the way in which findings were presented accessible to the reader?	***Strong*** Qu1, Qu2 = yes ***Moderate*** all other combinations not included in 'strong' or 'weak' scoring criteria **Weak** Qu1= c OR Qu1, Qu2 = b
(Q2) Were numerical descriptions of important outcomes given? a) Yes b) No c) Unclear	In cases where outcomes were found to be important to the research question and aims, were these reported both numerically and descriptively?	
Global Ratings:	**1 Strong** (no weak ratings) **2 Moderate** (one weak rating *or* two weak ratings, one being SELECTION domain) **3 Weak** (two or more weak ratings)	

Appendix Five: Quality Assessment Outcomes for the 19 Studies Meeting PICO Criteria

Authors of study	Year of study	Quality assessment outcome	Reason for Exclusion
Abeles & Morton	1999	Moderate	
Campbell, Edwards, Horswill & Helman, study 1 and 2	2007	Moderate	
Dijkstra & Moerman	2012	Strong	
Gobbo, Mega & Pipe, study 1 and 2	2002	Strong	
Hendrich	2019	Strong	
Itsukushima, Nishi, Maruyama & Takahashi	2006	Weak	Cross-study comparison did not reveal overall conclusion on the impact of modality manipulation on the MI effect. Sample was not adequately described and methodology was not able to answer the research question(s).
Kiat, studies 1, 2, 3 and 4	2018	Moderate	
Mitchell & Zaragoza ,study 1 and 2	1996	Moderate	
Roebers, Gelhaar, and Schneider	2004	Moderate	
Stoll	2021	Moderate	
Ulatowska, Olszewska and Hanson, study 1 and 2	2016	Strong	
Yamashita	1996	Moderate	Study 1 excluded due to relevance to topic area.

Appendix Six: Data Extraction Form

A) General Information

Paper Title:

Author(s):

Year of Publication:

How many studies contained within paper?

Country:

Study Aim(s):

B) Population

Sample Size:

Gender Information:

Age Information:

Ethnicity Information:

C) Modality Manipulations

What modalities were manipulated?

Within or Between Modality Manipulations?

D) Misinformation Paradigm

At which stage did modality manipulation take place?

Encoding	Misinformation	Test

Was the MI paradigm altered in some way?

E) Outcome(s)

Statistical Analysis Conducted:

Summary of modality effect on MI effect

Which modalities were most associated with the MI effect?

Was modality congruence explored? If so, what were the findings?

Study Limitations:

Final conclusions:

Suggestions for future research?

Additional Comments:

Appendix Seven: Overall Aims of Studies

Authors and year of study	Summary of overall aims
Abeles & Morton, 1999	To show how the reversed eyewitness misinformation effect, demonstrated by Lindsay and Johnson (1989), can be virtually eliminated if the test conditions encourage access to the critical memory.
Campbell, Edwards, Horswill & Helman, 2007	Study 1: To combine procedures employed in misinformation and semantic word list research, to present participants with restated, neutral, and misleading post-event information either in an auditory or written modality. Study 2: To assess the effects of post-event information on participant's ability to recall the original information.
Dijkstra & Moerman, 2012	To examine the role of modality in correct recognition and misinformation acceptance in a naturalistic cognitive task.
Gobbo, Mega & Pipe, 2002	Study 1: To examine how mode of event presentation impacts on memory and the acceptance of misinformation as a function of the type of to-be-recalled items (preparation and construction), age, and time. Study 2: To compare the suggestibility of children who had participated in an event and those who had heard a narrative about the same event when exposed to the event a single time or to a criterion of learning.
Hendrich, 2019	To further examine how modality influences false memory using the misinformation paradigm.
Kiat, 2018	Study 1: To contrast differences in the strength of the association between misinformation susceptibility and basic visual as well as verbal source monitoring errors. Study 2: To test the robustness and replicability of detected relationships between misinformation susceptibility and basic visual as well as verbal source monitoring errors. Study 3: To investigate potential links between neural activity associated with recollective processing in misinformation endorsements and basic visual as well as verbal source monitoring errors. Study 4: To investigate potential links between neural activity associated with event and narrative encoding on the misinformation effect and misinformation susceptibility

Authors and year of study	Summary of overall aims
Mitchell & Zaragoza, 1996	Study 1: To manipulate variability in order to assess whether it plays a role in the errors that result from repeated suggestion. To investigate whether increasing contextual variability of the repeated exposures to post event suggestions would increase subjects' suggestibility. Study 2: To replicate and extend the findings from study 1 and to better understand how contextual variability may come to affect source judgements.
Roebers, Gelhaar, and Schneider, 2004	To investigate the influence of presentation modality (live, video, slideshow) on children's memory, suggestibility, recognition, and metamemorial monitoring processes.
Stoll, 2021	To measure how misinformation modality and misinformation type affect misinformation acceptance.
Ulatowska, Olszewska and Hanson, 2016	Study 1: To examine the susceptibility to misinformation after encoding original information in 1 of 4 different formats. Study 2: To examine the effect of various formats on how the three stages of the MI paradigm interact in order to determine the strength of the encoding specificity effect.
Yamashita, 1996	Study 2: To investigate the influence of modalities (visual and verbal) in the recognition test and of levels of memorableness upon the misinformation effect.

Appendix Eight: Misinformation News Report Script

"On tonight's news, a suspect has been apprehended in relation to a recent handbag theft in the area where a substantial amount of cash was stolen during the incident. The suspect was apprehended after police reviewed CCTV footage of the crime, and believe that the culprit looked like a local resident. Police have also provided the following headshot of the suspect and are appealing to members of the public to come forward with any further information. At present, it is reported that the arrested suspect is maintain their innocence."

Appendix Nine: Proof of Ethical Approval for Research

Dear Miss Kara Deering & Dr Melissa Colloff,

Re: "Working title: Do differences in facial positioning influence the susceptibility of memory? An investigation into the Misinformation effect and eyewitness misidentification."

Application for Ethical Review ERN_19-0852

Thank you for your application for ethical review for the above project, which was reviewed by the Science, Technology, Engineering and Mathematics Ethical Review Committee.

On behalf of the Committee, I confirm that this study now has full ethical approval.

I would like to remind you that any substantive changes to the nature of the study as described in the Application for Ethical Review, and/or any adverse events occurring during the study should be promptly brought to the Committee's attention by the Principal Investigator and may necessitate further ethical review.

Please also ensure that the relevant requirements within the University's Code of Practice for Research and the information and guidance provided on the University's ethics webpages (available at https://intranet.birmingham.ac.uk/finance/accounting/Research-Support-Group/Research-Ethics/Links-and-Resources.aspx) are adhered to and referred to in any future applications for ethical review. It is now a requirement on the revised application form (https://intranet.birmingham.ac.uk/finance/accounting/Research-Support-Group/Research-Ethics/Ethical-Review-Forms.aspx) to confirm that this guidance has been consulted and is understood, and that it has been taken into account when completing your application for ethical review.

Please be aware that whilst Health and Safety (H&S) issues may be considered during the ethical review process, you are still required to follow the University's guidance on H&S and to ensure that H&S risk assessments have been carried out as appropriate. For further information about this, please contact your School H&S representative or the University's H&S Unit at healthandsafety@contacts.bham.ac.uk.

Kind regards,

Ms Sam Waldron
Research Ethics Officer
Research Support Group

C Block Dome (room 132)

Aston Webb Building

University of Birmingham

Edgbaston B15 2TT

Tel: [redacted]
Email: [redacted]