Enhancing police interviews: How age, verbal instruction, rapport and question type influence adult witness confidence and accuracy reports

Kaye N. Cooke[1] and Jacqueline M. Wheatcroft[2]

Abstract
Witnesses are pivotal to police investigations and rapport-building is considered crucial to interview effectiveness. Study findings have implications for operational interviewing, research, and training. In this online study, 198 adult participants were randomly allocated to control, rapport or verbal instruction (VI) conditions, separated by age. Results showed those aged 18–33 were less accurate in free recall and more susceptible to confabulation in VI. Confidence for directive leading questions was higher in VI than rapport. Outcomes suggest a statement strengthening the requirement not to make-up details should be implemented, with effective rapport-building essential to guard against the impact of questioning.

Keywords
rapport and context-reinstatement, confidence and accuracy, witness age, questioning

[1]School of Natural & Social Sciences, University of Gloucestershire, Francis Close Hall, Cheltenham, UK
[2]School of Psychology, Faculty of Health, Innovation, Technology & Science, Liverpool John Moores University, Liverpool, UK

Corresponding author:
Jacqueline M. Wheatcroft, School of Psychology, Faculty of Health, Innovation, Technology & Science, Research Centre for Brain & Behaviour (RCBB), Tom Reilly Building, Byrom Street, Liverpool L3 3AF, UK.
Email: j.m.wheatcroft@ljmu.ac.uk

Introduction

Accurate and confident eyewitness reports are fundamental to successful forensic investigation and the adversarial court system. In the early 1900s, Binet claimed that interviewer questioning can influence witness responses (Wells et al., 2006), and this was highlighted, as early as 1908, by Munsterberg for legal contexts. Subsequently, researchers have investigated techniques to facilitate more detailed and accurate accounts from witnesses (Gabbert et al., 2020; Wagstaff et al., 2014; Wells et al., 2006), with links demonstrated between the quality of accounts obtained in interview and the tools and techniques used (Wagstaff et al., 2014; Wells et al., 2006; Wheatcroft and Woods, 2010). In terms of the current paper, it is argued that definitions of rapport are inconsistent and the range of methodologies used in the research available make comparisons difficult. For example, much research on rapport has focused on measurement of non-verbal behaviours. However, Novotny et al. (2021) found that participants were more willing to talk about topics personal to them when verbal commonalities were used 'alone' compared to when combined with non-verbal mirroring. What is interesting about the outcomes of Novotny's study is that when different elements of rapport (as defined by researchers) are investigated together more precise and meaningful insights can be drawn. It is beneficial therefore to understand why one should not necessarily draw from each of the separate literatures that consider factors in isolation. Potential differential effects can be missed. Thus, combining factors with some fundamental features of interviewing allows for both main and simple comparisons. In addition, nuanced findings can support practical knowledge and application of investigative interviews.

There is good reason why one might expect different outcomes when looking at witnesses of different age groups. In relation to age, a general decline in memory accuracy is known (e.g., Gawrylowicz et al., 2014; Henry et al., 2020; Li et al., 2005; Walhovd et al., 2016) though whether this differs between free (i.e., open questions) and cued recall (i.e., targeted questions) accounts is contested by some (e.g. Caso et al., 2024). For enhancement tools, use of a context reinstatement type instruction has been found to improve accuracy (Wagstaff et al., 2011). Evidence-based guidance on which are the most effective memory retrieval tools to use during forensic interviewing is limited, with investigative interviewers' understanding of how best to exploit memory processes restricted (Howe and Knott, 2015). For example, even when interviewers have been appropriately trained in investigative techniques to obtain best evidence, often including special instructions, these are not always optimised (Mugno et al., 2018; Powell et al., 2010; Wagstaff et al., 2014), and witnesses may sometimes be held responsible for the inaccurate accounts obtained, rather than officers acknowledging personal limitations (Kebbell and Milne, 1998; Powell et al., 2010). However, without effective techniques enabling witnesses to confidently recount everything they remember, with no embellishment, resultant evidence may be of insufficient quality. As such, the importance of accurate evidence is paramount as a guilty person may go free, or an innocent person may be convicted. To the authors' knowledge, no research has investigated together whether rapport, age, interview enhancement tools and questioning impact the accuracy and confidence of eyewitness reports. This paper investigates these issues. First, however, the underpinning to some investigative techniques will be outlined and considered.

Protocols: Context reinstatement and rapport

The Cognitive Interview (CI; Fisher and Geiselman, 1992) is a complex procedure that requires substantial training to learn, and is lengthy to administer. Further, all police officers may not receive appropriate training, and even trained officers often deviate from the procedures specified in the training (Wagstaff et al., 2014). In less complex crimes, use of the full CI may be considered disproportionate. In addition, time pressures mean many officers do not consider the full CI to be cost-effective in everyday policing (Wheatcroft and Wagstaff, 2010); leading researchers to look for shortened versions of, or brief alternatives to, the full CI that can be used when time is at a premium (Dando et al., 2009). One such shortened version is the context reinstatement (CR; see Liverpool Interview Protocol (LIP) (Wagstaff et al., 2014; Wagstaff and Wheatcroft, 2012) which prompts witnesses to imagine themselves at the crime scene, encouraging them to use all their senses. This process recreates physical and psychological contexts in order to aid more detailed retrieval (Dianiska et al., 2019). CR has been found to prompt significantly better memory recall (Smith and Vela, 2001), greater detail (Dianiska et al., 2019; Memon et al., 2010; Wagstaff et al., 2011) and greater accuracy (Wilcock et al., 2007). CR also appears to be effective irrespective of age (Wagstaff et al., 2014). Wagstaff et al. (2011) found benefits of CR were most pronounced for free recall accounts with improved accuracy of retrieved information. One explanation for the outcomes is that local processing, the brain's consideration of smaller detail (e.g., facial features or a vehicle registration plate), is being activated (Huff et al., 2011); local processing will be returned to later in the paper.

The value of rapport as an element of police interview protocols is the subject of continued debate, particularly in investigative interviewing of children (e.g., Collins et al., 2014; Collins and Carthy, 2019; Giles et al., 2021; Vallano and Schreiber Compo, 2015). For example, rapport can minimise the social demands of event retrieval, thereby increasing cognitive capacity (Dando et al., 2023), and is often employed to encourage interpersonal connections between interviewer and interviewee (Foster et al., 2023; Gabbert et al., 2020). Effective rapport can also improve accuracy in adult witnesses (Vallano and Schreiber Compo, 2015) and reduce the reluctance of child witnesses (Gous and Wheatcroft, 2020; Hershkowitz et al., 2006; Saywitz et al., 2019) and offenders (Vallano and Schreiber Compo, 2015) to talk. Gabbert et al. (2020) reported that 91% of studies reviewed suggest rapport positively impacted disclosure (see also Magnusson et al., 2020). However, in a study of serving police officers only 54% were found to always use rapport when interviewing, albeit the participants in that study were young-in-service non-specialist interviewers (Dando et al., 2008).

There is also a lack of evidence-based guidance for rapport (Saywitz et al., 2015), with no standardised definition (Abbe and Brandon, 2013; Brouillard et al., 2024; Lavoie et al., 2021; Neequaye, 2023; Wheatcroft et al., 2014), and little agreement of how rapport is measured (Brouillard et al., 2024; Collins and Carthy, 2019; Neequaye and Mac Giolla, 2022; Vallano and Schreiber Compo, 2015). Lack of "consistency across studies in the way rapport is defined and measured" – therefore – "creates challenges for developing effective training" (Brouillard et al., 2024: 3). Further, measurement tends to be primarily focussed on interviewer, rather than the interviewee (Matsumoto and Hwang, 2021) with suggestion that interviewers themselves impact rapport efficacy. For example, personal

characteristics and experiences, such as expectations, perceptions, motives, and behaviour, including relevant communication (Bell et al., 2016). Such characteristics from in-person studies cannot be translated to an online study, as reported here, highlighting a need to address and/or define rapport in the context of communication rather than interpersonal behaviours.

While much existing research relates to children (Vroom et al., 2025) rapport is considered a fundamental interview skill valuable for investigative interviews with all witnesses, including vulnerable adult witnesses and suspects (Collins et al., 2014; Collins and Carthy, 2019; Kassin et al., 2007; Nash et al., 2016). Studies have shown value in maintaining rapport in a suspect interview, such as use of empathy and enhanced co-operation leading to fuller accounts (Brouillard et al., 2024; Walsh and Bull, 2012) and increased accuracy using effective rapport with witnesses, such as interviewers speaking more gently and referring to interviewees by name leading to significantly fewer pieces of incorrect information (Collins et al., 2002; Vallano and Compo, 2011). On balance, while some studies show rapport to be useful, any added value may be lost if scholars do not agree on how to measure rapport. Regardless, rapport remains under-researched as a tool with adult witnesses and across age groups.

Age

Some researchers propose no age differences exist for recall accuracy (Adams-Price, 1992; Chan et al., 2009). However, others have suggested there is an impact of witness age on accuracy (Kassin et al., 2001) with those over age 60 most affected (Wilcock et al., 2007), and most influenced by question type (Cohen and Faulkner, 1989; Prull and Yockelson, 2013). However, Prull and Yockelson used misinformation effects with 64 participants; students aged 18–22 years or self-referred residents aged 60–88 years, making comparison with the study reported difficult (i.e., adult groups aged 18–33, 34–49, 50–65). Overall, a decline in cognitive ability with age is recognised, perhaps representing inferior memory encoding quality (Li et al., 2005) with older witnesses remembering less detail (Aizpurua et al., 2009; Searcy et al., 2001). As noted earlier, one explanation is that local processing elements; that is, attending to the specific details of a stimulus or processing information in a narrower and more detail-oriented way (Kimchi, 1992) can distract from global processing elements; processing information in a more general and big-picture way (Navon, 1977). For example, the Navon task presents stimuli of large letters made up of smaller letters to examine whether one first sees detail (local) followed by the overall outlay (global). Findings have been generally associated to fewer details being remembered by older participants (Ebaid and Crewther, 2019; Insch et al., 2012; Oken et al., 1999). When Roux and Ceccaldi (2001) used a Navon paradigm word task with older participants bias towards global processing was found. Other research has explored both global and local processing using the Navon paradigm before participants viewed a reconstructed crime. Those in the global condition made significantly more correct identifications in subsequent line-up tasks (Darling et al., 2009; Perfect et al., 2007) supporting proposals that global processing influences identification (Macrae and Lewis, 2002). The complexities surrounding research into eyewitness reports highlight the difficulties in the inclusion of specific comparative data.

Conversely, younger adults have been found to give more complete or accurate accounts (Gawrylowicz et al., 2014; List, 1986). In List's study, larger participant pools across two studies with mean ages of 10, 20.1, and 72.3 years were used. Children's reports were as complete as, but less accurate than, younger adults. Older adult reports were less complete but as accurate as college students. The results, that college student accounts were more complete, is unsurprising given the disparate ages of participants, though it is still possible that processing may have been more relevant than age. Faber et al. (2023) used a similar methodology to the study reported here. In Faber et al.'s study, participants reported more unverifiable but not more accurate details in free recall, whereas they performed better in cued recall and delivered higher ratings of reliving, vividness, re-experience, and emotions, suggesting a richer recall experience. Whilst there are no specific ages given for the 234 participants, making comparison problematic, the mean of 42.46 years suggests the participant pool may be comparable to this study. Wilcock et al. (2007) established an effect of CR for age with a medium to large effect size (V = 0.46), in that it may be possible for older adults to perform at an equivalent level to younger adults using supporting photographic materials. This was one of a number of studies found using CR with older participants (Memon et al., 2002; Searcy et al., 2001; Wilcock et al., 2007). Nevertheless, the findings for links between age and witness performance are mixed. As noted, a decline in accuracy with age is recognised, though some memory enhancement tools have been found to improve performances (see Self-Administered Interview; Gawrylowicz et al., 2014). Thus, to allow systematic investigation and ensure a more consistent age spectrum, the study reported here used 198 participants equally spanned across age groups, comprising younger (18–33), middle (34–49) and older (50–65) adults.

In addition to the investigative protocols and the influence of age, there are a range of outcome measures used to detect differences between conditions. Some important measures in the investigative context are accuracy, confidence and the relationship between confidence and accuracy, together with question type and vulnerability to confabulation. It is important to outline the background to these aspects of the study.

Accuracy

In order to remember what has been witnessed or experienced, top-down processing draws on information previously known, to make sense of it. Bottom-up processing applies this knowledge to the current situation; the potential point of encoding (Turner, 2015). In cued recall, questions attempt to prompt access to stored information; guiding witnesses to access fine details of the event witnessed (Huff et al., 2011; Koriat and Goldsmith, 1996). Therefore, cued recall questions may allow witnesses to recover details that were encoded unintentionally. However, information accuracy can be impacted by attention paid at the time of encoding (Smith et al., 2018), and which is of particular relevance for eyewitness accounts. For example, accuracy may be impacted by whether it is a central detail being processed; a fact or detail integral to the event or story (e.g., description of an offender), or peripheral detail; a fact or detail irrelevant to what actually happened (e.g., description of a bystander) (Burke et al., 1992). Indeed, when recalling emotional events, as eyewitnesses may well be, poor memory for peripheral detail has been shown relative to central detail

(Christianson and Loftus, 1991; Lanciano and Curci, 2011). However, these latter studies used university undergraduate participants who were asked to remember emotional events from their own lives or watch static slides, rather than recalling a crime witnessed via a video clip, as used in this study. Other studies have shown that memory for peripheral detail is most impacted by perceptual load[1] (Murphy and Greene, 2016), and this would be relevant in an eyewitness context where witnesses' mental effort can be important in how accurately information can be recalled.

There is also potential for the original memory to be overlaid, corrupted or influenced by additional information (Loftus and Hoffman, 1989), and any gaps infilled (Wixted, 2023). Studies have shown that misleading questions or negative feedback can influence witnesses' responses (Gudjonsson and Clark, 1986; Polczyk et al., 2024). Indeed, the process of eliciting memories can change those memories (Wixted, 2023) making the first account critically important. Thus, while there are multiple points at which the accuracy of witness memory might be impacted, the study reported here is concerned with retrieval (i.e., the process by which stored memories are accessed; Kaye and Tree, 2016; Smith et al., 2018). Importantly, when real world incidents take place, witnesses are seldom aware of the relevance of what they are seeing at the time (Gudjonsson, 2003). Consequently, when retrieving memories, ambiguities may be supplemented using schema. Such cognitive shortcuts are activated to infill memory gaps using associated memories of what might normally happen in similar situations (Abelson, 1981; Ormerod and Adler, 2010; Robinson et al., 1997). It is also possible for confabulations to emerge and used to fill memory gaps (Gudjonsson, 2003; Gudjonsson and Clare, 1995). Confabulations may take the form of fabrication (i.e., introduction of new information), or distortion (i.e., an altered version of reality). It is proposed that doubt and expectation are key contributors to confabulations (Mercer et al., 1977), with links made between the use of leading questions by interviewers and resulting confabulations (Gudjonsson and Young, 2010).

Interestingly, whilst some research found fMRI used determined a decline in brain processing and memory occurs with age (e.g. Dennis et al., 2007, 2008; Fandakova et al., 2018), research available reporting the relevance of age on confabulation can be mixed. For example, one study found confabulation is more prevalent for older adults, whose accurate recall may be over-laid by previously learnt information, with confabulations representing disruption of executive processing at the point of retrieval (Attali and Dalla Barba, 2013). Dennis et al. (2007) propose that older adults not only forget past events they may fabricate memories. However, in the context of false confessions, Gudjonsson and Young (2010) argue that confabulation in immediate and delayed recall correlates negatively with memory and positively with suggestibility. The complexity around what creates confabulation means there is little to no literature that considers the impact of rapport on confabulation in the context of questioning, nor research linking age and confabulation other than that noted above (i.e., Attali and Dalla Barba, 2013; Dennis et al., 2007), representing clear and obvious gaps in the literature.

Moreover, it is known that inaccurate eyewitness identification remains a leading cause of wrongful conviction (R v Malkinson, 2023; Rakoff and Loftus, 2018) with misinformation or poor interview techniques at the point of memory retrieval potential contributors (Butler and Loftus, 2018; Wixted, 2023). Therefore, using interview techniques that do not take account of how information is processed will likely be detrimental to the

accuracy of information retrieved. To reiterate, global processing attends to a spatial or holistic perspective (e.g., reporting a person or a vehicle), whilst local processing considers details (e.g., facial features or a vehicle registration plate) (Huff et al., 2011). Some researchers maintain these two processes are distinct (Förster and Dannenberg, 2010), others believe an interaction is present (Guy et al., 2019).

Given the above, the current study will consider these aspects as, for example, a free recall account may involve more global processing, using open-ended questions to encourage a free narrative from participants. In contrast, focused cued recall questions, using questions that direct toward a response (Evans and Fisher, 2011), would more likely involve local processing (Koriat and Goldsmith, 1996; Perfect and Weber, 2012). In addition to the method of retrieval, question types and individual confidence to answers are key to interview outcomes.

Question type, confidence and within-subjects confidence-accuracy (W-S C-A)

The way that questions are presented to witnesses has been shown to negatively influence response accuracy (Loftus and Palmer, 1974). Links have been found between interview technique and account quality (Collins et al., 2002; Wells et al., 2006; Wheatcroft and Ellison, 2012; Wheatcroft and Woods, 2010) as well as between questioning type and witness confidence (Allwood et al., 2008; Gous and Wheatcroft, 2020; Wheatcroft and Ellison, 2012; Wheatcroft and Woods, 2010). Other work examining question effects has found lawyerese questions, i.e., those containing "leading and suppositional phrases" confused witnesses (Wheatcroft et al., 2004: 83). Where witness reports are often the only evidence available (Odinot and Wolters, 2006), reliable and effective interview techniques are central to accurate accounts.

The impact of question type on witness confidence is particularly relevant to forensic outcomes (Gous and Wheatcroft, 2020; Wheatcroft and Ellison, 2012; Wheatcroft and Woods, 2010), with one study in the USA reporting that evidence was dismissed by a jury because a witness was asked leading questions (Ruva and Bryant, 2004). However, it is important to discriminate between types of questions. Directive leading (DL) questions suggest, or assertively imply, how a question should be answered (e.g., The onlooker didn't call for help, did he?). While non-directive leading (NDL) questions are also a form of closed question, they include content without additional directive pressure (e.g., Was the jogger wearing a bracelet?) (Gous and Wheatcroft, 2020). Both question types are used regularly in legal contexts and research has demonstrated the impact of such questions on both witness confidence and accuracy (Gous and Wheatcroft, 2020; Wade and Spearing, 2023). DL questions are the most detrimental (Ramadhani et al., 2019; Wheatcroft et al., 2015a), least reliable (Henderson, 2016; Kebbell and Gilchrist, 2010), when used in cross-examination have the potential to cause witnesses to change previous accurate responses (Eades, 2012; Valentine and Maras, 2011), and undermine witness credibility (Plotnikoff and Woolfson, 2009). However, it is argued that the effects of leading questions may be different according to witness age and vulnerability (Wheatcroft and Ellison, 2012). For example, Erikson's Theory of Development (1950) suggests at early-middle age people are entering the stage of life where they have more confidence in their own ability (Brandau and Evanson, 2018; Parrish, 2014). Research that investigates what enables

witnesses to achieve the most accurate and confident reports, and the relationship between confidence and accuracy, continues to be essential. In the current study, both DL and NDL question types are used across age ranges to investigate the impact on response accuracy, confidence and within-subjects confidence-accuracy (W-S C-A).

Witness self-reported confidence levels have revealed mixed findings, with high levels of confidence in initial eyewitness reports raising positive perceptions around the validity of evidence (Wells et al., 2006; Wixted et al., 2015; Wixted and Wells, 2017). However, general caution in the confidence-accuracy (C-A) relationship is proposed (Berkowitz et al., 2021, 2022; Sauer et al., 2019; Wade et al., 2018) especially when that confidence level is self-reported (Perfect, 2004). Some argue that links between C-A can be unreliable (Wheatcroft and Woods, 2010), and the C-A relationship becomes most pertinent in forensic contexts where a witness' apparent confidence may be relied upon unduly by jurors assessing the credibility of their testimony (Berkowitz et al., 2021, 2022; Brodsky et al., 2010; Wheatcroft and Woods, 2010). Other research has examined confidence and accuracy in the context of information retention over time (Odinot and Wolters, 2006; Wheatcroft et al., 2015b) or with participant pools limited to university students (e.g. Luna and Martín-Luengo, 2011; Odinot and Wolters, 2006). Nevertheless, while outcomes remain mixed, dependent upon where researchers focus and how studies are operated, it is useful to assess C-A in the context of interviewing.

The within-subjects confidence-accuracy (W-S C-A) relationship is especially important in investigative interviews where, again, the credibility of witness testimony is judged relative to confidence levels (Fox and Walters, 1986; Lindsay et al., 1989; Plotnikoff and Woolfson, 2009; Sah et al., 2013) with a general perception that confident witnesses are more accurate (Berkowitz et al., 2022; Gous and Wheatcroft, 2020; Wixted et al., 2018). However, early research suggests the C-A relationship can be easily distorted after original identification has been made (Luus and Wells, 1994). It may be that those inherently lacking confidence are more likely to succumb to suggestion, especially in cross-examination; that is, when legally questioned (Wade and Spearing, 2023). However, research has established a strong C-A relationship in particular circumstances. For example, following simple questions in cued recall after viewing a mock crime (Kebbell and Giles, 2000; Kebbell and Johnson, 2000); and when giving a free recall account of a crime video (Caso et al., 2024).

One explanation for these disparate findings is that questioning techniques cause processing shifts (Antes and Mann, 1984; Huff et al., 2011; Pacheco-Unguetti et al., 2014) with both witness accuracy and confidence levels influenced (Kebbell and Giles, 2000; Sporer et al., 1995; Wheatcroft et al., 2004). In the latter study, W-S C-A effects became significant when 'difficult' questions were asked, influenced by increased confidence to DL questions. Robinson et al. (1997) propose that the requirement for a confidence score may simply increase cognitive load and impact accuracy, in the same way as complex questions. Caso et al. (2024) suggest that incompatible cued recall questions reduce the C-A calibration, with more reliable C-A calibrations elicited when free recall is followed by relevant probing cued recall questions. Finally, Wixted and Wells (2017) argue little regard is given to the relationship between C-A by the US legal system, with 70% of 349 overturned convictions having relied on eyewitness identification. In light of the varied research, mixed findings and judicial lack of focus on C-A, this study will also investigate whether rapport, question type

and age, when considered together, are influential to W-S C-A and which investigative techniques, if any, might improve the relationship.

Research aim and hypotheses

The aim of this study is to establish how interview protocols such as rapport and verbal instruction, question type, and age, influence adult witness accuracy and confidence. In light of the above considerations the following hypotheses were formulated:

H1: Older adults will be less accurate than younger adults.

H2: For free recall, there will be a difference in accuracy for interview protocol (i.e., rapport and verbal instruction, compared against the control) and age.

H3: For free recall, there will be a difference in the number of overall confabulations elicited (i.e., fabrications and distortions combined) for interview protocol and age.

H4: In cued recall, accuracy will increase in the verbal instruction condition compared against the rapport and control conditions.

H5: Directive-leading (DL) questions will lead to decreased accuracy.

H6: Directive-leading (DL) questions will lead to increased confidence.

H7: There will be a difference in within-subjects confidence-accuracy (W-S C-A) for interview protocol and age.

Method

Design and participants

A 3 (protocol: control, rapport, and verbal instruction) × 3 (age: 18–33 years, 34–49 years, 50–65 years) between-participant experimental design was used. 198 participants aged 18–65 years ($M = 41.05$, $SD = 13.62$) were recruited through Facebook and online exchange groups. An a-priori review of existing literature established an effect size of 0.28 (Smith and Vela, 2001). GPower*, using a conservative small effect size, alpha of 0.05 and beta 0.8, established a requirement for 215 participants (Cohen, 1992). 37 (19%) participants self-reported as male, 159 (80%) as female. One participant associated with neither gender, one preferred not to say. 99% of participants self-reported normal or corrected-to-normal eyesight. No incentives were provided for completion.

Materials and procedure

An advertisement was placed on Facebook with a link to the study. The participant information sheet (PIS) ensured participants were aware of the experimental procedure and that a video clip of a crime would be viewed. Having read the PIS, the start of the video was signalled, and participants informed there was no sound. The video clip (Wheatcroft, 2020) showed a man leaving a public house and walking to a car before reversing the car slowly into a passing jogger. The vehicle drove away, leaving the injured jogger on the ground. All participants observed the same stimulus clip.

Participants were asked to use a computer with a keyboard and screen, in accordance with internet ethics guidance (British Psychological Society, 2021a). On clicking the study link, participants were randomly allocated to protocol conditions using a free online tool (allocate.monster) as follows: control (word search distraction task); rapport (10 unrelated questions which replicated the types of question commonly used in the rapport building stage of investigative interviews); verbal instruction (i.e., pre-recorded verbal CR instruction) (Wagstaff et al., 2014). The CR is focused on mental context reinstatement (the impact of imagining the context of the experience). However, reference to this is termed context reinstatement in this paper. All participants experienced equivalent latency between viewing the video and the provision of a free recall account. After viewing the video clip, and according to condition, all groups were given a free recall instruction; 5 minutes to type what they remembered from the video into a free-text box. Following this, participants were asked to answer 20 cued recall questions (i.e., 10 DL, 10 NDL) and rate their confidence in each answer given. The cued recall questions included a mixture of positive and negative statements to avoid response bias. Finally, participants were directed to a debrief and thanked for their time.

Data

Free recall data was coded for overall accuracy against a framework consisting of 60 items (30 central, 30 peripheral). The framework ensured researcher bias in interpretation of the accounts given was minimised. One point was given for each correct item, with correct but incomplete responses allocated half a point, to a maximum score of 60. Inter-rater reliability was conducted by two researchers independently. Cohen's Kappa coefficient ('κ') reached 94% agreement on 10% of the data equally distributed across the conditions. The number of confabulations (i.e., fabrications and distortions) were coded following the approach of Gudjonsson (2003). For example, 1 point was scored for each fabrication (e.g., 'calling for an ambulance' when an ambulance was not present in the stimulus) and 1 point for each distortion (e.g., 'he found the keys' as opposed to 'he picked up the keys' - stimulus showed the keys had been dropped on the ground). Consideration was given to comparing word count but disregarded as research suggests information quality is best measured by the amount of detail, not the number of words (Elntib et al., 2015; Warmelink et al., 2019). In addition, when examining information accuracy in the context of format (i.e., written or oral), Elntib et al. (2015) concluded that written accounts tend to be more dense in information than oral ones.

Cued recall was coded for the number of correct answers from 20 questions (i.e., 10 DL, 10 NDL) with one point given for each correct answer. Participants rated their confidence in each answer using a Likert scale, where 1 represented 'not at all confident' and 6, 'absolutely certain'. The maximum score for overall confidence was 120; 60 each for DL and NDL.

Ethics

Participants were informed of the aims of the research, procedure, potential risk and how data would be used. The video was assessed as minimal risk of adverse effect. Relevant

privacy regulations were highlighted, contact information given and the right to withdraw without consequence made clear. As data was anonymised once uploaded, participants were asked to provide an eight-character unique reference to enable withdrawal by a specified date should they wish to do so. The study was approved by the Ethics Committee of the University of Gloucestershire (Approval no. FPY/20/022) on 15/02/21. All participants provided informed consent prior to enrolment in the study. This research was also conducted ethically in accordance with the British Psychological Society Code of Ethics (2021).

Results

A two-way 3 (Protocol: control, rapport, and verbal instruction) × 3 (Age: 18–33 years, 34–49 years, 50–65 years) between-groups analysis of variance (ANOVA) was performed on each dependent variable for free recall (i.e., overall accuracy, central accuracy, peripheral accuracy, overall confabulations, fabrications and distortions) and cued recall (i.e., overall accuracy, overall confidence, DL accuracy, DL confidence, NDL accuracy and NDL confidence, and W-S C-A). Residual analysis was performed to test for 2-way ANOVA assumptions and, where necessary, post-hoc tests conducted. Where normality was not established by Kolmogorov-Smirnov ($p < .05$) but Levene's test showed homogeneity of variances ($p > .05$), outcomes were considered reliable and sufficient for the analysis to be robust, particularly as group sizes were equal. Tests which failed to meet this criterion are reported.

Descriptive statistics are provided in Table 1 for free recall variables and Table 2 for cued recall variables.

Accuracy

Overall free recall accuracy

There was a significant main effect for age on overall free recall accuracy, $F_{(2,198)} = 9.38$, $p < .001$, $\eta p^2 = .09$, 1-β>.98. No effect for protocol was found, $F_{(2,198)} = 2.47$, $p > .05$, $\eta p^2 = .03$, 1-β = .50. No significant interaction was observed, $F_{(4,198)} = .29$, $p > .05$, $\eta p^2 < .01$, 1-β = .11. Tukey post hoc tests showed significantly lower free recall accuracy for those aged 18-33 ($M = 11.97$, $SD = 5.17$) than other age groups; 34-49 ($M = 14.73$, $SD = 5.63$) ($p = .01$), and 50-65 ($M = 15.97$, $SD = 5.50$) ($p < .001$). No other comparisons were significant ($p > .05$); see Table 1.

Overall central accuracy

There was a significant main effect for age on overall central accuracy, $F_{(2,198)} = 7.08$, $p = .001$, $\eta p^2 = .07$, 1-β>.93. No effect for protocol was found, $F_{(2,198)} = 2.94$, $p > .05$, $\eta p^2 = .03$, 1-β = .57. No significant interaction was observed, $F_{(4,198)} = .51$, $p > .05$, $\eta p^2 = .01$, 1-β = .17. Tukey post hoc tests showed significantly lower overall central accuracy for those aged 18-33 ($M = 11.14$, $SD = 4.57$) than other age groups; 34-49

Table 1. Free Recall: Means and Standard Deviations for: Protocol x Age for Overall FR Accuracy, Central Accuracy, Peripheral Accuracy, Overall Confabulations, Distortions and Fabrications (N = 198).

Protocol	Age	Overall FR accuracy		Central accuracy		Peripheral accuracy		Overall confabulations		Distortions		Fabrications	
		M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Control	18-33	11.52	4.56	10.68	3.99	0.84	0.90	2.00	1.66	0.59	0.85	1.41	1.50
	34-49	13.93	5.43	12.45	4.42	1.48	1.69	2.95	2.21	1.23	1.41	1.73	1.32
	50-65	14.68	6.05	12.95	5.20	1.73	1.56	1.91	1.77	0.95	1.36	0.95	1.13
	Total	13.38	5.47	12.03	4.60	1.35	1.45	2.29	1.93	0.92	1.24	1.36	1.34
Rapport	18-33	11.34	5.82	10.34	4.92	1.00	1.18	2.05	1.89	1.05	1.56	1.00	1.16
	34-49	13.93	5.58	12.52	4.75	1.41	1.33	2.09	1.54	1.23	1.15	0.86	1.13
	50-65	16.41	5.07	14.68	4.29	1.73	1.48	2.59	1.76	1.32	1.36	1.27	1.16
	Total	13.89	5.80	12.52	4.92	1.38	1.35	2.24	1.73	1.20	1.35	*1.05	1.14
Verbal instruction	18-33	13.05	5.11	12.41	4.70	0.64	0.93	*4.09	3.35	1.36	1.43	***2.73	2.62
	34-49	16.34	5.78	14.70	4.81	1.64	1.64	*1.55	1.74	1.55	1.41	***1.73	0.87
	50-65	16.82	5.34	14.59	4.26	2.23	1.72	*1.64	1.71	1.18	1.37	**1.46	1.24
	Total	15.40	5.59	13.90	4.65	1.50	1.60	3.03	2.47	1.36	1.39	*1.67	1.88
Total	18-33	***11.97	5.17	***11.14	4.57	***0.83	1.01	2.71	2.58	1.00	1.34	1.71	1.98
	34-49	**14.73	5.63	*13.23	4.71	*1.51	1.54	2.53	1.86	1.33	1.32	1.20	1.17
	50-65	***15.97	5.50	***14.08	4.60	***1.89	1.58	2.32	1.75	1.15	1.35	1.17	1.17
	Total	14.22	5.66	12.82	4.77	1.41	1.46	2.52	2.09	1.16	1.33	1.36	1.50

*p < .05, **p < .01, ***p < .001.

Table 2. Cued Recall: Means and Standard Deviations for Protocol X Age for Overall CR accuracy, Overall CR confidence, W-S C-A, NDL accuracy, NDL confidence, DL accuracy and DL confidence (N = 198).

Protocol	Age	Overall CR accuracy		Overall CR confidence		W-S C-A		NDL accuracy		NDL confidence		DL accuracy		DL confidence	
		M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Control	18-33	13.27	2.25	77.32	13.50	.22	.23	7.23	1.07	38.32	6.28	6.05	1.84	39.00	9.13
	34-49	13.50	2.54	78.68	13.04	.34	.19	7.23	1.31	38.27	6.22	6.27	1.70	40.41	7.56
	50-65	13.45	2.20	81.55	16.49	.20	.24	7.45	1.53	38.95	7.74	6.00	1.77	42.59	9.37
	Total	13.41	2.30	79.18	14.32	.25	.23	7.30	1.30	38.52	6.69	6.11	1.75	40.67	8.71
Rapport	18-33	13.18	1.84	75.73	12.75	.22	.29	7.00	1.23	36.09	6.10	6.18	1.62	39.64	7.74
	34-49	13.14	2.46	76.91	13.23	.26	.22	7.27	1.32	37.36	6.84	5.86	1.88	39.55	8.02
	50-65	12.82	2.08	78.86	16.00	.26	.25	7.00	1.07	37.59	8.06	5.82	1.68	41.27	9.38
	Total	13.05	2.12	77.17	13.91	.25	.26	7.09	1.20	37.02	6.97	5.95	1.71	*40.15	8.32
Verbal instruction	18-33	12.91	2.47	83.05	15.50	.21	.24	6.82	1.65	40.41	9.49	6.09	1.72	42.64	6.83
	34-49	14.64	2.65	89.09	13.71	.30	.22	7.50	1.47	**42.41	7.17	7.14	1.73	46.68	7.61
	50-65	12.59	2.42	75.86	16.92	.36	.22	6.68	1.04	**37.82	7.90	5.91	1.87	38.05	10.40
	Total	13.38	2.64	82.67	16.14	.29	.23	7.00	1.44	40.21	8.33	6.38	1.83	*42.45	9.02
Total	18-33	13.12	2.17	78.70	14.11	.22	.25	7.02	1.33	38.27	7.55	6.11	1.70	40.42	7.99
	34-49	13.76	2.59	81.56	14.20	.30	.21	7.33	1.35	39.35	7.01	6.42	1.82	42.21	8.26
	50-65	12.95	2.24	78.76	16.39	.28	.25	7.05	1.26	38.12	7.80	5.91	1.75	40.64	9.77
	Total	13.28	2.35	79.67	14.92	.26	.24	7.13	1.31	38.58	7.44	6.15	1.76	41.09	8.70

*p < .05, **p < .01.

(M = 13.23, SD = 4.71) (p = .03) and 50-65 (M = 14.08, SD = 4.60) (p = .001) respectively. No other comparisons were significant, p > .05; see Table 1.

Overall peripheral accuracy

Kolmogorov-Smirnov normality test and histograms showed residuals were only normally distributed for age 50–65 in control and verbal instruction (p > .05). However, Levene's test showed homogeneity of variance (p = .19). There was a significant main effect for age on overall peripheral accuracy, F $(2,198)$ = 9.67, p < .001, ηp^2 = .09, 1-β = .98. However, there was no effect for protocol, F $(2,198)$ = .21, p > .05, ηp^2<.01, 1-β = .08, and no significant interaction was observed, F $(4,198)$ = 9.67, p > .05, ηp^2 = .09, 1-β = .99. Tukey post hoc tests showed significantly lower overall peripheral accuracy for those aged 18–33 (M = .83, SD = 1.01) than other age groups; 34-49 (M = 1.51, SD = 1.54) (p = .02) and 50-65 (M = 1.89, SD = 1.58) (p < .001). No other comparisons were significant, p > .05; see Table 1.

Overall confabulations

There was a significant main effect for protocol on overall confabulations, F $(2,198)$ = 3.14, p = .046, ηp^2 = .03, 1-β = .60. Tukey post hoc tests were not significant due to marginal significance. However, a significant interaction between protocol and age was also observed F $(4,198)$ = 3.09, p = .02, ηp^2 = .06, 1-β = .81. Pairwise comparisons showed a significantly higher number of confabulations in the verbal instruction condition for those aged 18–33 (M = 4.09) than those aged 34-49 (M = 1.55), 95% CI [.07-3.02]) (p = .04) and 50-65 (M = 1.64) 95% CI [.16-3.11]) (p = .02), with a medium effect size (Cohen's d = 0.65). There was no effect for age, F $(2,198)$ = .62, p = .54, ηp^2<.01, 1-β = .15 and no other comparisons were significant, p > .05; see Table 1.

Distortions

There was no significant main effect of protocol for distortions, F $(2,198)$ = 1.82, p > .05, ηp^2 = .02, 1-β = .38. No effect for age was found, F $(2,198)$ = 1.03, p > .05, ηp^2<.01, 1-β = .23 and no significant interaction was observed, F $(4,198)$ = .44, p > .05, ηp^2<.01, 1-β = .15; p > .05; see Table 1.

Fabrications

There was a significant main effect for protocol on fabrications, F $(2,198)$ = 3.12, p = .047, ηp^2<.03, 1-β = .59. Tukey post hoc tests showed significantly higher fabrications for those in verbal instruction (M = 1.67, SD = 1.88) than in rapport (M = 1.05, SD = 1.14) (p = .04). A significant interaction was also observed, F $(4,198)$ = 4.17, p = .003, ηp^2 = .08, 1-β = .92. Pairwise comparisons showed a significantly higher number of fabrications in the verbal instruction condition for those aged 18-33 (M = 2.73) than those aged 34-49 (M = 1.73), 95% CI [.69-2.77]) (p < .001) and 50–65 (M = 1.46) 95% CI [.41-2.50]) (p = .003), with a medium effect size (Cohen's d = 0.65). There was no significant effect for age,

F (2,198) = 3.03, p = .05, ηp^2 = .03, 1-β = .58 and no other comparisons were significant (p > .05); see Table 1.

Overall cued recall accuracy

There were no significant main effects for protocol on overall cued recall accuracy, F (2,198) = .49, p > .05, ηp^2<.01, 1-β = .13, or for age group, F (2,198) = 2.17, p > .05, ηp^2 = .02, 1-β = .44. No significant interaction was observed, F (4,198) = 1.47, p > .05, ηp^2 = .03, 1-β = .45; see Table 2.

For a sensitive evaluation of the different question types a further analysis was undertaken on DL and NDL accuracy.

Directive leading question (DL) accuracy

Kolmogorov-Smirnov normality test and histograms showed residuals were normally distributed for control 50–65 and verbal instruction 34–49 (p > .05). Levene's test showed homogeneity of variance (p = .28). There was no significant main effect for protocol on DL question accuracy, F (2,198) = .92, p > .05, ηp^2 = .01, 1-β = .21. No effect was observed for age, F (2,198) = 1.18, p > .05, ηp^2 = .01, 1-β = .26 and no significant interaction was observed, F (4,198) = 1.56, p > .05, ηp^2 = .02, 1-β = .28; see Table 2.

Non-directive leading question (NDL) accuracy

Kolmogorov-Smirnov normality test and histograms showed residuals were not normally distributed for control 18–33 or verbal instruction 18-33 and 34-49 (p < .05). Levene's test showed homogeneity of variance (p = .94). There was no significant main effect for protocol on NDL question accuracy, F (2,198) = .99, p > .05, ηp^2 = .01, 1-β = .22, or for age, F (2,198) = 1.44, p=>.05, ηp^2 = .02, 1-β = .31. No significant interaction was observed, F (4,198) = 1.06, p=>.05, ηp^2 = .02, 1-β = .33; see Table 2.

Confidence

Overall cued recall confidence

There was no significant main effect for protocol on overall cued recall confidence, F (2,198) = 2.38, p > .05, ηp^2 = .03, 1-β = .48, nor for age, F (2,198) = .82, p > .05, ηp^2<.01, 1-β = .19. No significant interaction was observed, F (4,198) = 2.20, p > .05, ηp^2 = .05, 1-β = .64; see Table 2.

Pearson's correlation found a positive relationship between overall cued recall confidence and overall cued recall accuracy, r (198) = .18, p = .02. No relationship was found between free recall accuracy and cued recall accuracy, r (198) = .12, p = .10.

For a sensitive evaluation of the different question types a further analysis was undertaken on DL and NDL confidence.

Directive leading question (DL) confidence

There was a significant main effect of protocol for DL question confidence, $F (2,198) =$ 3.10, $p = .048$, $\eta p^2 = .03$, 1-β = .59. However, there was no main effect for age, $F (2,198) =$ .54, $p > .05$, $\eta p^2 < .01$, 1-β>.14, and no significant interaction was observed, $F (4,198) =$ .96, $p > .05$, $\eta p^2 < .02$, 1-β>.30. Tukey post hoc tests showed significantly higher DL question confidence scores in verbal instruction ($M = 42.45$, $SD = 9.02$) than rapport ($M = 40.15$, $SD = 8.32$) ($p = .048$). No other comparisons were significant, $p > .05$; see Table 2.

A Pearson's correlation found an overall positive relationship between DL question confidence and DL question accuracy, $r (198) = .24$, $p = .001$. On further investigation, a strong relationship was shown for DL question confidence and DL question accuracy in the verbal instruction condition, $r (198) = .71$, $p < .001$.

Non directive leading question (NDL) confidence

Kolmogorov-Smirnov normality test and histograms showed residuals were normally distributed ($p > .05$) except verbal instruction 18-33 ($p < .05$). However, Levene's test showed homogeneity of variances ($p = .58$). There was no significant main effect for protocol, $F (2,198) = 1.33$, $p > .05$, $\eta p^2 = .01$, 1-β = .29 and no effect was shown for age, $F (2,198) = .87$, $p > .05$, $\eta p^2 < .01$, 1-β = .20. A significant interaction for protocol and age was observed for NDL question confidence, $F (4,198) = 3.04$, $p = .02$, $\eta p^2 = .06$, 1-β = .8. Pairwise comparisons showed significantly higher confidence for verbal instruction in those aged 34-49 ($M = 42.41$, 95% CI [43.10-50.27]) than those aged 50-65 ($M = 37.82$, 95% CI [34.46-41.63]) ($p = .003$) with a large effect size (Cohen's $d = 0.95$). No other comparisons were significant ($p > .05$); see Table 2. Pearson's correlation showed a positive relationship between NDL question confidence and NDL question accuracy, $r(198) = .17$, p = .01. No other comparisons were significant, $p > .05$.

Within-subjects confidence-accuracy (W-S C-A)

No significant main effect for protocol on W-S C-A was found, $F (2,198) = .66$, $p > .05$, $\eta p^2 < .01$, 1-β = .16. No effect for age was found, $F (2,198) = 2.05$, $p > .05$, $\eta p^2 = .02$, 1-β = .42 and no significant interaction was observed, $F (4,198) = 1.31$, $p > .05$, $\eta p^2 = .03$, 1-β = .41; see Table 2.

Discussion

This study examined the impact age and interview protocols have on accuracy and confidence measures, using both free recall (FR) and cued recall (CR). In free recall accounts, confabulation measures were also investigated.

Free recall accuracy

In contradiction to previous research, which found younger adults gave the most complete and accurate accounts (Gawrylowicz et al., 2014; List, 1986), the finding in this study

rejects the hypothesis that fewer accurate responses would be observed for older participants (H1), with lower accuracy shown in the 18-33 condition (H2). It is possible that those aged 18-33 were less efficient at global processing, as Roux and Ceccaldi (2001) proposed, and that older adults are more biased towards a general bigger-picture approach (Navon, 1977) resulting in reasonably equivalent performance across the groups. Previous research suggests less detail will be recalled by older participants, not those who are younger, with distraction by local processing detail articulated as an explanation (Ebaid and Crewther, 2019; Insch et al., 2012; Oken et al., 1999). While short-term memory is thought to peak at around age 22 (Hartshorne and Germine (2015), the current study found fine detail associated with local processing may be best retrieved by those aged above 34. However, whilst age plausibly impacts accuracy, it appears an element of individual cognition may be involved. For example, as a result of no time pressure and good-quality images taken in good lighting, from the same viewpoint, on the same day, and so on (Megreya and Burton, 2006). In addition, context variations in how information is presented can also be relevant (Megreya and Burton, 2006; Searcy et al., 2010). In the study reported here, in those aged 18-33, overall central and peripheral accuracy was lower. Further, one would expect that peripheral accuracy will not be as accurate as central accuracy (Burke et al., 1992); the distinction between peripheral and central was not evident in the 18–33 group.

The study was conducted online rather than face-to-face which may provide a context variation influencing outcome. The finding that overall central and peripheral accuracy was lower may reflect a technological age where those aged 18–33 are not used to writing as much as older participants. However, it has been shown that information quality is best measured by details, not the number of words (Elntib et al., 2015; Warmelink et al., 2019); thus, this particular issue is unlikely to have made a significant difference to the study outcome. While some participants aged 18–33 did give detailed accounts, the opportunity to avoid any confusion regarding study instructions was not possible. For example, if the study had been carried out face-to-face any problems with instructions could have been clarified. Indeed, social interactions in interviews have been found to be important in obtaining detailed, complete and comprehensible accounts (North et al., 2008).

Confabulations: Fabrications and distortions

Significant differences were found for age on confabulations and fabrications, lending partial support for H3. As above, the interaction between age and protocol for overall confabulations and fabrications in respect of those aged 18–33 in the verbal instruction condition was unexpected and contradicts previous research (Attali and Dalla Barba, 2013). The findings may reflect, and also be explained by, inattention at the point of encoding by this younger age group. Some fabrications were embellishment of what had been seen; with participants developing explanations for occurrences, rather than simply reporting what was seen, e.g., "maybe cramp", "in their blind spot", "weighing up his chances of escape". However, a small number of participants reported that the driver must have been intoxicated, that the keys were found nearby, or the car was being stolen, albeit the spread across age and protocol suggest this was not an effect of the manipulation. One explanation is stereotyping based on behaviour interpretation, something crucial to

navigating the complexities of social existence (Westra, 2019). Top-down processing draws on pre-existing knowledge and schema to provide an explanation for the behaviour witnessed, bottom-up processing applies this to what is being seen (Turner, 2015). In a study by McGlothlin and Killen (2010), child participants were more likely to interpret something picked up as being stolen than someone helping, described by Westra as "divergent moral judgment" (2019: 2823). Similarly, there are interpretations of someone escaping the scene of the crime; thus, despite this study using only adult participants, Westra's findings may still have relevance.

Moreover, several people reported either a child, a dog or people playing football or sport; none of which were shown, again with no pattern for participant age or protocol. One cannot be sure, therefore, that participants followed the instructions. However, perhaps schemas were employed for grass areas or parks; adding in people playing football, dogs being walked and the presence of children to fill memory gaps with what is 'usually there' (Abelson, 1981; Ormerod and Adler, 2010; Rae Tuckey and Brewer, 2003). Another reported that the jogger wasn't wearing ear buds, maybe applying a jogging schema, in reverse. Schemas play an important role in the accuracy of eyewitness accounts, with some suggesting descriptions that do not fit stereotypes are more likely to be accurate (Rae Tuckey and Brewer, 2003). It is also argued that older people rely more on schema-based processing (Overman et al., 2013) albeit, as previously noted, no evidence was found that effects were age-related. These observations appear therefore to be individual participants employing schemas to infill memory gaps. Another consideration is that the study was conducted online due to COVID-19 restrictions and it is thus possible that engagement of the 18–33 age group may have been influenced by this contextual variation. For example, Romero-Rodríguez et al. (2023) found that, in University students, learning was affected by digital fatigue rated as medium-high; though extrapolating these variables is beyond the scope of this paper.

Cued recall accuracy

Contrary to hypotheses H4 and H5, no effects were found for interview protocol or age on cued recall accuracy scores; neither for directive-leading, or non-directive leading questions. It is unclear whether the method of delivering the context reinstatement in verbal instruction was relevant to H4. Whilst the pre-recorded context reinstatement added ecological validity, the lack of social interaction with participants made it impossible to assess attention at an early stage of retrieval. Effects may have been different if the study had been conducted face-to-face, where assessment of attention at the encoding and retrieval stages would have been possible. Integration of context reinstatement instructions within the rapport stage of investigative interviews might maximise the opportunity to reduce reluctance of witnesses (Gous and Wheatcroft, 2020; Hershkowitz et al., 2006; Saywitz et al., 2019). As discussed, Caso et al. (2024) advocate that free recall followed by relevant cued recall probing questions is likely to retrieve the most reliable information. Combining approaches may elicit the most thorough and accurate accounts possible.

In forensic settings cued recall questions (such as directive-leading and non-directive leading) are asked as part of prompting memory retrieval. Face-face questioning ensures

witnesses are fully engaged and focused before questions are asked and retrieval attempted. However, whilst face-face questioning ensures focus, it significantly increases pressure on the witness. In order to counter this, special measures in court allow vulnerable witnesses to give evidence indirectly via a video-link (Government, 1988). Whilst it is proposed that barristers dislike the video-link process because they are unable to test evidence directly (Davies and Westcott, 2018), particularly when asking leading questions (Valentine and Maras, 2011), Doherty-Sneddon and McAuley (2000) found that younger children became more accurate and resistant to leading questions when a video-link was used. Whilst no research has been found exploring the impact of video-link on adult witnesses, conducting this study remotely, with participants reading questions before typing their answers and with no social interaction with the researcher, conceivably could have impacted outcomes. It is important to recognise that this possibility could inform the emergence of the digital-legal space.

Cued recall confidence

Support was shown for H6. Higher confidence was found for DL questions in the verbal instruction condition compared to the rapport condition. Erikson's theory of development may, in part, explain the higher average confidence scores for those aged 34–49 (Brandau and Evanson, 2018; Parrish, 2014). As noted, DL questions are more likely reflective of 'lawyerese'; complex multi-faceted questions used in cross-examination to coerce or incite a desired response (Brennan, 1995; Wheatcroft and Woods, 2010). Whilst the content of the question can be confusing, the assertive nature of these questions can make people feel more confident in their answers. One explanation, in consideration of Cialdini's (2004) principles of compliance and conformity, is that individuals may fail to respond in accordance with their private judgements in relation to confidence in the face of pressure; that persuasive engagement can result in a positive sense of self, expressed through increased attribution of confidence to DL questions. Plus, the answers to these types of questions are likely to impact jurors' perceptions of witness credibility (Wheatcroft et al., 2004). Indeed, six participants later contacted the researcher asking whether a mistake had been made in the wording of DL questions as these participants considered the question to be confusing. One aim of the research was to explore confidence levels expressed by participants when they were answering different types of leading questions. Participants in the verbal instruction condition were significantly more confident in answering DL questions, reflecting previous commentary that such question forms can increase confidence (Gous and Wheatcroft, 2020; Wheatcroft, 2018). At first sight, it appears that a context reinstatement type verbal instruction produced higher levels of confidence. However, only if the relationship between confidence and accuracy is a positive one is this meaningful, because the greater the positive relationship between confidence and accuracy the more certainty one can have in the answers being correct. This is particularly important in legal contexts where the accuracy of evidence is paramount. An overall positive relationship between DL confidence and DL accuracy was found, suggesting those who expressed higher levels of confidence to these types of questions can also be more accurate. On further investigation, the verbal instruction condition appeared to create the conditions where higher levels of confidence expressed to

DL questions were more likely to be accurate. Overall, protocol and age saw no effects on W-S C-A specifically (H7).

In support of the DL finding, research has shown that using question types in court preparation can improve the confidence of witnesses when they respond to DL questions used during cross-examination (Wheatcroft and Woods, 2010). Therefore, though DL questions can cause confusion, it appears that context reinstatement type instructions can increase witness confidence whilst not significantly impairing the relationship between confidence and accuracy. This finding has real-world import, as jurors assess witness credibility against their confidence in answering questions (Brodsky et al., 2010; Maricchiolo et al., 2009; Sporer, 1993). Thus, the advantage of increased confidence appears to be from the way witnesses present themselves before juries and their resultant perceived credibility. Whilst this presentation does not seem to always accord with accuracy, in this study at least, the verbal instruction was helpful in this respect.

Limitations

Clear directions were given in the task completion instructions to use a computer with keyboard and screen and to complete the experiment in a quiet space. However, as there was little control over compliance, any inconsistency in image size and environment may have been a confound, as well as how much information could be typed with ease in the free recall account if a keyboard was not used. It is not certain that every detail remembered was recorded, and it is possible that some participants may have 'moved on' when they became bored or distracted. Equally, participants may have typed less because they remembered less, with no reflection on time spent or typing ability. It also cannot be dismissed that in forensic settings witnesses often do not comprehend the relevance of what they are seeing until later questioned, whereas in an experimental study they are perhaps intuitively aware that they need to remember something.

Whilst the time of day the experiment was completed was not recorded, this may have impacted on whether participants were fully alert. Lighting conditions at the time of viewing the video clip, for example, the use of artificial lighting or levels of natural light and how this reflected on the screen, may also have impacted results. Whilst visual acuity was requested and participants asked to only complete the experiment with corrected to normal vision, this was not verifiable, nor was the distance from which the screen was viewed. These features, that would be reported, for example, in witness identification, are difficult to operationalise outside of the laboratory setting.

Future research

Face-to-face replication would ensure consistency and allow the free recall account to be accurately timed, ensuring any differences found were more likely due to memory retrieval. It is also possible to add a note-taker condition or facility to record participant free recall, adding ecological validity by recording accounts in the same way as investigative interviewers might obtain statements from witnesses. In this study, a female voice read the verbal instruction. Replications might include using a male voice or for participants to read the context reinstatement themselves. In addition, study replication with the perceived

pressure of face-to-face questioning, and using a mix of male and female voices, would assess the impact on witnesses relative to real world settings.

Age has been found to be relevant in this study. Therefore, replication with children aged 6-17 would explore any additional differences in age related accuracy and confidence. Such a study would also provide evidence to discern whether inclusion of a context reinstatement type verbal instruction with age related children increases the quality and accuracy of evidential accounts, or whether rapport is more important. Indeed, additional factors may be identified that are only relevant to child witnesses. These aspects are currently under investigation.

Conclusion

The findings of this study have implications for operational interviewing, research, and training. In contradiction to previous research and our predictions, this study found the 18–33-year group to be the least accurate when providing free recall. As free recall is the recommended approach to obtaining a first account it would be helpful for research to corroborate the outcome using face-to-face methodology before firm recommendations can be made. Nevertheless, this group were also most susceptible to confabulation when VI was used, indicating a need for a statement strengthening the requirement not to make-up details be incorporated into context reinstatement instructions, and for this addition to be included in *all* investigative interview training.

Against predictions, interview protocol did not impact accuracy in free recall, VI did not increase accuracy in overall cued recall, W-S C-A remained unaffected by interview protocol and age, and directive leading questions did not decrease accuracy. In support of previous research, this study did show VI increased confidence for directive leading questions compared to the rapport condition. However, increased confidence to these types of questions is not a positive outcome in the context of forensic interviewing (as generally speaking high confidence does not necessarily accord with accuracy), highlighting the need to ensure directive forms of leading questions are avoided in investigative interviews. Clearly, ordinary adult witnesses are detrimentally affected by such problematic questions. Importantly, the findings demonstrate that rapport appeared to mitigate against inflated confidence to some extent. Finally, effective rapport-building is essential to guard against the impact of such questions on witness' expressed confidence.

ORCID iDs

Kaye N. Cooke https://orcid.org/0009-0004-2774-5632
Jacqueline M. Wheatcroft https://orcid.org/0000-0001-7212-1598

Consent to participate

Author contributions

Kaye Cooke (Conceptualisation; Data curation; Formal analysis; Methodology; Writing, original draft; Writing, review & editing). Jacqueline Wheatcroft (Conceptualisation; Methodology; Writing, review & editing).

Funding

Declaration of conflicting interests

Data Availability Statement

Data is available on request and may be shared privately for review: https://figshare.com/s/35ed0f0c45abe97c251a

Open science framework

This study was not registered prior to execution.

Note

1. In relation to mental workload, perceptual load is the mental effort required to process physical stimuli, while cognitive load is the effort placed on working memory, such as problem-solving or complex thinking (Liu et al., 2022).

References

Abbe A and Brandon SE (2013) The role of rapport in investigative interviewing: a review. *Journal of Investigative Psychology and Offender Profiling* 10(3): 237–249.

Abelson RP (1981) Psychological status of the script concept. *American Psychologist* 36(7): 715–729.

Adams-Price C (1992) Eyewitness memory and aging. Predictors of accuracy in recall and person recognition. *Psychology and Aging* 7(4): 602–608.

Aizpurua A, Garcia-Bajos E and Migueles M (2009) False memories for a robbery in young and older adults. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory & Cognition* 23(2): 174–187.

Allwood CM, Helene Innes-Ker Å, Homgren J, et al. (2008) Children's and adults' realism in their event-recall confidence in responses to free recall and focused questions. *Psychology, Crime and Law* 14(6): 529–547.

Antes JR and Mann SW (1984) Global-local precedence in picture processing. *Psychological Research* 46(3): 247–259.

Attali E and Dalla Barba G (2013) Confabulation in healthy aging is related to poor encoding and retrieval of over-learned information. *Aging, Neuropsychology, and Cognition* 20(3): 339–355.

Bell K, Fahmy E and Gordon D (2016) Quantitative conversations: the importance of developing rapport in standardised interviewing. *Quality and Quantity* 50(1): 193–212.

Berkowitz SR, Garrett BL, Fenn KM, et al. (2021) Eyewitness confidence may not be ready for the courts: a reply to Wixted et al. *Memory* 1(2): 1–2.

Berkowitz SR, Garrett BL, Fenn KM, et al. (2022) Convicting with confidence? Why we should not over-rely on eyewitness confidence. *Memory* 30(1): 10–15.

Brandau M and Evanson TA (2018) Adolescent victims emerging from cyberbullying. *Qualitative Health Research* 28(10): 1584–1594.

Brennan M (1995) The discourse of denial. Cross-examining child victim witnesses. *Journal of Pragmatics* 23(1): 71–91.

British Psychological Society (2021a) Ethics Guidelines for internet-mediated Research. *British Psychological Society*: Leicester. Available at: https://www.bps.org.uk/guideline/ethics-guidelines-internet-mediated-research.

British Psychological Society (2021) *Code of Ethics and Conduct.* British Psychological Society. Available at: https://www.bps.org.uk/guideline/code-ethics-and-conduct

Brodsky SL, Griffin MP and Cramer RJ (2010) The witness credibility scale: an outcome measure for expert witness research. *Behavioural Science & Law* 28(6): 892–907.

Brouillard C, Gabbert F and Scott AJ (2024) Addressing current issues in assessing professional rapport: a systematic review and synthesis of existing measures. *Applied Cognitive Psychology* 38(3): e4205.

Burke A, Heuer F and Reisberg D (1992) Remembering emotional events. *Memory & Cognition* 20: 277–290.

Butler BJ and Loftus EF (2018) Discrepancy detection in the retrieval-enhanced suggestibility paradigm. *Memory* 26(4): 483–492.

Caso A, Gabbert F and Dando CJ (2024) Eyewitness confidence in the interviewing context: understanding the impact of question type and order. *Applied Cognitive Psychology* 38(3): e4197.

Chan JCK, Thomas AK and Bulevich JB (2009) Recalling a witnessed event increases eyewitness suggestibility: the reversed testing effect. *Psychological Science* 20(1): 66–73.

Christianson S-Å and Loftus EF (1991) Remembering emotional events: the fate of detailed information. *Cognition & Emotion* 5(2): 81–108.

Cialdini RB and Goldstein NJ (2004) Social influence; compliance and conformity. *Annual Review of Psychology* 55(1): 591–621.

Cohen J (1992) A power primer. *Psychological Bulletin* 112(1): 155–159.

Cohen G and Faulkner D (1989) Age differences in source forgetting: effects on reality monitoring and on eyewitness testimony. *Psychology and Aging* 4(1): 10–17.

Collins K and Carthy N (2019) No rapport, no comment: the relationship between rapport and communication during investigative interviews with suspects. *Journal of Investigative Psychology and Offender Profiling* 16(1): 18–31.

Collins K, Doherty-Sneddon G and Doherty MJ (2014) Practitioner perspectives on rapport building during child investigative interviews. *Psychology, Crime and Law* 20(9): 884–901.

Collins R, Lincoln R and Frank MG (2002) *The Effect of Rapport in Forensic Interviewing. Psychiatry, Psychology and Law, 9(1), 69–78.*

Criminal Justice Act (1988). Retrieved from: https://www.legislation.gov.uk/ukpga/1988/33/contents

Dando C, Wilcock R and Milne R (2008) The cognitive interview: inexperienced police officers' perceptions of their witness/victim interviewing practices. *Legal and Criminological Psychology* 13(1): 59–70.

Dando C, Wilcock R and Milne R (2009) The cognitive interview: novice police officers' witness/victim interviewing practices. *Psychology, Crime and Law* 15: 679–696.

Dando C, Taylor DA, Caso A, et al. (2023) Interviewing in virtual environments: towards understanding the impact of rapport-building behaviours and retrieval context on eyewitness memory. *Memory & Cognition* 51(2): 404–421.

Darling S, Martin D, Hellmann JH, et al. (2009) Some witnesses are better than others. *Personality and Individual Differences* 47(4): 369–373.

Davies GM and Westcott HL (2018) Safeguarding vulnerable witnesses. In: Davies G and Beech A (eds) *Forensic Psychology: Crime, Justice, Law, Interventions*. 3rd edition. John Wiley, 399–425.

Dennis NA, Kim H and Cabeza R (2007) Effects of aging on true and false memory formation: an fMRI study. *Neuropsychologia* 45(14): 3157–66. doi: 10.1016/j.neuropsychologia.2007.07.003.

Dennis NA, Kim H and Cabeza R (2008) Age-related differences in brain activity during true and false memory retrieval. *Journal of Cognition & Neuroscience* 20(8): 1390–1402. doi: 10.1162/jocn.2008.20096.

Dianiska RE, Swanner JK, Brimbal L, et al. (2019) Conceptual priming and context reinstatement: a test of direct and indirect interview techniques. *Law and Human Behavior* 43(2): 131–143.

Doherty-Sneddon G and McAuley S (2000) Influence of video-mediation on adult–child interviews: implications for the use of the live link with child witnesses. *Applied Cognitive Psychology* 14(4): 379–392.

Eades D (2012) The social consequences of language ideologies in courtroom cross-examination. *Language in Society* 41(4): 471–497. https://www-jstor-org.glos.idm.oclc.org/stable/41682224?seq=1#metadata_info_tab_contents

Ebaid D and Crewther SG (2019) Visual information processing in young and older adults. *Frontiers in Aging Neuroscience* 11: 116.

Elntib S, Wagstaff GF and Wheatcroft JM (2015) The role of account length in detecting deception in written and orally produced autobiographical accounts using reality monitoring. *Journal of Investigative Psychology and Offender Profiling* 12(2): 185–198.

Erikson EH (1950) *Childhood and Society*. W. W. Norton.

Evans JR and Fisher RP (2011) Eyewitness memory: balancing the accuracy, precision and quantity of information through metacognitive monitoring and control. *Applied Cognitive Psychology* 25(3): 501–508.

Faber P, Nielsen NP and Berntsen D (2023) Effects of mental context reinstatement on accuracy and recollective experience. *Applied Cognitive Psychology* 37(5): 1004–1015.

Fandakova Y, Sander MC, Grandy TH, Cabeza R, Werkle-Bergner M and Shing YL (2018) Age differences in false memory: The importance of retrieval monitoring processes and their modulation by memory quality. *Psychology & Aging* 33(1): 119–133. doi: 10.1037/pag0000212.

Fisher RP and Geiselman RE (1992) *Memory Enhancing Techniques for Investigative Interviewing: The Cognitive Interview. Springfield, Ill.* Charles C. Thomas.

Förster J and Dannenberg L (2010) GLOMOsys: a systems account of global versus local processing. *Psychological Inquiry* 21(3): 175–197.

Foster I, Talwar V and Crossman A (2023) The role of rapport in eliciting children's truthful reports. *Applied Developmental Science* 27(3): 221–237.

Fox SG and Walters HA (1986) The impact of general versus specific expert testimony and eyewitness confidence upon mock juror judgment. *Law and Human Behavior* 10(3): 215–228.

Gabbert F, Hope L, Luther K, et al. (2020) Exploring the use of rapport in professional information-gathering contexts by systematically mapping the evidence base. *Applied Cognitive Psychology* 35(2): 329–341.

Gawrylowicz J, Memon A, Scoboria A, et al. (2014) Enhancing older adults' eyewitness memory for present and future events with the self-administered interview. *Psychology and Aging* 29(4): 885–890.

Giles S, Alison L, Christiansen P, et al. (2021) An economic evaluation of the impact of using rapport-based interviewing approaches with child sexual abuse suspects. *Frontiers in Psychology* 12: 778970.

Gous G and Wheatcroft JM (2020) Directive leading questions and preparation technique effects on witness accuracy. *Sage Open* 10(1): 2158244019899053.

Gudjonsson GH (2003) *The Psychology of Interrogations and Confessions: A Handbook.* John Wiley & Sons.

Gudjonsson GH and Clare ICH (1995) The relationship between confabulation and intellectual ability, memory, interrogative suggestibility and acquiescence. *Personality and Individual Differences* 19(3): 333–338.

Gudjonsson GH and Clark NK (1986) A theoretical model of interrogative suggestibility. *Social Behaviour* 1: 83–104.

Gudjonsson GH and Young S (2010) Does confabulation in memory predict suggestibility beyond IQ and memory? *Personality and Individual Differences* 49(1): 65–67.

Guy J, Mottron L, Berthiaume C, et al. (2019) A developmental perspective of global and local visual perception in autism spectrum disorder. *Journal of Autism and Developmental Disorders* 49(7): 2706–2720.

Hartshorne JK and Germine LT (2015) When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological Science* 26(4): 433–443.

Henderson E (2016) Best evidence or best interests? What does the case law say about the function of criminal cross-examination? *International Journal of Evidence and Proof* 20(3): 183–199.

Henry JD, Terrett G, Grainger SA, et al. (2020) Implementation intentions and prospective memory function in late adulthood. *Psychology and Aging* 35(8): 1105–1114.

Hershkowitz I, Orbach Y, Lamb ME, et al. (2006) Dynamics of forensic interviews with suspected abuse victims who do not disclose abuse. *Child Abuse & Neglect* 30(7): 753–769.

Howe ML and Knott LM (2015) The fallibility of memory in judicial processes: lessons from the past and their modern consequences. *Memory* 23(5): 633–656.

Huff M, Schwan S and Garsoffky B (2011) Recognizing dynamic scenes: influence of processing orientation. *Perceptual & Motor Skills* 112(2): 429–439.

Insch PM, Bull R, Phillips LH, et al. (2012) Adult aging, processing style, and the perception of biological motion. *Experimental Aging Research* 38(2): 169–185.

Kassin SM, Tubb VA, Hosch HM, et al. (2001) On the "general acceptance" of eyewitness testimony research. *American Psychologist* 56(5): 405–416.

Kassin SM, Leo RA, Meissner CA, et al. (2007) Police interviewing and interrogation: a self-report survey of police practices and beliefs. *Law and Human Behavior* 31(4): 381–400.

Kaye H and Tree J (2016) Investigating memory: experimental and clinical investigations of remembering and forgetting. In: Ness H, Kaye H and Stenner P (eds) *Investigating Psychology*. The Open University, Vol. 3, 48–99.

Kebbell MR and Gilchrist EA (2010) Eliciting evidence from eyewitnesses for court proceedings. In: Adler JR and Gray JM (eds) *Forensic Psychology Concepts, Debates & Practice*. Routledge.

Kebbell MR and Giles DC (2000) Some experimental influences of lawyers' complicated questions on eyewitness confidence and accuracy. *Journal of Psychology* 134(2): 129–139.

Kebbell MR and Johnson SD (2000) Lawyers' questioning: the effect of confusing questions on witness confidence and accuracy. *Law and Human Behavior* 24(6): 629–641.

Kebbell MR and Milne R (1998) Police officers' perceptions of eyewitness performance in forensic investigations. *The Journal of Social Psychology* 138(3): 323–330.

Kimchi R (1992) Primacy of wholistic processing and global/local paradigm: a critical review. *Psychological Bulletin* 112(1): 24–38.

Koriat A and Goldsmith M (1996) Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review* 103(3): 490–517.

Lanciano T and Curci A (2011) Memory for emotional events: the accuracy of central and peripheral details. *Europe's Journal of Psychology* 7(2): 323.

Lavoie J, Wyman J, Crossman AM, et al. (2021) Meta-analysis of the effects of two interviewing practices on children's disclosures of sensitive information: rapport practices and question type. *Child Abuse & Neglect* 113: 104930.

Li S-C, Naveh-Benjamin M and Lindenberger U (2005) Aging neuromodulation impairs associative binding: a neurocomputational account. *Psychological Science* 16(6): 445–450.

Lindsay RCL, Wells GL and O'Connor FJ (1989) Mock-Juror belief of accurate and inaccurate eyewitnesses: a replication and extension. *Law and Human Behavior* 13(3): 333–339.

List JA (1986) Age and schematic differences in the reliability of eyewitness testimony. *Developmental Psychology* 22(1): 50–57.

Liu JC, Li KA, Yeh SL, et al. (2022) Assessing perceptual load and cognitive load by fixation-related information of eye movements. *Sensors* 22(3): 1187.

Loftus EF and Hoffman HG (1989) Misinformation and memory: the creation of new memories. *Journal of Experimental Psychology: General* 118(1): 100–104.

Loftus EF and Palmer JC (1974) Reconstruction of automobile destruction: an example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior* 13(5): 585–589.

Luna K and Martín-Luengo B (2011) Confidence-accuracy calibration with general knowledge and eyewitness memory cued recall questions. *Applied Cognitive Psychology* 26(2): 289–295.

Luus CAE and Wells GL (1994) The malleability of eyewitness confidence: co-witness and perseverance effects. *Journal of Applied Psychology* 79(5): 714–723.

Macrae CN and Lewis HL (2002) Do I know you? Processing orientation and face recognition. *Psychological Science* 13(2): 194–196.

Magnusson M, Ernberg E, Landström S, et al. (2020) Can rapport building strategies, age, and question type influence preschoolers' disclosures of adult wrongdoing? *Scandinavian Journal of Psychology* 61(3): 393–401.

Maricchiolo F, Gnisci A, Bonaiuto M, et al. (2009) Effects of different types of hand gestures in persuasive speech on receivers' evaluations. *Language & Cognitive Processes* 24(2): 239–266.

Matsumoto D and Hwang HC (2021) An initial investigation into the nature and function of rapport in investigative interviews. *Applied Cognitive Psychology* 35(4): 988–998.

McGlothlin H and Killen M (2010) How social experience is related to children's intergroup attitudes. *European Journal of Social Psychology* 40(4): 625–634.

Megreya AM and Burton AM (2006) Unfamiliar faces are not faces: evidence from a matching task. *Memory & Cognition* 34(4): 865–876.

Memon A, Hope L, Bartlett J, et al. (2002) Eyewitness recognition errors: the effects of mugshot viewing and choosing in young and old adults. *Memory & Cognition* 30(8): 1219–1227.

Memon A, Meissner CA and 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(4): 340–372.

Mercer B, Wapner W, Gardner H, et al. (1977) A study of confabulation. *Archives of Neurology* 34(7): 429–433.

Mugno AP, Malloy LC and La Rooy DJ (2018) Interviewing witnesses. In: Davies GM and Beech AR (eds) *Forensic Psychology Crime, Justice, Law, Interventions*. 3rd edition. John Wiley & Sons, 201–253.

Murphy G and Greene CM (2016) Perceptual load affects eyewitness accuracy and susceptibility to leading questions. *Frontiers in Psychology* 7: 1322.

Nash RA, Nash A, Morris A, et al. (2016) Does rapport-building boost the eyewitness eyeclosure effect in closed questioning? *Legal and Criminological Psychology* 21(2): 305–318.

Navon D (1977) Forest before trees: the precedence of global features in visual perception. *Cognitive Psychology* 9(3): 353–383.

Neequaye DA (2023) Why rapport seems challenging to define and what to do about the challenge. Collabra. *Psychology* 9(1): 90789.

Neequaye DA and Mac Giolla E (2022) The use of the term rapport in the investigative interviewing literature: a critical examination of definitions. *Meta-Psychology* 6: 2808.

North AS, Russell AJ and Gudjonsson GH (2008) High functioning autism spectrum disorders: an investigation of psychological vulnerabilities during interrogative interview. *Journal of Forensic Psychiatry and Psychology* 19(3): 323–334.

Novotny E, Frank MG and Grizzard M (2021) A Laboratory Study Comparing the Effectiveness of Verbal and Nonverbal Rapport-Building Techniques in Interviews. *Communication Studies* 72(5): 819–833.

Odinot G and Wolters G (2006) Repeated recall, retention interval and the accuracy–confidence relation in eyewitness memory. *Applied Cognitive Psychology* 20(7): 973–985.

Oken BS, Kishiyama SS, Kaye JA, et al. (1999) Age-related differences in global-local processing: stability of laterality differences but disproportionate impairment in global processing. *Journal of Geriatric Psychiatry and Neurology* 12(2): 76–81.

Ormerod P and Adler JR (2010) Without fear or favour, prejudice or ill will: magistrates' sentencing decisions. In: Adler JR and Gray JM (eds) *Forensic Psychology Concepts, Debates & Practice*. Taylor & Francis Group, 123–144.

Overman AA, Wiseman KD, Allison M, et al. (2013) Age differences and schema effects in memory for crime information. *Experimental Aging Research* 39(2): 215–234.

Pacheco-Unguetti AP, Acosta A and Lupianez J (2014) Recognizing the bank robber and spotting the difference: emotional state and global vs. local attentional set. *Spanish Journal of Psychology* 17: E28.

Parrish M (2014) *Social Work Perspectives on Human Behaviour*. McGraw-Hill Education.

Perfect TJ (2004) The role of self-rated ability in the accuracy of confidence judgements in eyewitness memory and general knowledge. *Applied Cognitive Psychology* 18(2): 157–168.

Perfect TJ and Weber N (2012) How should witnesses regulate the accuracy of their identification decisions: one step forward, two steps back? *Journal of Experimental Psychology: Learning, Memory & Cognition* 38(6): 1810–1818.

Perfect TJ, Dennis I and Snell A (2007) The effects of local and global processing orientation on eyewitness identification performance. *Memory* 15(7): 784–798.

Plotnikoff J and Woolfson R (2009) *Measuring Up. Evaluating Implementation of Government Commitments to Young Witnesses in Criminal Proceedings*. National Society for the Prevention of Cruelty to Children.

Polczyk R, Szpitalak M, Kuczek M, et al. (2024) Interrogative suggestibility: the role of source monitoring, compliance, and memory in the context of minimally leading questions. *Personality and Individual Differences* 222: 112583.

Powell MB, Wright R and Clark SE (2010) Improving the competency of police officers in conducting investigative interviews with children. *Police Practice and Research* 11(3): 211–226.

Prull MW and Yockelson MB (2013) Adult age-related differences in the misinformation effect for context-consistent and context-inconsistent objects. *Applied Cognitive Psychology* 27(3): 384–395.

R v Malkinson (2023) *EWCA Crim 954*.

Rae Tuckey M and Brewer N (2003) The influence of schemas, stimulus ambiguity, and interview schedule on eyewitness memory over time. *Journal of Experimental Psychology: Applied* 9(2): 101–118.

Rakoff JS and Loftus EF (2018) The intractability of inaccurate eyewitness identification. *Daedalus* 147(4): 90–98.

Ramadhani R, Indrayani LM and Mahdi S (2019) Language power in attorney's leading questions to discredit witness's testimonies during court trial: a forensic linguistic study. *International Journal of English Literature & Social Sciences* 4(6): 2002–2009.

Robinson MD, Johnson JT and Herndon F (1997) Reaction time and assessments of cognitive effort as predictors of eyewitness memory accuracy and confidence. *Journal of Applied Psychology* 82(3): 416–425.

Romero-Rodríguez J-M, Hinojo-Lucena F-J, Kopecký K, et al. (2023) Digital fatigue in university students as a consequence of online learning during the Covid-19 pandemic. *Educación XXI* 26(2): 141–164.

Roux F and Ceccaldi M (2001) Does aging affect the allocation of visual attention in global and local information processing? *Brain and Cognition* 46(3): 383–396.

Ruva CL and Bryant JB (2004) The impact of age, speech style, and question form on perceptions of witness credibility and trial Outcome[1]. *Journal of Applied Social Psychology* 34(9): 1919–1944.

Sah S, Moore DA and MacCoun RJ (2013) Cheap talk and credibility: the consequences of confidence and accuracy on advisor credibility and persuasiveness. *Organizational Behavior and Human Decision Processes* 121(2): 246–255.

Sauer JD, Palmer MA and Brewer N (2019) Pitfalls in using eyewitness confidence to diagnose the accuracy of an individual identification decision. *Psychology, Public Policy, and Law* 25(3): 147–165.

Saywitz KJ, Larson RP, Hobbs SD, et al. (2015) Developing rapport with children in forensic interviews: systematic review of experimental research. *Behavioural Sciences & the Law* 33(4): 372–389.

Saywitz KJ, Wells CR, Larson RP, et al. (2019) Effects of interviewer support on children's memory and suggestibility: systematic review and meta-analyses of experimental research. *Trauma, Violence, & Abuse* 20(1): 22–39.

Searcy JH, Bartlett JC, Memon A, et al. (2001) Aging and lineup performance at long retention intervals: effects of metamemory and context reinstatement. *Journal of Applied Psychology* 86(2): 207–214.

Searcy J, Bartlett JC and Memon A (2010) Influence of post-event narratives, line-up conditions and individual differences on false identification by young and older eyewitnesses. *Legal and Criminological Psychology* 5(2): 219–235.

Smith SM and Vela E (2001) Environmental context-dependent memory: a review and meta-analysis. *Psychonomic Bulletin & Review* 8(2): 203–220.

Smith HMJ, Ryder H and Flower HD (2018) Eyewitness evidence. In: Davis GM and Beech AR (eds) *Forensic Psychology Crime, Justice, Law, Interventions*. 3rd edition. John Wiley & Sons, 173–199.

Sporer SL (1993) Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology* 78(1): 22–33.

Sporer SL, Penrod S, Read D, et al. (1995) Choosing, confidence, and accuracy: a meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin* 118(3): 315–327.

Tuckey MR and Brewer N (2003) How schemas affect eyewitness memory over repeated retrieval attempts. *Applied Cognitive Psychology* 17(7): 785–800.

Turner J (2015) Making sense of the world. In: Turner J and Barker MJ (eds) *Living Psychology: From the Everyday to the Extraordinary.* The Open University, Vol. 2, 3–46.

Valentine T and Maras K (2011) The effect of cross-examination on the accuracy of adult eyewitness testimony. *Applied Cognitive Psychology* 25(4): 554–561.

Vallano JP and Compo NS (2011) A comfortable witness is a good witness: rapport-building and susceptibility to misinformation in an investigative mock-crime interview. *Applied Cognitive Psychology* 25(6): 960–970.

Vallano JP and Schreiber Compo N (2015) Rapport-building with cooperative witnesses and criminal suspects: a theoretical and empirical review. *Psychology, Public Policy, and Law* 21(1): 85–99.

Vroom JR, Danby MC and Sharman SJ (2025) A systematic mapping review of the literature examining pre-substantive rapport-building techniques in investigative interviews with children. *Psychology, Crime and Law* 1–27.

Wade KA and Spearing ER (2023) The effect of cross-examination style questions on adult eyewitness accuracy depends on question type and eyewitness confidence. *Memory* 31(2): 163–178.

Wade KA, Nash RA and Lindsay DS (2018) Reasons to doubt the reliability of eyewitness memory: commentary on Wixted, Mickes, and Fisher (2018). *Perspectives on Psychological Science* 13(3): 339–342.

Wagstaff GF and Wheatcroft JM (2012) *The Liverpool Interview Protocol (LIP): Manual*. University of Liverpool.

Wagstaff GF, Wheatcroft JM, Caddick AM, et al. (2011) Enhancing witness memory with techniques derived from hypnotic investigative interviewing: focused meditation, eye-closure, and context reinstatement. *International Journal of Clinical & Experimental Hypnosis* 59(2): 146–164.

Wagstaff GF, Wheatcroft JM, Hoyle JD, et al. (2014) Enhancing memory with the liverpool interview protocol: is an association with hypnosis a problem? *Contemporary Hypnosis & Integrative Therapy* 30(3): 142–151.

Walhovd KB, Krogsrud SK, Amlien IK, et al. (2016) Neurodevelopmental origins of lifespan changes in brain and cognition. *Proceedings of the National Academy of Sciences* 113(33): 9357–9362.

Walsh D and Bull R (2012) Examining rapport in investigative interviews with suspects: does its building and maintenance work? *Journal of Police and Criminal Psychology* 27(1): 73–84.

Warmelink L, Subramanian A, Tkacheva D, et al. (2019) Unexpected questions in deception detection interviews: does question order matter? *Legal and Criminological Psychology* 24(2): 258–272.

Wells GL, Memon A and Penrod SD (2006) Eyewitness evidence: improving its probative value. *Psychological Science in the Public Interest* 7(2): 45–75.

Westra E (2019) Stereotypes, theory of mind, and the action-prediction hierarchy. Synthese. *An International Journal for Epistemology, Methodology & Philosophy of Science* 196(7): 2821–2846.

Wheatcroft JM (2020) *Hit and Run Video Clip*. Provided for research purposes from Cheshire Police, UK.

Wheatcroft JM, Caruso D and Krumrey-Quinn J (2015a) Rethinking leading. The directive, non-directive divide. *Criminal Law Review* 5(5): 340–346.

Wheatcroft JM and Ellison LE (2012) Evidence in court: witness preparation and cross-examination style effects on adult witness accuracy. *Behavioural Sciences & the Law* 30(6): 821–840.

Wheatcroft JM and Wagstaff GF (2010). UK police officers' perceptions of the cognitive interview (CI): usefulness, confidence and witness reliability. In: Paper presented to the *International Investigative Interviewing Research Group; 3rd Annual Conference. Stavern, Norway, June, 2010.*

Wheatcroft JM and Woods S (2010) Effectiveness of witness preparation and cross-examination non-directive and directive leading question styles on witness accuracy and confidence. *International Journal of Evidence and Proof* 14(3): 187–207.

Wheatcroft JM, Wagstaff GF and Kebbell MR (2004) The influence of courtroom questioning style on actual and perceived eyewitness confidence and accuracy. *Legal and Criminological Psychology* 9(1): 83–101.

Wheatcroft JM, Wagstaff GF and Manarin B (2015b) The influence of delay and item difficulty in criminal justice systems on eyewitness confidence and accuracy. *International Journal of Humanities and Social Science Research* 1: 1–9.

Wheatcroft JM, Wagstaff GF and Russell K (2014) Specialist police interviewer perceptions of the enhanced cognitive interview: usefulness, confidence and witness reliability. *Police Practice and Research* 15(6): 505–518.

Wilcock RA, Bull R and Vrij A (2007) Are old witnesses always poorer witnesses? Identification accuracy, context reinstatement, own-age bias. *Psychology, Crime and Law* 13(3): 305–316.

Wixted JT (2023) *Eyewitness Memory*. Oxford University Press.

Wixted JT and Wells GL (2017) The relationship between eyewitness confidence and identification accuracy: a new synthesis. *Psychological Science in the Public Interest* 18(1): 10–65.

Wixted JT, Mickes L, Clark SE, et al. (2015) Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *American Psychology* 70(6): 515–526.

Wixted JT, Mickes L and Fisher RP (2018) Rethinking the reliability of eyewitness memory. *Perspectives on Psychological Science* 13(3): 324–335.