An exploration into the reliability of child witness memory evidence: The role of standardised assessment, misinformation, and the perception of professionals

By

CATRIN ELERI WILLIAMS

A thesis submitted to the University of Birmingham

for the degree of

Doctorate in Forensic Psychology Practice (ForenPsyD)

Centre for Forensic and Criminological Psychology

School of Psychology

College of Life and Environmental Science

University of Birmingham

April 2021

Abstract

The Criminal Justice System (CJS) regularly calls upon child witnesses in legal proceedings. It is vitally important that the reliability of a witness's account is assessed to ensure that accurate evidence is being relied upon in legal decision-making. The overarching aim of this thesis is to better understand the practicalities of working with child memory evidence and how to support this by: (1) examining how a measurement of suggestibility to incorrect information (i.e., *misinformation*) could be incorporated into practice, to identify whether assessing suggestibility to misinformation could benefit forensic practitioners and child witnesses; and (2) examining how children are perceived by criminal justice professionals, and how professionals work with child memory evidence.

This thesis includes a meta-analysis, which found that child witnesses are more likely than adults to succumb to misinformation (Chapter 2). Next, a qualitative study with focus groups of UK professionals working with child memory evidence elucidated opinions about evidence from children and found that professionals were able to identify gaps and issues with current practice and did not appear to have access to a standardised assessment of reliability (Chapter 3). Therefore, the Bonn Test Statement of Suggestibility (BTSS; Endres, 1997) was analysed to assess its applicability to support professionals. It was identified as theoretically useful, though important ethical considerations were highlighted (Chapter 4).

Together, this work identifies gaps in the current literature, such as the need to better understand the impact that professionals have on child well-being and evidence, the perceptions that different professionals (e.g., lawyers) have of children, and further meta-analytical assessment of moderating factors of suggestibility to misinformation (e.g., retention interval, misinformation timing, cognitive factors). Practically, the work identifies a need for training and psychological resources to support professionals in the UK to assess the reliability of accounts from child witnesses.

Acknowledgements

I would like to first thank my supervisor, Dr Melissa Colloff who's ongoing support, attention to detail and commitment was integral to my journey in writing this thesis. I am extremely thankful for her kindness, patience and expertise and am grateful for her input. Also, to Dr Heather Flowe and Dr Christopher Jones who were also extremely supportive in the development of my second chapter. I would also like to extend my thank you to the rest of the faculty in the department who were continually supportive in with my academia and well-being.

Of course, I would like to thank the participants who were involved in my research and who trusted me to share their experiences with. I feel extremely grateful to have been provided the opportunity to talk with people who are such integral members of our society. Their ongoing commitment and support for the children they work with is admirable and I am thankful to be involved in a small part of your careers.

To my parents, I am eternally grateful. They are the ones who supported me both practically and emotionally throughout the entire doctorate and process of writing my thesis. Without them I would not have been able to realise a lifelong dream of developing this thesis, and I hope that they know how lucky I feel to have them as my loving parents. To the rest of my family too, including my sister and beautiful niece and nephew, thank you for continually reminding me of how fortunate and how loved I am, even on days when I was overwhelmed with graphs and numbers.

Finally, I would like to thank Lewis, my amazing fiancé. A *brave* man who proposed to me half-way through writing this thesis. I love you and your support throughout this has been unwavering – I could not have done it without you. Helped of course by our beautiful dog Sibi, who liked to remind me to take breaks by giving her regular walks and treats. I

think as a family, we should celebrate the completion of this thesis by planning a wedding and living happily ever after.

Declaration

This thesis is submitted to the University of Birmingham in support of my application for the degree of Doctorate in Forensic Psychology Practice. It has been composed by myself and has not been submitted in any previous application for any degree. The work presented (including data generated and data analysis) was carried out by the author.

Inclusion of Published work

Part of the thesis (Chapter 2 – Meta-analysis) has been submitted for review for publication by the author. As part of this work, Dr Melissa Colloff contributed to the planning of this research and provided feedback on drafts of the manuscript; Dr Christopher Jones provided the software that was utilised for the data analysis; and Dr Heather Flowe re-analysed an existing meta-analysis on this topic (Payne, Toglia & Anastasi, 1994) which can be found in the introduction of this piece of work (and referenced accordingly in the thesis).

Table of Contents

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER I

General Introduction

The Criminal Justice System (CJS) regularly calls upon child witnesses in various different legal proceedings, in both the magistrates court and crown court. The Citizens Advice Witness Service identified that in 2017-2018 there was a total of 12,318 child victims and witnesses that were required to attend trial-related hearings. In criminal proceedings, child witnesses might be expected to provide testimony to police officers or during a court case because they have been a victim of a crime (e.g., abuse) or observed a crime. Children as young as 2-years-old have been asked to provide testimony in the UK (e.g., Bowcott, 2017). For witnesses of any age, it is vitally important that the reliability of their account is assessed to ensure that correct evidence is being collected for investigations, which ultimately supports reliable decision making in legal contexts.

The following example demonstrates the importance of thoroughly assessing the reliability of accounts from witnesses (as discussed by Otgaar, et al., 2017). Otgaar et al. (2017) explained how 'Galileo' Elementary school children had reported to being sexually abused in school in 2009. The father of an alleged victim filed a charge with the police, and the police interviewed 20 children who also reported the abuse. It was identified that the first report made by a child was not sexual, rather the child merely told his parents he was punished in class. The child's therapist concluded that the child had trauma-related symptoms and used an anatomical doll to uncover potential trauma. Following a period of play with the doll, the child started to incorporate statements of sexual abuse into his account. Other parents were sent a letter of the abuse and informed of how to interpret trauma symptoms in their children. After this letter, many more children started to 'remember' being sexually

abused in school. These reports slowly started to become more bizarre in nature, such as reports of being made to look at crocodiles in a teacher's home during apparent sexual abuse. This example demonstrates how child witnesses can come to remember and believe in events and details that did not really occur; that is, to develop false memories. Fortunately, expert witnesses in the case attested to this, and thus the teachers were not prosecuted.

The CJS follows protocols when interviewing children in attempts to improve the reliability of evidence and ensure the well-being of the witnesses. Children and young people under the age of 18 automatically qualify for 'Special Measures'. These are a series of provisions and changes to usual court proceedings and criminal investigations that help a vulnerable witness to give their best evidence and help relieve some associated stress (see Youth Justice and Criminal Evidence Act, 1999). Moreover, in England and Wales, it is typical that a child be interviewed following the Achieving Best Evidence in Criminal Proceedings guidelines (ABE; Home Office, 2011). The ABE follows four general phases: a period of rapport building, a free narrative account, questioning, and interview closure and it should be conducted by an officer with training in interviewing children (Richardson et al.,). Interviewers are expected to follow guidelines to support with questioning, such as asking open-ended questions (i.e., questions with unrestricted, broad possible answers) and avoiding leading questions (i.e., questions that suggest an answer).

Prior to the questioning phase in the ABE, children will be asked to provide definitions of what a truth or a lie is and will be asked to judge this based on examples. For example, children could be told a story about a boy who smashes a window with a football and tells his mother that another boy kicked the ball. Children will subsequently be asked if the boy in the story was lying or telling the truth. It is hoped that this exercise will support in the assessment of reliability and understanding of 'truth telling' (Richardson et al.,2018). It is worth highlighting however that reliability and 'truth telling' are two slightly different

concepts, as some children may be attempting to tell the truth, but various factors may have influenced the reliability of their accounts (i.e., the accuracy of their memories) without their conscious knowledge. One factor that can influence the accuracy of witness memories is exposure and suggestibility to post-event misleading information—termed misinformation (e.g., Loftus et al., 1978)—this term will be discussed further below.

Overview and Definitions

For the purposes of this thesis, there are four concepts that are regularly discussed: misinformation, suggestibility, reliability and credibility. For ease of understanding, these terms are explained in this section. Post-event *misinformation* is inaccurate information that is provided to a witness after they have witnessed an event (Loftus et al.,1978). *Suggestibility* is the likelihood that the witness will 'accept' the misinformation that is provided; the more misinformation they incorporate into their subsequent memory accounts, the more 'suggestible' they are deemed to be (Gudjonsson & Clark, 1986). *Reliability* of the account is the likelihood someone's account is accurate and therefore true to the actual events; it is characterised by the consistency or repeatability of memory report. Finally, *credibility* of a witness is the quality of being believed or trusted by another individual (see London & Ceci, 2012).

Theories of memory

To better understand theories of misinformation and suggestibility, it is first necessary to have a basic understanding of common memory theories. There is a plethora of longstanding memory theories that outline how memories are encoded, stored and retrieved. For example, the 'Information Processing Theory' (Miller, 1956) posits the capacity of memory, including short-term and long-term memory. Atkinson and Shiffrin (1968) proposed the 'multi-store model' which suggests that sensory memory (i.e., information picked up from the senses), short-term memory (i.e., small amount of information for a short amount of

time) and long-term memory (i.e., stored memory that is kept for a long period of time) are encoded; being stored as long-term after a period of rehearsal or repetition. The 'levels of processing' theory (Craik & Lockhard, 1972) instead posits that the level of processing determines how well the memory is consolidated. Memories can either undergo 'shallow' processing (e.g., a person may ask themselves "what does an object look like?") or 'deep' processing (e.g., a person may ask themselves "what does it mean?"). Long and short term-memory differ in relation to the time the memories remain accessible in the brain. Short-term memory has been found to decay rapidly, after around 18 seconds (Peterson & Peterson, 1959), but long-term memories may never decay.

Since these theories were introduced, further memory processes have been introduced, such as the idea of 'working memory' (i.e., temporary storage to support cognitive tasks; Baddeley, 1992) and both 'explicit memory' (i.e., conscious retrieval of memory) and 'implicit memory (i.e., unconscious) which have their own sub-processes (Graf & Schacter, 1985). Neuroimaging has allowed for assessment of memory consolidation, retrieval, decay and the role of synapses in the hippocampus in memory consolidation (e.g., Dong et al., 2015). This is beyond the scope of the current thesis; however, a basic understanding of memory processes supports in understanding the theoretical underpinnings of this topic.

For instance, when a new memory (e.g., post-event information) interferes with recall, this is an example of *retroactive interference,* which interferes the usual storage and retrieval process from long-term memory (Underwood et al., 1960). *Proactive* interference on the other hand is when a new piece of information is interfered by previous beliefs, attitudes or memories (Underwood, 1960).

For the current research area, there is specific interest in how memories change, to develop false memories. Specifically, how they change in relation to misinformation and the impact this has on one's memory (e.g., encoding, storage and retrieval). This is an important

distinction to make, as false memories can develop both because of misinformation (e.g., Loftus, 1978) and can also develop *spontaneously* (see Reyna, 2016). The current thesis focuses on the role of misinformation, though research on spontaneous false memories will be discussed on occasion to support theoretical understanding of memory and how false memories develop.

Misinformation and Suggestibility

When a person begins to report the incorrect post-event information as if it were part of their original memory, this is called the 'misinformation effect' (Loftus, 1978). In research, the misinformation effect is measured using the misinformation paradigm. This includes first providing participants with a stimulus event (e.g., a video of a mock crime), then providing the participant with suggestive information or 'misinformation' about a detail in the stimulus event, and finally testing the witness's memory to identify if they have incorporated the misinformation into their subsequent recall (see Loftus, 2005). The memory test can take various forms, such as free-recall which involves asking participants to provide a full account of the event, cued-recall where participants are asked specific questions about the event, or a recognition test where participants are shown various items and asked if they remember them being in the event (yes/no). The misinformation effect has been documented in both adults and children (e.g., Chan et al.,, 2009; Loftus et al., 1978; Otgaar et al.,2018; Roebers & Schneider, 2000; Zaragoza & Lane, 1994).

There continues to be theoretical debate about what happens to memory when a person is exposed to the misleading information. For example, Loftus (1975) hypothesised that the misleading information *alters* the initial event memory representation. At this time, she suggested that some false memories develop as the post-event information or the misinformation replaces or combines with the original memory trace, and recollection is a culmination of both (Loftus, 1975). The concept of 'retrieval blocking' suggests that

memories of both the original and misinformation memories co-exist, and that retrieval of the original memory is blocked by stronger and more recent misinformation memory (Bekerian & Bowers, 1983).

To better understand the mechanisms underlying the development of false memories, the 'modified' recognition memory test was developed. In the modified test procedure, the misinformed item (e.g., a screwdriver) is not an option in the subsequent memory test, and instead participants choose between the original detail actually seen in the events (e.g., a hammer), and a brand new 'foil' item (e.g., a wrench; McCloskey & Zaragoza, 1985). It was hypothesised that if misleading information altered the original memory trace (as proposed by Loftus), it would do so even when the misinformed item was not available at test. Therefore, those participants who had been misinformed would select the correct original item less and the brand-new foil item more, compared to participants in the control group who had not been misled. It was found that participants in both groups were just as likely to remember the original item, which the authors concluded to mean that the original memory was not altered nor was it inaccessible (McCloskey & Zaragoza, 1985). It was suggested that perhaps strategic memory effects (e.g., strategic elimination of possible options or guessing) could account for the misinformation effect and as a result argues against any memory impairment hypotheses (Zaragoza et al.,1987). Strategic effects however do not account for the increased misinformation effects seen after longer retention intervals using the modified memory test (Belli et al.,1992) which again suggests genuine memory impairment.

Other memory-based theories have been put forward to explain how false memories develop. The Activation-based model (Ayers & Reder, 1998) hypothesises that the activation strength of the memories explain why misinformation items are reported, as the misinformation sources are more highly 'activated' as they have been encountered more recently. Source memory theories (e.g., source misattribution, Johnson, 1988) suggest that

poor discrimination between memories of two sources (i.e., the event source and the misinformation source) can lead to false memories. For example, one memory trace may be for the original memory event, and another could be from when a person was questioned about the event and provided with misinformation. Traces from one event (e.g., being asked about the colour of a car, even though there was no car) have the potential to be attributed to another (e.g., the belief that there was a car in the original event) which leads to a false memory (e.g., Wade et al., 2002).

Fuzzy trace theory (Reyna & Brainerd, 1995) is another trace memory-based account of false memories, which suggests that there are two different memory traces: *verbatim* traces that record basic sensory-information and details of the memory and *gist* traces that record the semantic/conceptual meanings and interpretations of a memory. Since the verbatim memory traces record detailed sensory information, they are thought to protect from spontaneous false memory development as it records the tangible details of an event (Reyna et al., 2016). However, verbatim traces may increase the likelihood of accepting false post-event information (i.e., misinformation), as the detail of the post-event information will be recorded as verbatim traces, and therefore be more likely to be incorporated into later accounts (Reyna et al., 2016).

It has been suggested that this explains the lack of correlation seen between post-event misinformation studies and spontaneous false memory studies (Reyna et al., 2016). However, both spontaneous and post-event misinformation false memories are theorised to develop if gist memories are relied on too heavily as this is characterised by a general overview or familiarity of context (e.g., knowing that it was a ‘tool’), allowing for the details to be changed.

Children and Adults

For a long time in the literature, evidence has shown that children are more suggestible than adults to post-event misleading information provided by external agents. However, children may be less likely to develop spontaneous false memories themselves (Otgaar et al., 2016). One explanation for the finding that children are less likely to internally generate spontaneous false memories is the associative activation theory. This theory hypothesises that internally generated false memories develop due to one's understanding and knowledge of the world. Knowledge of interrelated information becomes stronger and more automatic as we learn more, and therefore adults are more likely than children to develop false memories spontaneously, due to the stronger and more automatic connections between information that they have developed (Howe et al., 2009). This theory suggests then adults may be more susceptible than children are, as they have more knowledge of the world and therefore more associations are possible (Howe et al., 2009). FTT suggests a similar process, as gist memories become more relied on in adulthood due to the better understanding and meaning of the world, thus increasing the likelihood for spontaneous false memories (Reyna et al., 2016). Theoretically, children have fewer gist and verbatim traces, so may be less likely to develop false memories, but also less likely to protect themselves from developing them.

Other theories have also been suggested to explain why children may be more susceptible to misinformation than adults. For example, children have been identified as having more difficulty with source monitoring; the ability to identify where the source of one's memory originated (Thierry et al.,2001) which is associated with increased suggestibility (Lindsay et al.,, 1991). Moreover, other non-memory-based factors, such as social-demand influences, have been theorised as also explaining the difference between children and adults. Social demand accounts identify that children may accept misleading questions more as they are more socially compliant with perceived authority of the adult

questioning them (Ceci & Bruck, 1993; Lampinen & Smith, 1995). These theories are discussed in detail in Chapter 2.

Current Research Directions

In sum, the general consensus between academics appears to be that the impact of misinformation on memory is due to a mixture of the theories and memory mechanisms discussed above. Although children are susceptible to suggestion, more so than adults, the recent research appears to be that there are certain situations where children are not more suggestible than adults and that there are individual differences of suggestibility between children of the same age (Klemfuss & Olagues, 2020).

A recent review by Klemfuss and Olagues (2020) has found a plethora of individual differences that impact on the likelihood to be suggestible in both adults and children. The review identifies three broad categories: demographic factors, cognitive factors, and psycho-social factors. Demographic factors such as gender differences may be important, where males are more likely to comply with misleading questions (Gilstrap & Ceci, 2005). Cognitive factors such as lower intelligence (Bettenay et al.,, 2015) and lesser language skills (Curci et al., 2017) could increase suggestibility in children. Finally, psycho-social factors such as historical trauma (Vagni et al., 2015) and emotional arousal at the time of interview (Quas et al.,2014) also appear to be related to higher suggestibility in children, though the outcome of these factors yielded conflicting results.

Klemfuss and Olagues (2020) has allowed the research community to better understand the mechanisms behind suggestibility in children, so that recommendations can be made for professionals and practical forensic working. An important recommendation provided by the Klemfuss and Olagues (2020) review, was that cognitive factors like intellectual impairment and language skills be carefully considered prior to interview as these are highly related. The Psycho-social domains however have limited and conflicting results, so only tentative

suggestions can be made for practice. The conclusion of the review was that more research is needed to examine individual differences in suggestibility, and that more steps are required to ensure that forensic practitioners could accurately understand risk of suggestibility in children.

Thesis Aims

The overarching aim for this thesis is to better understand the practicalities of memory evidence from children. It will do that by: (1) examining how to incorporate measurement of suggestibility to misinformation into practice, to identify whether assessing suggestibility to misinformation could benefit forensic practitioners and child witnesses; and (2) examining how children are perceived by UK criminal justice professionals, and how professionals work with memory evidence from children. The thesis will collate relevant information to discern the presence and size of the misinformation effect in children compared to adults. It will also consider how suggestibility to misinformation can be measured for both theoretical advance in research and for practical use in the CJS by reviewing a potential psychometric. The thesis will also investigate perceptions of the reliability of child memory evidence and working with child memory information. By collating this information, it is hoped that the thesis might outline key areas where misinformation and suggestibility measures could support CJS processes.

Chapter Summaries

To achieve these aims, **Chapter Two** first collates and reviews the relevant data on child suggestibility by utilising meta-analytical processes. Although much of the research identifies that children are more suggestible than adults (e.g., Ceci & Bruck, 1993), sometimes adults have been found to be more suggestible than children (e.g., Otgaar et al., 2018), and from my review of the available literature there has been no meta-analytical estimates of the size of the misinformation effect in children compared to adults. Next, the

research in **Chapter Three** uses qualitative techniques to explore the perceptions of 12 professionals (police officers, social workers and sexual harm support staff) in the UK who currently work with memory evidence from child witnesses to understand the practicalities of working with child memory evidence. There is currently limited exploration of this topic with professionals in the UK, especially using a qualitative approach. Specifically, it was of interest to determine how professionals perceive working with children and their opinions on the reliability of accounts from child witnesses. It was of added interest to explore whether or not professionals were aware of the current research around child memory and reliability and whether or not this was of use to their practice. Next, **Chapter Four** reviews a measure of child suggestibility and critiques its potential practical use in the CJS. The Bonn Test Statement of Suggestibility (BTSS; Endres, 1997) was developed to measure individual differences in suggestibility in younger children between the ages of 4-10. The BTSS is available in various different languages, so the review compiles the research conducted across countries to provide an overall conclusion about the properties of the scale (e.g., validity and reliability), prior to reflecting on its potential practical use in CJS. Finally, **Chapter Five** summarises and connects the findings from the previous chapters. Both the theoretical and the practical implications of the outcomes are discussed and recommendations for both policy and future research are presented based on the findings.

CHAPTER II

The Misinformation Effect in Children and Adults: A Meta-analytical Review

Abstract

A prominent view in cognitive psychology and legal systems worldwide is that witnesses can provide inaccurate accounts because they are suggestible to misleading information, and children are inferior witnesses because they are more suggestible than adults. Yet, some studies investigating age-related differences yield conflicting conclusions. Further, many studies compare inaccuracy (i.e., the number of incorrect responses) to misinformation items, not the size of the misinformation effect (i.e., the difference in accuracy between control and misinformation items) in children and adults. We conducted a systematic search and found 17 papers comparing adults and children using a misinformation paradigm, yielding a total of 2,582 participants. We conducted two meta-analyses. The first compared the number of incorrect responses to misinformation items in children and adults. The second (including 10 papers from the original 17) compared the size of the misinformation effect in children and adults using sub-group analyses. In both analyses, children were more likely than adults to report incorrect responses when provided with misinformation. Moreover, there was clear evidence of a misinformation effect in children, but not in adults. A reanalysis of a meta-analysis by Payne et al. (1994) replicated these results. These findings have implications for understanding developmental changes in suggestibility, the reliability of adult witness memory, and methodology in the misinformation literature.

Introduction

Within Criminal Justice Systems (CJS), both child and adult witnesses are regularly requested to recall criminal events. During these investigations, witnesses may be subject to numerous interviews by different professionals attempting to obtain a complete and accurate account of the events in question. Unfortunately, each interaction could potentially introduce new information about the event into the witness's memory. For example, an interviewer might inadvertently provide information to the witness that the witness did not know (i.e., externally derived information), and the witness may subsequently incorporate the information into their own account of what they saw. This is problematic for at least two reasons. First, the witness's account should be independent from other evidence in the case. Second, the information incorporated into the witness's account may be inaccurate. Thus, the incorporation of post-event information can impede criminal investigations and potentially lead to miscarriages of justice.

The impact of inaccurate post-event information—termed 'misinformation'—on witness memory accuracy has been extensively studied experimentally using the misinformation paradigm (Loftus et al., 1978). In the misinformation paradigm, participants are shown a target event and are subsequently presented with misinformation about the event before their memory for the event is tested. It is widely accepted that the misinformation effect is observed in both children and adults, with many studies and reviews reporting that result (e.g., Chan et al., 2009; Loftus et al., 1978; Otgaar et al., 2018; Zaragoza & Lane, 1994). Notably, these findings have informed guidance for legal practitioners around the world for child and adult witnesses (e.g., National Research Council, 2014), and are cited by psychological scientists when giving expert witness testimony (e.g., New York vs. Weinstein, 2019). However, the question remains as to how suggestible children and adults are and what is the most appropriate analysis to conduct to

determine that. In this paper we use meta-analytic techniques to examine the size of the misinformation effect and investigate whether children are more suggestible to misinformation than adults.

Numerous cognitive and social theories have been put forth to explain the misinformation effect, and these theories differ in their predictions about age-related changes in susceptibility to misinformation. We begin by providing a brief overview of the theories that have been proposed to account for age-related changes in suggestibility.

Theories Predicting Children Are More Suggestible Than Adults

The predominant view in the CJS and psychological literature is that both children and adults are susceptible to misinformation, but children are more susceptible than adults (Cassel et al.,1996). Many studies have found that young children are more likely than older children and adults to change an initial account to incorporate post-event misinformation (e.g., Bruck & Ceci, 1999). For example, 3-year-olds reported more incorrect responses following misinformation than 7-year-olds when interviewed 2 weeks after an inoculation (Alexander et al., 2002), and a similar difference was found between 3.5-year-olds and 7-year-olds when interviewed 10 days after a 'Circus Day' (Geddie et al.,2000). Theoretical explanations for developmental decreases in susceptibility to misinformation include: Authority and Compliance, Language, Inhibitory control, and Source monitoring. We consider each of these next.

Authority and Compliance

If the source of misinformation is deemed to be proficient and hold expertise (i.e., have authority status), then this can increase an individual's likelihood of incorporating misinformation into their own account (Ceci et al.,1987; Smith & Ellsworth, 1987). Children may be viewed as low-credible communicators, whereas adult police officers are viewed as highly credible and having authority (Skagerberg & Wright, 2009). Due to the authority-

status difference that exists between children and adults, child witnesses are more likely to be influenced by social demands compared to adult witnesses (Ceci et al., 1987). Therefore, authority explanations of the misinformation effect hold that, compared to adult witnesses, child witnesses are more willing to please adults as authority figures and as a result are more likely to agree with suggestions from adults and succumb to misinformation (Bruck & Ceci, 1997).

Language

Other research suggests that children compared to adults are more susceptible to misinformation (at least in part) due to their poorer language skills (Clarke-Stewart et al.,2004). Simply understanding fewer words may increase the likelihood that an individual can be misled (Clarke-Stewart et al., 2004). For example, Kulkofsky and Klemfuss (2008) found that those 3- or 4-year-olds with poorer language abilities were more likely to accept misinformation. Therefore, children may be more suggestible than adults, because they are generally less likely to understand the words being used during questioning and are therefore more likely to be misled by suggestive questions (Clarke-Stewart et al., 2004).

Inhibitory Control

Children may be less likely than adults to control their impulses because the pre-frontal cortex is not yet fully developed (Casey et al., 1997), and therefore they may be more likely to thoughtlessly agree, or propose a narrative that agrees, with an interviewer's suggestions (Roberts & Powell, 2001). Alexander and colleagues (2002) found that, in children aged between 3 and 7, cognitive inhibition initiated during the Stroop Test and 'effortful control' on the Child Behaviour Questionnaire (CBQ; Rothbart, Ahadi, Hershey & Fisher, 2001) were associated with fewer incorrect answers to misleading questions about a witnessed event. Clarke-Stewart et al. (2004) also found that children who were able to

'wait' were less susceptible to misinformation. As such, the development of impulse control with age, may increase the ability to defend against misinformation.

Source Monitoring

Some researchers deem source monitoring errors to be the most integral memory-based mechanism explaining the misinformation effect (Polczyk, 2007). Source monitoring is a person's ability to discriminate where they learnt about information, for example determining if information was learnt while witnessing the event or being told about the event afterwards (Ackil & Zaragoza, 1995). Source-monitoring confusion increases the likelihood to succumb to misinformation, because people incorrectly remember the post-event misinformation as being part of the original event (Lindsay & Johnson, 1989). According to a source-monitoring account, children are more susceptible to misinformation than adults because children generally experience more source-confusion than adults (Lindsay et al.,1991).

Theories Predicting Children Are Not More Suggestible Than Adults

Other research suggests that children may not be more suggestible than adults and may even be less suggestible than adults in some circumstances (Duncan et al.,1982; Otgaar et al., 2018). For example, Otgaar et al (2016) showed adults and children (4-6-years old, 7-9-years old, and 10-12-year-old) a video of a bank robbery. Afterwards, they presented words that were associatively related to the video (e.g., pistol, vault) and misinformation (e.g., told participants there was a pistol at the event). The researchers found a developmental reversal effect; they found that adults were more susceptible to misinformation than children under circumstances where the misinformation was associatively related to the to-be-remembered information. The researchers concluded that younger children are less prone to suggestion than older children and adults in conditions

where misinformation is due to participants fostering associations between presented and non-presented details.

Associative Activation Theory

The developmental reversal effect in suggestibility can be explained by the associative activation theory (AAT; Howe et al.,2009). According to this theory, false memories arise due to associations between the network of interrelated information nodes in our brains. As we develop, throughout life, we gain more knowledge and more understanding, and mental associations become more automatic, frequent, and stronger (Howe et al., 2009). Therefore, the AAT suggests that, if the misinformation is associatively related to information in the event, adults are more likely to be susceptible to that misinformation than children. Indeed, this pattern of findings has been evidenced in research using word lists (Howe, 2008), stories (Howe & Wilkinson, 2011), and associative item recall tasks (Lyons et al., 2010).

Fuzzy Trace Theory

Another memory-based account of the misinformation effect is fuzzy trace theory (FTT; Reyna & Brainerd, 1995). According to fuzzy trace theory, there are two parallel memory stores: verbatim and gist. Verbatim traces contain the context and detail of the memory, whereas gist traces include information about the general meaning of a memory (Reyna & Brainerd, 1995). Regarding false memories, fuzzy trace theory assumes that verbatim traces will protect a memory from being manipulated, but gist traces will not as they allow for semantic associations to be made following memory consolidation (Reyna et al., 2016). Similar hypotheses using FTT believe that verbatim memory may promote false memory when subjected to misinformation, as details of the misinformation memory may be remembered more vividly, thus increasing likelihood to recount that information. However, the same verbatim memories may be protective of spontaneous false memories due to

having memory of the event details (Reyna et al., 2016). In adulthood, the better developed knowledge of the world may increase reliance on gist traces which might suggest that adults are more subject to false memory development (Reyna et al., 2016). Compared to adults, children are theorised to have not yet developed the same gist or verbatim memory capabilities. The FTT posits that children may remember less but that they are more 'faithful recorders' (Reinard et al., 2016) because they have less knowledge and information about the world at which to influence their gist memory traces. However, children may potentially have less protection against false memories should they develop, due to the less developed verbatim trace encoding.

Mental Representations

A final theoretical explanation of susceptibility to misinformation is concerned with mental representations. According to this account, there can be numerous versions of one event stored in memory, and these representations are the foundation for beliefs of any given witnessed event (e.g., Templeton & Wilcox, 2000). It has been found that representational and memory abilities account for a significant proportion of the variability in misinformation reporting (Templeton & Wilcox, 2000). That is, both children and adults who are more skilled at holding in mind multiple representations are more susceptible to misinformation (Templeton et al., 2000). Children demonstrate representational abilities similar to that of adults from the age of 6 (Templeton & Wilcox, 2000). Therefore, according to this account, age may not be the most important predictor of suggestibility, instead it is an individual's ability to hold different representations of a memory that is important.

Although much research has been conducted on the misinformation effect, and multiple theories have been proposed, what is currently unknown is the size and consistency of the misinformation effect in children compared to adults. This is important because

research on the misinformation effect is regularly used in practice to inform legal proceedings. Two key reviews have been published in recent years and warrant discussion (Otgaar et al., 2018; Payne et al., 1994).

Existing Reviews and Analytical Considerations

Recently, Otgaar et al (2018) published a narrative review of children and adult false memory studies. The authors concluded that children may be more or less suggestible than adults depending on certain conditions. For example, compared to adults, children may be less suggestible in situations that rely on forming associative meanings, because their lack of knowledge can protect them from suggestions; but children may be more suggestible in situations when the answer is insinuated. The authors included papers sampling a broad range of ages and labelled age groups in each study as either 'younger' or 'older', resulting in interesting general conclusions about suggestibility in children compared to adults, but no precise measurement of the size of this effect. For example, one of the papers included in the review (Ornstein, Gordon & Larus, 1992) sampled 3-year-olds (labelled as 'younger') and 6-year-olds (labelled as 'older'), while another paper included in the review (Cohen & Harrick, 1980) sampled 9- and 12-year-olds (labelled as 'younger') and college age students (labelled as 'older').

Moreover, and critically, in keeping with other research in the field, Otgaar and colleagues (2018) focused their review on participants' *incorrect answers to misinformed items*. Typically, in misinformation studies, control data are also collected. In within-participant experimental designs, participants are misled about some details from the original event (e.g., told that a balloon burst at the party when it had not; misled items) and not misled about other details (e.g., told that someone get kissed at the party when they had; control items). In between-participant designs, some participants are misled (misinformation condition), while others are not (control condition). In the review by Otgaar et al. (2018), the

key information discussed was the frequency with which children versus adults reported misinformation and provided an incorrect response to questions about misled details. The accuracy of answers to questions about non-misled details (i.e., control details) were not included in the analysis. Note that the approach taken by Otgaar et al. is not uncommon. Other reviews, for example comparing suggestibility to misinformation in young and older adults (e.g., Wylie et al.,2014), have also omitted control items and compared only misinformation items across the age groups. Notably, without considering control items when comparing children to adults, it is possible that the increased likelihood of children reporting misinformation is due to differences in memory maturation, not age-related differences in suggestibility per se. Put simply, although children might be more likely to incorrectly report misinformation than adults, children might also just be more likely to incorrectly report other information about which they have not been misled.

Control data have been considered in other reviews. In a prominent meta-analysis, Payne, et al (1994) examined the magnitude of the misinformation effect by comparing the difference in accuracy between control and mislead items. Payne et al. (1994) analysed studies (48 cases from 11 studies) that used the misinformation paradigm to study memory impairment across short and long retention intervals on the *modified recognition test*. The modified test was developed to better understand the memory mechanisms underlying the misinformation effect. In the original misinformation recognition tests, subjects are asked to choose between two items: the originally seen item (e.g., a hammer) and a misled item (e.g., a screwdriver). Whereas in the modified version, the misleading information (e.g., a screwdriver) is not included into the final test; instead participants are asked to choose between the original item (e.g., a hammer) and a new item (e.g., a wrench; McCloskey & Zaragoza, 1985). Payne et al. reported a significant misinformation effect, which was larger at long retention intervals compared to short retention intervals (based on 48 cases from 13

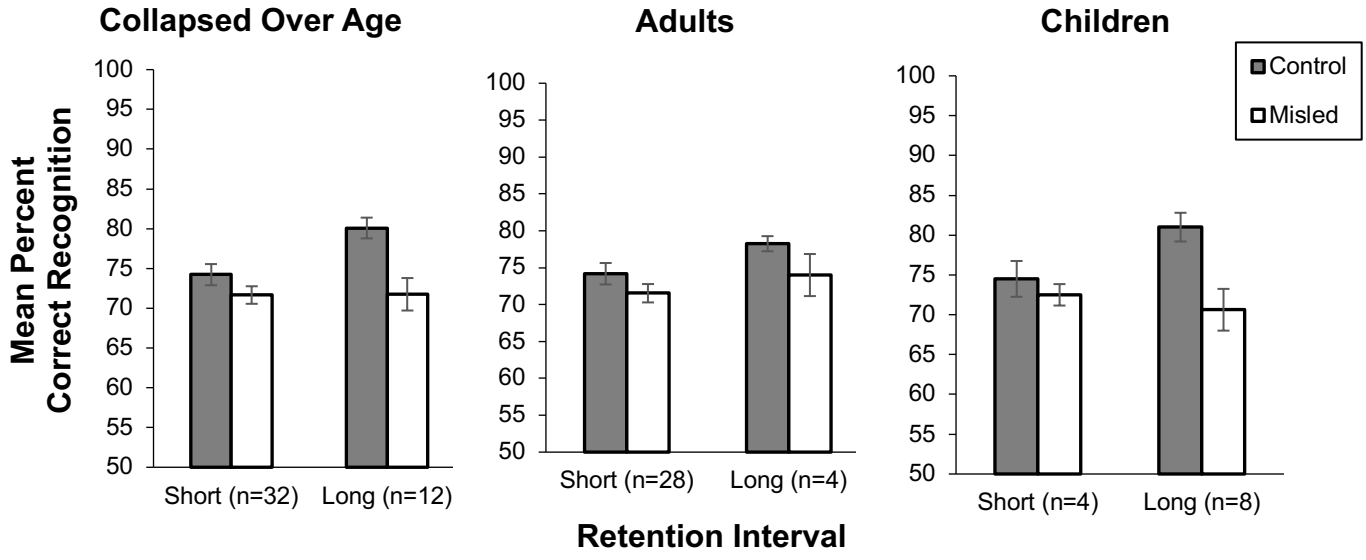
studies, but with 4 cases excluded due to ceiling effects). The analysis did not take into account methodological rigor of the individual studies, nor did it weight studies by sample size. Figure 1A shows the data reported by Payne et al. (1994) in Table 1 of their paper (Williams et al., under review). Across the 44 cases, mean recognition accuracy on the modified test was significantly lower in the misled conditions than in the control conditions (71.7% vs. 75.8% accurate, $t(43) = 4.65$, $p = .000016.$, $r^2 = .335$). There is a larger difference in accuracy between control and misled items on the modified test in the studies that used long compared to shorter retention intervals (long $M_{difference} = .08$ versus short $M_{difference} = .02$), as indicated by an item type x retention interval interaction effect, which is moderate in size, $F(1, 42) = 7.33$, $p = .01$, $\eta_p^2 = .149$. Payne et al. concluded from their analysis that "the misinformation effect is indeed a real memory phenomenon" (p. 381).

The data from Payne et al. (1994), however, deserves a second look because participant age, which was not accounted for in the analysis, may have influenced the results. Williams et al (under review) noted that child participants contributed to 67% (8 out of 12) of the data points in the long retention interval studies and to only 12% (4 out of 32) of the data points in the short retention interval studies. Conditioning the data reported in Payne et al. (1994) on age suggests that there could be differences between adults and children in the size of the misinformation effect depending on retention interval. Figure 1B-C show the data from Payne et al., this time separated for adults and children, and as a function of retention interval. The average mean difference in control and misled accuracy on the modified test for the adult studies using a short retention interval was .03 ($M_{control} =$ 73.81 versus $M_{misled} = 71.11$) and for children it was .02 ($M_{control} = 74.50$ versus $M_{misled} =$ 72.50). However, for studies using a long retention interval, the difference between control and misled accuracy on the modified test for adults was again .03 ($M_{control} = 78.50$ versus $M_{misled} = 75.00$); but, for children the difference was .10 ($M_{control} = 81.00$ versus $M_{misled} =$

70.62), over three times larger (Williams et al., under review). Therefore, when separating the data by age and retention interval, the misinformation effect (as measured by the modified test) could be smaller in adults compared to children, but only at longer retention intervals. Note, however, that this analysis relies on cross-study comparisons because some studies sampled adults and others children, using different stimulus materials, and the number of studies when conditioned on age and retention interval is small. Therefore, an updated meta-analysis is needed to analyse papers that include both child and adult groups within a single study.

Figure 1

Mean percent correct recognition (and standard error bars) for the control and misled conditions from the short- and long-interval conditions (A) as presented in the original Payne et al. (1994) review collapsed over age, and the same studies separated for (B) adult participants and (C) child participants.



The Current Study

To better understand suggestibility in children and adults, we conducted a systematic search of the post-event misinformation literature. We employed pre-defined eligibility

criteria for studies including both child and adult participant groups and assessed the validity of study findings (e.g., by assessing risk of bias). We focused on comparing two predefined age categories; children (<14 years) or adults (>18 years)—to prevent overlap of age categories (child or adult) across studies. We conducted two meta-analyses. First, we examined susceptibility to post-event misinformation by directly comparing the frequency of incorrect responses to misled items in children and adults (*misinformation item analysis*). Second, we compared the mean difference between control items and misinformation items (i.e., the size of the misinformation effect) in adults and children using sub-group analyses (*control - misinformation item analysis*).

Method

No approval of research ethics committees was required to accomplish the goals of this review because no experimental work was conducted.

Identifying Primary Studies

Search of Electronic Databases

Systematic searches of the literature were conducted in March and October 2019. Search terms were developed to identify studies examining the misinformation effect in children compared to adults. The search terms were *child* with adult* and suggestibility*, and synonyms of c*hild** (*children; develop*; development; school age*; young*/youth*, both achieved using 'you*'; juvenile**) with each synonym of *suggestibility* (*misinformation, false memories, memory, recall* and mislead**). We entered the search terms into 9 databases: Web of Science, Psych Info, Scopus Database, Nursing and Allied Health Database (Proquest), CINHAUL Plus Database (EBSCO), Criminal Justice Database (Proquest), Social Science Database (Proquest), Child Development and Adolescent Studies (EBSCO), and Sociology Database (Proquest). We also used other search methods, such as identifying relevant papers in reference lists.

Selection Criteria

The main criteria for a study to be included in the meta-analysis was that it employed suggestive techniques, misleading questions, or introduced misinformation, and directly compared memory accuracy between children (aged < 14) and adults (aged > 18). We operationalised children as any group of participants 14-years or younger. Currently in England and Wales, children under the age of 14 are not expected to give evidence under oath, and therefore cannot be held in contempt of court (Youth Justice and Criminal Evidence Act, 1999). A witness is only able to give evidence under oath if they are able to give 'intelligible testimony' (i.e., are competent to provide evidence), if they have sufficient appreciation of proceedings, and are able to understand the questions being asked of them. Conversely, we operationalised adults as any group of participants 18-years or older, which is considered the threshold of adulthood in most legal codes (e.g., UK, US, China, Australia, Greece). The full inclusion criteria are detailed in Table 1.

Table 1
Inclusion and Exclusion Criteria

Inclusion Criteria	**Justification**
Method	
Studies using the misinformation paradigm	
Participant's experience/witness an event and recall the event or recognise items from an event (no line-up studies, DRM paradigm studies)	Focusing the inclusion criteria on recall and recognition from an event only can reduce heterogeneity between studies.
Include both general memory research and forensic setting (e.g., crime event) research	Both are relevant for the current analysis, since both examine the misinformation effect.
Participants	
Papers comparing children and adults using the same experimental methodology	To permit direct comparison across the age groups.

Children defined as anyone under the age of 14	Any child under the age of 14 is unable to give sworn evidence in criminal proceedings in England and Wales as stated in the Youth Justice and Criminal Evidence Act 1999.
Adults defined as anyone over the age of 18	18 years is considered the threshold of adulthood in most legal codes UK, US, China, Australia, Greece).
Outcome data *Studies that report either means and standard deviations, F-test statistics, or Cohen's d effect size*	To ensure that the study outcome can be calculated into an effect size to use in the meta-analyses.
Type of article *Empirical research. Not meta-analyses, theoretical papers, reviews of research, commentaries on research, clinical guidance papers, case studies, qualitative papers, psychometric validation studies*	The excluded article types do not provide the outcome data that can be used within a meta-analysis.

The results of the systematic search are presented in Figure 2. Following the removal of duplicates, the search yielded 689 articles that appeared relevant by assessment of title alone. Next, 372 articles were excluded because they were not relevant, leaving 317 to be screened by abstract. Following this, 251 articles were excluded for various reasons as detailed in Figure 2. Finally, 66 full texts were read and 49 excluded as they did not fit the inclusion criteria. This resulted in 17 studies that were included in our meta-analysis. Table 2 displays a summary of the main features of each of the 17 studies. Note that the studies differ methodologically to the studies included in the Payne et al. (1994) meta-analysis which all used a modified recognition misinformation test and immediate testing (see Table 2).

Table 2
Relevant details of papers included in the current meta-analysis

Study	Event	Event Mode	Control Within- or Between-subjects	Control Type	Interval between encoding and test	Timing and method of misinformation presentation	Filler task between encoding and questioning	Memory test as reported in study results	Tests format	Rapport Phase
Ackil & Zaragoza (1995)	No crime	Video	Between	Consistent	1 week	Immediately after the event, before the filler and interval. Presented as a summary.	Sorting task after 1 week interval	Proportion of suggested items recognised as "old"	Children interview, Adults paper	Absent
Bjorklund et al. (2000)	Crime	Video	Between	Neutral	2 days	Immediately after the filler, before the interval. Presented as first set of interview questions, prior to final memory test.	Puzzles & magazines immediately after video	Percentage of incorrect recognition responses to items	Interview	Absent
Brackmann, Otgaar, Sauerland & Howe (2016)	Crime	Video	Within	Neutral	24 hours	After the interval and filler. Presented as another witness's account.	Tetris after event and after misinformation	Percentage of incorrect responses to cued recall questions about the observed items and the meaning of video content	Interview	Absent
Cassel & Bjorklund (1995)	Crime	Video	Between	Neutral	1 month	After the 1 week interval and again after the 1 month interval. Presented during first set of questioning and final memory questioning	Puzzles & magazines immediately after video	Percentage of correct responses to positive and negative leading cued-recall questions	Interview	Absent
Cassel, Roebers & Bjorklund (1996)	Crime	Video	Within	Neutral	1 week	After the 1 week interval. Presented during final questioning.	No filler task	Percentage of correct responses to cued-recall questions	Interview	Absent
Cohen & Harrick (1980)	Crime	Video	Within	Consistent	1 week	Immediately after the event, 1 week before testing. Presented as	No filler Task	Proportion of correct responses to cued-recall questions	Interview	Absent

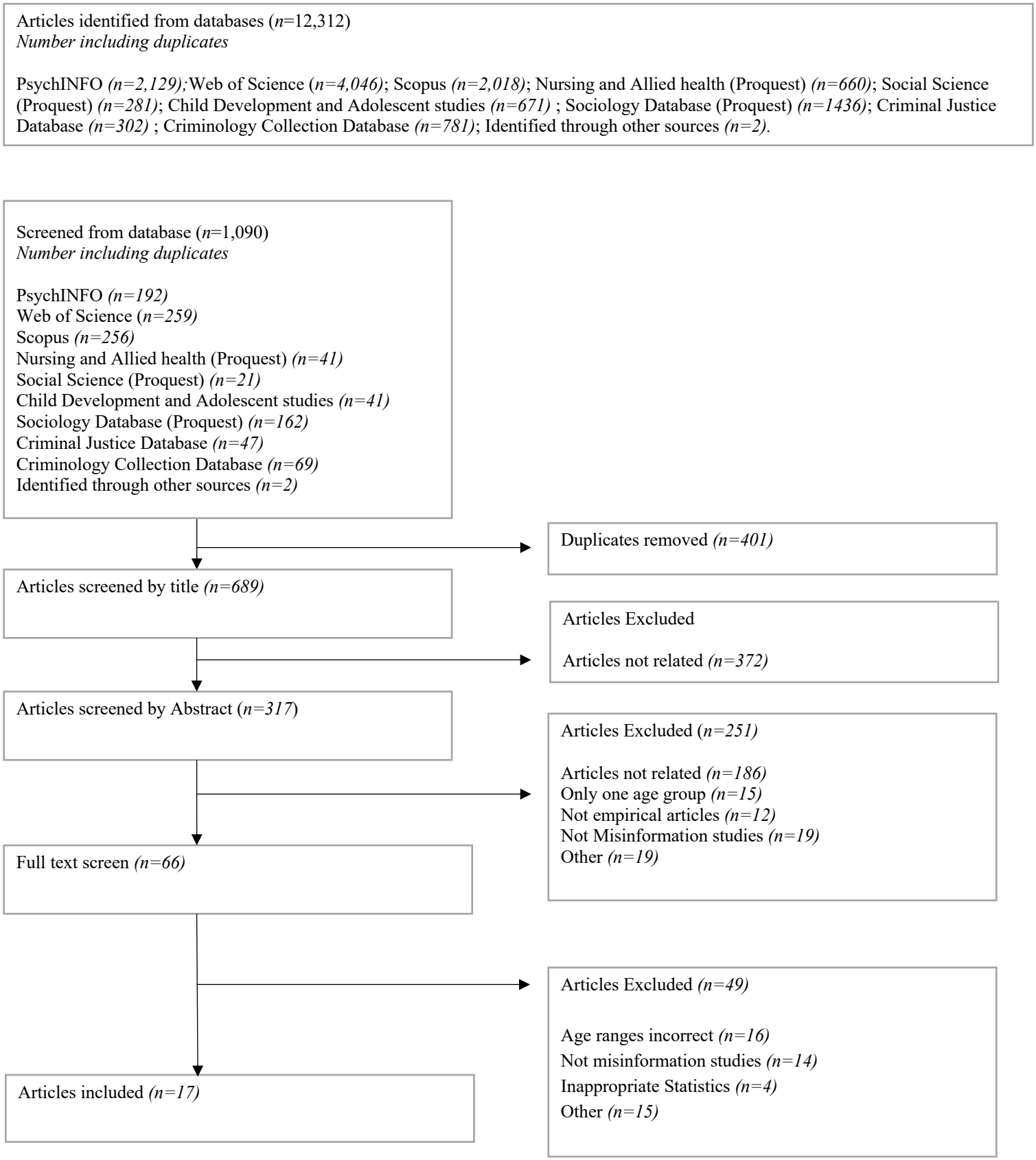
						first set of interview questions, prior to final memory test.				
Flin Boon & Knox (1992)	No crime	Live	Between	Neutral	5 months	One day or 5 months after the event, depending on experimental group. Presented during questioning.	No filler task	Mean suggestibility scores to cued-recall questions	Interview	Absent
Goodman & Reed (1986)	No crime	Live	Within	Neutral	4/5 days	After the interval. Presented during final questioning	No filler task	Mean correct answers to cued-recall questions	Paper	Absent
Hunt & Borgida (2001)	No crime	Video	Between	Neutral	6/8 days	Immediately after the video, before the interval. Presented as first set of interview questions, prior to final memory test.	No filler task	Mean accuracy of cued-recall questions	Interview	Present
Laumann & Elliot (1992)	No crime	Video	Within	Neutral	Immediate	Immediately after the filler activity. Presented during final memory test questioning.	Drawing a picture	Percentage of compliance with misleading cued-recall questions	Interview	Absent
Robinson & Briggs (1997)	No crime	Video	Between	Neutral	24 hours	After the interval. Presented during final memory test questioning.	No filler task	False positives made to misleading and objective cued-recall questions	Interview	Absent
Roebers & Fernandez (2002)	No crime	Video	Between	Neutral	3 weeks	After the interval. Presented during final memory test questioning.	No filler task	Mean correct answers to open-ended answerable cued-recall questions	Interview	Present
Roebers & McConkey (2003)	No crime	Video	Between	Neutral	2 weeks	1 week after the event, 1 week before the final memory test. Presented as first set of interview questions, prior to final memory test.	No filler task	Mean percentage of incorrectly answered misleading questions (Yes-No questions)	Interview	Present
Roebers & Schneider (2001)	Crime	Video	Within	No information	4 weeks	Presented 3 weeks after the event, 1 week before final memory test.	No filler task	Mean number of correctly reported	Interview	Absent

						Presented as first set of interview questions, prior to final memory test.		items to cued-recall questions		
Roebers & Schneider (2005)	No crime	Video	Within	Neutral	3 weeks	After the interval. Presented during final memory test.	No filler task	Percentage of correct answers to cued-recall questions	Interview	Present
Roebers, Bjorklund, Schneider & Cassell (2002)	Crime	Video	Between	Neutral	1 week	After the interval. Presented during final memory test questioning.	No filler task	Percentage of correct responses to cued-recall questions	Interview	Present
Templeton & Wilcox (2000)	No crime	Video	Within	Consistent	Immediate	Before the questioning. Presented as a summary read aloud.	No filler task	Mean percentage of correct responses to cued-recall questions	Children interview, Adults paper	Absent

Note. The control type is defined as "Neutral" when participants were provided with information or questioned using language that did not insinuate an answer; "Consistent" when participants were provided with information or questions that gave them correct information about the event; "No information" when participants were provided with no information prior to being questioned or the questions were open-ended. Rapport is coded as present when it was stated in the study methodology that interviewers spent time building rapport with participants prior to questioning, and absent if no rapport phase was described in the study methodology. The information in the Memory test column provides details of the type of data that was extracted from each paper for the purpose of the current meta-analysis. Many papers included multiple other analyses, but the meta-analysis was only interested in the portion of the data reported here.

Figure 2

Results of Systematic Search and application of the Inclusion Criteria



Study Quality Assessment

Quality criteria were developed to assess risk of bias in the 17 papers and each paper was rated as either 'low' or 'suspected' risk of bias. The quality criteria were adapted from existing frameworks including: Downs and Black (1998), The Cochrane Collaboration Risk of Bias Tool (Higgins et al., 2011), and the Risk of Bias Assessment Tool for Nonrandomised Studies (Kim et al., 2013). Table 3 summarises the quality criteria and Table 4 shows the risk of bias assessment and overall quality score for each paper. The overall quality score was computed by summing each of the area of risk of bias (low risk = 0, suspected risk = 1), expressed as a percentage of the potential maximum score. As described below, the overall quality score was used in part to determine the precision of a study effect in the random effects model.

Table 3

Quality Assessment Framework

Risk of Bias	Definition	Low risk of Bias	Suspected risk of bias
Selection Bias	Selection bias occurs in intervention studies when there are systematic differences between comparison groups in response to treatment or process. Intervention studies are especially susceptible to selection bias unless efforts are made to minimise it. The most effective method is random allocation to treatment and control items. However, for each of the current studies, purposive sampling (child vs adult) will need to occur before randomisation to different control and experimental (i.e., misinformation) conditions.	The characteristics of the study population are clearly reported and without evidence of bias.	The characteristics of the study population are not clearly reported.
Performance Bias	Performance bias refers to systematic differences in exposure to factors other than the interventions of interest. After enrolment into the study, blinding (or masking) of study participants and personnel may reduce the risk that knowledge of which intervention was received. Within the current analysis, blinding is only possible to a certain extent because participants are either children or adults. However, blinding can be used for misinformation and control items or conditions and in other methodological details.	Detailed reporting of the methodology and procedure. Use of blinding techniques.	No details of procedure. No attempted blinding.
Detection Bias	Detection bias refers to whether the design of the study is optimised to detect the effect in question. Ratings of design bias therefore reflect the position of the study design within the hierarchy of possible designs, with less optimal designs receiving some penalty. Also, blinding of outcome assessors can be especially important for assessment of subjective outcomes.	Clear coding methodology explained and attempts to reduce biases. Blind coding.	Possible subjective coding. No blind coding.
Statistical Bias	Bias resulting from the (inappropriate) statistical treatment of the data. Within the current analysis, the most appropriate statistical approach is an analysis of variance.	Use of analysis of variance.	Use of other analysis to measure mean difference.
Reporting Bias	Reporting bias refers to systematic differences between reported and unreported findings. Within a published report those analyses with statistically significant differences between intervention groups are more likely to be reported than non-significant differences. This sort of within-study publication bias is usually known as outcome reporting bias or selective reporting bias, and may be one of	Reported all results of measures as outlined in the method.	Not all descriptive and/or summary statistics are presented.

Risk of Bias	Definition	Low risk of Bias	Suspected risk of bias
	the most substantial biases affecting results from individual studies (Chan & Douglas, 2005).		
Generalisability	Generalisability describes the extent to which research findings can be applied to settings other than that in which they were originally tested.	Sufficient sample for generalisation.	Idiosyncratic features of sample and study design

Table 4
Quality Assessment Results for the 17 Studies

Study	***Selection Bias***	***Performance Bias***	***Detection Bias***	***Statistical Bias***	***Reporting Bias***	***Generalisability***	***Quality Index***
Ackil & Zaragoza (1992)	Suspected risk	Low risk	Low risk	Low risk	Suspected risk	Suspected risk	75%
Bjorklund, Cassel, Bjorklund, Brown, Park, Ernst & Owen (2000)	Suspected risk	Low risk	Suspected risk	Low risk	Suspected risk	Suspected risk	67%
Brackmann, Otgaar, Sauerland & Howe (2016)	Suspected risk	Suspected risk	Suspected risk	Low risk	Low risk	Low risk	75%
Cassel & Bjorklund (1995)	Suspected risk	Suspected risk	Low risk	Low risk	Suspected risk	Suspected risk	67%
Cassel, Roebers & Bjorklund	Low risk	Low risk	Low risk	Low risk	Low risk	Suspected risk	92%
Cohen & Harrick (1980)	Suspected risk	Low risk	Suspected risk	Low risk	Low risk	Suspected risk	75%
Flin Boon & Knox(1992)	Low risk	Suspected risk	Suspected risk	Low risk	Suspected risk	Suspected risk	67%
Goodman & Reed(1986)	Suspected risk	Low risk	Suspected risk	Low risk	Suspected risk	Suspected risk	67%
Hunt & Borgida (2001)	Suspected risk	Suspected risk	Low risk	Suspected risk	Suspected risk	Suspected risk	58%
Laumann & Elliott(1992)	Low risk	Low risk	Low risk	Low risk	Suspected risk	Low risk	92%
Robinson & Briggs(1997)	Suspected risk	Suspected risk	Suspected risk	Suspected risk	Suspected risk	Suspected risk	50%
Roebers & Fernandez (2002)	Low risk	Suspected risk	Low risk	Low risk	Low risk	Low risk	92%
Roebers & McConkey(2003)	Suspected risk	Low risk	Suspected risk	Low risk	Low risk	Suspected risk	75%
Roebers & Schneider (2001)	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	100%
Roebers & Schneider (2005)	Low risk	Low risk	Suspected risk	Low risk	Low risk	Low risk	92%
Roebers, Bjorklund, Schneider & Cassel (2002)	Low risk	Suspected risk	Low risk	Low risk	Low risk	Low risk	92%
Templeton & Wilcox (2000)	Suspected risk	Suspected risk	Suspected risk	Low risk	Suspected risk	Suspected risk	58%

Data Extraction

For the first analysis (misinformation item analysis), we extracted information about the mean number or percentage of correct or incorrect responses to misinformation items, for children and adults. For the second analysis (control – misinformation item analysis), we extracted information about the number or percentage of correct or incorrect responses to control items (i.e., no misinformation) and misinformation items, for children and adults. Data were transformed into mean or percentage correct (from mean or percentage incorrect) where necessary, consistent with the approach taken in the Payne et al. (1994) meta-analysis. This approach enabled us to examine the size of the misinformation effect in adults versus children, where a larger difference in percent correct between the control and misinformation items indicated a larger misinformation effect. Note that for the control - misinformation item analysis, only 10 of the 17 studies included in the misinformation item analysis were identified as being appropriate to use because the 7 other papers did not report the statistics required for this analysis (e.g., were missing control data or descriptive statistics; see Appendix A for details).

For both analyses, we extracted the mean (number or percentage), the standard deviation (*SD*) and sample size for the child and adult groups. If *SD* was not reported, the pooled *SD* could be calculated. If means, *SD*'s and sample sizes were not reported, then the *F* statistic was used to estimate Cohen's *d*. If neither summary statistics (mean, *SD* and *n*) nor *F* statistics were reported, then effect sizes reported in the primary studies were extracted. The effect sizes as reported in the primary studies are frequently calculated from data that has been adjusted for the association with one or more covariates, which may result in dissimilarity with the effects reported in other primary studies.

Multiple reporting of outcomes for children or adults within one paper (e.g., having multiple groups of children at different age rages) were combined into a single quantitative outcome (i.e., into one child group) using the procedure described by Borenstein et al.

(2009). Three studies differentiated between questions on the memory test that asked about central or peripheral information in the event (e.g., details about the treasure, or dialogue about the treasure, respectively; Roebers & McConkey, 2003). For two studies, the overall combined mean data (over central and peripheral items) were extracted (Roebers & McConkey 2003; Cassel & Bjorklund, 1995). For one study that did not report an overall combined mean (Roebers et al.,2002) outcomes for the peripheral information was extracted because there were more questions asking about peripheral information in this study.

Data Analysis Strategy

Data that violates assumptions

Cohen's *d* has been shown to systematically overestimate the value of mean difference in small sample sizes (Borenstein et al., 2009). This bias can be removed by transforming Cohen's *d* into an unbiased estimate called Hedges g (Hedges, 1981) for the calculations and back-transforming into Cohen's *d* for interpretation and reporting. We took this approach.

The Omnibus Tests

The omnibus tests were calculated using random effects models, which aims to estimate the mean distribution of possible effects, rather than identify one true effect size. The goal of our random effects models was to obtain a mean effect from a range of studies. This prevents the overall estimate being overly influenced by larger studies, studies of poorer methodological quality or under-influenced by smaller studies. Therefore, in the random effects models, the precision of an effect is estimated as a function of the sample size, and also the rating of methodological quality (i.e., the overall quality score reported in Table 4). Therefore, the random effects models represent the synthesis that would have been obtained if all of the studies were as methodologically rigorous as the best-rated study included (e.g., see Doi & Thalib, 2008). We used the DerSimonian and Laird method for

calculating the between study variation for fitting the random effects models. The DerSimonian and Laird method assumes that the random effects are normally distributed and that therefore the effect sizes reported in the primary studies included into the analysis are also approximately normally distributed. In both analyses, we checked, and no substantive deviation from a normal distribution was observed.

Results

Misinformation Item Analysis

For this analysis, we compared children and adults' incorrect answers to misinformation items. Focusing on incorrect answers to misinformation items is consistent with the methodology in other similar reviews (e.g., Wylie et al., 2014).

Heterogeneity

Higgins I^2 was used to measure heterogeneity across studies. A larger Higgins I^2 value indicates greater heterogeneity, and values over 75% indicate problematic levels of heterogeneity (Higgins & Thompson, 2002). A substantial level of heterogeneity (87%) was observed (I^2 = 86.9%, 95% CI [80.6, 91.2], tau^2 =.25, Q = 122.44; df = 16; $p < .0001$). This suggests that the Standardised Mean Difference (SMD) estimates of the primary studies are biased by the presence of confounding factors. To better understand the unacceptable level of heterogeneity, the impact of disproportionately influential studies was assessed using a "leave-one-out" analysis, in which the random effects model was calculated with each of the primary studies removed. None of the papers were considered influential enough to warrant removing from the analysis (See Table 5 for details).

Table 5

Leave-one-out analysis for the misinformation item analysis, to demonstrate the impact of removing each study on the standardised mean difference (SMD) effect size and heterogeneity

Omitting	**Random effects model**		**Heterogeneity**	
	SMD	**SMD 95% CIs**	**I^2 (%)**	**Tau^2**
Ackil & Zaragoza (1992)	1.11	[0.83; 1.40]	87.6	0.28

Bjorklund, Cassel, Bjorklund, Brown, Park, Ernst, Owen (2000)	1.10	[0.82; 1.36]	86.8	0.25
Brackmann, Otgaar, Sauerland & Howe (2016)	1.28	[0.85; 1.41]	87.7	0.27
Cassel & Bjorklund (1995)	1.13	[0.85; 1.41]	87.7	0.28
Cassel, Roebers & Bjorklund (1996)	1.16	[0.90; 1.43]	86.0	0.24
Cohen & Harrick (1980)	1.12	[0.84; 1.39]	87.7	0.26
Flin Boon & Knox (1992)	1.17	[0.92; 1.43]	85.2	0.22
Goodman & Reed (1986)	1.12	[0.85; 1.40]	87.7	0.26
Hunt & Borgida (2001)	1.09	[0.82; 1.36]	87.3	0.26
Laumann & Elliott (1992)	1.10	[0.82; 1.37]	87.4	0.26
Robinson & Briggs (1997)	1.14	[0.86; 1.40]	87.7	0.26
Roebers & Fernandez (2002)	1.09	[0.82; 1.36]	86.9	0.26
Roebers & McConkey (2003)	1.05	[0.80; 1.30]	85.4	0.21
Roebers & Schneider (2001)	1.14	[0.86; 1.42]	87.3	0.27
Roebers & Schneider (2005)	1.08	[0.81; 1.33]	85.4	0.23
Roebers, Bjorklund, Schneider & Cassell (2002)	1.15	[0.88; 1.42]	87.2	0.25
Templeton & Wilcox (2000)	1.17	[0.91; 1.43]	85.9	0.23

The Omnibus Test

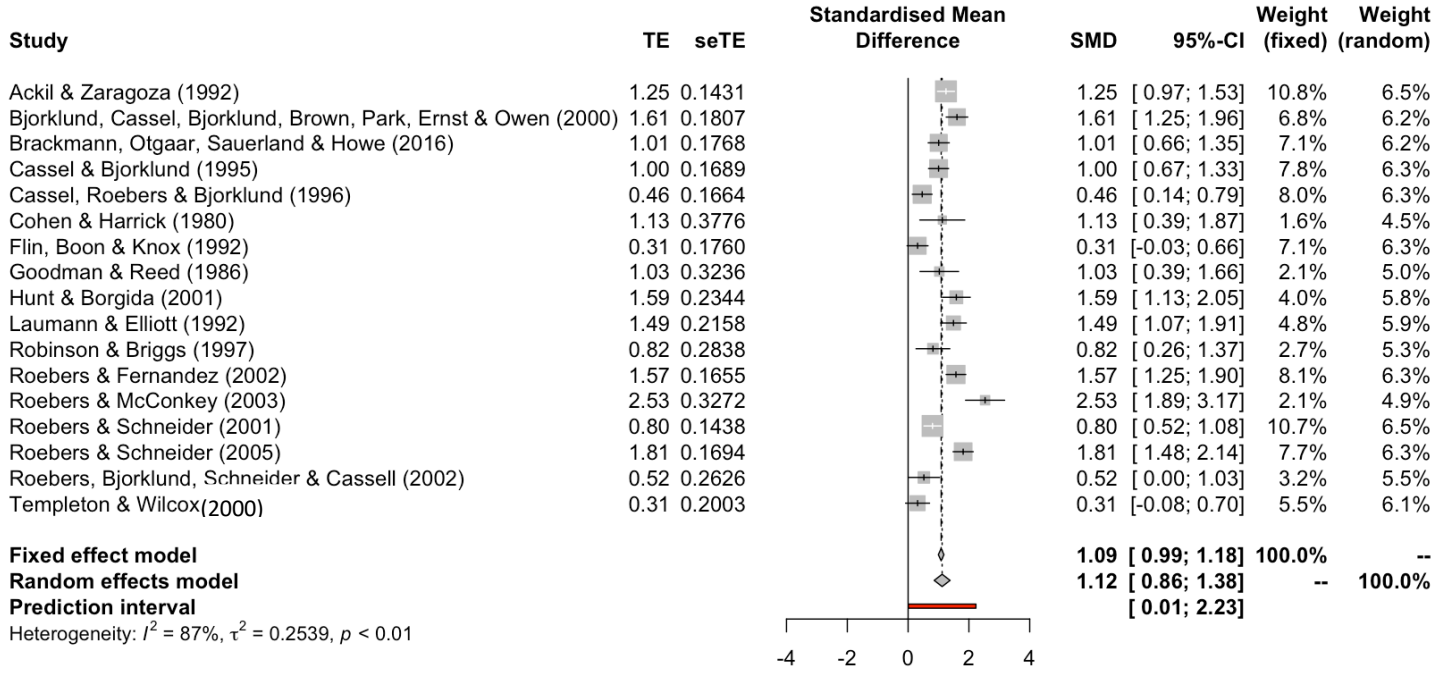
Figure 3 displays the forest plot for the misinformation item analysis. Each study compared children's and adults' incorrect answers to misinformation. Each row in the forest plot depicts one study and displays the SMD for that study. The SMD shows the difference between children and adults on the mean score outcome for misinformation items. The vertical line at 0 depicts a null result. Any outcome to the right of the vertical line (a positive

SMD) indicates that children made more incorrect answers to misinformation items than adults. Any outcome to the left (a negative SMD) indicates that children made fewer incorrect answers to misinformation items than adults. As can be seen, children made more incorrect answers to misinformation items than adults in every study.

The overall SMD across studies is depicted by the diamond at the bottom of Figure 3. The random effects model suggested an overall weighted average SMD of 1.12 between the child and adult groups ($z = 8.39$, $p < .0001$) and a 95% confidence interval of between 0.86 and 1.38. This indicates that there was a large effect, and that children made more incorrect answers to misinformation items than adults. Therefore, although there appears to be other confounding variables that are influencing on the overall result (as shown by the heterogeneity analysis), the outcome of the meta-analysis illustrates that children are more likely to make incorrect responses to misinformation items than adults.

Figure 3

Forest plot showing the misinformation item meta-analysis, where the Standardised Mean Difference (SMD) indicates differences in the proportion of incorrect answers to misinformation items in children and adults.



Note. The SMD shows the difference between children and adults on the mean score outcome for misinformation items, along with 95% CIs around the SMD for each study. The vertical line at 0 depicts a null result. Any outcome to the right of the vertical line (a positive SMD) indicates that children made more incorrect answers to misinformation items than adults. Any outcome to the left (a negative SMD) indicates that children made fewer incorrect answers to misinformation items than adults. The weights for the fixed effects model and the random effects model are also illustrated on the forest plot and represent the contribution of each study to the fixed and random effects models. The 95% prediction interval gives the range in which the point estimate of 95% of future studies will fall, assuming that true effect sizes are normally distributed through the domain.

Control - Misinformation Item Analysis

The question remains as to why children more likely to provide incorrect responses to misinformation items than adults. Is it because the misinformation effect is larger in children than in adults? Comparing incorrect answers to misinformation items directly between children and adults cannot help to answer that question. A larger number of incorrect answers on misinformation items for children compared to adults may reflect age-related differences in memory ability more generally, not necessarily that children are more likely than adults to report misinformation. To address this possibility, a measure of baseline memory performance, wherein memory for control items about which participants have not been misled, is required. Therefore, we conducted a further meta-analysis on the studies that tested participants on control and mislead items, thereby allowing us to calculate the mean difference between control and misled items. We then compared the size of the mean difference (i.e., the size of the misinformation effect) in adults and children. We first analysed the adult and child groups separately to examine the heterogeneity in each group to identify any influential studies or outliers that required omitting, then we compared the size of the misinformation effect in the adult and child groups.

Heterogeneity

Adult group. The heterogeneity in the adult group was above what was considered to be acceptable (I^2 = 85.4%, 95% CI [75.0, 91.5], tau^2 = .42; Q = 61.80; df = 9; $p < .0001$). Heterogeneity was explored using a 'leave-one-out' analysis (Table 6), and no papers were identified as needing removal. Flin, Boon and Knox (1992) had the largest impact, but as the 95%CI still spans over 0, it remained in the analysis.

Table 6

Leave-one-out analysis for the adult group in the control - misinformation item analysis, to demonstrate the impact of removing each study on the standardised mean difference (SMD) effect size and heterogeneity

Omitting	Random Effects Model		Heterogeneity	
	SMD	SMD 95% CI	I^2 (%)	Tau^2
Bjorklund, Cassel, Bjorklund, Brown, Park, Ernst & Owen (2000)	-.16	[-0.67; 0.35]	86.1	.49
Brackmann, Otgaar, Sauerland & Howe (2016)	-.17	[-0.65; 0.32]	86.2	.45
Cassel & Bjorklund (1995)	-.18	[-0.66; 0.31]	85.6	.43
Cohen & Harrick (1980)	-.094	[-0.55; 0.40]	87.0	.45
Flin Boon & Knox (1992)	.072	[-0.32; .46]	80.8	.28
Hunt & Borgida (2001)	-.11	[-0.61; 0.38]	87.1	.46
Roebers & Fernandez (2002)	-.068	[-0.56; .42]	86.5	.47
Roebers & Schneider (2001)	-.13	[-0.59; 0.98]	87.0	.50
Roebers & Schneider (2005)	.0047	[-0.41; 0.42]	79.6	.32
Roebers, Bjorklund, Schneider & Cassel (2002)	-.20	[-0.66; 0.26]	84.7	.40

Child Group.

In the child group, the level of heterogeneity was still higher than considered acceptable (I^2 = 86.7%, 95% CI [77.4, 92.1], tau^2=.0.22, Q=67.57; df=9; $p < .0001$). Therefore, we again evaluated the impact of omitting each paper using a leave-one-out analysis (Table 7). Given that the heterogeneity is considered problematic regardless of if any papers are removed are not, and the SMD did not change substantially when any of the papers were removed, we retained all papers in the analysis.

Table 7

Leave-one-out analysis for the child group in the control - misinformation item analysis, to demonstrate the impact of removing each study on the standardised mean difference (SMD) effect size and heterogeneity

Omitting	Random effects model		Impact on Heterogeneity	
	SMD	SMD 95% CI	I^2 (%)	Tau^2
Bjorklund, Cassel, Bjorklund, Brown, Park, Ernst & Owen 2000	.57	[0.21; 0.93]	87.5	.26
Brackmann, Otgaar, Sauerland & Howe 2016	.54	[0.21; 0.88]	85.2	.22
Cassel & Bjorklund 1995	.49	[0.19; 0.79]	83.4	.17
Cohen & Harrick 1980	.61	[0.27; 0.95]	88.2	.23
Flin, Boon & Knox 1992	.67	[0.54; 1.00]	86.1	.20
Hunt & Borgida 2001	.59	[0.24; 0.93]	88.1	.23
Roebers & Fernandez 2002	.61	[0.25; 0.97]	88.1	.25
Roebers & Schneider 2001	.68	[0.37;0.99]	82.1	.17
Roebers & Schneider 2005	.57	[0.22; 0.93]	87.8	.25
Roebers, Bjorklund, Schneider & Cassel (2002)	.65	[0.29; 1.00]	87.1	.24

The Omnibus Test

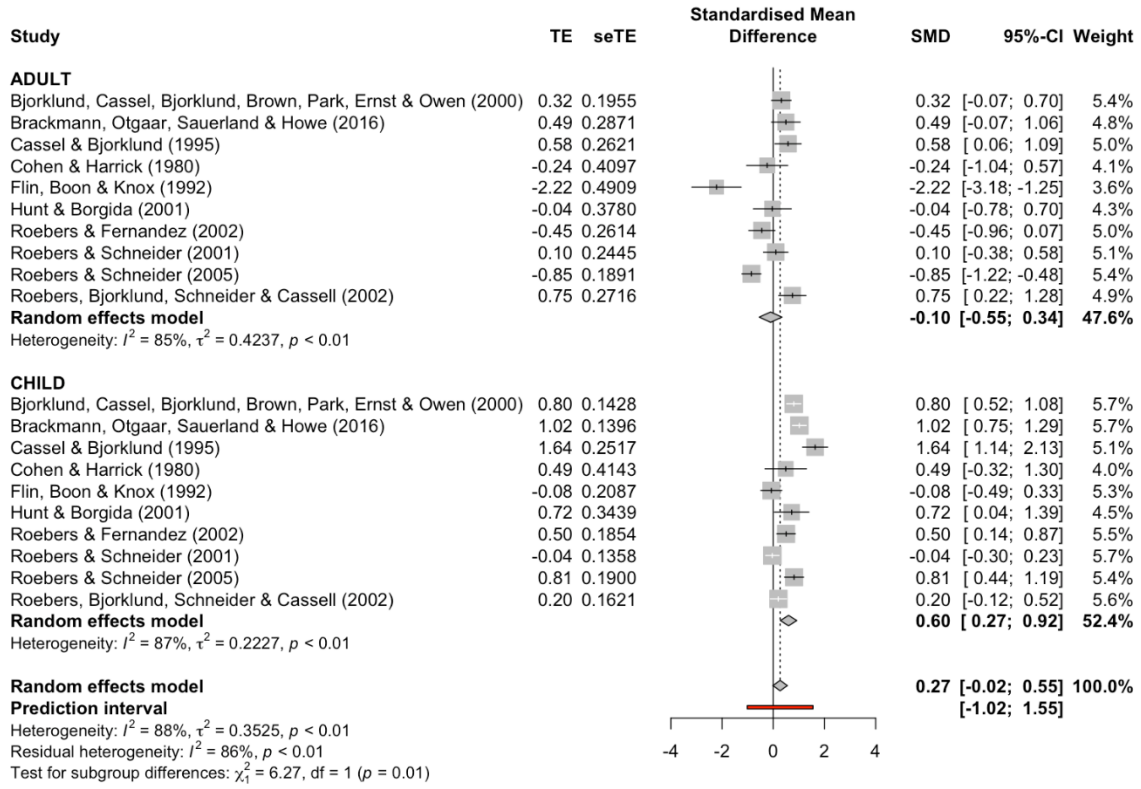
Figure 4 displays the forest plot for the control - misinformation item analysis. Again, each row in the forest plot depicts one study and the SMD for that study. The SMD is the difference between the proportion of correct answers in the control and misinformation groups. The vertical line at 0 depicts a null result, suggesting that participants made a similar proportion of correct answers in the control condition and the misinformation condition. Any outcome to the right of the vertical line (i.e., a positive SMD) indicates that participants gave a higher proportion of correct answers to control items than misinformation items. Any outcome to the left of the vertical line (a negative SMD) indicates that participants gave a higher proportion correct answers to the misinformation items than the control items. The adult results are displayed at the top of Figure 4 and the child results are displayed at the bottom. Across all

studies (across both adult and child groups) heterogeneity was still above what is considered acceptable (I^2 =88.4%; 95% CI [83.6, 91.9]) which suggests that the misinformation effect may be influenced by confounding factors that were not considered in the current analysis. The overall weighted average SMD across all studies is depicted by the large diamond at the bottom of Figure 4. The overall weighted average SMD was 0.27 (95% CI [-0.017; 0.55], $tau^2 = 0.35$; Q = 164.46; df = 19; $p < .0001$). This indicates that combined over adults and children there is a small misinformation effect, and a larger proportion of correct answers were made to control items than misinformation items. However, the 95% CIs of the SMD overlap with zero, indicating that there is a possibility that there is no misinformation effect, and therefore, it cannot be stated with confidence that a positive SMD will be observed in every study.

It is clear from Figure 4 that the size of the misinformation effect was different in adults and children. Indeed, the subgroup analysis confirmed that the misinformation effect was significantly different in the adult and child groups. First, considering adults, the SMD result for each study is variable, with some studies yielding a positive SMD (a higher proportion correct in the control condition) and others yielding a negative SMD (a higher proportion correct in the misinformation condition). Overall, the SMD weighted average for the adults was -0.10 (95%CI [-0.55, 0.34]; Q = 61.80, $tau^2 = 0.42$, $I^2 = 85.4\%$). Because the 95% CI substantially overlaps 0, this suggests that adults were no less accurate on the misinformation items in comparison to the control items, and therefore, no misinformation effect was observed overall. Next, considering children, the misinformation effect is apparent, with most of the studies yielding a positive SMD. Overall, the SMD weighted average for the children was 0.60 (95% CI [0.27; 0.92], Q = 67.57, $tau^2 = 0.22$, $I^2 = 86.7\%$). This demonstrates that children were less accurate on the misinformation items in comparison to the control items; that is, in children, a statistically significant misinformation effect was observed.

Figure 4

Forest Plot showing the control-misinformation item meta-analysis, where the Standardised Mean Difference (SMD) indicates differences in the proportion of correct answers to misinformation items and control items in both adults and child participants.



Discussion

In this paper we used meta-analytic techniques to examine the size of the misinformation effect and determine whether children are more suggestible to misinformation than adults. We conducted a systematic search and collated results from 17 studies testing adults (aged 18+) and children (aged <14) on a misinformation paradigm, including a total of 2,852 participants. In a first meta-analysis, we compared the frequency adults and children provided incorrect answers to misinformation items, and found a large effect that children were more likely to yield to suggestions than adults. In a second analysis (including 10 papers and 1,557 participants), we compared the size of the misinformation effect (i.e., the difference between proportion correct to control and misinformation items) in children and adults. We found a

significant misinformation effect in children, but not in adults. These findings have important implications for understanding age-related changes in suggestibility, and also for refining methods used to study the influence of misinformation on memory.

Considering age-related changes in suggestibility and both of our analyses, the overall conclusion is that children are more likely to report misinformation than adults, and they are more likely to report incorrect answers to questions about misinformation than control details. This effect could be explained through retrospective interference, in which the new information has interfered with the encoding, storage or retrieval of the original memory (Underwood et al., 1960) and is consistent with ongoing research that finds that suggestibility to post-event misleading information decreases with age in children (e.g., Ceci & Bruck, 1993, 1995; Cassel et al., 1996; Poole & White, 1991). A number of theoretical mechanisms that have been proposed to explain why children are more likely to report incorrect answers to misinformation compared to control items than adults including: authority and compliance (Ceci et al., 1987); inhibitory control (Clarke-Stewart et al., 2004); source monitoring (Lindsey et al., 1989); and language acquisition (Clarke-Stewart, 2004). Critically, it is still not clear what theoretical mechanism(s) can account for our findings. The misinformation effect that we observed could reflect genuine memory impairment (e.g., source monitoring deficit), or a response bias effect (e.g., being influenced the experimenter as an authority figure) or a strategic effect (e.g., guessing or process-of-elimination). Differences on the modified recognition test could more clearly reflect memory impairment in children (McCloskey & Zaragoza, 1985). Future research should continue to examine the mechanisms underlying to misinformation effect in children for both theoretical development and to inform appropriate practice recommendations for obtaining accurate accounts from children.

Moreover, it is important to note that under some conditions, children may be less susceptible to report misinformation than adults, such as in situations when the misinformation is associatively related to the original event (e.g., associative activation theory; Howe et al.,

2009) or when children are less skilled at holding in mind multiple representations (Templeton & Wilcox, 2000). Although the associative activation theory cannot explain the pattern of findings we observed, it is possible that this account may be able to explain susceptibility to *internally generated* false memories, which we did not test here. Future meta-analyses could compare children and adults' susceptibility to internally generated false memories, by conducting a meta-analysis on studies that employed the Deese–Roediger–McDermott (DRM; Roediger & Mc Dermott, 1995) paradigm, which is designed to induce internally generated false memories.

Although both of our analyses indicated that children are more suggestible than adults, our work highlights an important methodological point. In reviews examining suggestibility in children and adults, the mean difference between control and misinformation scores are not often directly compared. Often, the number of incorrect responses to misinformation items are compared between children and adults. Results from an analytical approach like that are ambiguous and difficult to interpret theoretically, as more incorrect answers to misinformation items could be due to age-related differences in memory ability more generally, rather than age-related differences in suggestibility per se. Children may make more errors even to non-misled items, reflecting poorer memory ability compared to adults. In our view, reviews examining the misinformation effect should at least calculate the mean difference between control and misinformation (i.e., experimental) conditions, rather than focusing on the experimental condition alone. In doing so, differences in performance across the control and misinformation conditions would reflect differences in the size of the misinformation effect, not other factors. We compared the mean difference between control and misinformation conditions (i.e., the misinformation effect) in adults and children. Note that this analysis approach still does not elucidate the mechanisms underlying the misinformation effect (i.e., a memory impairment vs. response bias), and therefore, in the future, an even better way of measuring and testing the mechanisms that underlie the misinformation effect could be used and developed. Nevertheless,

an analysis that considers answers to both control and misinformation items is superior to focusing on answers to misinformation items alone. Using the control – misinformation analysis, we found evidence of a misinformation effect in children, but not in adults.

The finding that there was no clear and consistent misinformation effect in the adult sample may be surprising. The misinformation effect has been studied in adults since the early 1970's (Loftus & Palmer, 1974), and many studies have concluded that the misinformation effect is prevalent in adults (e.g., Ackil & Zaragoza, 1995; Sharman & Powell, 2012; Wylie et al., 2014), and this research has been influential in guiding procedures in the legal system (e.g., 2012 New Jersey Supreme Court jury instructions). Payne et al. (1994) conducted a meta-analysis of the misinformation effect using a modified recognition test and concluded that the misinformation impaired memory. However, the Williams and colleagues (under review) reanalysis of the Payne et al. (1994) meta-analysis, separately considering child and adult groups, suggested that performance on misled and control items varied in relation to retention interval, with children showing a larger difference than adults at long retention intervals. Together, this suggests that the size of the misinformation effect in adult witnesses deserves further scrutiny, particularly in relation to retention interval.

Our work suggests that, at least in the papers in our meta-analysis and the papers in the meta-analysis by Payne et al. (1994), there is no clear trend to suggest that adults are any less correct if they are provided with misinformation. The outcome of the current analysis could indicate that adults are able to recognise the errors in the proposed misinformation and therefore do not incorporate the misinformation into their accounts when recalling their memories.

It is worth noting that the heterogeneity is high in both of our analyses, suggesting that other methodological differences and factors that have not been considered in the current meta-analyses are influencing the study outcomes. It may be that the outcome of both analyses can be explained by the various confounding factors that were not considered in detail in the current analysis. These are important to consider in future research. First, for example, research has

found that there is a difference in suggestibility of adults depending on what type of misinformation is used. The misinformation effect is more likely when the misinformation manipulation provides additional information compared to contradictory information about the witnessed event (Umanath et al.,2018). As another example, misinformation about peripheral compared to central details is more likely to be accepted (Dalton & Daneman, 2006).

Moreover, the studies included in the current review used a range of delays (i.e., retention intervals; See Table 2). There are three key intervals in misinformation research that need to be considered: (a) between the event and the misinformation, (b) between the misinformation and the memory test, and (c) between the event and the memory test. The effect of different intervals on adult and children's acceptance to misinformation is not currently clear. For example, some studies have found that misinformation acceptance increased with delay between event presentation and questioning in adults (Loftus et al., 1978; Swire et al., 2017) and children (Pipe et al.,2007). The longer intervals between witnessed event and questioning are theorised to weaken the original memory trace, which then leads to larger compliance with suggested information (Reyna & Brainerd, 1995), supporting the trace-strength models of memory (Reyna et al.,2002).

However, the story is likely more complex, as suggestibility to misinformation can also be dependent on when the misinformation is presented between the witnessed event and the questioning about that event. In adult studies, adult recall of event details has been found to be better if the misinformation was presented after a longer interval compared to immediately after the event (Roberts et al.,1999). The misinformation is thought to serve as a reminder for the original information and supports retrieval after a delay (Schmidt & Bjork, 1992). On the other hand, in child studies, some research finds that misinformation is accepted more readily by children if presented after a longer interval (e.g., 2 weeks) compared to if it is presented immediately after the event (e.g., Powell & Roberts, 2002). Trace-altering theories suggest that children may be more willing to accept misinformation if they doubt their own memories after

the natural forgetting that occurs with longer delays (Brainerd & Reyna, 1995). However, other research with children has found that misinformation presented closely after the witnessed event, a while (approximately 20 days) before interview is more likely to be accepted (Roberts et al., 1999). This may suggest that when the misinformation and event are presented in close succession, similar temporal information is encoded, and the memories decay at a similar rate, which reduces both adults and children's ability to correctly attribute misinformed details and true event details to their respective sources (Lindsay, 1990).

Overall, there are limited studies that have addressed manipulation of event-misinformation delay and misinformation-test delay (while controlling for the other delays) in children. One such study found that the misinformation effects were not determined by the delays alone but by a combination of the different delay types, the types of details requested, and the time exposed to the event (Roberts & Powel, 2007). The outcome of this study, combined with the mixed outcomes in other research outlined above, means that there is currently no clear conclusion about the effect of delay or misinformation presentation time on children's misinformation acceptance. Through using the Information Processing Model (Miller, 1956), the varied delay types of misinformation presentation time may be impacting on different processes along the encoding, storage, retrieval of memory: Misinformation provided immediately after event will likely impact the encoding of that memory, misinformation provided later on might be impacting on the retrieval of that information. Because of the limited number of studies that directly compared children and adults, our analysis combined over studies that employed different methodologies (e.g., how misinformation was manipulated, centrality of the details, delays). But as the literature continues to grow, various methodological differences could be considered and assessed using meta-analytical methods to better understand variations in the size and prevalence of the misinformation effect in children and adults.

Practical implications

Our analysis provides further quantitative support for the notion that children are more suggestible to misinformation than adults, and therefore, interviewing children requires specific skills and knowledge by the person undertaking the task. Already, a substantial body of research has developed and recommended evidence-based interviewing techniques for practitioners (Lamb et al.,2008) and led to the development of standard protocols for child interviewing (Lamb et al., 2008). The cognitive interview (Fisher et al.,), the stepwise interview (Yuille et al. 1993) and the National Institute of Child Health and Development (NICHD) interview protocol (Lamb et al., 2008) were all created to permit child-friendly evidence collection and minimise the possibility that memory is contaminated by external influences. Empirically based recommendations for interviewers typically outline that interviewers should ensure that the questions are open-ended and do not lead children into answering or providing detail if it was not already first mentioned by the child themselves (Hershkowitz, 2001). Moreover, policy and practice recommendations that place importance on preventing misinformation in children's accounts do exist. For example, the "Achieving Best Evidence in Criminal Proceedings" document contains a detailed description of good practice in working with witnesses and victims in the CJS and how to appropriately interview them to obtain good quality statements (ABE, 2011). The 'Brief Introductory Guide on Investigative Interviewing' is another example of police guidance provided to police officers in Europe under the framework of the European Code of Police Ethics. Our analysis further highlights the importance of evidence-based practice and policy guidance such as these.

Despite the availability of empirically based recommendations for interviewing child witnesses, a review in 2016 highlighted that many forensic interviewers do not obtain the necessary knowledge and training for employing these appropriate interviewing techniques (Poole, 2016). Furthermore, a large-scale study in the USA found that only half the sample reported to having formal interrogation training and that the majority was informal "on the job" style training (Cleary & Warner, 2016). Similarly, some officers used a confrontational style

and coercive techniques when questioning (Cleary & Warner, 2016). Moreover, research into Scottish court investigations has found that both prosecution and defence lawyers ask closed questions, with defence lawyers asking more suggestive questions, more attempts to accuse children of lying or being coached (Skinner et al.,2018) and repetitive questions (Andrews & Lamb, 2017). Results like this suggest that psychological research in this area may not be reaching practitioners. Our work further highlights that empirically based techniques and non-suggestive interview techniques are of paramount importance. Ultimately, specific and tailored training for police personnel and interviewers is imperative to ensure that children have a fair chance at providing reliable and informative accounts. Reliable memory accounts are essential in enabling high-quality investigations, valid decision-making, and ensuring the integrity of the CJS.

Given that reliable witness memory accounts are essential in ensuring the integrity of the CJS, it may be tempting for some readers to contemplate the results of this meta-analysis and other similar studies and come to dis-trust child testimony. There is a danger that professionals could cease the regular interviewing of relevant witnesses, and legal decision-makers will disregard evidence from children, simply because of their age. Indeed, research shows that younger children are less likely to be believed than older children and adults (e.g., Kassin et al., 2001; Knutsson & Allwood, 2014; Newcombe & Bransgrove, 2011). It is our view that practitioners should continue to interview children to ensure complete evidence collection and that children's voices are heard. Under the right conditions (i.e., methods that preserve and protect memory evidence), children are able to provide accurate accounts of their memories (e.g., Brown & Lamb, 2015; Otgaar et al., 2016). Moreover, even if reports are collected using gold-standard interview techniques, honest witness memory reports from both adults and children may contain a mixture of accurate and inaccurate information (Ball & O'Callaghan, 2001; Brown et al.,2013; Memon, Meissner, & Fraser, 2010). On meta-analyses of free recall tasks such as the self-administered interview, accuracy is 92% on average across studies (Pfeil,

2018), which shows that participants make errors, but are more likely to be accurate than inaccurate. Therefore, instead of disregarding evidence from children, tools should be developed that can help legal decision-makers discriminate between accurate and inaccurate information within an individual witness's testimony, regardless of their age. Metacognition refers to an individual's ability to monitor when their own memories are and are not accurate. Research shows that metacognition in adults (e.g., confidence judgments) is informative about memory accuracy (e.g., Colloff et al., 2017; Roberts & Higham, 2002; Wixted & Wells, 2017). Our lab has begun to examine how metacognitive measures in children (e.g., uncertainty behaviours) could be harnessed in the CJS to determine likely memory accuracy and to promote better legal decision-making (e.g., Winsor et al., in press).

Limitations

One overarching limitation across the primary studies is the variation in how memory is measured following misleading information. Many studies do not investigate if succumbing to misinformation is due to a genuine change in the underlying memory (Windschitl, 1996). Future studies should examine the relative contribution of social compliance, which impacts misinformation acceptance, or changes in response bias (i.e., willingness to endorse misled items), versus changes in the underlying memory trace, to enhance theoretical understanding of misinformation effects. Another limitation in the primary studies analysed is that many did not capture how anxiety or trauma may influence memory accuracy and suggestibility to misinformation in forensic contexts. Trauma may lead to more errors in memory (Vandermass, Hess & Baker-Ward, 1993), and traumatic memories seem to be vulnerable to suggestive questioning (Segovia et al.,2017). It is difficult (if not impossible) to ethically create a traumatic environment, especially for child participants. Therefore, the size of the misinformation effect in the real-world contexts is difficult to currently estimate.

Another limitation is that our meta-analyses included children between the ages of 3 and 14, which covers a substantial timeframe for developmental changes in cognition and

suggestibility. Various studies have found that older children (e.g., aged 12) are less suggestible than younger children (aged 4-5; Cassel & Bjorklund, 1995), and even middle-aged children (aged 8) are less suggestible than younger children (aged 3; Poole & Lindsay, 2002). However, some studies have not identified such developmental changes (e.g., between 7-9-year-olds and 10-12-year-olds; Ceci et al., 1987). Our meta-analyses also included adults over 18 years old, but a meta-analysis compiling 39 independent effect sizes demonstrated that older adults (aged around 70) provided more incorrect answers to misinformation than younger adults (aged around 25; Wylie et al.,2014). Therefore, there may also be changes within adulthood that are not sensitively addressed in the current analysis. Our analyses did not consider detailed age-related changes in childhood or adulthood, as this inclusion criteria would have greatly decreased the number of studies that could be included in the analysis. Moreover, we were interested in an overarching and direct comparison between children and adults, considering also how the misinformation effect should be measured. A more detailed understanding of developmental changes over age should continue to be addressed in future studies.

Conclusion

In conclusion, our meta-analysis suggests that children are more suggestible than adults across the various methodologies used in misinformation paradigm studies. When control information was considered in the analysis, the outcome still indicated that children were likely to report misinformation; however, there was little evidence of a misinformation effect in adults. Moving forward, it is suggested that the size of the difference between control and misinformation items (i.e., the size of the misinformation effect) is considered when comparing the susceptibility of different participant groups (e.g., young and old) to misinformation. By using this methodology, researchers can more precisely specify why differences in error rates occur, and control for other impacting factors that might affect overall memory accuracy or performance, such as general memory improvements with age. Furthermore, research should continue to determine the internal and external factors that influence susceptibility to

misinformation given the high heterogeneity that was identified in our meta-analyses. Research should also continue to examine if the misinformation effect is due to a memory impairment or a change in response bias. Ultimately, future research can further our theoretical understanding of the misinformation effect and help to improve and sustain police interviewing practices. This will help to ensure that children and adults in the CJS are provided with the support, understanding, and interviewer expertise that they deserve to enable them to provide reliable and complete accounts of their experiences.

CHAPTER III

Child Witness Reliability: A qualitative assessment of professional perception

Abstract

In Chapter 2, we found that children are more likely to succumb to the misinformation effect than their adult counterparts. Outcomes such as this have the potential to negatively impact on how child witness memory is perceived by others and a child's credibility in investigative proceedings. There are few qualitative studies investigating the perceptions of professionals in the UK on this subject area. The research in the current chapter explored perceptions of child witness reliability held by social workers, sexual harm support staff such as independent sexual violence advisors, and police officers. The main aim was to identify common patterns in professionals' opinions, but similarities or differences between the groups of professionals were also of interest. Focus groups were used to encourage discussion and debate, and Thematic Analysis was used to detect commonalities and patterns within the content of the discussion. Three main themes emerged, with sub-themes also being identified. The main themes were: an acknowledgement that case specific factors impact on the reliability of a child's account; that the CJS is not appropriate for children; the global perception of 'child memory' is perceived as both reliable and unreliable. The results suggest that there are changes that could be made to the CJS to improve the quality of evidence collected and to ensure the well-being of the child. Future research should examine the potential impact of professional perceptions on child witnesses during CJS proceedings; both on the quality of evidence and on the wellbeing of the child.

Introduction

In the Criminal Justice System (CJS), evidence is regularly collected from witnesses to support in decision-making in legal proceedings. Increasingly, child witnesses are being called upon to provide testimony, especially in cases where children are the victims of crime. For example, in March 2016, around 227,500 offences against children were recorded by police. Around half (49%) of child abuse offences do not proceed further through the CJS because of 'difficulties' with the evidence (ONS, 2019). At times, children are the only witness to the crime in question and therefore the accuracy of their account is paramount in building a case. Children are often accused of having unreliable memories and for some time it was accepted as 'common knowledge' that children's memories are malleable which could lead to false reports (Cashmore & Bussey, 1996; See Chapter 2). Yet, there has been recent research to suggest a child's memory can be reliable; when appropriate interview strategies are employed, children as young as 3 years old can recall accurate information (Lamb et al., 2015). Research with child witnesses has supported the implementation of accommodations for children in court (e.g., Goodman et al., 1999) such a closed-circuit television (Drizin & Colgan, 2004) and improved methods of questioning (Poole & Lamb, 1998). For example, the recommended Cognitive Interview (Memon et al.,2010) and 'achieving best evidence' (ABE; Ministry of Justice, 2011) procedures, both identify that open questioning (e.g., "Tell me about…)", rather than closed questioning (e.g., "Was the car red?") is the best means of achieving accurate accounts.

Although recommended practices can support the accuracy of children's accounts, as can be seen by the ONS data, many cases involving children are not taken forward. Part of the reason for this may be related to the perceptions that professionals have about children. Theories of 'people perception' in social psychology explain that the conclusions that we draw about others are derived from the implicit perceptions that we have about that person

(Brooks & Freeman, 2019). These perceptions are influenced partly by 'social categorisation' in which people are perceived or judged based on their social category (e.g., age, race, gender; Young & Bruce, 2011). Moreover, the 'implicit personality' theorists suggests that certain beliefs or judgements about a person influence subsequent perceptions about that person (Bargh et al.,1996). For example, if one was perceived as being 'happy', they might also be judged as 'generous' or 'kind' as these may be associated with 'happy'. Both of these theories on people perception hold that people often use mental short-cuts (e.g., timesaving strategies in making judgements and decisions), which can lead to stereotyping and prejudice (Nogushi et al.,2014).

The information derived from these perceptions is influenced by the groups that we are in and the environments that we frequent, which leads to the sharing and development of new ideas, values, beliefs and knowledge. Therefore, the memory reconstruction is developed by our social world. This is the basis of the 'social representation theory' (SRT; Moscovici, 1998) which posits those collections of cognitions are representations of social and cultural connections.

SRT states that social representations of something are 'anchored' in all communications where individuals draw upon already-known information to inform new ideas and references. In the current topic, anchoring of previously 'known' or understood information (e.g., children are unreliable) are likely to influence or transform new information, thus increasing the likelihood that the new information might be interpreted in a way that is prejudiced or biased.". Moreover, 'objectification' in SRT (Moscovici, 1998) posits that a person may try to better understand the unknown or abstract by transforming it into a concrete and clear phenomenon. Therefore, it is worth noting that the concept of 'child reliability' or 'child memory' may be transformed into something that can be objectively seen or heard (e.g., a child's emotional state during questioning or the amount of detail they

provide). Thus, the objectification of the social representation may impede the accurate assessment of reliability, which again will lead to a bias perception.

According to the aforementioned theories, professional's perceptions about child reliability could potentially lead to a variety of different judgments that could influence or impede the perceived credibility and competency of the child witness. Moreover, these perceptions can have consequences. The perceptions that people have about children in in the CJS have the potential to determine the outcome of a verdict or how a case proceeds through the CJS. For example, when a child is perceived as being less credible, the defendant is less likely to be deemed guilty (Goodman-Delahunty et al.,2010).

There is ongoing scientific research into the role of professional perceptions about witnesses and how these translate in court (e.g., Melinder et al., 2004; Granhag et al., 2005). Frameworks on *credibility* can support with understanding of how reliability might be perceived in the courtroom. For example, the Dangerous Decisions theory (DDT; Porter & Brinke, 2009) posits that professionals make flawed decisions on the credibility of witnesses using interpersonal judgements about trustworthiness that are associated with perceptions of threat (Adolphs, 2002). Decisions are influenced by one's internal schemas and perception of the world, including their experiences with past witnesses (Porter et al., 2009) leading to a 'tunnel vision' assimilation of ambiguous evidence to support their initial perception (Granhag, 2007). Ultimately, this perception may be unreliable, and then lead to 'dangerous decisions' (Porter et al., 2009). In this case a 'dangerous decision' might be deeming a child to be an unreliable witness or decide to not collect testimony. Therefore, understanding people's perceptions of child witness reliability is important to gain insight into how this might be impacting the progress and success of their involvement in court and how their perception may influence later attitudes, beliefs and stereotypes.

Previous Quantitative Research

There is psychological research published that attempts to gauge people's perceptions of children providing evidence in the CJS (e.g., Cutler & Penrod, 1995; Memon et al., 2003). For example, studies have found that jury-eligible laypeople are prone to believe children can be accurate witnesses (Quas et al., 2005), but that children are more suggestible than adults (Benton et al., 2005). Lay people have been assessed to have limited understanding and knowledge about factors that can impact the accuracy of a child's memory reports, such as interviewing techniques (McAuliff & Kovera, 2007) and repetitive questioning (Quas et al., 2005).

Professional's knowledge and understanding about the reliability of eyewitness memory is assumed to be more accurate than laypeople's (Buck et al., 2014). Mainly, quantitative research (e.g., using surveys and questionnaires) has been conducted and that research suggests there are varying perceptions between professionals that work within the CJS concerning how they view child witnesses. Melinder et al. (2004) used a 56-item questionnaire to measure professional's beliefs about child witnesses, including the credibility of child witnesses and the reliability of evidence from children, the impact of stress on child witnesses, their views on the importance of witness age, their confidence in their own evaluations of children as witnesses, and their own knowledge about the prevalence of child sexual abuse. In total, the authors analysed completed questionnaires from 478 professionals (including Norwegian judges, police detectives, psychiatrists and attorneys) and found that psychiatrists and police officers appeared to have more trust in a child witness's reliability than judges and attorneys (Melinder et al., 2004). Attorneys were most likely to be sceptical of children and their reliability, which was closely followed by the critical opinions from the judges within the Norwegian sample (Melinder et al., 2004). The findings were

supported by a Swedish study, which also found that attorneys and judges were more sceptical of a child witness's account in comparison to police officers (Granhag et al., 2005).

In an American study, however, police officers, as well as judges, were more likely to be sceptical of the reliability of an account from a child witness than mental health practitioners and child protection workers (Everson et al., 1996). The differences in findings across studies, could suggest a potential cultural difference in the belief system of police officers or differences in the training they receive. These differences could also be attributed to the comparison groups across studies, as Everson et al. (1996) compared their police officer and judges against mental health practitioners and child protection workers, whereas Melinder et al. (2004) compared police officers against prosecutors and lawyers. Since the studies used different measurement scales and comparison groups, it is difficult to directly compare overall attitudes across the studies. Nevertheless, the studies appear to demonstrate that perceptions differ with profession, and that judges seem to be relatively sceptical regarding reliability of evidence from children regardless of whether they live in Scandinavia or the United States of America.

Previous Qualitative Research

Although measuring beliefs and perceptions through the use of questionnaires can provide a general understanding regarding opinions about the reliability of evidence from children and the factors influencing reliability, very little is currently known about professional's explanations, experiences and justifications for their perceptions. These important insights provide context to perceptions and a more in-depth understanding for research to adequately capture true experiences. To date (to our knowledge), there have been few studies that have examined professional perceptions, knowledge and understanding using more exploratory qualitative techniques, where professionals are asked to explain their perceptions and opinions in their own words.

In one study, Aarons et al (2004) interviewed police officers who had specific training in interviewing child witnesses and investigative techniques with children. These participants were asked about their experience and perceptions surrounding children with intellectual disabilities. One theme identified in this research was concerned with the cultural 'battles', which identified that the norms and skills for working with children were developed from experience with colleagues, rather than through formal guidelines and training that was offered. Similarly, within the culture of their organisation, the police officers perceived that outside of the child abuse unit, there was a distinct lack of understanding and that others perceived them to be more like social workers and "touchy-feely" (Aarons et al., 2004, p. 272) in comparison to their co-workers. The police officers reported that there were investigative priorities which ultimately means that there was a significant importance placed on the number of closed cases (Aarons et al., 2004). The police officers demonstrated their frustration upon passing the cases to the Criminal Investigation Unit to be continued by detectives for further investigation. Common experiences were that the participants believed cases were then oversimplified and the detectives were more focused on 'facts' with little focus on the welfare of the witness, thus on many occasions the trial did not reach court as the child may not be deemed a reliable witness. This highlighted a disparity between professionals, in relation to their goals and how they wanted to collect evidence. Finally, the authors identified within the interviews that training and skill development was lacking with regard to opportunities to improve on their interviewing skills (Aarons et al., 2004).

More recently, Cassidy et al (2020) interviewed police officers in the UK to understand how they conceptualise the credibility of a child witness. Officers considered interviewing techniques to be important in eliciting the information that they needed from child witnesses, though they simultaneously held individual beliefs about what cues from a child might suggest deceit (i.e., decreased credibility) that were not in the official training. The police

officers appreciated that poor interviewing decreased the accuracy of the information in cross-examinations and that a lack of reliability impacted on the evaluation of credibility (Cassidy et al., 2020). They also discussed the importance of empowering the witness through individually tailored interviewing, but ultimately, there was an importance placed upon the 'end product' (i.e., the collected evidence) and how this was presented to the Crown Prosecution Service (CPS). It was identified that there was confusion around what the CPS considered to be credible and the lack of communication with the CPS about the reliability of the account. This demonstrates some interagency differences which provides some justification to interagency exploration of logistics, practicalities and perception of child reliability and credibility.

Current Research

The current research design is motivated by the paucity of qualitative studies and limited information on how professionals within the United Kingdom perceive child witnesses, along with their justifications of their opinions. As in other previous research (e.g., Melinder et al., 2004), there will be comparison between groups of different professionals to gauge if any differences exist between them, thus indicating if there are perhaps any differences in training or in the organisational culture.

Aims of the study

The aim of the present study is to use a qualitative approach to aid in understanding how UK professionals understand and assess the reliability of child witness memory and their justifications, opinions and experiences in working with child witnesses during criminal investigations. This will be achieved by encouraging a reflective conversation in focus groups about the reliability of child witness memory with professionals that have experience working with children, namely with social workers, sexual harm support staff and police officers. In a focus group, a group of individuals are encouraged to discuss a specific topic to

gather their opinions, perceptions and attitudes (Cornwall & Jewkes, 1995). It is envisaged that the insights gained will provide us with a better understanding of how professionals work with child memory reliability. This could support understanding into why many cases with child witnesses are not convicted and to inform policy or practice changes that could improve how professionals work with children in the CJS and other situations on which such evidence is relied upon to make important decisions, such as social care and safeguarding. This may also highlight possible differences that exist between professional agencies, which would encourage future research to consider this.

Method

Ethical Protocol

Participants were approached to take part in the research, and all received a participant information sheet before they could give their informed consent to join the focus groups. On the information sheet, participants were made aware that the goal of the focus group was so that they could discuss their experiences working with children with other professionals. Participants were made aware that confidentiality could not be achieved between the researcher and participants, as the focus groups were being held in person or over video conferencing software. The focus groups were recorded and were transcribed within 24 hours with the voice data being destroyed after transcription. Participants were able to withdraw at any time throughout the focus group and withdraw their data up to 10 weeks after the focus group.

Topics including children may sometimes be sensitive and difficult to discuss, as child trauma experiences and criminal activity could be involved in conversations. As a result, participants were provided with the phone numbers of well-known crisis helplines, as well as the contact information of the researchers involved should they raise a need. All participants were also informed that it would be imperative for the researchers to highlight to the relevant

authorities (e.g., relevant police personnel) or initiate safeguarding procedures (e.g., raise a safeguarding concern with social care) if there was any evidence of abuse or misconduct that arose within the discussions in the focus group. These ethical considerations were communicated clearly with the professionals via the participant information sheet. Ethics was granted by the Science, Technology, Engineering and Mathematics Ethical Review Committee at the University of Birmingham Ethics Committee in February 2019.

Participants

Purposive sampling was used to recruit participants. Participants were required to be currently employed or have experience employed in a field that worked directly with children going through criminal justice proceedings (e.g., in court proceedings) or other similar proceedings in other settings (e.g., safeguarding procedures). The participants were approached via email, utilising contacts and professional connections already established and also recruitment during an academic conference. The number of years of professional experience was not stipulated.

It was planned for between 5-8 participants to be recruited, in line with literature recommendations for non-commercial topics (e.g., Krueger & Casey, 2014). Following the recruitment process, each group of professionals yielded 4 respondents who could join, and therefore smaller focus groups were conducted. In total, there were eight female participants, and four male participants.

Group 1 - Sexual harm support staff

The sexual harm support staff (SHS) group contained two independent sexual violence advisors' (ISVA) and two forensic examiners. ISVA's receive specialist training and provide advice and practical support for, and address safety needs of, victims of sexual violence and work in partnership with other agencies in the CJS. ISVAs undertake risk assessments and implement various support packages and generally support clients through the CJS process.

Forensic examiners work in the collection of evidence during the investigations (e.g., collection of DNA evidence from victims of abuse). Participants were recruited and interviewed during their attendance at a workshop on memory in the CJS and each of the group members had no previous contact with each other prior to the focus group being conducted.

Group 2 – Police officers

The police officer (PO) group contained four officers with varying degrees of experience working with children. All of the officers were constables; one of whom was specially trained in investigative interviewing of children, two had extensive experience in working with and interviewing children and one worked in the domestic abuse unit. In general, the officers are tasked with maintaining the law and order in their force area. They are involved in the arrest process, investigations of both offenders and victims, and also the court process should this be necessary. Participants were recruited from their place of work, so were colleagues within the same police precinct and therefore were known to each other prior to the focus group being conducted.

Group 3 – Social Workers

The social worker (SW) group contained four individuals, three of whom were qualified and practicing social workers and one of whom had social work training and was currently working as a youth offending officer, which was identified as a similar job role. In general, social workers support clients during crisis to provide resources to aid in problem solving. Social workers specialise in either adult or child social care and tend to work with local authorities or voluntary organisations. For the current group, the participants worked directly with children in a variety of different contexts, including young victims and perpetrators of crime. The group were recruited through their places of work, and two of the group members knew each other prior to the focus group being conducted, as they shared

some clients. The social workers focus group was conducted via an online conference/meeting platform to adhere to nationwide lockdown procedures that were implemented at the time of data collection for this focus group.

Data Collection

The data were collected using a semi-structured discussion in focus groups which intended to openly investigate professionals' opinions about the reliability of memory evidence from children, and their experience of working with child memory evidence. A focus group is effective for collecting a wide range of views in a short space of time, and within the current research it was used to establish if there were any disagreements within the same group or 'profession' by encouraging a discussion between the group members.

The research was conducted through a constructivist lens (Hoffman, 1990), where reality is a construct of the human mind and opinions and perceptions are built from one's experiences, relating to their already possessed knowledge. This position posits that perceptions are subjective and constructed through their reflections on the world, thus is useful for the current research design which is specifically interested in those subjective and unique perceptions built from professional experiences. Focus groups were recorded using a pre-approved voice recorder and the researcher made written interview notes throughout to support them in recording points of interest and to note down any seemingly important non-verbal behaviours of the participants.

Interview Protocol

The focus groups were moderated by a facilitator (the researcher) and conducted either at a pre-arranged conference room (sexual harm support staff and police officer group) or using an online video-conferencing software (social worker group) and sessions were around an hour in length. The process began by reminding the participants that they were

being asked to discuss their experiences working with child memory evidence in the CJS or in similar area (e.g., social care and safeguarding).

The semi-structured interview protocol focused on key areas such as (1) how professionals assess the reliability of memory evidence from children, (2) discussion of policies or practices they may have in their place of work for working with children, (3) discussion of training in relation to working with child memory evidence, (4) their opinion on the reliability of reports from children and how age influences the reliability of memory.

All questions were created to be open-ended as to facilitate discussion with potential follow-up questions being utilised to clarify something, if needed. The facilitator was able to prompt a participant to expand on a point and also ensured that the discussion remained on topic. The questions were developed to be inquisitive of the professional's own experiences and so that participants were not influenced by the facilitator's knowledge on the research into this topic.

Data Analysis

In order to analyse this dataset, thematic analysis was used in accordance with the method outlined by Braun and Clarke (2006). Thematic analysis is used to identify patterns within conversation, and a 'theme' is a captured aspect of the data that emerges on different occasions. The process of thematic analysis was conducted in five stages. First, familiarisation with the data set was required and initial notes and thoughts were recorded. This was achieved through reading the data, following with transcription of the data into NVIVO software. Next, the initial codes and ideas were generated from the transcripts and, following this, the codes were organised into groups and combined where appropriate using the NVIVO software. Finally, the organised groups were arranged to identify themes, which are broad and encompassing of many codes.

A reflexive thematic analysis was felt to be most appropriate for analysing the data. This approach aims to encourage the subjectivity and the role of the researcher during interview and analysis, with the developed themes being the final outcome of the research (Braun & Clarke, 2020). The role of reflection within this approach is paramount (see Appendix B for reflexive account) which supports understanding the subjective role of the researcher who is also a staff member within the CJS.

The themes were identified through an indicative lens which means that the data determined the development of the themes and that themes are not developed based on a pre-existing theory (Patton, 1990). Sub-themes were identified where groups of codes were associated with another group of codes, but where the two groups were also varied enough to not be considered as a single group. Potential relationships between themes were considered based on their similarities and their influence on one another, prior to the final review and refinement.

Results

The results identify three broad themes that encapsulate what the participants believed to be most important with regards to child reliability. These three themes emerged from all groups of professionals, but any discrepancies between groups will be discussed when they occur. Following this, a thematic map will be discussed which highlights the links between the sub-themes. See Appendix C for details on the codes, examples from the transcription and an outline of the global themes.

Theme 1: Case specific factors impact on the reliability of a child's account

One consideration that each of the professionals either mentioned, or agreed with, was the idea that there were unique, case-specific factors that could impact on the reliability of a child's memory evidence. This concept encapsulates the external influences from other

people, and other contextual factors that impact on the reliability of a child's account. Common factors that were discussed included the influence of the family and caregivers, and also the impact of trauma and anxiety that was related to the crime or experience in question.

Subtheme: The influence of the family and caregiver on the child's account

All three groups of participants highlighted that the child's family or a caregiver is likely to influence the reliability of a child's account. As expected, all cases involving children, also involves family members, which was identified as a specific barrier to collecting reliable evidence from the children. The professionals were of the opinion that input from family members or caregivers could decrease the accuracy of the information that the child reports. In many cases, it was explained that children were led to change their answers based on the parents/family members motives or needs. In relation to crime, this was usually to prevent a child from reporting a crime that they (the parent) commited. Reliability of the account then was expected to be influenced, as there was purposeful interjection from parents to change the child's account or memory: "*I had it particularly with asylum seekers in that their memory was influenced by agents and family members back home*" (SW1).

> I spoke to her on her own without her mum because I know what her mum is like and she can lead her into things. So, like you said, you can tell when someone is leading them into saying something (PO3)

This theme was more prevalent in the police officer sample in comparison to the social worker sample. The sexual harm support staff group approached this topic with more concern regarding the emotional impact that the caregiver had on the child, rather than their impact on the reliability of the witness report. Much of their job role appears to revolve around both the

child welfare, and the parent welfare, so their approach to this topic was very different to the other groups: "*Because at the end of the day, perhaps the child is still going to look their adult or guardian to help them to get them through it all. Pretty much isn't it*" (SHS1). The ongoing discussion was characterised by concern about the child's well-being and the influence that their parents might have on this, rather than the impact they might have on evidence:

> Very much to be about kind of managing the parent's emotions and you know [inaudible] … and you know reassuring them that we are at this point now and just let the legal system take its time. It does not always have the outcome that they are expecting (SHS2)

Subtheme: The impact of trauma and anxiety

There was a clear pattern among all participants that they believed trauma and anxiety to have a significant impact on a child's wellbeing, as well as on the quality of their memory and evidence. As a witness/victim to a crime, it is assumed that a child has experienced considerable trauma. Most professionals approached this topic with compassion, with specific concern on the well-being of the child and empathy regarding this emotional time. Also, it was identified that this trauma may impact the quality and reliability of the memory evidence. This topic led to some apparent frustration in some cases, as some participants were of the opinion that trauma and well-being was not approached appropriately by other professionals, due to their apparent focus on evidence:

> …it's one of the most frustrating aspects of my role. Seeing how crucial it is... the necessity for their memory to be so good when they are going through the most traumatic experience of their life. I think it's really unfair and I think it's really

problematic. I think it has a massive impact on investigations and massive impact on the child's sort of... general wellbeing. (SW3)

One police officer however, was of a conflicting opinion to this, identifying that trauma might have a consolidation effect on memory, and improve the likelihood that a child will remember details of an event: "*But at the time I think the level of trauma makes your brain remember. No matter how old you are*" (PO1). This opinion was not outwardly shared by the other police officers in the group.

Theme 2: The global perception of 'child memory' being perceived as both reliable and unreliable.

Although Theme 1 was focused on the contextual factors and the external influences on a child's account, this theme relates to the global perception that professionals have about the psychology of children and the concept of a child being a witness. This theme encapsulates the perceived internal factors (e.g., cognition), and has less to do with the context in which the child is in. This theme is developed by a mixture of both compassionately motivated child-believing perceptions (e.g., all children are believable), and skill-focused apprehension in relation to cognitive skills and memory.

Subtheme: Perception of reliability relates to perception of believability

There was a tendency for participants to associate the idea of 'reliability' with 'believability' and 'honesty'. That is, all three concepts were regularly used interchangeably to mean the same thing, without appreciation that these concepts can be different. For example, a child might present as believable, but their memory might not be reliable. The perception of reliability seemed to be directly related to the participants understanding of

'lying' and the belief that children are unable to lie: "*There is that kind of... that innocence of youth. You know when -SW1- was saying, children tend not to make things up and that is kind of generally truthful isn't it*" (SW3). This ultimately led to a more positive perception of the child memory credibility, and therefore their perceived reliability: "*And … children tend not to make things up. Especially at a young age. It is around them being able to repeat that*" (SW3).

There were ongoing contradictions to this line of thought however, especially in the police officer group. Some participants agreed with the idea that children are unable to lie: "*And I think that she would have found it hard to lie because children I think do find it hard to lie*" (PO1). However, the same officers proceeded to share experiences of when children they had worked with were assessed as lying, and as being unbelievable and not credible.

> So, if they are going to come into your interview and tell a pack of lies it can throw the whole case out…
> ... But it just sounded like a pack of lies, it didn't sound believable at all. (PO2)

The above trajectory of thought was supported by the perception that some children are antisocial, and this seems to decrease the perceived credibility, and thus reliability, of the child's account: "*But, she didn't want to give a statement anyway cause this girl is a real shit bag*" (PO4). Again, this notion was only identified in the police officer group and was not outwardly shared in the social worker or sexual harm support staff groups. The following quote illustrate these perceptions of antisociality being shared in the police officer group: "*it gets to the point where we'll kind of go in and speak to them and they'll tell us to fuck off and we'll go back and they say the same thing*" (PO1).

Conversely, the sexual harm support staff and social workers were of the opinion that *all* young people that approached them were telling the truth and will "*just believe every single person that is coming through the door"* (SHS1). Some participants in these groups were of the opinion that children were believable, but not always believed by others working in the CJS: "*Because if we don't believe kids then ... I don't know who will..."* (SW3).

For the police officers, the believability of the child was also related to their ability to focus and answer questions during investigations. The lack of focus in the interview decreased how professionals perceived their ability to be a witness: "*You just can't keep them on track. They start talking about trampolining. You can't keep them focused on one point and you are constantly bringing them back to the track because they're going off the track*" (PO2).

Subtheme: Children have limited understanding and knowledge of their experiences

The knowledge that a child holds about the world, and their understanding of the context and situation that they are experiencing (e.g., context of the crime) was discussed regularly amongst professionals. It was assumed that their limited understanding of their experiences may decrease the reliability of their account. First, it was identified that their limited knowledge of the CJS and the practical aspects of this would be detrimental to the quality of the evidence they provided. Specifically, participants considered that children are sometimes unaware of the seriousness of the process and the consequences of their statements.

> And I don't think he knew the implications of what he was saying. Didn't have a clue. He didn't know the consequences of saying something like that. He could have put his aunty in prison (PO2).

Similarly, it was identified that children may struggle in understanding the crime they have witnessed, or experienced (e.g., sexual assault) in comparison to adults who may have more knowledge or understanding about these events. Therefore, professionals believed that children may be less likely to correctly recount their experiences and may have less information to give than perhaps an adult counterpart.

This developed to suggest that children are therefore less inclined to report details due to their lack of understanding with regard to the information that would be important to disclose: "*So... you know I suppose then giving information when they were not aware that it was wrong or shouldn't happen ...”* (SHS4).

The idea that children often lacked contextual understanding about events was illustrated by the participant's discussions surrounding language development. Specifically, this related to the language used when questioning, especially with sexual assault cases, as children may have their own unique words for genitalia or sexual acts: "*this word is something that I will continue to use with the child instead of going into technical phrases. I think that language is very important*" (SHS2). Participants were of the opinion that the language and child's limited experience or ability to describe the events could hinder reliable evidence from being collected: "*A three-year-old? Their speech hasn't even developed yet”* (PO3).

Subtheme: The child's memory is vulnerable to change

Professionals were aware that child memory is vulnerable to change. They were of the opinion that age was a determinant of the vulnerability of memory, with younger children showing a higher likelihood that their memories could deteriorate or alter over time. The police officers in this cohort had further training in child interviewing techniques, so

appeared to have more technical and practical information to support this discussion, whereas the SW and SHS group had fewer experiences in this regard. For example, police officers discussed the idea that children are suggestible to external information: "*...they have seen it's something that they've heard. So, it's not an accurate from their memory, its more accurate from what they've been told*" (PO3). Similarly, that they may succumb to pressures of social desirability, "*Any child, you know, you want to please people and you want... and children just want to say the right thing*" (PO3).

Overall, all professionals appeared to appreciate that child memory might be subject to changes, that would then impact on the reliability of the account. Some of the professionals (across all groups) placed ownership on the 'system' (i.e., the CJS) to accommodate for these difficulties, rather than to 'blame' or place ownership on the child for having them. However, some professionals (two of which were police officers, and two of which were social workers) were more likely to identify how child memories might be a hindrance to the collection of good evidence, because the memories are vulnerable to change. As a result, this perception suggests that some may not believe children to have memories reliable enough to support evidence collection for a criminal case.

Theme 3: The Criminal Justice System is not appropriate for children

There was a clear and consistent perception from all participants that the CJS was not appropriate to gather reliable accounts from children. There was an agreement that the CJS is only appropriate for adults, and that children were attempting to 'fit' in the adult's system. This was further evidenced by the discussions regarding the limited resources and funding available in these services, "*some services are very good, but we are all so stretched and there is no money anywhere and all of these services are getting tighter and tighter*" (SHS2).

Participants specified that the CJS was inappropriate for children because they deemed the stress and the anxiety of being involved in the criminal justice proceedings as having a detrimental effect on both the quality of the child's evidence and their well-being: "… *I think it puts a massive pressure on the children when they are going through a really difficult process already…*(SW4)". There was a sympathetic stance taken from all groups and many mentioned how the system may be re-traumatising the children, "*someone is so young... I think... is it fair on a child that young? To relive something that is obviously traumatic…"* (PO3).

Specifically, many of the social workers and the sexual harm support staff were concerned with the lack of support and resources for a child's well-being when they might need it during the investigation or trial. "…. *You've opened up this Pandora's box, and they're left to deal with it. It doesn't sit well with me* (SW2)".

Subtheme: Means of gathering evidence and questioning is not appropriate

The matter of interview techniques for children highlighted a divide between professionals. Social workers appeared to have a pessimistic opinion of questioning children and were of the opinion that current questioning was not conducted fairly. There was a consensus that interviewing expected too much of children and that some professionals, mainly barristers, were using questioning techniques to generate an answer that was preferable for their defence case rather than the child: "*and then they might ask the same question more than one time because they don't get the answer that they were expecting*" (SHS3). Frustration was evident during discussions on this topic, which was centred around interviewing being unjust due to tactical questioning in court, "*and it's almost as if it's… trying to catch kids out sometimes*" (SW2).

The interagency discrepancies seemed to be mirrored during discussions with police, through admittance that the child is expected to say the 'right' thing (i.e., what the police need for evidence). It is unclear if this was intended to suggest that they had expectations from the child, though the following quote from a police officer suggests that police interviewers may attend interviews with pre-conceived ideas of the information they need: "… *Very easily undermine the case. If the child says the wrong thing*" (PO2).

One of the social workers was particularly vocal about their opinion that both the system and some of the professionals within it treated children with little sensitivity. They believed that professionals were likely to 'forget' that the child was a child and stated that evidence collection was their only concern:

> …the police have been so focused on gathering the evidence and getting specific answers out of the children that they've forgotten that the child is going through a really traumatic time. (SW4)

Interestingly, one of the police officers was also concerned with the unfair treatment of children during investigations. However, the majority of the police discussions on this topic were focused on the practicalities of getting good, reliable evidence: "*All about the open questions. So, questions starting with what, when, why, who and using all the 'tell me', 'explain to me', 'describe to me'* (PO2)", "*If you're asking a child, you don't want to lead them into saying things they're thinking, oh, is this what they want me to say at this point is this what they want to hear*" (PO3).

One clear difference between groups was their understanding and perception of the importance of evidence. Each group discussed evidence, with police officers demonstrating the greatest concern in this regard. Some participants were conflicted as they understood that

evidence collection is very important and therefore needs to be collected, "*everything, every visit, every discussion, everything is recorded, documented. It can all potentially be put into evidence*" (SW3). Participants were simultaneously frustrated with the process needed to achieve evidence collection and also the impact it had on the wellbeing of the child.

Again, within this theme, social workers and sexual harm support staff explained similar opinions and believed that children were very easily undermined by the 'system' and evidence collection and as a result may be deemed less credible or less reliable than they are.

> when I was in discussions with the police and the solicitors and everybody that was involved with that process, they were all saying, 'well she hasn't been clear' or 'she hasn't been able to explain' or she hasn't been able to do 'this that or the other'. And to me, I just thought that all of that was irrelevant because we ... she … she had said what had happened, but she wasn't able to repeat it in the same way or repeat it in the way that they needed her to. That completely skewed the investigation and I think that it was a massive miscarriage of justice. (SW3)

Finally, this theme also covers concerns regarding training in child memory. All groups of professionals believed that they require more training in child memory and the reliability of child memory evidence.

> … they try and cram so much training in and it's difficult to teach things like this that we need to know, and we need to be experts in um… because they don't have enough time to do it. (PO1)

Subtheme: Inter-agency working can be problematic

The impact of interagency collaboration on a child witness was a concern for every one of the professionals. For example, "*some social workers work really well with us and some don't*" (SHS4). All groups demonstrated on-going frustrations and concerns with lawyers and barristers and how they make prosecution decisions and their behaviour during court proceedings. Also, some of the social workers demonstrated irritation with the police and how they worked with children: "*I think the police often work with children will come into it from an adult's perspective rather than we all go in with a child perspective right from the go*" (SW3). Although there was frustration targeted towards other groups of professionals, there was no discussion or understanding of how these differences and interagency working could be improved, or how they might impact child memory reliability.

Thematic Map

To better understand and visualise how these themes interact, and how they influence each other, a thematic map was created. A thematic map does not intend to suggest a model, but instead is used as a method of capturing the shape of a thematic analysis to support with exploration (Braun & Clarke, 2006). The thematic map (Figure 5) demonstrates how the themes and subthemes relate to one another. Firstly, the global perception of child memory (Theme 2) and the case specific influences on memory (Theme 1) are thought to be related (as indicated by the horizontal line connecting these themes on Figure 5). Ultimately, the perception of the concept 'child memory' will influence the perceived vulnerability to case specific factors. Both concepts were discussed interchangeably between the participants, and both themes likely influence each other. For example, should someone have a perception that 'child memory' is vulnerable, then the case specific factors (e.g., influence of parents) will be

a greater concern in comparison to someone who believes child memory is not vulnerable, and therefore less likely to be influenced by case specific factors.

The majority of relationships appear within the subthemes (See Figure 5). Firstly, the impact of trauma and anxiety is thought to relate with the means of gathering evidence as the participants were of the opinion that the evidence collection in the CJS had a re-traumatising effect, thus further influencing the child's experiences of trauma and anxiety and then further influencing the evidence. Furthermore, the professional's approach to this appeared to differ, with most of the interagency frustration relating to the perception that other professionals were not conscious or compassionate to the well-being of the child. Moreover, the relationship between Theme 3 subthemes identifies the discrepancy between the professional roles and their approach to gathering evidence.

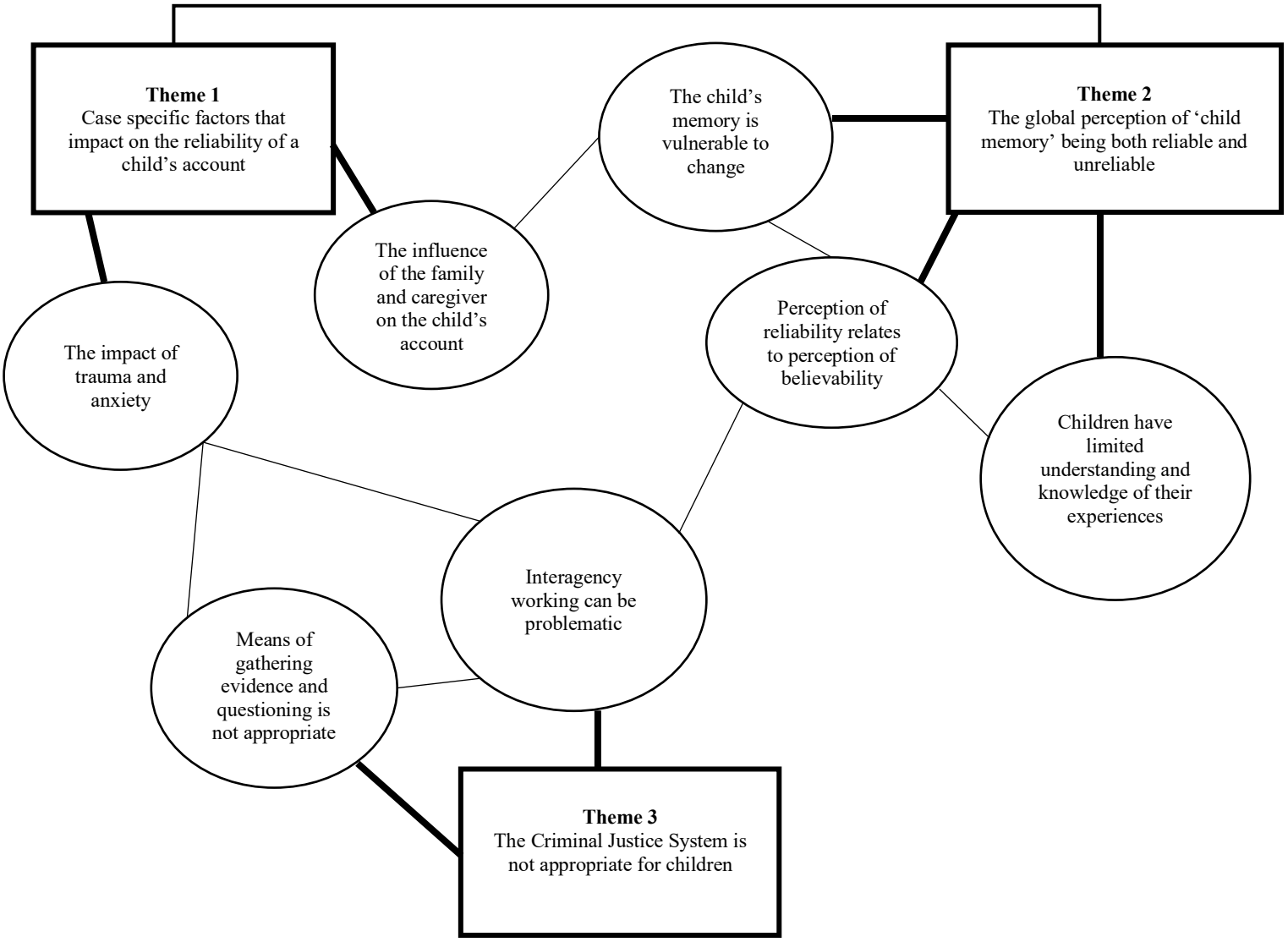
The professional roles and responsibilities subtheme also then relates to the perception of believability of the child subtheme, as there were differences between professionals in their perception and analysis of child believability. In summary, the police personnel were more cautious in their assessment of believability due to experiences of antisociality and deception among the children that they had worked with. Relationships between the sub-themes in Theme 2 exist as each of the factors influence and relate to one another. For example, perception that the child's memory is vulnerable to change, and perception that they have limited understanding or knowledge, then influences the perception of believability.

The final between-theme relationship is between the subtheme 'the influence of the family and caregiver on the child account' and 'the child's memory is vulnerable to change'. The professionals were of the opinion that the parents/caregivers would either inadvertently change the witness account through misinformation or purposely change it through coaching.

Ultimately, professionals believed this was a problem because the child witness memory is vulnerable to change.

Figure 5

Thematic map, showing three main themes, subthemes and the relationship between them.



Discussion

The aim of the present study was to use a qualitative approach to aid in understanding how UK professionals understand and assess the reliability of child memory and their justifications, opinions and experiences in working with child witnesses during criminal investigations. Three focus groups were arranged to facilitate discussion, one with social workers, one with police officers, and one with sexual harm support staff. These groups were selected to explore a range of opinions over a spectrum of people who work with child witnesses in various ways in the CJS.

Thematic Analysis revealed that three common themes emerged. First, *case specific factors that impact on the reliability of a child's account,* encompassed life experiences and specific influences that can decrease the reliability of a child's account. The influence of family members as well as the impact of trauma and anxiety were identified as sub-themes. Another theme that emerged was the *global perception of 'child memory' being both reliable and unreliable* and the opinion that the professionals held generally about children. The 'believability' and 'truthfulness' of children in the CJS was a common discussion point within this theme as there was mixed opinions on whether a child was 'believable' and whether or not children were truthful or able to lie. There was a perception that children have less understanding and knowledge of the world and their experiences than adults, which may mean they are less likely to report things that may be understood as being wrong or illegal by adults (e.g., not understanding that sexual assault is wrong). Finally, within this theme was the perception that child memory is more vulnerable to change than adult memory. Specifically, it was perceived that children compared to adults are more suggestible, have a lesser ability to focus, and are influenced more by social desirability and

therefore their memory is more malleable than adults. Lastly, the discussions were also focused on the idea that *the Criminal Justice System is not appropriate for children* and not set-up in a way to appropriately collect reliable memory evidence from children. This opinion was evident within each of the focus groups, with some participants demonstrating frustration and irritation when discussing this topic. A prominent sub-theme here included the interviewing and evidence collection being inappropriate and traumatising for children, with many professionals showing concern for the child's well-being. Secondly, it was evident that there were problems in inter-agency working, with conflicts between professional roles both in terms of opinions and practice. See below for a discussion of themes in relation to the broader research and theory.

Case specific factors that impact on the reliability of a child's account

Professionals were understanding that there were case specific factors that are different for every child that will impact on that child's memory, and thus reliability. Specifically, in relation to trauma and memory, participants identified that trauma could be detrimental to the reliability of the collected evidence and could be exacerbated due the criminal justice process 're-traumatising' the child. One of the police officers in the current research identified trauma as perhaps having a more consolidating and beneficial impact on recalling an event. This is in keeping with some historical opinions about trauma, where it was previously thought that overstimulation during the event was related to better consolidation of the memory (Pitman, 1989). However, recent research has found that trauma memories are subject to alteration and can be inaccurate just as other nontraumatic memories can (Strange & Takarangi, 2015).

Moreover, specific trauma memories are associated with psychological problems such as post-traumatic stress disorder (PTSD; Engelhard et al.,2008). It has been

identified that poor treatment and abuse in the home can lead to poor memory encoding of specific traumatic events (Gordon et al.,2001) and that this leads to less details being retrieved or recalled during a free recall interview (Eisen et al.,2007) which supports the view that there are individual differences in relation to how memory is encoded, stored and retrieved (Miller, 1956). The current cohort were aware or accepting of these differences in individual cases and children.

Another common discussion point was that changes to memory accounts were related to the caregivers and the family unit. All groups were of the opinion that parents could change children's accounts and influence children in court. Warren and Peterson (2014) identified that parents were more likely to communicate with children using closed questions and therefore had the potential to lead them and change the memory unintentionally. The current group, however, tended to voice that parents often intentionally changed the account. Cassidy et al. (2020) identified that their police groups shared a similar understanding, though this was framed as 'loyalty' of the child to the parents. The groups in the current study, however, framed this issue as if the child was a victim of the intentions of a deceitful parent.

The perception of children

A regular discussion in the focus groups was concerning professionals' perceptions of the credibility and 'believability' of children. Professionals were mixed in their opinions, with the majority of both the social workers and sexual harm support staff stating that they believe any victim who shares their story. However, the police approached the topic with more hesitance and a more analytical means of assessing truthfulness. The police illustrated their hesitance by providing some examples and

experiences when a child did not present as truthful when being questioned and explained how this negatively impacted on the evidence that they were able to collect.

All groups discussed the idea that children are 'unable' to lie. However, there is limited research to suggest any links between understanding the concept of lying, and actually telling a lie (Talwar & Lee, 2002). Therefore, the typical 'truth and lies' competency questions (see Chapter 1) may be a potential unneeded barrier to including reliable witnesses in court as they may be unable to distinguish between a truth and a lie (Talwar & Lee, 2002). Overall, all professionals agreed that children were less able to lie than adults. Cassidy et al. (2020) found a similar trend in police officers where 'children become more capable of lying as they get older' and that they are 'incapable lie-tellers'. Research into this topic has found that children as young as 2 years old are able to lie spontaneously (Evans & Lee, 2013), but that younger children are less inclined than older children to lie (e.g., Polak & Harris, 1999; Talwar et al., 2002). Early motivations to lie tend to be unsophisticated and related to self-protection (e.g., avoiding punishment; Talwar & Crossman, 2011).

Another common finding within literature in this area, is the idea that children have less knowledge and understanding of both the crime-related event they experienced and the CJS. A child's knowledge of the world and expectations of how the world works increases with age, and this effects how the memory is consolidated because it determines how the memory is interpreted (Principe et al.,2000). As supported by other research (e.g., Cassidy et al., 2020), children are believed to report on things that they do not understand, perhaps because they are being coached by others, which ultimately could impact on their truthfulness or their reliability.

The current focus groups identified that lesser knowledge in childhood is likely to impact on the amount and reliability of memory evidence from children. The Implicit Personality theories (Bargh et al., 1996) would suggest that these perceptions might be developed due to their link with other beliefs or attitudes towards children. For example, it is likely a stereotyped belief that children, due to age, have lesser knowledge, which subsequently will impact on other attitudes in relation to that child (e.g., reliability or credibility). Many previous papers have identified that although children tend to recall *less* information and offer more brief accounts than adults and older children (Vandermaas et al., 1993), this information is usually no less accurate when the accounts are collected without suggestive influences (e.g., Oates & Shrimpton, 1991; Otgaar et al., 2018).

The consensus among professionals in the current research was that a child's memory is more vulnerable to change than an adult. In terms of suggestibility, this was a concern of the police officer group who identified that children can be led during questioning and that they can get 'caught up' in the story. These beliefs are generally consistent with the existing literature. Research has shown that children are more suggestible in many contexts than adults and are especially vulnerable to leading questioning (Ceci & Bruck, 1993; See also Chapter 2). Age is considered to be a strong predictor of suggestibility to misinformation (Cassel et al., 1996) and experimental memory research shows that young children are more likely to change an initial account by incorporation of post-event false information in comparison to older children and adults (Bruck & Ceci, 1999). It is clear that the police officers in this sample were aware of the impact of misinformation as this was discussed as a subject in their training. However, the other professionals did not discuss misinformation or how it

might impact reliability. This identifies inter-agency differences and a lack of interagency cohesion in relation to how they may approach and manage child witness memory, and a lack of standardised or understood means of measuring reliability or memory.

The findings from the current research are supported by the Social Representation Theory (Moscovici, 1998) which posits that through processes of 'anchoring', individuals feel better able to understand novel situations by drawing upon already known information. Generalised perceptions about child memory and reliability likely draw upon societal understanding and experiences of similar events. These representations can ultimately lead to the production of stereotypically ideal defendants and victims (Lindholm & Cederborg, 2016). Deviations from this can change how one perceives a narrative of an event, which could impact on perceived reliability or credibility of a witness. For example, older sexual assault victims have been deemed to be more responsible for their assaults than younger victims (Rogers et al.,) as adolescents are perceived as being more capable of lying due to having more general knowledge (Rogers & Davis, 2007). Each profession could be influenced by their own unique social representations influenced by the culture of the organisation and context of their job roles and due to the differences of their experiences. For example, police may experience more antisociality when working with children compared with sexual harm support staff who usually work with victims. Therefore, we see a difference in what 'anchors' they may have, thus a difference in the social representations that derive from this.

The role of social representations has been utilised in more specific theories that address witness-based assessments. For example, the Dangerous Decisions Theory

(DDT; Porter et al.,, 2009) posits that credibility judgements are made based on one's beliefs, attitudes and experiences of the person being judged. This threat-based judgement (Adolphs, 2002) ultimately leads to 'tunnel vision' (Granhag, 2007) and influences what evidence is accepted. This has the potential to lead to decisions based on 'intuition' which could be guided by incorrect stereotype (Porter et al., 2009). In relation to the current cohort, we have some evidence to suggest that the police officers have experiences of young people being antisocial towards them, which may be influencing their perception and thus could be influencing the judgements that they are making on their credibility (Porter & Brinke, 2009). Conversely, the perceptions of vulnerability and attitudes that all children are truthful (i.e., as reported by the social worker and sexual harm support staff group) might have a similar but opposite impact on their perception of reliability. According to DDT (Porter & Brinke, 2009), both of these attitudes have the potential to lead to incorrect or 'dangerous' verdicts and decisions.

The appropriateness of the CJS

The CJS was designed to operate with consideration of adults and is often inappropriately managed to cope with the demands of children (Malloy et al., 2007). The CJS is sometimes considered to be traumatic, and often the interviewing process is repeated more than once which can be stressful (Quas et al., 2005). The participants in the current study were of the opinion that it is not fair for children to be put through the England and Wales CJS, because it is too stressful, lacks support for children, and is not conducive to meeting the needs of the children. These opinions do not reflect the key considerations as outlined in policy of working with vulnerable witnesses in the CJS in England and Wales which states that it is a requirement for the court and the tribunal to

adopt a more flexible approach when accommodating children (Youth Justice and Criminal Evidence Act, 1999). For example, children should be afforded the opportunity to provide their witness account during live link (YJCEA, s24). However, the current research suggests that the guidelines for accommodating children may either be insufficient, or that the UK based professionals in these focus groups may have limited experience of seeing these special measures be effective.

Specifically, the cross-examination process was a concern for many of the professionals, characterised by frustration towards barristers and also acknowledgments that this process was inappropriate for younger children. The purpose of cross-examination is to 'test' the evidence and the reliability of the memory evidence, including testing the credibility of the witness being questioned. Research examining child performance within these settings, however, has identified that the high-pressure process of cross examination leads to loss of credibility in court for children, regardless of their cognitive ability, because children feel pressured to change answers under cross-examination (Zajac & Hayne, 2003). The police officers in the focus groups conducted by Cassidy and colleagues (2020) also had a pessimistic view of cross examination and its use with children. Relatively recently, it was suggested that vulnerable witnesses are provided the option to pre-record cross examination to avoid addressing an entire court (YJCEA, s.28; 2019). Although implementation of this scheme has been slow and still ongoing nationally, the results of the current research suggest that changes such as this are important to improve the experience for child witnesses.

Differences emerged across professionals, in their opinions and beliefs with regard investigative interviewing and the support of children. It was evident throughout

discussion that there was frustration towards other professionals and how they manage their roles. For example, social workers demonstrated frustration towards police personnel and how they managed questioning. There was suggestion that police and lawyers utilised questioning as a means to collect the 'correct' or preferred answer to be consistent with the evidence that was already provided, and necessary for their case. Social workers and sexual harm support staff reported that the police and lawyers used repeated questions until the 'right' answer was reached or alluded to children being used as 'jigsaw puzzles' such that their accounts were used to support evidence that was already collected, or to develop the narrative that police had already decided.

Defence and prosecution lawyers have been found to approach witness questioning differently, depending on their goals in court. For example, defence lawyers have been found to focus more on peripheral details of an event (47%) than prosecutors (36%) (Andrews & Lamb, 2018), and that defence lawyers asked more repeated questions than prosecutors did and as a result were more likely to elicit self-contradictions from the child witness (Andrews et al., 2015; Andrews & Lamb, 2017). This demonstrates that there are differences in conduct depending on the professional's role, which is partly echoed by the current cohort.

Interestingly, one of the police officers in the current cohort made note to a child saying the 'wrong thing'. It is unclear if they meant 'wrong' as unreliable or 'wrong' as inconsistent with their expectations with what would be useful from an evidentiary perspective. Empirical research has historically evaluated the investigative interviewing strategies of police officers in practice to be poor, due to insufficient use of open-ended questions (Cederborg et al.,2000; Clarke & Milne, 2001). Although the police in the sample appeared to be aware of their training on appropriate interview techniques with

children (i.e., the use of open questions), considering the views of the sexual harm support staff and social workers and the other empirical evidence (e.g., Cederborg et al., 2000; Clarke & Milne, 2001), it is unclear if this training is borne out in practice. Future research might benefit from addressing the need to close a case, or have consistent evidence, and how this might influence how police officers approach the questioning of children, or impact on the perceived reliability and useability of child memory evidence.

Benefits and Limitations

Qualitative research is focused on exploring the experiences and perceptions of the participants. It allows for discovery and reflection on the complexity of reality and exploration of debates among professionals (Braun & Clarke, 2013). The current study benefits the field by removing pre-emptive reduction of experience as can be seen within questionnaires and similar quantitative assessment (Atieno, 2009). Here, the professionals were able to spend more time justifying and exploring their perceptions, which allowed the researcher to gain a richer understanding. Thematic analysis was used to explore the data as this is currently an underdeveloped area within qualitative research, with only one recent published study in the area (e.g., Cassidy et al., 2020). Future qualitative research may be conducted to consider the nuanced meaning of a specific theme. The use of qualitative research does not attempt to assign frequencies or draw distinct conclusions from the data. Therefore, should future research wish to generalise the outcome from the current research, quantitative methods could be used.

One of the focus groups was conducted online through an online video conferencing call system, because of unavoidable social distancing measures that were in place during data collection. Although this allows for geographical diversity and simpler recruitment (Callejo, 2001), the richness of the non-verbal signs was lost. As a

result, the thematic analysis focused only on the content of the discussion without the analysis of non-verbal factors such as body language and non-verbal participant interaction. This may have decreased the richness of the information collected in the sample of the social workers. Therefore, future research should attempt face-to-face style interviews with this cohort.

Future research considerations

Research on perceptions of child witnesses have found an association between the gender of the observer and opinions and perceptions of children (Quas et al., 2005). Research on juries, for example, has found that women are more supportive of a child witness, whereas men are more sceptical and more likely to provide a non-guilty verdict at trial (Najdowsky & Bottoms, 2015). Some research with professionals has similarly found that women are more likely to hold attitudes consistent with believing children than men (Quas et al., 2005) which may reflect the more nurturing and care-taking role that females may adopt in these situations (Quas et al., 2005). In the current research, there were more females than male participants, and therefore there might be a higher likelihood for believing children regardless of professional role. It may be of interest for similar research in the future to reflect with professionals about the role of gender and how this might relate to the perceptions that they hold about child witnesses, and then whether or not this impacts how they relate with children in their professional responsibilities.

The use of focus groups to address the research aims will have been subject to limitations such as the dominance effect (i.e., one person is more dominant in discussion), the 'groupthink' bias (i.e., the group may agree to maintain cohesion) and the 'halo effect' (i.e., a member influences discussion due to their perceived

'importance'; Mukherjee et al., 2015). Although these were not overtly recognised in the current analysis, the use of other qualitative data collection (e.g., individual interviews or written accounts) may also aid in supporting the outcomes.

From the outcomes of this work, future research should consider exploring the opinions and perceptions of lawyers and judges in the UK CJS. The current professionals all demonstrated some concern or frustration towards the barristers within the CJS, so exploring the perceptions held by barristers would be of interest. Practically, organisations and agencies may benefit from opening discussions with their staff about their frustrations of inter-agency working. Moreover, it would be of interest to examine how different ways of working and the differing perception of a child between agencies is likely to impact on the welfare of the child, as well as the impact on the reliability of a child's evidence. For example, a child may first work with a social worker, then an ISVA, then a police officer, then a lawyer, and then a therapist. That is, children are subjected to a range of different professionals with a range of different goals and approaches to their work which could ultimately impact on their well-being and memory account. Previous research has found that children are more self-contradictory with defence lawyers compared to prosecution lawyers (Andrews et al., 2015). Outcomes such as this suggest that witness memory evidence is likely to be dependent on the professional asking the questions. There are a range of reasons why these differences might be seen, such as the type of questioning used or opportunity to build rapport.

The current research indicates that at least three groups of professionals have different opinions, different professional goals, and different methods of working with children. On a few occasions, specifically looking at the 'believability' of a child and

the importance of evidence collection, there were some conflicting ideas between professionals. Therefore, there may be corresponding differences in the witness account and memory reliability based on the professional with whom the child is communicating. It is currently not known what impact these differences have on the welfare of a child and what children feel they are able to share as evidence. Future research might also ask children how they perceive the process, and their opinions on different professionals, and how those professionals relate to them as a witness.

Importantly, this research has identified that there is varied and subjective opinions of child reliability and their competence to give evidence in court. None of the participants were particularly forthcoming about methods they use to assess reliability and their personal opinions on the matter were diverse despite many of them having to make regular decisions on a child's reliability. This suggests that there is currently no standardised means of measuring reliability that is understood or utilised by professionals. Perhaps the use of a standardised measure could support professionals with this decision making. One such measure it the Bonn Test of Statement Suggestibility (Endres, 1997) which has been developed to examine suggestibility in children which may be useful to meet this practical need but requires reviewing in terms of its reliability and validity (See Chapter 4).

Practical Considerations

The current research has implications for practice. First, it is apparent that more training about child memory and the reliability of child memory evidence is desired by agencies that are required to work with children (see Chapter 5). The police group appeared to have more training than the social workers or sexual harm support staff in eliciting accurate memories from children, again however, it is not clear how efficiently

the training was executed in practice. Moreover, since the social workers or sexual harm support staff had conflicting opinions on the impact of trauma on memory, more training is required in that regard.

Within the CJS in the England and Wales, there are currently special measures that have been put in place to support child witnesses. The Youth Justice and Criminal Evidence Act (1999) introduced these to help vulnerable witnesses give evidence in court. Examples include screens (s23, YJCEA), live link (s24 YJCEA), removal of wigs and gowns (s26 YJCEA). However, a common pattern of discussion within the current focus groups was that these special measures were not adequate as the professionals continued to be dissatisfied with how children were managed. It was identified that the system is both traumatising for children and not set up in a way that promotes their wellbeing. Both future practice and research should consider gathering more opinions and experiences working under these measures and identify areas that require improvement. Changes to the CJS and supporting children could improve not only the mental health of the witness, but also the quality of evidence that is being collected.

Conclusion

The aim of the present study was to use a qualitative approach to aid in understanding how UK professionals perceive and assess the reliability of child witness memory and their justifications, opinions and experiences in working with child witnesses during criminal investigations. Through using focus groups and thematic analysis, the study identified three key themes that emerged from the dataset. From these themes, there are crucial considerations for both future practice and research. Namely, future assessment of inter-agency cohesion, the influence of professionals on child well-being and child witness evidence, and investigation on how these might be

improved. Also, research should continue to consider CJS changes and how to better improve the system for children, by continually assessing the impact of procedural changes (e.g., special measures). Robust and on-going development of this research area could potentially create justification for large-scale changes to the CJS and how children are managed within it. Overall, the opinions posited by professionals in the current group is varied and subjective, which ultimately suggests that the assessment of child memory reliability is not standardised or necessarily consistent across professional groups.

CHAPTER IV

Psychometric Critique - The Bonn Test Statement of Suggestibility (BTSS)

Introduction

In Chapter 3, it was found that professionals did not appear to have a consistent means of assessing the reliability of memory evidence from children. In Chapter 4, The Bonn Test Statement of Suggestibility (BTSS) will be critiqued. The BTSS was developed by Endres (1997) to measure individual differences in suggestibility of children, between the ages of 4 and 10. Put simply, the measure is designed to assess how suggestible a child is and how likely the child is to succumb or 'yield' to post-event information. The rationale of this measure was to resemble similar aspects of the Gudjonsson Suggestibility Scale (GSS; Gudjonsson, 1984) which is the most widely used measure for suggestibility in adults. The BTSS was developed to be appropriate for use with children, for example it includes simplified stories and coloured pictures. A story is read aloud to the child to facilitate remembering of the story and then the child is asked questions about the story content. The stories used are neutral events (i.e., not crime related) as to be relevant for different areas of child and developmental research. The BTSS is available in different languages, including English, Italian, Dutch, Swedish and German. There have been a number of empirical research articles assessing the BTSS (see Appendix E), but there has not yet been collation or review of this research. The current review aims to collate relevant research and analysis of the BTSS (see Appendix E for details) to allow for a comprehensive overview. Examination will include a detailed outline of the scale's use, its research contributions, the properties of the measure, and finally a discussion regarding its applicability within the CJS.

Theoretical Foundation of the BTSS

Within a witness interview, the aim is to gather information about a suspected crime. Unfortunately, interviewing witnesses can lead to the potential incorporation of incorrect post-event information, termed 'misinformation', into a witness's account (Loftus, 1979). Misinformation can be delivered to witnesses from interviewers via implied descriptions and expectation (i.e., hints), through leading questions (i.e., a question that leads them to a certain answer) or through misinformation (i.e., incorrect information communicated to the participant). Prior to the development of appropriate psychometrics, there was no unified way to measure one's suggestibility or likelihood to succumb to misinformation. However, research began to conceptualise the idea of 'suggestibility' as an individual trait (Gudjonsson et al., 1986) meaning that a psychometric measure was required for its assessment.

Research from various psychological disciplines highlight that there are individual differences in suggestibility that need to be measured. Measures of suggestibility can be used for research in both cognitive psychology (e.g., memory development) and in forensic psychology (e.g., witness reliability). The leading view is that although both adults and children can be suggestible, children are more prone to suggestion (Bruck & Ceci, 1999; Cassel et al.,1996; See Chapter 2). While age was previously considered to be one predictor of suggestibility (Bruck & Ceci, 1999), more recent research began to assess other variables that are associated with suggestibility, such as lower intelligence (Bettenay et al., 2015) and poorer language skills (Curc et al., 2017). Cognitive functions (e.g., intelligence, short term memory and language) have been regularly linked to increased suggestibility in young children for both neutral events (Chae, Goodman, Eisen & Qin, 2011) and negative or upsetting (i.e., crime-related) events (Eisen et al., 2007). In addition, psycho-social processes are found to be linked with suggestibility. For example, emotional arousal (Eisen et al., 2007), negative temperament such as sadness or anger (Gilstrap & Papierno, 2004), and

mental health factors (Chae et al., 2011) have all been found to increase suggestibility. Research finding individual differences in susceptibility to suggestibility demonstrates the importance of a reliable suggestibility measure.

Other Suggestibility measures

Previous measures of suggestibility have been identified as having various limitations which motivated the development of the BTSS (See Appendix D for details). First, the *Test of Statement Suggestibility* (TAS; Burger, 1971) is a measure of suggestibility developed for adults. In the TAS, 30 slides are presented to participants for half a second, and then suggestive and non-suggestive questions are asked about the slide contents. The number of incorrect answers to suggestive questions is computed as a measure of suggestibility. Despite its high reliability (r = .90), Endres (1997) identified that the measure has not been utilised in forensic services since the development in 1971. Similarly, the *Wurzburg Suggestibility Test* with test norms for children aged 12-13 years-old (WST; Bottenberg & Wehner, 1971) presents pictures to participants followed by 20 short statements referring to the pictures, asking participants to identify if the statement was true or false. However, this does not appropriately represent CJS procedures (i.e., the questioning techniques used) and subjects may easily be able to guess the purpose of the test or believe it to be a discrimination task (Endres, 1997). Another similar assessment is the *Suggestibility Test* for 9–10-year-olds and 12–16-year-olds (SET-S; Zimmermann, 1979). This also utilises pictures and leading questions about the details of the events depicted in the picture. However, unlike the BTSS, the SET-S is focused on assessing older children only (9-years-old to 16-years-old).

Finally, the Gudjonsson Suggestibility Scale (GSS; Gudjonsson, 1984) was considered to be the superior measure for suggestibility (Endres, 1997). It includes a short story followed by questions that contain both true suggestions and misleading information. The scale collects both 'yield' information (i.e., a measure of how much someone would succumb to

misinformation) and a 'shift' score (i.e., changes made to the account). The GSS has been found to be a reliable and valid measure in test-retest reliability (r = .92; Gudjonsson, 1984) and with a Cronbach's alpha of .75 meaning that there is a small variation around the true score obtained by a participant (Merckelbach et al.,1998) and continued support has been found through ongoing use in both forensic research and practice (e.g., Gignac & Powell, 2008). Although the GSS appears to be a reliable and valid measure for suggestibility, it was developed for use with adults, and therefore might not be appropriate for use with children.

The limitations with the previous measures (i.e., the AS, WST and SET-S) motivated the development of the BTSS for children. A measure for younger children (approximately 4 to 10-years-old) was deemed important, because younger children have been shown to be particularly suggestible and professionals perceive child memory evidence to be poor in forensic contexts (Cassel et al., 1996; see also Chapter 3).

Research using the BTSS

As identified, a measure of suggestibility is beneficial for research purposes. Past research has utilised the BTSS to assess the relationship between suggestibility and other variables, which demonstrates the scale's ongoing applicability and usefulness within suggestibility research. For example, Otgaar and Candel (2009) aimed to assess performance on two false memory paradigms—the Deese–Roediger–McDermott paradigm (DRM; Roediger & McDermott KB, 1995) and the misinformation paradigm—and their relationship with age. Otgaar and Candel (2009) found that development of false memories for non-presented words increased with age, but that suggestibility and acceptance of misinformation decreased with age. Although this study did not address the effectiveness of the BTSS as a measure per se, research such as this illustrates how the BTSS is being used in research on suggestibility and false memory, and the potential benefit of its use for quantifying suggestibility.

The BTSS has added to important information about suggestibility, memory and interview procedures. Knowing how children react to suggested information is integral for adapting and addressing these issues in practice. The BTSS research can therefore aid with ongoing theories and hypotheses of internal mechanisms for suggestibility through its application in research on this topic.

Psychometric Characteristics

The BTSS comprises of two parallel versions aimed at assessing the suggestibility of children aged 4-10 years. The measure contains a story (approximately 330 words), with four coloured illustrations and a set of 31 questions. Administration is separated into four phases:

1. Presentation of the stimulus including the story and the illustrations (*Stimulus Story*).
2. Free report where children are encouraged to retell the story content (*Opening Questioning).*
3. 15-minute interval where a non-verbal test is administered (*Interval Phase).*
4. Questioning of the child about the stimulus story (*Questioning Phase).*

The *stimulus story* is presented so that the story and illustrations are shown together, allowing verbal and visual sensory information to be processed. The stories were taken from a thesis written by Bader (1993), with minor modifications (Endres et al., 1997). The protagonist of the story is the same sex as the child being questioned. In one of the stories, a toy duck is lent to a friend for the weekend, the friend breaks it, and it is repaired. In the other story, friends have an accident while roller-skating on the pavement, and a third child is hurt.

The following phase of *opening questioning* immediately follows the stimulus story and allows for the child to recapitulate the narrative to aid with encoding. The information provided by the child can be used as a memory performance control measure. That is, the

free-recall allows for assessment of insufficient understanding or insufficient attention characterised by the child's inability to engage with this stage; if so, the test would not be applicable. The opening questioning (and later questioning phase) are administered verbally so that it is in keeping with what might be expected in 'real-world' witness questioning and therefore is deemed to be ecologically valid and forensically relevant.

Next, the *interval phase* aims to weaken the memory trace for the stimulus story, allowing more space for the influence of suggestion. The Culture Fair Test Scale 1 (Cattell, 1966) is administered to divert the attention away from the story. Finally, the *questioning phase* occurs, which involves 31 questions about the content of the stimulus story. All of the questions are suggestive and are assigned to three question classes or are distractor questions (See Table 8); these are summed to provide the overall score of suggestibility.

Table 1

The four different classes of question used within the BTSS questioning phase

Question Class	Explanation	Example	Subscale
Misleading Yes-No	Incorrect suggestion with affirmation expected	"Oliver was on his way to school when it happened, wasn't he?"	yield
Alterative questions	Two non-correct options to choose	"Did he want to buy apples or bread?"	
Repeated questions	Immediate repetition of a question following previous answer	"Are you sure? Did he want to buy apples or bread?"	shift
Distractor question	Correct answer is suggested	"The boy's name was Oliver, wasn't it?"	

Scoring

The maximum score is 9 on the Yes-No questions, 8 on the Alternative Questions, and 8 on the Repeated Questions. The measure is split into two subscales: yield and shift. These subscales underline two types of suggestibility proposed by Gudjonsson (1992): the tendency

to give into questions (yield) and the tendency to change responses under conditions of social pressure (shift). The Yes-No questions and the Alternative Questions are combined to create the yield subscale *(maximum:17 points)* which measures acquiescence to the misinformation provided. The repeated questions make up the shift subscale (*maximum: 8 points*) that measures how children change their answers to negative feedback. The results of total suggestibility are obtained by the sum of the yield and shift subscales and varies between 0 and 25 points (Endres, 1997) with higher scores representing higher suggestibility. It should be noted that the 'Distractor Question' (See Table 8) are not included into the score on the scale, but are there to prevent participants from being aware of the assessment aims (i.e., to deliver misleading information via suggestive questions) during testing.

Psychometric Properties

Reliability

Internal Consistency across the whole scale

Within a measure that studies one construct, high internal consistency assumes that all of the items within the measure are univariate (i.e., measure the same construct). Many researchers argue that .70 is a generally accepted benchmark for psychometric testing (e.g., Cronbach, 1984) and utilise the Cronbach's alpha (Cronbach, 1984) to assess the correlation between test items (i.e., questions in the psychometric). In the Portuguese version of the BTSS, Dafflon (2012) measured a sample of 122 subjects between 6-8 years old, using the Toy duck story and found a good internal consistency across all of the items of the BTSS (α = 0.74). Similarly, Costa et al. (2008) tested 145 children ages 8 and 9 using the Portuguese BTSS and also found a good Cronbach's alpha (α = 0.714).

Internal Consistency across the three question classes

Other researchers have assessed the internal consistency of the BTSS by assessing correlations between each of the Question Classes (Table 8). For example, Endres (1997) found that Cronbach's alpha for each of the Question Classes (Table 9) ranged from .70 to .77 and increased to .85 when all of the items were considered together (total items; Table 9).

Table 9

Reliability estimates and correlations as reported by Endres, 1997.

Subscale	Cronbach's α	α for parallel version	Retest t	*r* with free recall	*r* with CFT	*r* with age
Yes-No	.74	.84	.67	-.52**	-.71**	-.72**
Alternative	.77	.73	.65	-.27	-.36	-.44**
Repeated	.70	.65	.32	-.16	-.16	-.28
Total Score	.85	.85	.66	-.41**	-.53**	-.62**

Note. Vocabulary scale of the WIS and non-verbal intelligence was measured by the Culture Fair Test (CFT; Cattell, 1966). N=62, **p < 0.001, *p < 0.1

Further studies used the story about the toy duck and supported the findings by Endres (1997), with Repeated Questions ($\alpha = 0.671$) and Alternative Questions ($\alpha = 0.680$) showing good internal consistency, though poorer internal consistency was found for the Yes-No questions (α=.510; Dafflon, 2012). The alpha value for the Yes-No component is lower than would usually be deemed appropriate, however further analysis here identified that exclusion of any item within the Yes-No questions did not bring any significant changes in internal consistency. Therefore, Dafflon (2012) concluded that each of the Question Classes (Yes-No; alternative; repeated) were appropriate to remain in the assessment.

Internal Consistency across the subscales (yield and shift)

Further researchers have considered internal reliability in the yield and shift subscales. Candel et al. (2000) tested children aged 5 to 10 and found that the Cronbach's alpha to be

.78 for the yield scale, .82 for the shift scale, and .87 for the total suggestibility score (Dutch version). Candel and colleagues (2000) assessed again the three subscales of yield, shift and total suggestibility in another study with another group of children, finding Cronbach's alpha of .75, .71 and .82 respectively.

Caffo et al. (2016) also assessed the Italian translation of the measure. They used the Kuder-Richardson coefficient (K-R) to examine internal consistency, finding .78 for the total scale, .78 for the yield scale and .77 for the shift scale. The internal consistency here is within the excellent range as stipulated by Landis and Koch (1977) in that it is above .75. Caffo et al. (2016) also removed one item at a time to identify potential changes to the internal consistency using Cronbach's alpha. Two items reduced the internal consistency: one within the yield subscale, and one within the shift subscale. However, inclusion of these two items was not determined to be severe enough to discredit the measurement and it was concluded that the internal consistency was good.

Test-retest reliability

Test-retest reliability is an essential characteristic of a respectable measure as an individual should obtain the same (or similar) score regardless of when they are tested, or in what context they are tested. The test-retest reliability of a scale is easily measured by testing participants on more than one occasion, with a delay between tests. A correlation coefficient of 0.7 (Kline, 1998) between the scores obtained at the two time points is often considered to be the minimum acceptable outcome.

BTSS test scores have been found to be stable over time, with Candel et al. (2000) finding test-retest reliability at .90 ($p < .05$) for the Total Suggestibility score. This was measured in 48 primary school children (3 age groups: average 5.7, 7.4 and 9.5 years of age) measured using the same story on two occasions, with a 6-week period in between.

Alternate Form Reliability

This reliability measure is similar to test-retest reliability; however, it is measured by the correlation between scores on a psychometric and another version of the same test. The original authors of the BTSS created two different versions. Endres (1997) compared outcomes on the two parallel versions of the measure in a sample of 62 children between the ages of 4 and 10. The two versions of the BTSS were used, with a gap of several weeks in-between with the same group of children. Overall, the two parallel versions of the measure were found to correlate (r = .66; Endres, 1997). Some question classes (Table 8) have also been found to correlate with one another across the two parallel versions: The Yes-No questions and alternative questions correlated with r values of .67 and .65, respectively. However, the correlation for the repeated questions across the two parallel version was only .32 (Endres et al., 1996). These correlations are below what we would expect to demonstrate good alternate form reliability for a test. However, in comparison to other psychometric measures and personality scales, these scores are higher than usually observed, especially considering that the BTSS is being administered to children (Endres, 1997).

Validity

Face Validity

Face validity addresses whether or not a test appears to measure what it claims to. The BTSS appears to appropriately use the misinformation paradigm procedure (e.g., Loftus, 1979). In the misinformation paradigm, participants are first presented with a target event and then provided with misinformation, in this case, in the form of leading questions.

Criterion Validity

The criterion validity of a psychometric assesses how closely the outcome of the measure corresponds with a different, similar test. The BTSS was developed as a child-

friendly version of the GSS. Therefore, comparison between these two scales would be helpful in examining criterion validity.

Roma et al. (2011) compared the Italian version of the BTSS to the GSS. On examining 84 children aged between 8 and 10 years old, they found that there was a strong correlation between the total suggestibility scores on the GSS and the BTSS ($r = .72$, $p < .001$). This outcome suggests that the two measures are assessing similar, if not the same, construct (i.e., suggestibility). The authors also compared subscales, yield and shift. They found that there was a strong correlation between yield subscale scores ($r = .71$, $p < .001$), though a lesser correlation in shift subscale scores ($r = .33$, $p < .05$) across the two measures. The lower shift score correlation can be attributed to the fact that the BTSS and GSS measure their shift variables differently. The BTSS repeats questions immediately after the participant has provided a response to 8 questions, whereas the GSS repeats questions after the participant has responded to all 30 questions. Children were therefore more likely to change their answers in the BTSS (34%) than in the GSS (18%) in response to the negative feedback (Roma et al., 2011). Also, Table 10 demonstrates that there is a relatively large difference of immediate recall scores obtained on the BTSS and GSS; this is likely due to the BTSS including pictures which make it easier for children to remember the content of the stories (Roma et al., 2011).

Table 10

Descriptive Statistics for the corresponding variables in the GSS2 and BTSS in the 84 subjects tested in Roma et al. (2011)

	GSS2				BTSS-1			
	IR	yield	shift	TS	IR	yield	shift	TS
Mean	14.95	6.07	3.62	9.69	24.86	6.02	2.71	8.74
SD	4.78	3.65	2.46	4.94	4.03	3.31	1.76	4.59
Min.	7	0	0	0	15	1	0	1
Max.	29	13	11	22	32	12	6	17

Note. This table shows the statistics for the mean correct items in Immediate Recall (IR), the mean yield and shift scores and the Total Suggestibility (TS).

Construct Validity

Construct validity refers to the quality of a measurement and the appropriateness of the claims made on the basis of the measure's outcome and whether the test measures its intended construct. Construct validity can be assessed by observing other variables and items that are empirically associated with the intended variable being measured. To do this, the suggestibility literature and its relationship with the BTSS will be discussed. As already highlighted within this paper (under 'Theoretical Foundations'), there are a number of individual differences that have been found to be related to suggestibility. The following papers have examined if scores on the BTSS are associated with these individual differences variables to determine if the BTSS is measuring suggestibility (or at least a similar construct).

Memory Strength

There is a theoretical basis to suggest that stronger memories are more resistant to suggestion (King & Yuille, 1987; Pezdek & Re, 1995). Free recall accuracy is negatively correlated with total suggestibility scores (r = -.67), yield subscale scores (r = -.57) and shift subscale scores (r = -.65) in the Dutch BTSS (Candel et al., 2000). This means that better memory accuracy is associated with lower suggestibility scores. The Italian BTSS supports this, as a significant negative correlation between memory score and total suggestibility scale has also been found (r = -.027, p< .001; Benatti et al., 2012). Finally, using the Portuguese measure, moderate, but significant, negative correlations between memory and BTSS total suggestibility were also found (r = -.507, p = .001; Dafflon, 2012).

Age

Moreover, suggestibility declines with age (See Chapter 2). As predicted by the broader literature, age has been found to be negatively correlated with total scores on the Italian

BTSS (r= -.307, p < .001; Bennatti, 2012). Moreover, there is a significant decrease in mean BTSS scores in older age groups on the Dutch measure (Candel et al., 2002), Portuguese (Costa & Pinho, 2008) and Swedish (Finnila et al., 2003) BTSS. Therefore, the results obtained by the BTSS in different age groups show the same pattern as other developmental research (Bruck et al., 1999).

Language

Research has found that lesser language skills are related to increased suggestibility (Clarke-Stewart et al., 2004). Dafflon (2012) found negative correlations between suggestibility subscales in the BTSS and constructs of language (see Table 11). A review by Costa et al. (2008) also found skills in vocabulary to be negatively related to BTSS scores (Kulkofsky & Klemfuss, 2008), again illustrating that the results obtained from the BTSS correspond to the predictions made by the broader suggestibility literature.

Table 2

Pearson's Correlations between BTSS subscales with language and comprehension, as reported by Dafflon (2012)

BTSS Scale	Vocabulary	Comprehension
Total suggestibility	-.421 **	-.346 *
Yes-No Scale	-.321 *	-.365 *
Repeated Scale	-.354 *	-

Note. Figures in parentheses show p values, **p < .01, *p < .05

IQ

Intelligence has been found to be negatively and linearly associated with suggestibility (Bruck & Melnyk, 2004; Gignac & Powell, 2006). As seen in Table 12, the BTSS has been found to have a negative correlation with IQ, measured using the Coloured Progressive Matrices and Standard Progressive Matrices. Further support for this finding has been

observed in both the Italian (Benatti et al., 2012) and the Portuguese (Costa et al., 2008) reviews of the measure.

Table 3

Pearson's correlations between free recall and IQ and BTSS-subscales, as reported by Candel et al. (2000).

BTSS Scale	Free Recall	IQ
yield	-.57*	-.35*
shift	-.65*	-.30*
Total suggestibility	-.67*	-.36*

Note: *p < .05

Overall, BTSS scores appears to correlate with individual difference variables that have previously been shown to be associated with suggestibility. Put simply, the measure seems to produce patterns of scores that would be predicted by the broader literature.

Content Validity

Content validity is concerned with whether a test is representative of all aspects of the construct being measured. In the BTSS, suggestibility is made up of the yield and shift subscales and also the question classes (Table 8). Each of these building blocks of the BTSS aim to measure different aspects of suggestibility.

The validity of the BTSS has been addressed by utilising Exploratory Factor Analysis (EFA). This technique identifies relationships between items and aims to factor a large matrix of correlations into terms of smaller numbers. This identifies important constructs that explain the variance (Norris, 2009). In relation to the BTSS, EFA has been used to identify how many factors load on to total suggestibility; the original authors finding that overall, there were two main factors, yield and shift (Endres, 1997), suggesting that the yield and shift subscales were meaningfully different.

Caffo et al. (2016) utilised the Kaiser-Meyer-Olki test and the Barlett's spherical test to conduct EFA. By analysing the relationship between all items in the measure, they found 8 different factors/constructs that loaded onto total suggestibility and explained 62% of the variance in scores. To analyse this outcome further, a scree plot was used to identify that the constructs could be simplified into two factors to account for this variance (Caffo et al., 2016). From these two factors, the authors identified that all items (apart from question 29) could be loaded into yield and shift subscales, as previously identified by the authors of the BTSS (Endres, 1997). The results found by Caffo et al. (2016) suggest that the original authors of the BTSS correctly identified two psychometric constructs within the measure that explained the variance of suggestibility. A Spearman's correlation coefficient has further found no significant correlation between the yield subscale and the shift subscale (r = .135, p =.123; Caffo et al., 2016) suggesting that they are independent. They concluded that the individual differences in the yield subscale appear to be linked to the cognitive abilities of the subject as scores are associated with age, memory and intelligence. The shift subscale however appears to be situational, and scores are not associated with age, memory and intelligence.

Dafflon (2012) conducted a similar assessment of the Portuguese BTSS by using EFA to assess the applicability of the question classes (Table 8). The items extracted loaded onto three factors, explaining 30% of the variance. The first extracted component explained 11.46% of the variance and contained 8 items (repeated questions), though Item 11 was assessed as belonging to another component. The second component (alternative questions) explained 9.68% of the variance contained the same 8 items as in the original scale. Lastly, the third component (Yes-No) explained 9.18% of the variance and contained 9 items from the original distribution. Again, this EFA appears to support the author's idea of three question classes to explain the variance in suggestibility scores.

Predictive Validity

Predictive validity refers to the ability of a test to predict a relevant criterion. If a psychometric can reliably predict something related to the variable it is intending to measure, then it can be considered to be a predictive measure. Importantly, suggestibility is a multifaceted construct and is impacted by both individual differences (e.g., intelligence, age, memory) and external or social factors (e.g., stress, social desirability). Therefore, it is important to note that the BTSS will only predict an element of an individual's overall suggestibility—their individual trait-like proclivity towards suggestibility—and could not assess the external factors that might also be impacting on their likelihood to accept misinformation in a particular situation.

Finnila et al. (2003) examined the predictive validity of the BTSS by assessing children on a crime-related witnessed event. Two realistic interview structures were used, varying in the pressure the children were under to comply with suggestions. It was found that those who scored highly on the BTSS were more likely to accept the suggestive interview questions than those who scored lower. This demonstrates that there is at least some predictive validity. However, as noted above the BTSS cannot account for the various external variables (e.g., differences in interviewer and interview conditions) that are likely to have also impacted on the suggestibility because it does not examine those factors (Finnila et al., 2003).

Similarly, as discussed earlier in this review, Candel and colleagues (2005) used the BTSS to measure the impact of misinformation, finding that the BTSS was not predictive of whether or not a child was likely to confess to a non-committed act (i.e., not predictive of acceptance to suggested information). This, along with the Finnila et al. (2003) study, provides mixed evidence for the predictive validity of the BTSS. Without confirmation that the BTSS has good predictive validity, it is difficult for practitioners to use this measure

confidently as there is no assurance that it can accurately predict suggestibility in an interview.

Test Norms

The outcome of a psychometric measure can be plotted to develop test norms to support with the interpretation of results through comparison against a larger sample. The scale authors did not make test norms available in their original paper (Endres, 1997), though Benatti and colleagues (2012) have offered comprehensive norms for the Italian version of the BTSS. Any studies or reviews on the BTSS prior to these norms being published did not include same age norms in which to compare their outcome. Without these norms it is very difficult to draw conclusions, such as when a participant is considered to be highly suggestible. For example, research conducted by Candel et al. (2005) used the BTSS to measure the relationship between suggestibility and likelihood to falsely confess, but the authors did not have information on what the BTSS outcome actually meant on its own (i.e., if a BTSS score was considered to indicate a child who was low, moderate, or highly suggestible). Therefore, they were unable to make judgements based on the outcome, without correlating it with another variable. Consequently, should this measure try to be used practically (e.g., in court), the outcome would be difficult to interpret.

Forensic Application

Benefits of the BTSS

According to this review, the BTSS has overall been identified as a relatively good measure as it has been found to be a reliable assessment. Therefore, theoretically, it could be suitable to use in the CJS to assess child witness's suggestibility. The outcome of the measure could aid professionals in determining the reliability of a child witness's account, and also the credibility of the statements made. This may decrease erroneous testimony,

incorrect convictions of innocent people or guilty people going free, and also encourage further training and research.

More specifically, the BTSS could be utilised as a baseline for suggestibility, to inform professionals on how best to approach a case in relation to questioning and management of evidence. For example, if a child is determined as being highly suggestible, it may change how the questioning is approached. Furthermore, the BTSS could be utilised in cases of mass accusation where many children make various accusations varying in content and consistency. If those children that provide more impractical or fantastical stories are rated having high suggestibility scores, it could support the defence in being able to argue that the witness memory may be contaminated with misinformation should there be limited tangible evidence.

Finally, the BTSS could benefit CJS proceedings by negating the effects of confirmatory bias in both investigative interviewing and also during court proceedings. Confirmatory bias is the tendency to be biased to information that confirms one's own beliefs and reject anything on the contrary (Goodman & Melinder, 2010). An investigative interviewer might have pre-established beliefs about the child which may influence their questioning and assessment of the child's suggestibility and reliability. The BTSS can prevent the attempted measurement of suggestibility using non-standardised means (e.g., professional opinion), by providing a standardised and ostensibly more reliable outcome on which to base decisions, rather than relying on the investigator's opinion alone.

Limitations of the BTSS

However, there are some important limitations of the BTSS which must be considered. First, there is currently limited research into the predictive validity of the BTSS, which is integral to its use in the CJS. If the suggestibility measure may impact on the outcome of the court proceedings, then it is extremely important for the suggestibility measure to accurately

predict how suggestible this child is going to be or has been during questioning. Without more research and data on the predictive validity, it will not be appropriate for use in proceedings. Furthermore, there are two parallel versions of the BTSS, but no data on the predictive validity of each version. Therefore, there is no information on which version of the measure would be best for practice (i.e., should an investigator use version one or version two?). Other than for research, there is no obvious reason for having two measures, as suggestibility is unlikely to be measured twice during criminal proceedings.

Very few studies on the BTSS or research on suggestibility more generally have considered other important factors that might influence a child's likelihood to be misled by suggestion, such as race or socioeconomic status (Bruck & Melnyk, 2004). Individuals who live in lower socioeconomic households are more likely to be a victim of violence, burglary and sexual assault (Cuthbertson, 2018). Therefore, practically it is expected that these individuals will be more prevalent in the CJS, and more likely to be questioned as witnesses to crime. The lack of race and socioeconomic research in relation to the BTSS then limits its applicability in the CJS and its use with all witnesses. Similarly, there are limited cross-cultural considerations for test-norms. Without test-norms, the outcome of the measure has no meaning for criminal proceedings as it cannot determine if a child is highly suggestible or not, as practitioners have limited norm data to use to interpret the outcome score. Currently, the BTSS is better suited to correlation or prediction studies within social and developmental sciences where it can be utilised to assess the impact of suggestibility on different outcomes, or to assess how individual differences can impact suggestibility.

The witnesses who are involved in the CJS are also likely to be experiencing a highly emotional and stressful time. In the majority of forensic literature, replicating this level of stress and anxiety is not ethical, especially in children. Therefore, the BTSS falls victim to the lack of ecologically valid testing to replicate the mental state of a child witness, though

research has previously identified that high levels of stress can impact on the child's susceptibility to false information (Chae et al., 2011). Similarly, other external influences, such as the time delay between an event and the interview, could affect suggestibility (see Chapter 2). Importantly, the BTSS measures the individual factors and individual differences that can impact the suggestibility of a child but does not measure the external factors that may also be important in determining suggestibility. Use of the BTSS could impact on professional and jury perception of the child witness (based on individual factors that the scale assesses), however legal decision-makers may fail to consider other important influences on a child's suggestibility (e.g., external factors).

Ethical Considerations

Regardless of the BTSS's reliability, validity or applicability, there are ethical considerations that impact its suitability in forensic applications. If a child is determined to be highly suggestible, or not suggestible, should this determine how reliable their account is deemed to be, or how seriously practitioners take the account? When discussing how psychometrics such as this could be useful within the CJS, one must continue to address what the outcome could mean for a young witness. Suggestibility is but one of the many factors that impact a witness's account, though the outcome of the measure could change the perception that professionals and potential jury members have of the witness and thus change the outcome of proceedings. Again, this is why it so critical that any measure implemented is reliable and valid for the populations on which it is being used.

Conclusion

This review has collated different translations of the BTSS to assess the scale's properties and applicability in the CJS. Each piece of research assessing different translations of the BTSS has prioritised different areas of reliability and validity which, when combined, allowed for a comprehensive review. Overall, the BTSS has been found as a reliable

assessment to measure individual suggestibility in the age range of children for whom it was developed, as it has good internal consistency, its scores correlate well with similar measures and its scores are consistent with predictions that come from the broader suggestibility literature. However, the applicability of the BTSS in practice is currently not encouraged given the lack of control over other external factors that are not assessed by the measure, the lack of predictive validity data, the lack of test-retest data, and also the ethical considerations. It appears that suggestibility measures should only be considered as a second-best option. Primarily, as practitioners, we should be focusing on how the processes in the CJS can support children in providing accurate statements and refrain from using techniques that increase the likelihood of them being discredited in court.

CHAPTER V

Discussion

This chapter provides a summary of the work that has been achieved throughout this thesis, including a review of the main findings and recommendations of how this work might have implications for research, policy and practice.

Overall Aims

The aim for this thesis is to better understand the reliability of memory evidence from children. This was achieved by: (1) examining how to incorporate measurement of suggestibility to misinformation into practice, to identify whether assessing suggestibility to misinformation could benefit forensic practitioners and child witnesses; and (2) examining how children are perceived by criminal justice professionals, and how practitioners work with memory evidence from children. The thesis collated relevant information to discern the presence of a misinformation effect in children compared to adults, reviewed the BTSS (Endres, 1997) as a potential means of measuring suggestibility, and explored perceptions of reliability and working with child memory information.

Summary of findings

Chapter 1 introduced the key concepts around misinformation, suggestibility and how this might impact on the reliability of a child's account. This included a summary of some of the leading theories that can explain the misinformation effect including changes to original memories that were encoded (Loftus, 1975), strategic memory effects (Zaragoza et al., 1987), activation-based theories (Ayers et sl., 1998), source misattribution (Johnson, 1998), and fuzzy trace theory (Reyna et al., 2016).

Chapter 2 was intended to first identify the presence and extent of the misinformation effect in children in comparison to adults. The meta-analysis was motivated by two reviews:

Otgaar, et al (2018) and Payne et al(1994). The former, because of the non-systematic analytical technique used, and the latter because of their failure to separate children from adults in examining the misinformation effect. In total, 17 papers met the inclusion criteria yielding a total number of 2,582 participants. Two analyses were conducted using a random effects model. First, the susceptibility to post-event misinformation was measured by directly comparing the frequency of incorrect responses to misled items in children and adults (*misinformation item analysis*). Second, the mean difference between control items and misinformation items (i.e., the size of the misinformation effect) in adults and children was analysed using sub-group analyses (*control - misinformation item analysis*).

In the first analysis, there was a large effect showing that children were significantly more likely to yield to suggestions than adults, which supports much of the previous research in the area (e.g., Ceci & Bruck, 1993). In the second, the meta-analysis found that there was a significant misinformation effect in children, but not adults.

Chapter 3 addressed the practicalities of child witness reliability, by interviewing practitioners that work with child witness memory. Specifically, this piece of research used a qualitative approach to explore how UK professionals understand and assess the reliability of child witness memory and their justifications, opinions and experiences in working with child witnesses during criminal investigations. The research encouraged a reflective conversation about the reliability of child memory with professionals, namely with social workers, police officers and sexual harm support staff (i.e., ISVA's and forensic examiners/researchers). Using thematic analysis, three main themes emerged, with sub-themes also being identified. The first theme was concerned with the contextual factors that differed among cases, such as the various external influences (e.g., parents) that had the ability to impact on the reliability of the child's account. The practitioners were considerate of these external influences and reported that they include these into their assessment of reliability. The second theme

identified that there were held perceptions about the global idea of 'children' and 'child reliability' which influenced professional perception on all cases. For example; social desirability being higher in children compared to adults, lesser cognitive skills in younger children, suggestibility and children's limited knowledge about the world. Finally, the professionals believed that the CJS in the UK is not appropriate to obtain reliable memory evidence from children, and that regardless of their abilities or experiences, the CJS is potentially traumatising for children and does not effectively support their well-being.

Overall, the research in Chapter 3 provided insight into how child memory evidence is perceived by this cohort of professionals working in the UK CJS. Ultimately, scientific research into this topic aims to support and improve the practicalities of working with child memory in practice. Therefore, involving practitioners in research is very important because it supports understanding of how research is translating into practice, and informs how research can support practice. The findings of the study in Chapter 3 suggest that further improvement of the CJS is needed in relation to how child witnesses are supported. In relation to both well-being and evidence collection, the current cohort were of the opinion that there was lack of guidance or training in this regard, and that perhaps guidelines were not being adhered to. For example, previous research has found that suggestive questions continue to be used in court (Andrews et al., 2015). There appears to be strained interagency cohesion in relation to how children are perceived, questioned and how reliability is understood or measured.

To support with establishing better means of assessing reliability in criminal justice settings, Chapter 4 focused on reviewing the Bonn Test Statement of Suggestibility (BTSS) developed by Endres (1997). The BTSS was measured to assess suggestibility in individuals between the ages of 4 and 10-years-old. The BTSS was designed similarly to the Gudjonsson Suggestibility Scale (GSS; Gudjonsson, 1984), where the likelihood to yield (i.e., succumb to

misinformation) and shift (i.e., change statements following repeated suggestion) was identified using a standard misinformation paradigm. Various reviews of the BTSS written in different languages, including English, Italian, Dutch, Swedish and German, were collated. It was concluded that the BTSS, overall, is a reliable measure of suggestibility. For example, the internal consistency was high, and the test-retest reliability scores were considered acceptable. Moreover, the scale appeared to be a valid measure in terms of its face validity, criterion validity, content validity and construct validity. Based on this information alone, theoretically the assessment could benefit the CJS and be used as one of a variety of methods employed to assess the reliability of a child witness's statement. However, various limitations of the BTSS were identified which mean that implementation of this measure in practice should be approached cautiously. First, the predictive validity of the scale is largely unknown, there are no useable test norms, and the measure has not been researched in diverse socioeconomic contexts. The two parallel versions of the BTSS also complicate matters, as there is no information on which version of the test should be used in practice, and the alternate form reliability of the two versions was inconsistent across reviews.

Moreover, this Chapter raised an ethical debate which was crucial in the assessment of its applicability in court. The aim of the BTSS would be to prevent incorrect convictions and support the furtherance of justice. It has the potential to improve a child's credibility through obtaining a good score and therefore may be supportive in a prosecution's case. However, might a measure such as this influence which children are believed, and which are not? Having a 'poor' score on the measure might severely impact a child witnesses experience in court and harm their credibility. As identified, the reliability of memory is influenced by many factors, with suggestibility being only one of them. Therefore, the potential for harming credibility may not be beneficial.

Practical Implications

Policy and Procedures

Research is relatively conclusive regarding children being vulnerable to misinformation, as identified by the meta-analysis (Chapter 2), which appears to be translating into the policies and procedures of working with children. For example, the Achieving Best Evidence in Criminal Proceedings (ABE; Ministry of Justice, 2011) guidelines identify how to approach and interview child witnesses to collect the most reliable evidence. Not only do guidelines such as this outline information on suggestibility and the inappropriateness of leading questions but they also support practitioners in handling a range of factors that can impact on both memory and well-being.

However, from interviewing professionals (Chapter 4), it is unclear (at least with this cohort) whether or not these guidelines are being adhered to or even if these policies are practically useful, as there were various issues identified. Some of the professionals reported that they were regularly asked to identify if a child's account of events was reliable, without appropriate training or information. Similarly, they were all regularly involved in interviewing procedures and evidence collection, but again were not aware of any training or policies on the topic (excluding the police officers who had received training). Other research in this area has found similar reports, with professionals identifying that more child-appropriate approaches are needed during investigative interviewing (Brubacher et al., 2011). Even within the police group, there were differences among the participants in relation to their knowledge of child memory and the appropriate approach when questioning them. The differences in their knowledge mostly stemmed from the fact that some officers had completed further training in investigation techniques compared to others. Regardless, all of the professionals (including the police officers) believed that training provided by child witness memory experts would be useful for their professional roles.

Training

In Chapter 3, a variety of different factors were discussed in relation to child memory, with suggestibility only being identified by the police officers, and not the social workers or sexual harm support staff. This outcome suggests that more rigorous training programmes might be useful for practitioners to gain more psychological understanding in this area. For example, training could focus on basic understanding of leading cognitive memory theories (e.g., memory models; Atikinson & Shiffin, 1968; see Chapter 1) and how this might translate for their jobs. Training could focus on the misinformation effect, and the various factors that might protect someone from succumbing to misinformation. Such training might support practitioners in feeling more confident about their appraisal of the reliability of a child's account and up-skill them to feel more competent in being involved in these important conversations. Research has found that training can reduce employees fear and worry about job expectations (Yang & Hsieh, 2006) and prevent feeling overwhelmed or pressured in the workplace (Cavalluzzo & Ittner, 2004).

Given the reported difficulties in inter-agency working found in Chapter 3, training might also support professionals in finding more common ground or consistency in how they approach cases. Literature in this area has found that any changes or updates to policy or guidelines will be more successful if training is offered because it can raise support and endorsement of changes (Kroll, 2015). Therefore, child memory training could change the narrative from 'this way of working is a legal obligation' to 'this is the right or best way to approach it because of XYZ' (Kroll, 2015). If training is consistent across agencies and everyone receives the same or similar training, professionals will be more aware of the 'bigger picture' and how their roles fit in as part of a wider system and then how this benefits the child.

Measuring suggestibility

In Chapter 4, the Bonn Test Statement of Suggestibility Measure (BTSS; Endres, 1997) was concluded as being reliable and *theoretically* (after further research into aforementioned reliability and validity constructs) may be useful in criminal trials to determine a child's suggestibility to the misinformation. For example, a high BTSS suggestibility score might encourage professionals to be more cautious about their questioning, or to approach questioning using different techniques, compared to children who might score lower on the suggestibility scale. Specifically, a child who scores highly could be afforded further protection from repeated or suggestive questioning during cross-examination, which continues to be used by lawyers in court and results in children changing their answers (Andrews et al., 2015). Moreover, the BTSS could support with cases of mass accusation (e.g., Galileo' Elementary school; Otgaar et al., 2017). If children that provide fantastical stories have high suggestibility scores, it could support the defence in being able to argue that misinformation is impacting on the witness memory should there be limited tangible evidence.

An alternative to the BTSS approach, is to use expert witnesses in decision making to potentially help to assess a child's suggestibility. The role of psychological expert testimony on reliability has been addressed by a meta-analysis, incorporating 23 studies with 4,669 psychology expert participants. There was an agreement rate of 68% on 16 factors that are related to reliability (Desmarais & Read, 2011), which demonstrates that there is still room for disagreement across experts in factors that are important for reliable accounts. However, the role of the psychological expert could support in encouraging reflection on factors that impact reliability of victim or witness accounts and to provide psychoeducation to jurors during decision making (Nicolson et al., 2017). Currently, it is believed that the research on reliability continues to lack ecological validity and, as such, may not be useful to incorporate

into every case. Instead, it might be more appropriate to call on the literature via expert witnesses for specific cases where a reliability concern is identified (e.g., evidence of misinformation during questioning; Nicolson et al., 2017). The role of psychological experts in court is beyond the scope of this current thesis, however research should continue to address how and when this might be useful in supporting child witnesses.

Ethics in practice

There are ongoing ethical concerns that require practitioner attention when working with child witnesses. Although understanding the potential shortcomings of the reliability of memory evidence is required to better support child witnesses, the research outcomes could potentially lead professionals and jurors to discredit children. For example, as discussed in Chapter 4, there are ethical difficulties in including suggestibility measures into assessment in court. Importantly, it is currently unethical to use the BTSS (Endres, 1997) in practice because of the lack of predictive validity measures and data norms. But also, the measure has the potential to further the victim-blaming culture in the UK CJS which is already under scrutiny. For example, the CJS in England and Wales has been criticised for responding poorly to survivors of sexual violence, where conviction rates are as low as 7.5% (HMICFRS, 2016). In conclusion, there may be certain cases and situations where a suggestibility measure might be useful (e.g., mass accusation and improbable reports). But ultimately, the question remains: should the outcome on any measure determine how we approach a child who is reporting a crime?

Child witnesses should be provided with adequate support to provide their account, as their input into a case can be extremely useful. For example, ongoing research (e.g., Otgaar et al., 2018) suggests that children may be more resistant to spontaneous false memories than adults, which is explained by the associative activation theory (Howe et al., 2009). Perhaps in instances where there are conflicting reports from different adults witnesses due to potential

spontaneous false memories, child witness reports can support in discerning which version is most probable.

Research and Theory Implications

Interviewing Professionals

As was clear in this thesis, more research is needed to better understand the perceptions and methods utilised in practice. Having access to published policies and procedures is not sufficient to understand how children are managed practically in the CJS. Information that is written and documented in official records and guidelines (e.g., ABE) appears to be different to what was reported by the professionals (Chapter 3). For example, the ABE states that vulnerable witnesses should receive support during the interview, pre-trial, court and post-trial, whereas the professionals interviewed in Chapter 3 were of the opinion that there was limited support and that the CJS was re-traumatising for the children involved. In all areas of scientific research, the practitioners who are expected to utilise and follow scientific guidance could be involved in exploration of the topic and research on its benefits and its practical use. This is helpful because they are ultimately the individuals who put research into practice and make the practical and logistical changes relevant to improve the experiences of the children progressing through the CJS.

As identified in Chapter 3, theories have been developed about the role of personal perception on decisions that are made about witnesses in the court room and research has identified that professional perception and interpersonal judgements guide judgements and outcomes (e.g., Dangerous Decision Theory; Porter & Brinke, 2009). Theories such as this identify that the perception and opinion of professionals should be an ongoing research priority and that there should be importance placed on finding non-subjective means of assessing reliability. Chapter 3 identified that the perceptions of child witnesses might differ depending on the profession, which has been supported by other research (e.g., Melinder et

al., 2004; Granhag et al., 2005). It would be interesting to further this research topic by interviewing different professionals that were not included into the current thesis, such as lawyers and judges.

This thesis has identified the presence and strength of the misinformation effect and that a measure of suggestibility could theoretically support assessment of reliability. It might be of use then to reflect with these professionals more specifically on the role of misinformation and the extent in which professionals understand and are sensitive to misinformation when working with children in the CJS. The more professionals that are included into research similar to this would ultimately provide a better understanding of child witness experiences.

It may also be of value to interview children who have experienced CJS procedures and their parents. The outcome could be compared with the perception of professionals about similar issues. This can be used to establish where the improvements or changes might be needed and perhaps highlight new issues that need to be addressed. Currently, there is limited research that focuses on asking child witnesses about their perceptions, and instead research focuses on the perception of professionals and jurors (e.g., Melinder et al., 2004). Ultimately, the goal with this research would be to gain further insight into the experiences of child witnesses to support with improving the CJS.

The misinformation effect

This thesis aimed to explore the reliability of child memory evidence, with a specific interest in the role of the misinformation effect and suggestibility. Chapter 2 addressed this, finding that the misinformation effect was observed in the child cohort, but not in the adult cohort. This was an interesting finding, as the misinformation effect in adults has been studied since early 1970's (Loftus & Palmer, 1974) and procedures in the legal system have been implemented to reduce the possible negative impacts of misinformation in adult witness

testimony (e.g., 2012 New Jersey Supreme Court jury instructions). Further support was found for this outcome by reanalysis of the Payne et al. (1994) study and separating the result from children and adults, which also found no strong evidence of a misinformation effect in the adult cohort (Williams et al., under review).

The null findings in adults may suggest that the memory theories put forward to explain false memories may need to be updated or may even be unnecessary when trying to understand misinformation in adults. Currently, the general consensus among researchers on this topic, is that there is no single memory process that is responsible or explains false memory development, and that it is likely a culmination of various proposed theories, such as activation-based theories (e.g., Ayers et al., 1998), source monitoring theories (e.g., Lindsay et al., 1989), trace theories (e.g., Brainerd & Reyna, 1995) and strategic memory effects (Zaragoza et al., 1987). At least in the adult cohort in the studies included in the meta-analysis, no support for these hypotheses have been found. However, these theories continue to provide explanation for why and how both spontaneous and misinformation-based false memories can develop in various contexts and in various areas of psychological research; not only eyewitness/forensic research (e.g., false memories of political events and their influence on political orientation; Frenda et al.,2013).

In the child cohort in the meta-analysis in Chapter 2, there was a clear misinformation effect, which was an expected trend and already understood by the trained police officers in Chapter 3. These findings support various theories that hypothesise potential reasons for the differences in suggestibility between children and adults. For example, higher likelihood for source confusion (Lindsay et al., 1989), less developed inhibitory control skills (Alexander et al., 2002), less developed language skills (Clarke-Stewart et al., 2004) and being more influenced by compliance to authority (Ceci et al., 1987).

However, an important outcome for Chapter 2 was the high levels of heterogeneity observed in both child and adult cohorts. This suggests that there are various methodological and individual differences that influence suggestibility other than age alone. A review by Klemfuss and Olagues (2020) identified a plethora of individual differences that increase suggestibility, including various demographic factors, cognitive factors and psycho-social factors. Future meta-analyses into this topic might benefit from including various sub-group analyses to identify what factors explain the heterogeneity and improve the ongoing understanding of factors influencing misinformation acceptance.

Research into how to decrease the misinformation effect in children is already underway. For example, techniques such as 'warning' participants about the misinformation (Blank & Launay, 2014) and 'training' the children about memory vulnerabilities before questioning (Blank & Launay, 2005); both of which have found to be effective. Theoretically, it is thought that misinformation warnings might prevent memory impairments because the source discrimination is interrupted and the response bias (e.g., recency effects) are undermined (Oeberst & Blank, 2012). In relation to memory mechanisms, the usual 'search-and-accept' (i.e., search for the memory, and accept the memory as fact) will theoretically be replaced with a 'search-and-discriminate' strategy (i.e., search for the memory and critique the memory to assess if it is correct or due to misinformation) (Oeberst & Blank, 2012). Ongoing work into reducing susceptibility to misinformation in children is imperative for improving CJS procedures, which were identified as lacking by the professionals in Chapter 3.

Conclusion

Using the information from this thesis, there is further evidence to support that children are suggestible to misinformation using the standard misinformation paradigm, but that further exploration of adult suggestibility may benefit the advancement of this topic (Chapter

2). One of the main aims of this thesis was to explore how a suggestibility measure might support practitioners in assessing reliability, and the outcome identifies that the BTSS (Endres, 1997) *may* be useful if research is conducted that evidences the scale's predictive validity and develops test norms (Chapter 4). The second aim for this thesis was to examine how child reliability is currently being perceived in the CJS. Chapter 3 demonstrates that the current cohort is understanding of the external case-factors that influence reliability, but also hold beliefs about all children; a mixture of both compassionately motivated child-believing perceptions, and skill-focused apprehension in relation to cognitive skills and memory. Moreover, all professionals were of the opinion that the CJS may not currently be appropriate to foster reliable evidence and may be traumatising for the child witnesses. Both of these outcomes demonstrate that there is still an ongoing need to continue research into how CJS services may be improved for young people and children. Using the expertise from practitioners in the field, research communities should further reflective exploration of the current special measures used in the CJS and identify where these could be improved upon. Moreover, there should be continued attempts to provide training for the professionals working with child witnesses to support with general understanding, approach to working, and interagency cohesion. Finally, research should be establishing the benefit of using child memory evidence and how a child witness can be an asset to building a case, rather than focusing on how they might hinder a case.

References

Aarons, N. M., Powell, M. B., & Browne, J. (2004). Police perceptions of interviews involving children with Intellectual Disabilities: A qualitative inquiry. *Policing and Society, 14*(3), 269-278. https://doi.org/10.1080/1043946042000241848

Ackil, J. K., & Zaragoza, M. S. (1995). Developmental differences in eyewitness suggestibility and memory for source. *Journal of Experimental Child Psychology, 60,* 57–83. https://doi.org/10.1006/jecp.1995.1031

Adolphs, R. (2002). Trust in the brain. *Nature Neuroscience, 5*, 8–9.

Alexander, K. W., Goodman, G. S., Schaaf, J. M., Edelstein, R. S., Quas, J. A., & Shaver, P. R. (2002). The role of attachment and cognitive inhibition in children's memory and suggestibility for a stressful event. *Journal of Experimental Child Psychology, 83*, 262–290. https://doi.org/10.1016/s0022-0965(02)00149-2

Alonso, L. E. (1998). *Le Mirada Cualitativia en Sociologia. [The qualitative gaze in sociology].* Madrid: Editorial Fundamentos. https://doi.org/10.2307/40184282

Andrews, S. J., Lamb, M. E., & Lyon, T. D. (2015). Question types, responsiveness and self-contradictions when prosecutors and defense attorneys question alleged victims of child sexual abuse. *Applied Cognitive Psychology, 28(2),* 253-261. https://doi.org/10.1002/acp.3103

Andrews, S. J., & Lamb, M. E. (2017). Lawyers' question repetition and children's responses in Scottish Criminal Courts. *Journal of interpersonal violence, 36*(1), 276-296. https://doi.org/10.1177/0886260517725739

Andrews, S. J., & Lamb, M. E. (2018). Lawyers' question content and children's responses in Scottish Criminal courts. *Psychology, Crime, & Law, 25*(10), 1008-1027. https://doi.org/10.1080/1068316X.2019.1611829

Atieno, P. O. (2009). An analysis of the strengths and limitation of qualitative and quantitative research paradigms. *Problems of Education in the 21st Century, 13,* 13-18.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human Memory: A Proposed System and its Control Processes. *Psychology of Learning and Motivation, 2,* 89-195. https://doi.org/10.1016/S0079-7421(08)60422-3

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

Bader, S. (1993). Die Entwicklung der zivilrechtlichen Deliktsfähigkeit gem. Empirisch-psychologische Überprüfung einer normativen Altersgrenze [The development of civil liability: Empirical validation of an age norm]. Unpublished diploma thesis, Psychological Institute, Bonn.

Baddeley, A. D. (1992). Working memory. *Science, 255*, 556–559. https://doi.org/10.1126/science.1736359

Ball, C. T., & O'Callaghan, J. (2001). Judging the accuracy of children's recall: A statement-level analysis. *Journal of Experimental Psychology: Applied, 7,* 331–345. https://doi.org/0.1037/1076-898x.7.4.331

Bargh, J. A., Chen, M., & Burros, L. (1996). Automaticity of social behaviour: Direct effects of trait construct and stereotype-activation on action. *Journal of Personality and Social Psychology, 71*(2), 230-244. https://doi.org/10.1037//0022-3514.71.2.230

Bekerian, D. A., & Bowers, J. M. (1983). Eyewitness testimony: Were we misled? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(1), 139–145. https://doi.org/10.1037/0278-7393.9.1.139

Belli, R. F., Windschitl, P. D., McCarthy, T. T., & Winfrey, S. E. (1992). Detecting memory impairment with a modified test procedure: Manipulating retention interval with centrally presented event items. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(2), 356–367. https://doi.org/10.1037/0278-7393.18.2.356

Benatti, F. (2012). La suggestionabilità [Suggestibility] (paragraph 4.1. to l 4.5), in Codognotto, S., Magro, T., La testimonianza del minore [The testimony of the child], Maggioli Editore, Santarcangelo di Romagna (RN).

Benton. D. F., Ross, E., Bradshaw, W.N., & Thomas, G.S. (2006). Eyewitness memory is still not common sense: Comparing jurors, judges and law enforcement to eyewitness experts. *Applied Cognitive Psychology, 20,* 115-129. https://doi.org/10.1002/acp.1171

Bettenay, C., Ridley, A. M., Henry, L. A., & Crane, L. (2015). Changed responses under cross- examination: The role of anxiety and individual differences in child witnesses. *Applied Cognitive Psychology, 29*(3), 485–491. https://doi.org/10.1002/acp.3125

Bjorklund, D. F., Cassel, W. S., Bjorklund, R. B., Brown, D. R., Park, C. L., Ernst, K. & Owen, F. A. (2000). Social demand characteristics in children's and adults' eyewitness memory and suggestibility: The effect of different interviewers on free recall and recognition. *Applied Cognitive Psychology, 14*(5), 421-433. https://doi.org/10.1002/acp.3125 10.1002/1099-0720(200009)14:5<421: aid-acp659>3.0.co;2-4

Blank, H., & Launay, C. (2014). How to protect eyewitness memory against the misinformation effect: A meta-analysis of post-warning studies. *Journal of Applied*

Research in Memory and Cognition, 3(2), 77-88. https://doi.org/10.1016/j.jarmac.2014.03.005

Borenstein, M., Hedges, L.V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Complex Data Structures. In Introduction to meta-analysis.* Chichester, U.K: John Wiley & Sons.

Bottenberg, E. H., & Wehner, E. G. (1971). Terminological communicability between psychiatrists and psychologists. *Archivfür Psychiatrie und Nervenkrankheiten, 215*(3), 33–45. https://doi.org/10.1007/BF00342821

Bowcott, O. (2017). Two-year-old girl gives evidence in UK abuse case. Retrieved from https://www.theguardian.com/law/2017/oct/10/two-year-old-girl-gives-evidence-in-uk-abuse-case.

Brackmann, N., Otgaar, H., Sauerland, M., & Howe, M. L. (2016). The impact of testing on the formation of children's and adults' false memories. *Applied Cognitive Psychology, 30*(5), 785-794. https://doi.org/10.1002/acp.3254

Brainerd, C. J., & Poole, D. A. (1997). Long-term survival of children's false memories: A review. *Learning and Individual Differences, 9*(2), 125–151. https://doi.org/10.1016/s1041-6080(97)90003-0

Brainerd, C. J., & Reyna, V. F. (1988). Memory loci of suggestibility development: Comment on Ceci, Ross & Toglia (1987). *Journal of Experimental Psychology: General, 117,* 197-200. https://doi.org/10.1037/0096-3445.117.2.197

Brainerd, C. J., & Reyna, V. F. (1998). Fuzzy-trace theory and children's false memories. *Journal of Experimental Child Psychology, 71*(2), 81–129. https://doi.org/10.1006/jecp.1998.2464

Brainerd, C. J., Reyna, V. F., Howe, M. L., & Kevershan, J. (1990). The last shall be first: How memory strength affects children's retrieval. *Psychological Science, 1*, 247–252. https://doi.org/10.1111/j.1467-9280.1990.tb00208.x.

Braun, V., & Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative research in psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2020). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology, 1*(1), 1-25. https://doi.org/10.1080/14780887.2020.1769238

Brigham, J. C., & Spier, S. A. (1992). *Opinions held by professionals who work with child witnesses*. In H. Dent & R. Flin (Eds.), *Children as witnesses* (pp. 93–111). Chichester, West Sussex, England: Wiley.

Brooks, J. A., & Freeman, J. B. (2019). Neuroimaging of person perception: A social-visual interface. *Neuroscience Letter, 693*(4), 40-43. https://doi.org/10.1016/j.neulet.2017.12.046

Brown, D. A., & Lamb, M. E. (2015). Can children be useful witnesses? It depends how they are questioned. *Child Developmental Perspectives, 9*(4), 250–255. https://doi.org/10.1111/cdep.12142

Brown, D. A., Lamb, M. E., Lewis, C., Pipe, M., Orbach, Y., & Wolfman, M. (2013). The NICHD investigative interview protocol: An analogue study. *Journal of Experimental Psychology: Applied, 19*, 367–382. https://doi.org/10.1037/a0035143

Brubacher, S. P., Roberts, K. P., & Powell, M. (2011). Effects of practicing episodic versus scripted recall on children's subsequent narratives of a repeated event. *Psychology, Public Policy, and Law, 17*(2), 286–314. https://doi.org/10.1037/a0022793

Bruck, M., & Ceci, S. J. (1997). The suggestibility of young children. *Current Directions in Psychological Science, 1,* 75-79. https://doi.org/10.1111/1467-8721.ep11512660.

Bruck, M., & Ceci, S. J. (1999). The suggestibility of children's memory. *Annual Review of Psychology*, *50*, 419–439. https://doi.org/10.1146/annurev.psych.50.1.419

Bruck, M., & Melnyk, L. (2004). Individual differences in children's suggestibility: A review and synthesis. *Applied Cognitive Psychology, 18*(8), 947-996. https://doi.org/10.1002/acp.1070

Buck, J. A., Warren, A. R., Bruck, M., & Kuehnle, K. (2014). How common is "common knowledge" about child witnesses among legal professionals? Comparing interviewers, public defenders, and forensic psychologists with laypeople. *Behavioral Sciences and the Law, 32,* 867-883. https://doi.org/10.1002/bsl.2150

Burger, H. (1971). *Die suggestive Beeinflußbarkeit von Aussagen über Beobachtungen. Entwicklung und erste Überprüfung eines Tests zur Aussagesuggestibilität (TAS) bei der Glaubwürdigkeitsbeurteilung kindlicher und jugendlicher Zeugen* [The suggestive malleability of statements about observations. Development and first validation of a Test of Statement Suggestibility (TSS) in the credibility assessment of child and juvenile witnesses]. [Doctoral dissertation, University of Freiburg].

Butler, B. J., & Loftus, E. F. (2017). Discrepancy detection in the retrieval-enhanced suggestibility paradigm. *Memory*, *26*, 483–492. https://doi.org/10.1080/09658211.2017.1371193.

Caffo, E., Rossi, L., Benatti, F., & Rigateli, M. (2016). *Suggestibility: Research and applications in the Forensic Field. Pilot study of the Bonn Test of Statement*

Suggestibility. [Doctoral Dissertation, University of Modena and Reggio Emilia, Italy].

Callejo, J. (2001). *El Grupo de Discusion: Introduccion a Una Pratica de Investigacion. [The Discussion Group: Introduction to An Investigation Practice].* Barcelona: Ariel.

Candel, I., Merckelbach, H., & Muris, P. (2000). Measuring interrogative suggestibility in children: Reliability and validity of the Bonn test of statement suggestibility. *Psychology, Crime & Law, 6*(1), 61-70. https://doi.org/10.1080/10683160008410832

Casey, B. J., Trainor, R. J., Orendi, J. L., Schubert, A. B., Nystrom, L. E., & Giedd, J. N. (1997). A developmental functional MRI study of prefrontal activation during performance of a Go-No-Go task. *Journal of Cognitive Neuroscience, 9,* 835–847. https://doi.org/10.1162/jocn.1997.9.6.835

Cashmore, J., & Bussey, K. (1996). Judicial perceptions of child witness competence. *Law and Human Behavior, 20*, 313–334. https://doi.org/10.1007/bf01499026

Cassel, W. S., & Bjorklund, D. F. (1995). Developmental patterns of eyewitness memory and suggestibility: An ecologically based short-term longitudinal study. *Law and Human Behaviour, 19*(5), 507. https://doi.org/10.1007/bf01499341

Cassel, W. S., Roebers, C. E. M., & Bjorklund, D. F. (1996). Developmental patterns of eyewitness responses to repeated and increasingly suggestive questions. *Journal of Experimental Child Psychology, 61,* 116-133. https://doi.org/10.1006/jecp.1996.0008

Cassidy, H., Akehurst, L., & Cherryman, J. (2020). Police interviewers' perceptions of child credibility in forensic investigations. *Psychiatry, Psychology and Law, 27*(1), 61-80. https://doi.org/10.1080/13218719.2019.1687044.

Cattell, R. B. (1966). *Handbook for the Culture Fair Intelligence Test, Scale 1*. Champaign, IL: IPAT.

Cavalluzzo, K, S., & Ittner, D. C. (2004). Implementing performance measurement innovations: Evidence from government. *Accounting, Organizations and Society, 29*(3–4), 243–67. https://doi.org/10.1016/s0361-3682(03)00013-8

Ceci, S. J., & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin, 113,* 403–439. https://doi.org/10.1037/0033-2909.113.3.403

Ceci, S. J., & Bruck, M. (1995). *Jeopardy in the courtroom.* Washington, D.C: American Psychological Association.

Ceci, S. J., Fitneva, S. A., & Williams, W. M. (2010). Representational constraints on the development of memory and metamemory: A developmental-representational theory. *Psychological Review*, *117*, 464–495. https://doi.org/10.1037/a0019067

Ceci, S. J., Papierno, P. B., & Kulkofsky, S. (2007). Representational constraints on children's suggestibility. *Psychological Science*, *18*, 503–509. https://doi.org/10.1111/j.1467-9280.2007.01930.x

Ceci, S. J., Ross, D. F., & Toglia, M. P. (1987). Suggestibility of children's memory: Psycholegal implications. *Journal of Experimental Psychology: General*, *116*, 38–49. https://doi.org/0.1037/0096-3445.116.1.38

Cederborg, A. C., Orbach, Y., Sternberg, K., & Lamb, M. (2000). Investigative interviews of child witnesses in Sweden. *Child Abuse & Neglect, 24*, 1355-1361. https://doi.org/10.1016/s0145-2134(00)00183-6

Chae, Y., Goodman, G. S., Eisen, M. L., & Qin, J. (2011). Event memory and suggestibility in abused and neglected children: Trauma-related psychopathology and cognitive functioning. *Journal of Experimental Child Psychology, 110*, 520-538. https://doi.org/10.1016/j.jecp.2011.05.006

Chan, A., & Douglas, A. (2005). Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *British Medical Journal, 330*(7494), 753-756. https://doi.org/10.1136/bmj.38356.424606.8f

Chan, J. C., Manley, K. D., & Lang, K. (2017). Retrieval-enhanced suggestibility: A retrospective and new investigation. *Journal of Applied Research in Memory and Cognition, 6,* 213–229. https://doi.org/10.1016/j.jarmac.2017.07.003

Chan, J. C. K., Thomas, A. K., & Bulevich, J. B. (2009). Recalling a witnessed event increases eyewitness suggestibility: The reversed testing effect. *Psychological Science, 20*(1), 66-73. https://doi.org/10.1111/j.1467-9280.2008.02245.x

Clarke, C., & Milne, R. (2001). Interviewing suspects of crime: The impact of PEACE training, supervision and the presence of a legal advisor. *Journal of Investigative Psychology and Offender Profiling, 8*(2), 149-162. https://doi.org/10.1002/jip.144

Clarke-Stewart, A. K., Mallow, L. C., & Allhusen, V. D. (2004). Verbal ability, self-control, and close relationships with parents protect children against misleading suggestions. *Applied Cognitive Psychology, 18,* 1037-1058. https://doi.org/10.1002/acp.1076

Cleary, H. M. D., & Warner, T. C. (2016). Police training in interviewing and interrogation methods: A comparison of techniques used with adult and juvenile suspects. *Law and Human Behavior, 40*(3), 270-284. https://doi.org/10.1037/lhb0000175

Cohen, R. L., & Harnick, M. A. (1980). The susceptibility of child witnesses to suggestion. *Law and Human Behaviour, 4*(3), 201-210. https://doi.org/10.1007/bf01040318

Colloff, M., Wade, K. A., Wixted, J., & Maylor, E. A. (2017). A Signal-Detection Analysis of Eyewitness Identification Across the Adult Lifespan. *Psychology and Aging, 32*(3). https://doi.org/ 10.1037/pag0000168.

Cornwall, A., & Jewkes, R. (1995). What is participatory research*? Social Science and Medicine, 41,* 1667-1676. https://doi.org/10.1016/0277-9536(95)00127-s

Costa, A., & Pinho, M. S. (2008). Sugestionabilidade interrogativa em crianças de 8 e 9 anos de idade. [Interrogative suggestibility in children aged 8 and 9 years]. *Analise Psicologica. 28*(1), 193-208. https://doi.org/10.14417/ap.266

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671-684. https://doi.org/10.1016/S0022-5371(72)80001-X

Cronbach, L. J. (1984). *Essentials of psychological testing* (4th ed.). New York, NY: Harper & Row. https://doi.org/10.1002/sce.3730350432

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.

Cuthbertson, P. (2018). Poverty and Crime: Why a new war on criminals would help the poor most. *Civitas,* 1-11. Retrieved from https://www.civitas.org.uk/content/files/povertyandcrime.pdf

Cutler, B. L., & Penrod, S. D. (1995). *Mistaken identification: The eyewitness, psychology, and the law.* New York: Cambridge University Press.

Dafflon, C. (2012). *Bonn Test of Statement Suggestibility (BTSS): efeitos das variáveis inteligência, memória, vinculação, autoconceito e desenvolvimento operatório em crianças dos 6, 7 e 8 anos.* [Bonn Test of Statement Suggestibility (BTSS): effects of the variables intelligence, memory, binding, self-concept and operative development in children aged 6, 7 and 8 years.]. [Doctoral dissertation, University of Coimbra].

Dalton, A. L., & Daneman, M. (2006). Social suggestibility to central and peripheral Misinformation, *Memory, 14*(4), 486-501. https://doi.org/0.1080/0965821050049507.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*, 17–22. https://doi.org/10.1037/h0046671

Desmarais, S. L., & Read, J. D. (2011). After 30 years, what do we know about what jurors know? A meta-analytic review of lay knowledge regarding eyewitness factors. *Law and Human Behavior, 35*(3), 200–210. https://doi.org/10.1007/s10979-010-9232-6

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, *7*(3), 177–188. https://doi.org/10.1016/0197-2456(86)90046-2

Doi, S. A. R., & Thalib, L. (2008). A quality-effects model for meta-analysis: *Epidemiology*, *19*(1), 94–100. https://doi.org/10.1097/EDE.0b013e31815c24e7

Dong, Z., Han, H., Li, H., Bai, Y., Wang, W., & Tu, M. (2015). Long-term potentiation decay and memory loss are mediated by AMPAR endocytosis. *Journal of Clinical Investigations, 125*, 234–247 https://doi.org/10.1172/jci77888

Doris, J. (1991). *The suggestibility of children's recollections: Implications for eyewitness testimony.* Washington, DC: American Psychological Association.

Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomized and non-randomized studies of health care interventions. *Journal of Epidemiology Community Health, 52,* 377-384. https://doi.org/10.1136/jech.52.6.377

Drizin, S. A. & Colgan, B. A. (2001). Let the cameras roll: mandatory videotaping of interrogations is the solution to Illinois' problem of false confessions, *Loyola University Chicago Law Journal, 32*(3), 337.

Duncan, E. M., Whitney, P., & Kunen, S. (1982). Integration of visual and verbal information in children's memories. *Child Development, 53*(5), 1215-1223. https://doi.org/10.2307/1129008

Eisen, M. L., Goodman, G. S., Qin, J., Davis, S., & Crayton, J. (2007). Maltreated children's memory: Accuracy, suggestibility, and psychopathology. *Developmental Psychology, 43*(6), 1275–1294. https://doi.org/10.1037/0012-1649.43.6.1275

Ellison, L., & Munro, V. (2014). A 'Special' delivery? Exploring the impact of screens, live-links and video-recorded evidence on mock juror deliberation in rape trails. *Social Legal Studies, 23*(1), 3-29. https://doi.org/10.1177/0964663913496676

Endres, J. (1997). The suggestibility of the child witness: The role of individual differences and their assessment. *Journal of Credibility Assessment and Witness Psychology,* 1, 44-67.

Endres, J., Poggenpohl, C., & Scholz, O. B. (1996). Preschool children's statement suggestibility: effects of memory trace strength and of warning against misleading

question [Paper presentation]. 6th European Conference, Psychology and Law 1996, Siena, Italy.

Endres, J., Scholz. O. B., & Summa, D. (1997). Statement suggestibility in children: Presentation of new assessment method and first results, in M. Stadler, T., Fabian, L., Greuel (a cura di), *Psychology of testimony,* Munchen, PVU.

Engelhard, I. M., Van Den Hout, M. A., & McNally, R. J. (2008). Memory consistency for traumatic events in Dutch soldiers deployed to Iraq. *Memory, 16,* 3-9. https://doi.org/10.1080/09658210701334022

Evans, A. D., & Lee, K. (2013). Emergence of lying in very young children. *Developmental Psychology, 49*, 1958–1963. https://doi.org/10.1037/ a0031409

Evans, A. D., & Lyon, T. D. (2012). Assessing children's competency to take the oath in court: The influence of question type on children's accuracy. *Law and Human Behavior, 36*, 195–205. https://doi.org/10.1037/h0093957.

Everson, M. D., Boat, B. W., Sherries, B., & Robertson, K. R. (1996). Beliefs among professionals about rates of false allegation of child sexual abuse. *Journal of Interpersonal Violence, 4*, 541-55. https://doi.org/10.1177/088626096011004006

Finila, K., Mahlberg, N., Santtila, P., Sandnabba, K., & Niemi, P. (2003). Validity of a test of children's suggestibility for predicting responses to two interview situations differencing in their degree of suggestiveness. *Journal of Experimental Child Psychology, 85,* 32-49. https://doi.org/10.1016/s0022-0965(03)00025-0

Fisher, R. P., McCauley, M. R., & Geiselman, R. E. (1992). *Improving Eyewitness Testimony with the Cognitive Interview*. In D. Ross, J. D. Read and M. Toglia (Eds).

Adult Eyewitness Testimony: Current Trends and Developments. New York: Cambridge University Press.

Flin, R., Boon, J., Knox, A., & Bull, R. (1992). The effect of a five-month delay on children's and adults' eyewitness memory. *British Journal of Psychology, 83,* 323-336. https://doi.org/10.1111/j.2044-8295.1992.tb02444.x

Frenda, A., Knowles, E., William, A., & Loftus, E. (2013). False Memories of Fabricated Political Events. *Journal of Experimental Social Psychology, 49(2),* 280-286. https://doi.org/10.1016/j.jesp.2012.10.013

Geddie, L., Fradin, S., & Beer, J. (2000). Child characteristics which impact accuracy of recall and suggestibility in pre-schoolers: is age the best predictor? *Child Abuse & Neglect, 24*, 223–235. https://doi.org/10.1016/s0145-2134(99)00133-7

Gignac, G. E., & Powell, M. B. (2006). A direct examination of the nonlinear (Quadratic) association between intelligence and suggestibility in children. *Applied Cognitive Psychology, 20*, 617-623. https://doi.org/10.1002/acp.1213

Gilstrap, L. L., & Ceci, S. J. (2005). Reconceptualizing children's suggestibility: Bidirectional and temporal properties. *Child Development, 76*(1), 40–53. https://doi.org/10.1111/j.1467- 8624.2005.00828.x.

Gilstrap, L. L. & Papierno, P. B. (2004). Is the cart pushing the horse? The effects of child characteristics on children's and adult's interview behaviours. *Applied Cognitive Psychology, 18*(8), 1059-1078. https://doi.org/10.1002/acp.1072

Goodman, G., & Melinder, A. (2010). Child witness research and forensic interviews of young children: A review. *Legal and Criminological Psychology, 12,* 1-19. https://doi.org/10.1348/135532506X156620.

Goodman, G. S., Quas, J. A., Bulkley, J., & Shapiro, C. (1999). Innovations for child witnesses: A national survey. *Psychology, Public Policy, and the Law*, *5*, 255–281. https://doi.org/10.1037/1076-8971.5.2.255

Goodman, G. S., & Reed, R. (1986). Age differences in eyewitness testimony. *Law and Human Behavior, 10*, 317–332. https://doi.org/10.1007/bf01047344

Goodman-Delahunty, J., Cossins, A., & O'Brien, K. (2010). Enhancing the credibility of complainants in child sexual assault trials: the effect of expert evidence and judicial directions. *Behavioral Sciences & the Law, 28*(6), 769–783. https://doi.org/10.1002/bsl.936

Gordon, B. N., Baker-Ward, L. E., & Ornsttein, P. A. (2001). Children's Testimony: A review of research on memory for past experiences. *Clinical Child and Family Psychology Review, 4*(2), 157-181. https://doi.org/10.1023/A:1011333231621

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology, Learning, Memory and Cognition, 11*, 501–518. https://doi.org/10.1037/0278-7393.11.3.501

Granhag, P. A., & Strömwall, L. A. (2004). *The detection of deception in forensic contexts*. Cambridge: Cambridge University Press.

Granhag, A. P., Strömwall, L, A., & Hartiwg, M. (2005). Eyewitness testimony: Tracing the beliefs of Swedish legal professionals. *Behavioral Sciences and the Law, 23,* 709-727. https://doi.org/10.1002/bsl.670

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. (1992). *The psychology of interrogations, confessions and testimony*. Chichester: Wiley. https://doi.org/10.1080/00029157.1994.10403080

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

Gudjonsson, G. H., & MacKeith, J. A. C. (1982). False confessions, psychological effects of interrogation. A discussion paper. In A. Trankell, *Reconstructing the past: The role of psychologists in criminal trials* (pp. 253–269). Deventer: Kluwer.

Hedges, L. V. (1981). Distribution theory for glass's estimator of effect size and related estimators. *Journal of Educational and Behavioural Statistics, 6*(2), 107-128. https://doi.org/10.3102/10769986006002107

Henderson, H. & Andrews, S. J. (2017). Assessing the veracity of children's forensic interviews. In *Finding the Truth in the Courtroom*. New York: John Wiley & Sons, Ltd. https://doi.org/10.1093/oso/9780190612016.003.0006

Hershkowitz, I. (2001). Children’s responses to open-ended questions: A field study of forensic interviews. *Legal and Criminological Psychology, 6,* 49–63. https://doi.org/10.1348/135532501168190

Higgins, J. P. T., Altman, D. G., Gotzsche, P. C., Juni, P., Moher, D., Oxman, A. D., … Cochrane Statistical Methods Group. (2011). The Cochrane Collaboration’s tool for

assessing risk of bias in randomised trials. *BMJ, 343*, 5928–5928. https://doi.org/10.1136/bmj.d5928

Higgins, J. P. T., & Thompson, S. G (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21,* 1539-1558. https://doi.org/10.1002/sim.1186

HM Government. (2018). *Victim Strategy.* Cm-9700. Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/746930/victim-strategy.pdf

HMICFRS. (2016). *Rape Monitoring Group: Digests and Data, 2015/16*. London. HMICFRS.

Hoffman, L. (1990). Constructing Realities: An Art of Lenses. *Family Process, 29*(1), 1-12. https://doi.org/10.1111/j.1545-5300.1990.00001.x

Home Office. (2001) *Achieving best evidence in criminal proceedings: Guidance for vulnerable or intimidated witnesses, including children.* London: Home Office.

Howe, M. L. (2008). Visual distinctiveness and the development of children's false memories. *Child Development*, *79*, 65–79. https://doi.org/10.1111/j.1467-8624.2007.01111.x

Howe, M. L., & Wilkinson, S. (2011). Using story contexts to bias children's true and false memories. *Journal of Experimental Child Psychology*, *108*, 77–95. https://doi.org/10.1016/j.jecp.2010.06.009

Howe, M. L., Wimmer, M. C., Gagnon, N., & Plumpton, S. (2009). An associative-activation theory of children's and adults' memory illusions. *Journal of Memory and Language*, *60*, 229–251. https://doi.org/10.1080/09658210802438474

Hunt, J. S., & Borgida, E. (2001). Is that what I said? Witnesses' responses to interviewer modifications. *Law and Human Behavior, 25*(6), 583-603. https://doi.org/10.1023/a:1012754207074

Johnson, M. K. (1988). Discriminating the origin of information. In: T.F. Oltmanns, & B. A. Maher (Eds.), *Delusional beliefs* (pp.34-65). New York: Wiley.

Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the "general acceptance" of eyewitness testimony research: A new survey of the experts. *American Psychologist, 56,* 405–416. https://doi.org/10.1037/0003- 066X.56.5.405

Kim, I, K., Kwon, E. S., & Ceci, S. J. (2017). Developmental reversals in report conformity: Psycho-legal implications. *Applied Cognitive Psychology, 31,* 128–138. https://doi.org/10.1002/acp.3309

Kim, S. Y., Park, J. E., Lee, Y. J., Seo, H. J., Sheen, S. S., Hahn, S., Son, H. J. (2013). Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. *Journal of Clinical Epidemiology, 66*(4), 408–414. https://doi.org/10.1016/j.jclinepi.2012.09.016

Kimberley, M., & Govern, J. (2004). Conformity to misinformation and time delay negatively affect eyewitness confidence and accuracy. *North American Journal of Psychology, 6*(2), 227-238.

King, M. A., & Yuille, J. C. (1987). Suggestibility and the child witness. In S. J. Ceci, M. Toglia, & D. Ross (Eds.), *Children's eyewitness memory* (pp. 24-35). New York: Springer-Verlag. https://doi.org/10.1007/978-1-4684-6338-5_2

Klemfuss, J. Z. (2015). Differential contributions of language skills to children's episodic recall. *Journal of Cognition and Development, 16*(4), 608–620. https://doi.org/10.1080/15248372.2014.952415.

Klemfuss, J. Z., & Olaguez, A. P. (2020). Individual Differences in Children's Suggestibility: An Updated Review. *Journal of child sexual abuse*, *29*(2), 158–182. https://doi.org/10.1080/10538712.2018.1508108

Kline, R. B. (1998). *Principles and practice of structural equation modelling*. New York, NY: The Guilford Press.

Knutsson, J., & Allwood, C. M. (2014). Opinions of legal professionals: Comparing child and adult witnesses' memory report capabilities. *The European Journal of Psychology Applied to Legal Context, 6,* 79–89. https://doi.org/10.1016/j.ejpal.2014.06.001

Kroll, A. (2015). Drivers of performance information use: Systematic literature review and directions for future research. *Public Performance and Management Review, 38*(3), 459-486. https://doi.org/10.1080/15309576.2015.1006469 https://doi.org/10.1080/15309576.2015.1006469

Krueger, R. A., & Casey, M. A. (2014). *Focus Groups, A practical guide for applied research.* Sage Publications, Inc: USA.

Kulkofsky, S., & Klemfuss, J. Z. (2008). What the stories children tell us can tell about their memory: Narrative skill and young children suggestibility. *Developmental Psychology, 44,* 1442-1456. https://doi.org/10.1037/a0012849

Lamb, M. E., Hershowitz, I., Orbach, Y., Esplin, P. W. (2008). *Tell me what happened: structured investigative interviews of child victims and witnesses*. England: Wiley-Blackwell.

Lamb, M. E., Malloy, L. C., Hershkowitz, I., & La Rooy, D. (2015). Children and the law. In *Handbook of Child Psychology and Developmental Science.* Hoboken, NJ: Wiley. https://doi.org/10.1002/9781118963418.childpsy312

Lampinen, J. M., & Smith, V. L. (1995). The incredible (and sometimes incredulous) child witness: Child eyewitnesses' sensitivity to source credibility cues. *Journal of Applied Psychology, 80*(5), 621–627. https://doi.org/10.1037/0021-9010.80.5.621

Landis, R. J., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174. https://doi.org/10.2307/2529310

Laumann, L. A., & Elliot, R. (1992). Reporting what you have seen: Effect associated with age and mode of questioning on eyewitness reports. *Perceptual and Motor Skills, 75,* 799-818. https://doi.org/10.2466/pms.1992.75.3.799

Lindsay, D. S. (1990). Misleading suggestions can impair eyewitnesses' ability to remember event details. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 1077-1083. doi.org/10.1037/0278-7393.16.6.1077.

Lindsay, D. S., Johnson, M. K., & Kwon, P. (1991). Developmental changes in memory source monitoring. *Journal of Experimental Child Psychology, 52*(3), 297–318. https://doi.org/10.1016/0022-0965(91)90065-z

Lindholm, J., & Cederborg, A. C. (2016). Legal Assessment of child victims of human trafficking for sexual purposes. *Behavioural Sciences and the Law, 34*(1), 218-233. https://doi.org/10.1002/bsl.2228

Loftus, E. (1975). Leading questions and the eyewitness report. *Cognitive Psychology, 7,* 560-572.

Loftus, E. (1979). *Eyewitness Testimony.* Cambridge, MA: Harvard University Press.

Loftus, E. (1994). The repressed memory controversy. *American Psychologist, 49*(5), 443–445. https://doi.org/10.1037/0003-066X.49.5.443.b

Loftus, E., 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*, 19-31. https://doi.org/10.1037/0278-7393.4.1.19.

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

London, K., & Ceci, S. J. (2012). Competence, credibility, and reliability of children's forensic reports: Introduction to special issue on child witness research. *Developmental Review, 32*(3), 161-164. https://doi.org/10.1016/j.dr.2012.06.001

Lyons, K. A., Ghetti, S., & Cornoldi, C. (2010). Age differences in the contribution to recollection and familiarity to false- memory formation: A new paradigm to examine developmental reversals. *Developmental Science*, *13*, 355–362. https://doi.org/10.1111/j.1467-7687.2009.00889.x

Malloy, L. C., Mitchell, E., Block, S., Quas, J. A., & Goodman, G. S. (2007). Children's eyewitness memory: Balancing children's needs and defendants' rights when seeking the truth. In Toglia, M. P., Read, J. D., Ross, D., & Lindsay, R. C. L. (Eds.), *The handbook of eyewitness psychology, Vol 1: Memory for events* (pp. 545-574). Mahwah, NJ: Erlbaum.

McAuliff, B. D., & Kovera, M. B. (2007). Estimating the effects of misleading information on witness accuracy: Can experts tell jurors something they don't already know? *Applied Cognitive Psychology, 21*(7), 849–870. https://doi.org/10.1002/acp.1301.

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,* 1-16. https://doi.org/10.1037/0096-3445.114.1.1

Melinder, A., Goodman, G. S., Eilertsen, D. E & Magnussen, S. (2004). Beliefs about child witnesses: A survey of professionals. *Crime & Law, 10*(4), 347-365. https://doi.org/10.1080/10683160310001618717

Memon, A., Holliday, R., & Hill, C. (2006). Pre-event stereotypes and Misinformation Effect in young children, *Memory, 14*(1), 104-114. https://doi.org/10.1080/09658210500152641

Memon, A., Meissner, C. A., & Fraser, J. (2010). The cognitive interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy, and Law, 16*, 340–372. https://doi.org/10.1037/a0020518

Memon, A., Vrij, A., & Bull, R. (2003). *Psychology and law: Truthfulness, accuracy and credibility.* Wiley: Chichester. https://doi.org/10.1002/acp.1017

Merckelbach, H., Muris, P., Wessel, I., & van Koppen, P. J. (1998). The Gudjonsson Suggestibility Scale (GSS): Further data on its reliability, validity, and metacognition correlates. *Social Behavior and Personality*, *26*(2), 203–210. https://doi.org/10.2224/sbp.1998.26.2.203.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Mukherjee, N., Huge, J., Sutherland, W. J., McNeill, J., Van Opstal, M, Dahdouh-Guebas, F., & Koedam, N. (2015). The Delphi technique in ecology and biological conservation: applications and guidelines. *Methods in Ecology and Evolution, 6*(9) 1097-1109. https://doi.org/10.1111/2041-210X.12387

Moscovici, S. (1998). Social consciousness and its history. *Culture & Psychology, 4*, 449-467.

Najdowski, C., & Bottoms, B. (2015). Effects of jurors' gender and attitudes toward intellectual disability on judgments in cases involving disabled juvenile defendants. *Journal of Forensic Psychiatry and Psychology, 26*(3), 407-424. https://doi.org/10.1080/14789949.2015.1017592

National Research Council. (2014). *Identifying the Culprit: Assessing Eyewitness Identification.* Washington, DC: The National Academies Press.

Newcombe, P. A., & Bransgrove, J. (2011). Perceptions of witness credibility: Variations across age. *Journal of Applied Developmental Psychology, 28*, 318–331. https://doi.org/10.1016/j.appdev.2007.04.003

Nicolson, D., Auchie, D., Duff, P., & Ferguson, P. (2017). *Assessing witness credibility and reliability: engaging experts and disengaging gage?* In *Scottish Criminal Evidence Law*. Edinburgh University Press, Edinburgh.

Norris, M., & Lecavalier, L. (2010). Evaluating the use of exploratory factor analysis in developmental disability psychological research. *Journal of autism and developmental disorders, 40*(1), 8–20. https://doi.org/10.1007/s10803-009-0816-2

Oates, K., & Shrimpton, S. (1991). Children's memories for stressful and non-stressful events. *Medicine, Science, Law, 31*(1), 4-10. https://doi.org/10.1177/002580249103100102

Oeberst, A., & Blank, H. (2012). Undoing suggestive influence on memory: the reversibility of the eyewitness misinformation effect. *Cognition*, *125*(2), 141–159. https://doi.org/10.1016/j.cognition.2012.07.009

Office for National Statistics. (2016). *Focus on violent crimes and sexual offences: Year ending March 2015*. London: Office for National Statistics.

Office for National Statistics. (2018). *Crime in England and Wales: year ending March '18*. London: Office for National Statistics.

Ornstein, P. A., Gordon, B. N., & Larus, D. M. (1992). Children's memory for a personally experienced event: Implications for testimony. *Applied Cognitive Psychology, 6*(1), 49–60. https://doi.org/10.1002/acp.2350060103

Otgaar, H., & Candel, I. (2009). Children's false memories: different false memory paradigms reveal different results. *Psychology, Crime and Law, 17*(6), 513-528. https://doi.org/10.1080/10683160903373392

Otgaar, H., De Ruiter, C., Howe, M. L., Hoetmer, L., & van Reekum, P. (2017). A case concerning children's false memories of abuse: Recommendations regarding expert witness work. *Psychiatry*, *Psychology and Law, 24*(3), 365-378. https://doi.org/10.1080/13218719.2016.1230924.

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*, 31–55. https://doi.org/10.1037/xge0000127

Otgaar, H., Howe, M. L., Brackmann, N., & van Helvoort, D. (2017). Eliminating age differences in children's and adults' suggestibility and memory conformity effects. *Developmental Psychology, 53,* 962–970. https://doi.org/10.1037/dev0000298

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

Payne, D. G., Toglia, M. P., & Anastasi, J. S. (1994). Recognition performance level and the magnitude of the misinformation effect in eyewitness memory. *Psychonomic Bulletin & Review, 1*(3), 376–382. https://doi.org/10.3758/BF03213978

Peterson, C., Morris, G., Baker-Ward, L., & Flynn, S. (2014). Predicting which childhood memories persist contributions of memory characteristics. *Developmental psychology*, *50*(2), 439–448. https://doi.org/10.1037/a0033221

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*(3), 193–198. https://doi.org/10.1037/h0049234

Pezdek, K., & Roe, C. (1995). The effect of memory trace strength on suggestibility. *Journal of Experimental Child Psychology, 60*(1), 116–128. https://doi.org/10.1006/jecp.1995.1034.

Pfeil, K. (2018). The effectiveness of the Self-Administered Interview—A meta-analytic review and empirical study with older adult witnesses. [Doctoral Dissertation, The University of Cambridge]. https://doi.org/10.17863/CAM.22056

Pipe, M. E., Lamb, M. E., Orbach, Y., & Cederborg, A. C. (2007). *Child sexual abuse: Disclosure, delay, and denial.* Mahwah: Lawrence Erlbaum Associates Publishers.

Pitman, R. K. (1989). Post-traumatic stress disorder, hormones, and memory. *Biological Psychiatry, 26*(3), 221-223. https://doi.org/10.1016/0006-3223(89)90033-4

Plotnikoff, J., & Woolfson, R. (2011) *Young witnesses in criminal proceedings: a progress report on Measuring up*? London: Nuffield Foundation.

Plotnikoff, J., & Woolfson, R. (2019) *Falling short? A snapshot of young witness policy and practice.* London: NSPCC.

Polczyk, R. (2007). *The mechanisms of the misinformation effect in the context of eyewitness testimony.* Cracow: WUJ.

Poole, D. A. (2016). *Interviewing children: The science of conversation in forensic contexts*. Washington, DC: American Psychological Association. https://doi.org/10.1037/14941-000

Poole, D. A., & Lamb, M. E. (1998). *Investigative interviews of children: A guide for helping professionals*. Washington, DC: American Psychological Association. https://doi.org/10.1177/109861101129197770

Poole, D. A., & Lindsay, D. S. (2002). Reducing child witnesses' false reports of Misinformation from parents. *Journal of Experimental Child Psychology, 81,* 117–140. https://doi.org/10.1006/jecp.2001.2648

Poole, D. A., & White, T. L. (1991). Effects of question repetition on the eyewitness testimony of children and adults. *Developmental Psychology, 27*(6), 975-986. https://doi.org/10.1037/0012-1649.27.6.975

Polak, A., & Harris, P. L. (1999). Deception by young children following noncompliance. *Developmental Psychology, 35,* 561-568.

Porter, S. & Brinke, L. (2009). Dangerous decisions: A theoretical framework for understanding how judges assess credibility in the courtroom. *Legal Criminological Psychology, 14,* 119-134. https://doi.org/10.1348/135532508X281520

Powell, M. B., & Roberts, K. P. (2002). The effect of repeated experience on children's suggestibility across two question types. *Applied Cognitive Psychology*, *16*, 367 – 386. https://doi.org/10.1002/acp.801

Principe, G. F., Ornstein, P. A., Baker-Ward, L., & Gordon, B. N. (2000). The effects of intervening experiences on children's memory for a physical examination. *Applied Cognitive Psychology, 14*(1), 59–80. https://doi.org/10.1002/(SICI)1099-0720(200001)14:1<59::AID-ACP637>3.0.CO;2-4

Quas, J. A., Rush, E. B., Yim, I. S., & Nikolayev, M. (2014). Effects of stress on memory in children and adolescents: Testing causal connections. *Memory, 22(*6), 616–632. https://doi.org/10.1080/09658211.2013.809766

Quas, J. A., Thompson, W. C., & Clarke-Stewart, K. A. (2005). Do jurors "know" what isn't so about child witnesses? *Law and Human Behavior*, *29*, 425–456. https://doi.org/10.1007/s10979-005-5523-8

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, *7*, 1–75. https://doi.org/10.1016/1041-6080(95)90031-4.

Reyna, V. F., Corbin, J. C., Weldon, R. B., & Brainerd, C. J. (2016). How Fuzzy-Trace theory predicts true and false memories for words, sentences, and narratives. *Journal of Applied Research Memory and Cognition, 1*(5), 1-9. https://doi.org/10.1016/j.jarmac.2015.12.003

Reyna, V. F., Holliday, R., & Marche, T. (2002). Explaining the development of false memories. *Developmental Review*, *22*, 436-489. https://doi.org/10.1016/s0273-2297(02)00003-5

Reyna, V. F., Lloyd, F. J., & Brainerd, C. J. (2004). How people make decisions that involve risk: A dual process approach. *Current Directions in Psychological* Science, *13*, 60–66. https://doi.org/10.1111/j.0963-7214.2004.00275.x

Richardson, E., Stokoe, E., & Antaki, C. (2018). Establishing intellectually impaired victims, understanding about 'truth' and 'lies': Police interview guidance and practice in cases of sexual assault. *Applied Linguistics, 40*(5), 773-792. https://doi.org/10.1093/applin/amy023.

Roberts, W. T., & Higham, P. A. (2002). Selecting accurate statements from the cognitive interview using confidence ratings. *Journal of Experimental Psychology: Applied, 8,* 33–43. https://doi.org/10.1037/1076-898x.8.1.33.

Roberts, K. P., Lamb, M. E., & Sternberg, K. J. (1999). Effects of the timing of postevent information on preschoolers' memories of an event. *Applied Cognitive Psychology, 13,* 541 – 560. https://doi.org/10.1002/(sici)1099-0720(199912)13:6<541::aid-acp618>3.0.co;2-5

Roberts, K. P., & Powell, M. B. (2001). Describing individual incidents of sexual abuse: a review of research on the effects of multiple sources of information on children's reports. *Child Abuse and Neglect, 25*, 1643–1659. https://doi.org/10.1016/s0145-2134(01)00290-3.

Roberts, K. R., & 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

Rogers, P., Lowe, M., & Reddington, K. (2016). Investigating the victim pseudomaturity effect: How a victim's chronological age and dress style influences attributions in a depicted case of child sexual assault. *Journal of child sexual abuse, 25,* 1-19. https://doi.org/10.1080/10538712.2016.1111964.

Robinson, J., & Briggs, P. (1997). Age trends in eyewitness suggestibility and compliance. *Psychology*, *Crime and Law, 3*(3), 187-202. https://doi.org/10.1080/10683169708410812

Roebers, C. M., Bjorklund, D. F., Schneider., & Cassel, W. S. (2002). Differences and similarities in event recall and suggestibility between children and adults in Germany and the United States. *Experimental Psychology, 49*(2), 132-140. https://doi.org/10.1027//1618-3169.49.2.132

Roebers, C. M., & Fernandez, O. (2002). The effects of accuracy motivation on children's and adults' event recall, suggestibility, and their answers to unanswerable questions. *Journal of Cognition and Development, 3*(4), 415-443. https://doi.org/10.1080/15248372.2002.9669676

Roebers, C. M., & McConkey, K. M. (2003). Mental reinstatement of the misinformation context and the misinformation effect in children and adults. *Applied Cognitive Psychology, 17,* 477-493. https://doi.org/10.1002/acp.886

Roebers, C. M., & Schneider, W. (2000). The impact of misleading questions on eyewitness memory in children and adults. *Applied Cognitive Psychology*, *14*, 509- 526. https://doi.org/10.1002/1099-0720(200011/12)

Roebers, C. M., & Schneider, W. (2001). Memory for an observed event in the presence of prior Misinformation: Developmental patterns of free recall and identification accuracy. *British Journal of Developmental Psychology, 19,* 507-524. https://doi.org/10.1348/026151001166227

Roebers, C. M., & Schneider, W. (2005). The strategic regulation of children's memory performance and suggestibility. *Journal of Experimental Child Psychology, 91*, 24-44. https://doi.org/10.1016/j.jecp.2005.01.001

Roediger, H. L., & McDermott, K. (1995). Creating false memories: Remembering words not presented in a list. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Roma, P., Sabetello, 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.

Rossi, L., Benatti, F., Oppo, A., Pesce, S., & Avato, F. M. (2011). Suggestionbilita interrogative: Proprieta psicometriche della versione Italiana del Bonn Test of

Statement Suggestibility (BTSS). [Suggestible Questions: Psychometric properties of the Italian version of the Bonn Test of Statement Suggestibility]. *Psichiatria dell'infanzia e dell'adolescenza, 78*(3), 684-696.

Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at 3-7 years: The Children's Behavior Questionnaire, *Child Development*, *72*, 1394–1408. https://doi.org/10.1111/1467-8624.00355

Saywitz, K. J., & Goodman, G. S. (1996). Interviewing children in and out of court: Current research to aid practice implications. In: Briere, J., Berliner, L., Bulkley, J. A., Jenny, C., Reid, T., editors. *The APSAC handbook on child maltreatment*. Thousand Oaks, CA: Sage; 297–318.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualization of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207 – 217. https://doi.org/10.1111/j.1467-9280.1992.tb00029.x

Schollum, M. (2017). Bringing PEACE to the United States: A framework for investigative interviewing. *The Police Chief,* 30-37.

Segovia, D., Strange, D., & Takarangi, M. (2017). Trauma memories on trial: Is cross-examination a safeguard against analogue traumatic memories? *Memory*, *25*(1), 95-106. https://doi.org/10.1080/09658211.2015.1126608

Sharman, S. J., & Powell, M. B. (2012). A comparison of adult witnesses' suggestibility across various types of leading questions. *Applied Cognitive Psychology, 26*(1), 48–53. https://doi.org/10.1002/acp.1793

Skagerberg, E. M., & Wright, D. B. (2009). Susceptibility to post identification feedback is affected by source credibility. *Applied Cognitive Psychology, 23*(4), 506-523. https://doi.org/10.1002/acp.1470

Skinner, G. M., Andrews, S. J., & Lamb, M. E. (2018). The disclosure of alleged child sexual abuse: an investigation of criminal court transcripts in Scotland. *Psychology, Crime, & Law, 25*(5), 458-481. https://doi.org/10.1080/1068316X.2018.1538415

Steller, M., & Keohnken, G. (1989*). Content based statement analysis.* In D. Raskin (Ed.), *Psychological Methods for Criminal Investigation and Evidence.* Springer: New York.

Strange, D., & Takarangi, M. K. T. (2015). Memory distortion for traumatic events: The role of mental imagery. *Front Psychiatry, 6*(27), 1-4. https://doi.org/10.3389/fpsyt.2015.00027

Swire, B., Ecker, U. K. H., & Lewandowsky, S. (2017). The role of familiarity in correcting inaccurate information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43,* 1948-1961. https://doi.org/10.1037/xlm0000422

Talwar, V., & Crossman, A. (2011). From little white lies to filthy liars: The evolution of honesty and deception in young children. *Advances in Child Development and Behavior, 40,* 139-179.

Talwar, V., & Lee, K. (2002). Development of lying to conceal a transgression: Children's control of expressive behavior during verbal deception. *International Journal of Behavioral Development, 26,* 436-444.

Templeton, L. M., & Wilcox, S. A. (2000). A tale of two representations: The Misinformation Effect and children's developing theory of mind. *Child Development, 71*(2), 402-416. https://doi.org/10.1111/1467-8624.00153

Thierry, K. L., Spence, M. J., & Memon, A. (2001). Before misinformation is encountered: Source monitoring decreases child witness suggestibility. *Journal of Cognition and Development, 2*, 1–26. https://doi.org/10.1207/s15327647jcd0201_1.

Thomas, A. K., Gordon, L. T., Cernasoz, P. M., & Bulevich, J. B. (2017). The effect of testing can increase or decrease Misinformation susceptibility depending on the retention interval. *Cognitive Research: Principles and Implications, 2*(1), 1-10. https://doi.org/10.1186/s41235-017-0081-4

Umanath, S., Ries, F., & Huff, M. J. (2018). Reducing suggestibility to additive verses contradictory Misinformation in younger and older adults via attention and/or explicit error detection. *Wiley Online Library, 33,* 793-805. https://doi.org/10.1002/acp.3522

Underwood, B. J., & Postman, L. (1960). Extra experimental sources of interference in forgetting. *Psychological Review, 67*(2), 73–95. https://doi.org/10.1037/h0041865

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.

Vandermass, M. O., Hess, T. M., & Baker-Ward, L. (1993). Does anxiety affect children's reports of memory for a stressful event? *Applied Cognitive Psychology, 7*, 109–128. https://doi.org/10.1002/acp.2350070204

Wade, K. A., Garry, M., Read, D. J., & Lindsay, S. D. (2002). A picture is worth a thousand lies: Using false photographs to create false childhood memories. *Psychonomic Bulletin and review, 9,* 597-603.

Warren, A. R., Hulise-Trotter, K., & Tubbs, E. C. (1991). Inducting resistance to suggestibility in children. *Law and Human Behaviou*r, *15*, 273-285. https://doi.org/10.1002/acp.2350070204

Warren, K. L., & Peterson, C. (2014). Exploring parent-child discussion of crime and their influence on children's memory. *Behavioural Sciences and the Law, 32,* 686-701. https://doi.org/10.1002/bsl.2144

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior*, *22,* 603–647. https://doi.org/10.1023/A:1025750605807

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103–128. https://doi.org/ 10.1016/0010-0277(83)90004-5

Windschitl, P. D. (1996). Memory for faces: evidence of retrieval-based impairment. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22*(3), 1101-1122. https://doi.org/10.1037/0278-7393.22.5.1101

Winsor, A., Flowe, H. D., Seale-Carlisle, T. M., Killeen, I. M., Hett, D., Jores, T., Ingham, M., Lee, B., Stevens, L., & Colloff, M. F. (in press). Children's expressions of certainty are informative. *Journal of Experimental Psychology: General*.

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,* 10–65. https://doi.org/10.1177/1529100616686966

Wylie, L. E., Davis., G. D., Patihis, L., Brank, E. M., Loftus. E., & Borenstein, M. (2014). Misinformation effect in older verses younger adults: A meta-analysis and review. In Toglia, M. P., Ross, D. F., Pozzulo, D., & Pica, D. (Eds) *The Elderly Eyewitness in Court.* UK: Psychology Press.

Yang, K., & Hsieh, Y. (2006). Managerial effectiveness of government performance measurement: Testing a middle-range model. *Public Administration Review, 67*(5), 861-879. https://doi.org/10.1111/j.1540-6210.2007.00774.x

Young, A. W., & Bruce, V. (2011). Understanding Person Perception. *British Journal of Psychology, 102*(4), 959-974. https://doi.org/10.1111/j.2044-8295.2011.02045.x

Yuille, J. C. (1993). We must study forensic eyewitnesses to know about them. *American Psychologist, 48*(5), 572–573. https://doi.org/10.1037/0003-066x.48.5.572

Zajac, R., & Hayne, H. (2003). I don't think that's what really happened: The effect of cross-examination on the accuracy of children's reports. *Journal of Experimental Psychology, 9,* 187-195. https://doi.org/10.1037/1076-898X.9.3.187

Zimmermann, W. (1979). Zu einigen Problemen und Ergebnissen der Suggestibilitätsuntersuchung im Rahmen der Glaubwürdigkeitsdiagnostik (Entwicklung und erste Standardisierung eines Testverfahrens) [On some problems and results of suggestibility assessment in the context of credibility evaluation (development and first standardization of a test instrument)]. *Kriminalistik und forensische Wissenschaften*, *37*, 25-58.

Zaragoza, M. S. (1987). Memory, suggestibility and eyewitness testimony in children and adults. In Ceci, S. J., Toglia, M. P., & Ross, D. F. (Eds.), *Children's eyewitness memory* (pp. 53-78). New York: Springer-Verla. https://doi.org/10.1007/978-1-4684-6338-5.

Zaragoza, M. S., & Lane, S. M. (1994). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(4), 934–945. https://doi.org/10.1037/0278-7393.20.4.934

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), 36–44. https://doi.org/10.1037/0278-7393.13.1.36

Appendices

Appendix A

Meta-analytical papers removed from second analysis

Table 1

Table outlining the explanation of why papers included in the misinformation item analysis were excluded from the control – misinformation item analysis.

Study	Explanation
Ackil & Zaragoza (1995)	This study compared performance in the misinformation and control conditions and found a misinformation effect in each age group (First grade, Third grade, Fifth grade and college). However, this study was not included into the second analysis (control-misinformation analysis) as the relevant *F* and *SD* statistics were not available.
Cassel, Roebers & Bjorklund (1996)	Participants were asked unbiased-leading questions (e.g., "Can you tell me who owned the bike"; i.e., no information), positive leading questions (i.e., "The bike was the girls, wasn't it?"; i.e., information that is correct) and misleading questions (i.e., "The mother owned the bike, didn't she"; i.e., information that is incorrect). The 'unbiased-leading' condition can be considered to be the control condition. However, the *F* statistic reported was concerned with comparing all three conditions rather than just the control and the misinformation. Moreover, the *SD*s were not reported.
Goodman & Reed (1986)	Participants were asked about a live event they experienced. This study did include a control condition; however, the study does not present the *F* statistics necessary for children's groups (control vs experimental) and adult group (control vs experimental) and therefore did not have the information that was needed.

Laumann & Elliott (1992)	Participants were asked to free recall, and then non-leading cued-recall questions and mildly misleading cued-recall questions. The relevant means and *SD*'s for the non-leading cued-recall questions were reported, but the means and *SD*'s for the misleading questions were not.
Robinson & Briggs (1997)	Participants were asked a series of questions based on a video clip, five of these being misleading questions. The study was interested in the role of the 'be complete' mnemonic, so the control items were in relation to that. Therefore, this paper did not present the *F* statistics or the *SD* statistics necessary to compare control to misinformation in the adult and the child groups.
Roebers & McConkey (2003)	This study did include a control condition, but it was not intended as a control for misleading questions, but rather as a control for another variable examined in the study (i.e., mental reinstatement). Therefore, the study could be included in our first analysis (misinformation item analysis) as the mean number of incorrect responses to misinformation items was reported for the child and adult groups. However, there was not a relevant control condition to use in our second analysis (the control – misinformation item analysis).
Templeton & Wilcox (2000)	Both children and adults were given a standard misinformation task, though the study was interested in the impact of having a modified memory test, or a standard memory test, so the control information was concerned with this. The required F statistic was not available, and no SDs reported.

Appendix B

Prospective and Retrospective reflection

During data collection (Group 1 – Sexual Harm Support Staff)

The sexual harm support staff was the first of the focus group in the series planned for this thesis. As a result, I had no predetermined ideas of what to expect in the conversations. It was clear throughout that there were two professionals who had much more experience working with children than the other two, as they were much more vocal and forthcoming with information that they provided. It was difficult on occasion to open this up to the quieter group members, but through prompts (e.g., what do you think about the comment that Participant X made?) I was able to achieve this. I felt that due to there being more comment from two of the participants, they developed some professional rapport with one another. As a result, my input was required less after a while as they were able to converse with one another and share experiences enthusiastically.

During data collection (Group 2 – Police Officers)

The police officers were from the same area, and therefore knew each other before attending the meeting. This did not seem to impact the content of the interview, other than that they were much more relaxed at the beginning of the focus group, and as a result 'ice breakers' were not needed. The officers were very open and detailed in the information that they provided. I think because they have shared professional experiences, they were much more forthcoming about opinions and experiences of their roles.

During data collection (Group 3 –Social Workers)

The most prevalent reflection that I had from the social worker group, was that two of the participants were very frustrated and angry at the system in which they worked. This provided me with rich detail about the more emotive aspects of working with child reliability and also shared some very important insights into the running of the CJS and the apparent shortcomings.

Prospective (before data analysis)

I believe that in working with older, more experienced professionals, I was able to gather useful content. I was under the impression that (especially the older participants) were

attempting to ‘teach’ me. As a result, I was met with rich and honest reflections as though I was being ‘warned’ about what working in this field was ‘actually like’. I think if I was older, with more experience I would have perhaps been met with stiffer and censored opinions.

Similarly, in being a student or a ‘researcher’, I was more neutral (i.e., I was not a colleague) and was therefore more able to gather their perhaps more personal emotions and views. I did notice that none of the professionals identified or discussed Psychologists in any capacity. Though I’m unclear as to whether this was because I was considered to be a psychologist, or instead if it something that they had simply not considered important to discuss.

During data analysis

During the data analysis, I noticed that there were many more similarities between the professionals than I was expecting. This alluded me to the fact that I likely had a preconceived belief before attending the interviews that differences were going to arise.

Retrospective (After completing data analysis)

I am relatively new to research, with only around 2 to 3 years’ experience of on-going qualitative research experience. As a result, these focus groups came with anxiety and apprehension, and only by the third group was I considering myself to be relatively confident. I was under the impression that the professionals would consider me young and under qualified, though reflection allowed me to realise this was the perception I was placing on myself.

During the research I also realised my own frustrations that I held with the criminal justice system and how we manage children within it. Previously, this was not a principal concern of mine, though hearing the stories and examples was a harrowing experience. It took some time for me to contemplate that this was a current issue that is currently impacting children in the system that I currently work in. Prior to undertaking this research, the problem seemed so ‘far away’. Following digestion of this information, I realised that the problems raised are much too close to my profession for comfort. I feel both motivated to attempt to improve the system where I can, and also a sense of hopelessness at the problem being much bigger than myself, and the fact that my research is probably limited in it will be able to achieve. I have yet to tackle with those two contradictions in my own mind, though this research has at least allowed the internal battle.

Appendix C

Initial Codes, organising themes and global themes

Table 2A

Table outlining Theme 1 with two subthemes, including supportive example quotes and initial codes

Theme 1	Example Quote	Initial codes
Global Theme 1 **Case specific factors that impact on a child's account reliability**	*SW4: I think recognising every child is different.*	• Each case is unique
Subtheme The influence of family and caregivers on the child's account	*SHS1: The parents must have a massive impact on the child and how they deal with the impact that this has on them as they grow up.* *PO1: So, when a child is trying to tell us the truth, they think the truth is something that their mum or their dad has told them to tell us.* *SHS2: For me they are different because I've worked on the troubled family agenda for 5 or 6 years and that is very targeted support and it's about poor parenting and the kind of toxic trios and all of that mix. Whereas the children and young people that I work with now come from all walks of life and um my background revolves around parents and its around working with the parents to improve child outcome. Children and young people who are survivors of rape or sexual violence come from all walks of life* *SW1: I had it particularly with asylum seekers in that their memory was influenced by agents and family members back home.*	• Caregivers viewed as vulnerable • Caregivers influence the child's account • The caregiver does not believe the child • Parents can change the account of the child through suggestibility or manipulation • Children need consent to testify on many occasions • Anti-social parents can impact on evidence (impact on child account) • Parents impacted by the outcome of the child account • Child less likely to provide accounts against their own parents

Subtheme The impact of trauma and anxiety	*SW2: she was then very traumatised, and her childhood had a massive impact on her sort of... retention and her ability to express herself clearly.* *PO1: But at the time I think the level of trauma makes your brain remember. No matter how old you are…* *SHS4: Sometimes children don't know… and the emotional state at the time sometimes contributes to it ... and some of the young people only end up in that predicament because of the life experiences… so yeah... I don't know about you but like... like if they are autistic or... learning disability. The trauma impacts on all of the children in the same way… it is very rare that you will see a child faking the effects of trauma... so… its quite generic regardless of their learning disability...*	• The child is likely traumatised • Children dissociate • Trauma influences memory • Trauma makes memory more vulnerable to change • Trauma consolidates memory

Table 2B

Table outlining Theme 2 with two subthemes, including supportive example quotes and initial codes

Theme 2	Example Quote	Initial codes
Global Theme 2 - **Global perception of 'child memory' being both reliable and unreliable**	*PO2: He shouldn't have been interviewed. He was too young. Too vulnerable. Not stable enough.*	• Children are more vulnerable than adults
Subtheme The 'believability' of a child	*SW3: And … children tend not to make things up.* *PO1: And I think that she would have found it hard to lie because children I think do find it hard to lie…* *SHS4: I don't know… we have teenagers that aren't believed because there are a lot of school sexual assaults at the moment and a lot of videos being shared and... children meeting at the park and for some reasons just... just because you meet a boy at the park that you were aware of what was going to happen or in control of it.* *SHS1: um… inevitably... you might be slightly more... you might find it slightly harder as a jury member to believe a 4-year-old compared to a 40-year-old... maybe…*	• Children are not believable • The memory of a child is a concern • Children are antisocial • Children tell the truth • Children are believable • Deception • Non-compliance with police proceedings • Non-compliance with CJS • Defiance of police requests • Intimidation of personnel • Vulnerable to suggestibility when questioning • Social desirability factors impacting on reliability • Jury do not believe • Children cannot lie
Subtheme Children have less understanding and knowledge of their experiences than adults	*PO2: And I don't think he knew the implications of what he was saying. Didn't have a clue.* *SHS1: They might not even understand what has happened so won't be able to put it into words or context.*	• Children don't understand the practicalities • No understanding of the consequences • Get less information from a child compared to an adult • Children have less knowledge

Subtheme The foundation of a child's memory is more vulnerable to change	*PO2: You can't keep them focused on one point and you are constantly bringing them back to the track because they're going off the track.* *PO3: I don't think that…. I don't think that they necessarily think that they've done something wrong and that they're trying to change the truth. I think they just get caught up in the… story.*	• Age decreases reliability of a memory • The memory will change • Children lack focus • Children attempt to be more socially desirable • Children are more suggestible • Children have less language abilities • Lack of understanding of consequences • Memory is changeable • Suggestibility • Use of leading questions should be avoided due to higher levels of suggestibility

Table 2C

Table outlining Theme 3 with two subthemes, including supportive example quotes and initial codes

Theme 3	Example Quote	Initial codes
Global Theme 3- The Criminal Justice System is not appropriate for children	*SHS2: I mean even being a grislily adult. Going to court and giving evidence and stuff is bloody horrible really. So being a kid going into this scary and serious environment…* *PO3: And so… I just think in court it's not fair on a child of that age… well… a lot of ages really to go to court.* *SHS4: A lot of them say that they regret reporting it because of the court process. It's sort of like... re-traumatizing them again.* *PO1: The other things Is as well... only… say if we went to an incident on a night shift say and we want to obtain a child's evidence we can't go dragging them out of bed all hours of the night. And then even if we do it the next day and we pass it on to a day shift to say that this child needs video interviewing or in a few days' time. We've then got to take that child out of school quite often or out of clubs or whatever they're going to. So, we're messing with they're…* *PO3: Day to day normal life.* *SW3: But then when going through the court process, that child then doesn't have access to counselling services. Because they have to wait till after.*	• The system is not fair on children • It is a high-pressure situation • System is not meeting the child's needs • The process is difficult • There is no clear guidance • There is high anxiety • Lack of funding for agencies • Reluctance to work with children • No support for children • Importance of evidence • Discouraging them from discussing the case • Treatment of children is insensitive • Frustration with the system • Children are undermined • Jury is biased • Expectations are too high
Subtheme Gathering evidence from children is not appropriate	*SW4: From an asylum seeker point of view, there's problems with the length of interview… so I was in an interview [Location redacted] home office for 7 hours... and this was an 11-year-old who had just come across s from [Location redacted]..* *SHS3: And then you have a child going through that highly stressful situation and having repeatedly asked the same question over and over again. You know. I think that*	• Questions are often repeated • Decisions in interview are subjective • Importance of leading questions and their impact • Tactical questioning is used • Interviewing techniques are problematic • Concerns about how interviews are carried out

	anyone in that situation, let alone a child would question their confidence. *PO2: Oh yeah, leading questions and all of that. What they tell you about leading questions is 'Don't do it!'* *SW4: And... I don't feel as if they ... the interview has been prepared well or that even the child has been prepared well.* *SHS1: I guess also what must affect the child and those aspects of it is the skill and the knowledge of the barristers as well... to tease out the information in the right way...* *SW3: I think that the younger the child is, the less they get listened to. Not by social workers, but by other professions and around using that as evidence. There is that focus on that evidence. A child of a young age can say what's happened, it's about repeating it in the same way... umm... it can cause a lot of difficulties.* *SW1: So, they can evidence the route. If the child lies about the route and says he hasn't been, or she hasn't been fingerprinted then straight away that lie is put down into the court. And they're not believed for what they say. So, they might be saying a lot of truths, but if they make that one mistake then they can be in a lot of trouble. And that one mistake can be evidence that shows easily.*	• Questioning techniques cannot be standardised • Children not being listened to • Children having to repeat the same things in the same way • Children are not believed during evidence collection
Subtheme Professional roles and responsibilities and inter-agency working	*SHS4: I guess also what must affect the child and those aspects of it is the skill and the knowledge of the barristers as well... to tease out the information in the right way...* *SW3: I think the police often work with children will come into it from an adult's perspective rather than we all go in with a child perspective right from the go because that's the job we're in* *PO1: We would still gather all of the evidence and send it to the CPS but unfortunately, it's not us that makes the decision. It's the CPS lawyer… and it's like… it's like what - PO2- said... we put the pocket note entry and put on*	• Lack of understanding of other professional's roles and responsibilities • Shared professional biases • Limits to professional roles and what can be achieved • Negative perception of police from ISVAs and social workers • Problems with inter-agency working • Frustration with other professionals and agencies • Lack of inter-agency cohesion

the file, on the case file, that the child was a witness and that he was in the room or was upstairs and this is what they have told us. However, uh... we haven't been able to obtain that evidence because the parents don't want us to do it... or the child is too young, too vulnerable, whatever. And then often the CPS lawyers would throw it out 'cuse they're no other… its one word against another... Um... theres no other evidence, so it wouldn't make the court room anyway.

SW3: As a social worker I have done the join investigation so around... questioning children

- Personal and moral Opinions
- Negative opinions shared about barristers in court

Appendix D

Details of suggestibility tests

Table 3

Psychological Tests of Suggestibility and their Characteristics. This table copied directly from Endres et al., (1997) development of the BTSS measurement

Test	Authors	Stimuli	Types of Questions	Sub-scales	Target Age	Test Norms	Remarks
Wuzburg Suggestibility Test (WST)	Bottenberg & Wehner, 1971	Table #5 of the TAT	Affirmative sentences	none	12 to 13	N=169, only girls	group test, written form
Test of Statement Suggestibility (TAS)	Burger, 1971	30 slides, scenes of everyday life	120 Yes-No questions	none	7 to 14	N=200	
Suggestibility Test (SET-S)	Zimmermann, 1978, 1982a, b, 1988	4 photos	18 Yes-No (assertive sentences)	sexual content suggested	9 to 10 and 12 to 16	N=110 (younger) N= 225 (older)	forms for younger and older children
Gudjonsson Suggestibility Scale (GSS)	Gudjonsson, 1984, 1987	Short story	Yes-No and alternative	yield, shift	Adults	N=195, children and adults	parallel form
Bonn Test of Statement Suggestibility (BTSS)	Endres & Scholz, 1995	Illustrated short story	Yes-No, alternative, repeated	3 (different Question formats)	4 to 10	N=62	parallel form

Appendix E
Summary of articles included into Psychometric critique

Table 4
Summary of all articles assessed and included in the review

Review Article	Language	Reliability Assessments	Validity Assessments	Conclusion	Notes
Endres (1997)	German	1. Internal Consistency 2.Parallel Form Reliability	1.Face Validity 2.Construct Validity	Reliable assessment	Author of the psychometric
Candel, Merckelback & Muris (2000)	Dutch (BTSS-NL)	1. Internal Consistency 2.Test-retest	1.Concurrent Validity 2.Correlation with researched variables	'Satisfactory' and reasonable test-retest stability & high internal reliability	
Candel, Merckelbach, Loyen & Reyskens (2005)	Dutch (BTSS-NL)	1. Internal Consistency 2. Test-Retest	N/A	Not a review paper of the BTSS but measured internal consistently to strengthen their methodology	
Benatti (2012)	Italian	1. Internal Consistency	1. Test Norms 2.Correlation with researched variables	Concluded that more research is needed in the Italian version	Italian Test norms
Dafflon (2012)	Portuguese	1. Internal Consistency 2.Factor Analysis	1. Correlation with researched variables	Portuguese version is 'acceptable'	
Costa & Pinho (2008)	Portuguese	1. Internal Consistency	1.Correlation with researched variables	Satisfactory number of theoretically related variables	Author who translated to Portuguese
Roma (2011)	English	N/A	1. Concurrent Validity	Good correlations between BTSS and GSS	
Caffo, Rossi, Benatti & Rigtelli (in press)	Italian	1. Internal Consistency 2. Factor Analysis	1.Normal Distribution (i.e. SEM)	Very positive and supportive of the test	