Confidence malleability in the interviewing setting, and its effect on subsequent memory monitoring and regulation

Alessandra Caso
Department of Psychology
Goldsmiths University

For the award of PhD

Acknowledgments

To those that have assisted, thank you.

Fiona and Gordon, thank you for your constant support through this incredible journey.

To Sam and my family for their unlimited patience, thank you!

Abstract

Eyewitness evidence is often essential for the outcome of an investigation. However, research has shown that memory is not perfect, and eyewitnesses can make mistakes. In the past forty years, researchers have developed evidence-based techniques to interview eyewitnesses effectively and to maximise the amount and accuracy of the information elicited in an interview. Despite the wealth of research on best practice interviewing techniques, little is known about how these might affect eyewitness confidence. This is important because confidence plays a pivotal role in the regulation of memory output. For example, research in metamemory showed that confidence judgements underpin eyewitnesses' decisions to report or withhold information. This PhD aims to fill this gap by investigating confidence changes within the context of an investigative interview and testing the hypothesis that confidence shifts following an interview might affect subsequent memory regulation. Study 1 showed that memory confidence can change after an interview. Study 2 built on this finding and showed that when the interview promotes free and undirected retrieval, confidence remains stable. On the contrary, when the interview promotes a directed retrieval via presenting Cued Recall questions, confidence decreases. A further investigation of the metacognitive processes that underpin these results (Study 3a and 3b) showed that different types of question lead to confidence shifts depending on the difficulty experienced when answering them. Finally, drawing upon these results, Study 4 investigated the conditions of an interview likely to lead to confidence shifts and those likely to promote confidence stability. Across the studies, no evidence was found that changes in confidence following an interview impact subsequent memory regulation. Overall, the results confirm and further the existing evidence in support of good practice in eyewitnesses interviewing. As such, evidence-based techniques are compatible with confidence stability, while deviations from them are likely to lead to decreased confidence.

Table of contents

List of tables

Chapter 4

Chapter 5

Chapter 6

Chapter 7

List of figures

Chapter 3

Chapter 6

Chapter 7

Overview

Eyewitness evidence is pivotal at all stages of the investigative processes. As such, investigators strive to extrapolate complete and accurate accounts each time an eyewitness is asked to recall the event witnessed. Psychological research has investigated the most effective interviewing methods by developing and testing different techniques to achieve this goal. Although best practice interviewing techniques are widely recognised as effective, we do not yet know how they impact eyewitnesses' metacognitive processes, despite research on metamemory showing that metacognitive processes, such as confidence, are used to regulate the amount and quality of the information reported. Thus, it is important to understand how best practice interviewing techniques influence confidence and how this in turn is used to regulate the information subsequently reported.

Chapter 1 of this PhD provides an overview of the existing literature in support of the best practice interviewing techniques currently available, and outlines the current recommendations relating to eyewitness interviewing. Here, particular attention is given to the information gathering stage of the interview, the effectiveness of different types of question asked, and their use at different stages of the interview.

Chapter 2 focuses on eyewitness confidence. Here, I outline the role of memory confidence in the regulation of memory output. Further attention is given to the literature investigating confidence within the context of an investigative interview.

An investigation of the factors likely to lead to shifts in memory confidence is reported in Chapter 3. Here, I present a systematic review of the literature with the aim to isolate the factors likely to cause changes in memory confidence. Of these factors, particular attention is given to those relevant to the interviewing context.

Based on the literature discussed in the introductory chapters, in Chapter 4 I present my first study, investigating how different quality interviews affect eyewitness confidence and the quality of information reported subsequently. Here, Study 1 tested the hypothesis that a Poor Practice Interview (but not a Best Practice Interview) leads to decreased confidence and poorer quality of information reported in a subsequent recall test.

In Chapter 5, a similar study is presented; however, a Free Recall and a Cued Recall interview are used as proxies of a Best Practice and a Poor Practice Interview. Here, Study 2 tests the hypothesis that confidence decreases only following a Cued Recall Interview (but not following a Free Recall Interview). Furthermore, in this study I investigate whether decreased confidence following an interview is correlated with the amount and accuracy of information reported in a subsequent recall attempt.

Based on the results on Study 2, Chapter 6 presents an investigation of the metacognitive processes that underpin the decrease in confidence following a Cued Recall Interview. Here, across two studies (Study 3a and Study 3b) I test the hypothesis that confidence is likely to decrease as a function of the difficulty experienced when answering different types of question. Further, I investigate whether decreased confidence leads to demonstrations of under-confidence and poor confidence-accuracy calibration in the information reported in a subsequent recall attempt.

The last experimental Chapter (Chapter 7) builds on the results of the previous studies and investigates the conditions of an interview that promote confidence stability and those likely to lead to decreased confidence.

Finally, in Chapter 8 I discuss the findings of the PhD as a whole, within the context of the existing literature. I highlight the contributions that this research has made to the field of investigate interviewing, and to the current debate on the relationship between confidence and accuracy.

Chapter 1: Best Practice in Eyewitness Interviewing

Introduction

In the absence of CCTV footage, the only way to gather information about a crime is by asking what happen to those that were on the crime scene. Often the police have little more than the eyewitness or victim's memories to reconstruct an incident and ensure the perpetrators to justice (Fisher et al., 1994; Geiselman & Fisher, 1989; Sanders, 1986). Therefore, it is important that the information gathered from eyewitnesses is as accurate as possible, since any error that eyewitnesses report represents a potential wrong clue that the police will pursuit. However, eyewitnesses' recollections are not perfect, and witnesses can make mistakes (Loftus, 1979).

It is well documented that memory is fallible and that numerous factors can impact on the accuracy of the account eyewitnesses report. Some of these factors are largely beyond the control of the Criminal Justice System. For example, research has found that contextual characteristics at the time of encoding, such as a short exposure to the criminal event, viewing distance or poor lightening conditions can decrease the accuracy of eyewitnesses' reports (e.g., Granhag et al., 2014; Meissner et al., 2007; Memon et al., 2004). Eyewitnesses' characteristics such as age or level of stress experienced during the event can also negatively impact the accuracy of the memory reported (Deffenbacher et al., 2004; Meissner et al., 2007; Yarmey, 1993).

Factors that decrease eyewitness accuracy can also stem from poor practice within the Criminal Justice System, including during the information gathering stage of the investigation. For example, the more time elapses between the incident and the first interview the less accurate eyewitnesses' reports are (e.g., Dysart & Lindsay, 2006; Odinot & Wolters, 2006). Or the more opportunities there are for eyewitnesses to encounter misleading post-event information from co-witnesses, the more likely they are to report incorrect information (e.g., Gabbert et al., 2003; Gabbert et al., 2012; Paterson et al., 2009; Paterson et al., 2010). However, these 'system' variables can be controlled, and their negative effects can be attenuated. Thus, in order to avoid forgetting and memory contaminations good practice guidelines suggest interviewing eyewitnesses as soon as possible.

Other system variables found to strongly impact on eyewitness' accuracy are related to the way in which the information is extrapolated from eyewitnesses. Perhaps one of the most compelling examples relating to the influence of

interviewing practice on memory reporting can be found in the literature on leading questions. In a classic study Loftus and Palmer (1974) showed participants a video of a car accident and asked them to estimate the speed of the cars. Participants who received the question suggesting that the cars had *smashed* into each other reported significantly higher speed estimates than those that received the question suggesting that the car *contacted* each other (see also Loftus, 2005). The literature on leading questions clearly shows that factors and variables at play during an interview can strongly influence the accuracy of the information reported by eyewitness and it is therefore pivotal to ensure that investigative interviews are conducted by following the available evidence-based practice.

A framework of best practice in eyewitness interviewing

A significant attempt to standardise the professional practice relating to eyewitness interviewing in the UK was carried out in 1992, when the Home Office commissioned a training programme aimed to upskill investigators in England and Wales and improve the quality of the information gathered. The programme led to the development of the PEACE model (acronym of *Planning and Preparation*, *Engage and Explain, Account, Closure, Evaluation*) – which is currently the framework of best practice in investigative interviewing. The PEACE recommendations are grounded into psychological research and represent the best evidence-based techniques currently available. Each stage is outlined briefly below.

Planning the interview

The *Planning and Preparation* phase includes recommendations relating to the planning of the upcoming interview, such as revising the investigation and the information already available and organising any special arrangement the eyewitness might require. Evidence of the utility of this phase can be found in a work by Griffiths and Walsh (2018). In their study investigators reported reflective comments on their practice and highlighted several benefits of this stage, including identifying the areas to be further probed, and deciding a common interviewing plan and cohesive strategy with a potential co-interviewer. Further benefits were highlighted by investigators in Howes's (2020) study. Here, practitioners reported discussing potential arrangements for booking and debriefing interpreters during this initial stage. As well as highlighting the benefits of this stage of the interview, research has

also shown that the lack of preparation at this stage of the interview is among the causes that lead to obtaining poor and incomplete accounts from eyewitnesses (Cherryman & Bull, 2001; Walsh & Milne, 2008). Although police officers often report to have no time to extensively prepare for the interview (e.g., Cherryman & Bull, 2001; Walsh & Milne, 2008), researchers strongly suggest that the Planning and Preparation stage is a vital part of the interviewing process (Clarke & Milne 2001; Walsh & Milne, 2008).

Interviewing instruction and rapport

The *Engage and Explain* phase recommends taking time to establish a positive relationship with the interviewee, ensuring that they are comfortable and understand why they are being interviewed, what the objectives of the interview are, and how it is going to be conducted. The importance of explaining the aim of the interview has been outlined by Walsh and Milne (2008). Investigators are encouraged to explicitly outline the interview process and to cover the ground rules including the expectation that the eyewitness will take control of the interview during the free report phase. Failure to clarify these aspects of the interview could result in poor understanding of the interviewer's expectations (Griffith & Milne, 2010).

Regarding the *Engaging* aspect of this phase, researchers generally agree that the interpersonal relationship between the interviewer and the interviewee can impact the quality of eyewitnesses' account. For example, Collins et al. (2002), found evidence that the interviewer's aptitude can influence the amount of information reported. When an effort was made by the interviewer to build rapport, participants reported more correct units of information at no cost of accuracy, compared with when the interviewers adopted a neutral or abrupt demeanour. This was observed especially in the free report phase of the interview. Evidence that rapport building also increases the accuracy of the information reported comes from Kieckhaefer et al. (2014) who found a high rapport interview to increase the accuracy of the account via inoculating against the subsequent exposure to misinformation (see also Nash et al., 2016; Vallano & Compo 2011). Furthermore, in a recent systematic review Gabbert et al. (2020) reported that the majority of the studies evaluated in the review found a positive influence of rapport on both amount and accuracy of the information reported.

Despite the consensus that building rapport increases the quality of information reported from witnesses, the elements that underpin an effective rapport are not yet clear. Consequently, rapport has been operationalised with both verbal and non-verbal components (St-Yves, 2006), or with a combination of both. Fisher and Geiselman (1992) originally suggested to ensure rapport by mean of verbal techniques, such as by using the interviewee's name, or by showing interest through the use appropriate questions or active listening. St-Yves (2006) also highlighted the importance of active listening and minimal verbal (e.g., uh hum, okay) and non-verbal encouragements (i.e. nodding, friendly facial expression), (see also Kieckhaefer et al., 2014; Vallano & Compo, 2011).

A further insight into the effectiveness of the use of non-verbal rapport techniques comes from the literature on social support for child eyewitness which showed that a supportive interview can increase the accuracy of the children's accounts. In this line of research, the non-verbal components of a supportive interview included conveying warmth, dressing casually, using eye contact, smiles, open body posture and a friendly tone of voice (e.g., Carter et al., 1996; Davis & Bottoms, 2002; Quas et al., 2005).

Gathering information

The *Account* phase of the PEACE model refers to the information gathering stage and includes recommendations about how to extrapolate accurate and complete accounts from witnesses. The information gathering process has received much attention and researchers have developed and tested various interviewing protocols tailored to both the type of interviewee and the contextual situation of the interview. Some of the best known interviewing tools are the Cognitive Interview (CI, Fisher & Geiselman, 1992; Fisher et al., 1989) used especially with cooperative witnesses, the Conversation Management (Shepherd, 1986; 1991) better suited to interview less cooperative witnesses (Dando et al., 2009). Some protocols are especially tailored for the child witness, such as the National Institute of Child Health and Human Development (NICHD, Lamb et al., 2007; Orbach et al., 2000). Others are especially designed for gathering information on the crime scene, such as the Structed Interview Protocol (SIP, Gabbert et al., 2015a; 2015b), or in the event in which multiple witnesses need to be interviewed as soon as possible, such as the Self-Administered Interview (SAI, Gabbert et al., 2009; Gabbert et al., 2012). Since these protocols

address specific issues, they vary in the form they are presented, for example the SAI is a written form while the others are guidelines of interviewing. Further differences can be found in the way the instructions are conveyed, for example the NICHD Protocol includes specific initial questions to assess if the interviewee understands the difference between telling the truth or lying; a practice phase and more detailed instruction about the interview that might not be needed when interviewing adults. However, the best practice interviewing protocols share many similarities, in particular relating to the structure of the interview and the types of question used.

The structure of an interview. The vast majority of best practice interviewing protocols recommend initiating an interview with a free report phase followed by a questioning or probing phase. This structure enhances the likelihood of obtaining detailed and accurate accounts from witnesses. In particular, ensuring an eyewitness has the opportunity to freely report the whole event in an uninterrupted free recall allows for the memory to be activated, and for the details that are not immediately accessible to be reached. This is in line with the spreading activation theory of memory (Anderson, 1983) which conceives memory as in a network of related details (called nodes) that are connected to each other. When a detail is recalled, an activation-signal is sent to all the related details, and when the signal reaches a sufficient strength the related details can also be activated and recalled. Allowing an eyewitness to focus on the recollection task and activate the memory at their own pace increases the chances of recalling more details, including those that are not immediately accessible (for a review see Wheeler & Gabbert, 2017).

Besides increasing the opportunity to provide detailed accounts, initiating an interview with a free recall is likely to increase the accuracy of the information reported. Research has shown that details reported in the free report phase of the interview are more likely to be correct than those reported during the questioning phase. In a recent study Kontogianni et al. (2020) found that the accuracy of the information reported in the follow-up questioning phase was significantly lower than that of the information reported in the initial free report phase. This result persisted even when participants were encouraged to be as accurate as possible before the questioning phase began. Thus, in order to enhance the opportunity to provide correct and detailed accounts, it is important that eyewitnesses are allowed to engage in an uninterrupted account of their memory.

Only after the free report has been completed should the questioning phase begin. The primary aim of this phase is to clarify and expand on the details provided in the freely reported account. During the probing phase a large proportion of new and correct information can be reported by eyewitnesses (e.g., Dando, 2013; Dando et al., 2009; Kontogianni et al., 2020; Memon et al., 1997), showing that the follow-up questions asked during this phase can be necessary to reach details that have not been reached during the free recall phase. However, during this stage of interview eyewitness memory is more likely to be influenced by the interviewer and their utterances, hence it is important that the probing phase is carried out by using appropriate and effective techniques.

Considering that each eyewitness is likely to store and retrieve their memory is a unique way, best practice guidelines recommend for this phase to be compatible with the structure and the content of the eyewitness's free report, including their linguistic style (e.g., Fisher et al., 2011; Gabbert et al., 2015a; 2015b; Kebbell et al., 2001). Here, the eyewitness's account should be broken down into separate topics and each topic should be probed with appropriate questions, by following an order and a language style that closely match the eyewitness's account. This procedure ensures that (i) only the content reported by the eyewitness is addressed in the probing phase, and (ii) the eyewitness memory is probed by using the eyewitness's own cues, such as their own words or the unique order with which they remember the event. While probing only the interviewee's account reduces the risk for the interviewer to contaminate the eyewitness' memory by minimising the opportunities to suggest misleading information, providing an eyewitness with self-generated cues increases the likelihood to trigger details that were not accessed during the free report phase (e.g., Kontogianni et al., 2020; Wheeler & Gabbert, 2017).

In summary, recommendations relating to interviewing structure are rooted within memory theories and are compatible with memory functioning. Initiating an interview with a free recall followed by a questioning phase is designed to facilitate and enhance the retrieval of detailed and correct accounts.

The types of question. A further important aspect of the information gathering process relates to the questions asked during the probing phase. As such, researchers have investigated the types of question that are more likely to elicit accurate and complete accounts from witnesses. An initial distinction can be drawn between open and closed questions; however, several different subtypes of questions can be clustered within these two broader categories (Oxburgh et al., 2010). For example, open questions allow for unrestricted answers, however further subtypes of question are included within this category. One type is the free recall or free invitation (e.g., Aldridge & Cameron, 1999; Hershkowitz, 2001) which allow for interviewee's complete control over the answer. Further open questions are the TED questions starting with the words "Tell"," Explain", or "Describe" (e.g., Griffiths & Milne, 2006), these require an elaboration of content already mentioned. Similarly questions that request to elaborate the answer further have been referred to as open-ended breadth questions (e.g., "Tell me what happened"), while those requesting more information on a detail already mentioned have been referred to as open-ended depth questions (e.g., "You mentioned x; tell me more about x"), (e.g., Powell & Snow, 2007).

On the contrary, closed questions are defined as those requesting specific or targeted answers can include *probing* questions which commence with "wh" - often referred to as *5Wh* (i.e., "Who", "What", "Where", "When", "Why", and "How") or *closed-specific* questions (Aldridge & Cameron, 1999; Davies et al., 2000; Griffiths & Milne, 2006; Loftus, 1982). Other subtypes include questions that require a yes/no answer (e.g., Price & Roberts, 2011) often referred to as *option-posing*, and questions that target specific details, known as *focused* or *directive* questions (e.g., Sternberg et al., 1996). Focused questions are also further divided in *leading* or *suggestive* questions (e.g., Cederborg et al., 2000; Davies et al., 2000).

Research has shown that open questions tend to elicit longer answers and more complete accounts from older children and adults compared to closed questions (e.g., Davies et al., 2000; Hamilton et al., 2016; Lamb et al., 2003; Lamb et al., 2007; Hershkowitz, 2001; Snook et al., 2012; Sternberg et al., 1996). Furthermore, open questions yield more accurate information (e.g., Ibabe & Sporer, 2004; Orbach & Lamb, 2001; Oxburgh et al., 2010; Shapiro, 2006; Sharman & Powel, 2011), and consistent accounts than closed questions (e.g., Orbach & Lamb, 2001; Poole & White, 1991). Among the different types of closed questions, the use of leading

questions is highly discouraged. These questions are not neutrally phrased and rather suggest an answer or contain a detail that was not mentioned by the eyewitness, thus they are likely to influence the eyewitness's response (Lamb et al., 2011; Loftus & Palmer, 1974; Loftus & Zanni, 1975). In answer to this type of question eyewitnesses tend to either comply and report the suggested information even if unsure about it, or to incorporate the suggested detail in their original memory independently of its accuracy (Roebers & Schneider, 2000). The influence of leading questions is stronger when the suggested details is related to a peripheral rather than a central information (Roebers & Schneider, 2000; Sutherland & Hayne, 2001), and if the question is asked after a long delay (Shapiro et al., 2005). For similar reasons, researchers suggest limiting the use of questions that rely upon recognition rather than recollection memory, such as forced choice or multiple-choice questions. These types of question are more likely to yield incorrect information because they encourage eyewitness to choose an answer even if they might be unsure about its accuracy (e.g., La Rooy et al., 2015).

However, despite responses to closed questions being more likely to be inaccurate, researchers do recognise that they can be useful when a specific piece of information is sought, because they narrow down eyewitnesses' attention to individual aspects of their memory. Thus, in order to minimise the impact of closed questions on the accuracy of an eyewitness's account, researchers suggest using them only when the information sought has not been elicited by open questions. For example, Griffiths and Milne (2006) introduced the distinction between *appropriate* and *inappropriate* closed questions depending on the time of the interview in which the questions are asked. For example, a yes/no question is defined as "appropriate" when it is asked in the conclusive probing phase of a topic area, after open (i.e., TED questions) and focused questions (e.g., 5Wh questions) have been asked. However, if the yes/no question is asked before the open or focused questions it is considered "inappropriate". Similarly, Gabbert et al. (2015a; 2015b) developed a traffic-light coloured map to flag the types of question that are more likely to yield incorrect information. Here, the authors advice using the closed questions associated with high risk only if open questions have failed to trigger the specific detail sought.

Closing the interview

The PEACE model recommends conducting a *Closure* phase in order to summarise the account given by the eyewitness, to check that it is as complete as possible and that it has been clearly understood. At this stage the summary of the information reported can act as a further activation-signal and jog the witness's memory for additional details (Anderson & Pichert, 1978; Wheeler & Gabbert, 2017). Thus, it is important that the interviewee is provided with the opportunity to report any additional information they might remember. Beside summarising the account, confirming that this has been understood and providing the opportunity to report additional details, two further components should be included at this point of the interview: (i) providing an overview of the following stages of the investigation and (ii) leaving a contact detail for the interviewee to get in touch should they remember any other detail (i.e., Clark & Milne, 2001; Walsh & Milne, 2008). Researchers have suggested that a comprehensive closure is pivotal to enhance cooperation and to leave the interviewee in a positive state should they be asked to continue collaborating with the investigation (Bull, 2010). Despite the importance of this stage, research analysing real life interviews has highlighted that interviewers do not often cover the components of the Closure stage in sufficient details (Walsh & Milne, 2008). In particular, evidence exists that interviewers often fail to (i) include a summary of the account, (ii) seek confirmation that the account was understood, and (iii) leave contact details (Scott et al., 2015; see also Clark & Milne 2001).

Evaluating own practice

Finally, the *Evaluation* phase recommends taking note of any inconsistencies with the information previously held that might have emerged during the interview. This stage should serve to integrate the new information into the existing investigation. It is also an opportunity for interviewers to monitor their own interviewing practice. Walsh and Milne (2008) highlighted that this stage is pivotal in ensuring self-reflections and continuous improvement and it should be carefully conducted as part of the interviewing process. However, they also noted that evaluation should be paired with further training in order to ensure investigators can rely upon up to date skills.

Gap between best practice recommendation and real-life interviews

Despite a large body of research that has consistently highlighted the risk related to inappropriate interviewing practice, deviations from best practice recommendations are not infrequent. Early research flagged several important issues related to common practice in investigative interviewing, in particular relating the way in which information is elicited from eyewitness. Fisher et al. (1987) assessed tape-recorded forensic interviews conducted in the US by experienced detectives and found that after initiating a free recall they tended to interrupt it after an average of 7.5 seconds. The interviews were dominated by rapid-fire closed questions, asked in an unstructured sequence, which did not match the interviewees' account. The lack of retrieval support was paired with the use of misleading questions and negatively phrased questions. The style was often difficult to understand due to the excessive use of jargon and technical terms. Similar findings were also reported by George (1991) in interviews conducted by British investigators. He found that interviews were dominated by the use of closed questions often misleading. Furthermore, pauses were rare which suggested questions were asked in a rapid sequence leaving little time for the interviewee to elaborate detailed answers (see also Milne & Bull, 1999).

Many of the issues in a typical forensic interview highlighted by Fisher et al. (1987) and George (1991) suggested that investigators were not eliciting detailed and reliable information from witnesses. In particular, the interruption of a free recall, the rapid-fire questioning, and the used of closed questions suggested that the interviewee had little opportunity to recall the entire event at their own pace, and that most of the information was provided in response to interviewer's specific questions. This interviewing style can only elicit the information targeted by the questions and it is unlikely to reach all the details the witnesses could potentially remember.

Consistent with early research, more recent studies documented similar issues in relation to current interviewing practice. For example, interviewers' interruption and excessive use of rapid-fire yes/no questions were found in Wright and Alison (2004), and Myklebust and Alison (2000). While a large number of studies reported that closed questions are still broadly used in real-life interviews (Carson & La Rooy, 2015; La Rooy et al., 2013; La Rooy et al., 2011; Loftus & Greenspan, 2017; Otgaar et al., 2018; Oxburgh et al., 2012; Wade et al., 2018). In addition, studies that analysed field interviews showed that the proportion of closed questions compared to open questions can be surprisingly high. For example, Myklebust and Bjørklund (2006) sampled 100 field interviews and found that interviewers tended to use 10

times more closed than open questions. More recently Luther et al. (2015) reviewed 45 interviews with children victim of sexual abuse conducted in Canada between 2006 and 2012 and found that only 8% of the interviewers' question were open invitations. In contrast 36% were option posing type of question and 61% were directive (see also Verkampt et al., 2019; Waterhouse et al., 2016; Wolfman et al., 2016). Concerning is also the possibility that the quality of the types of question asked might not be consistent over repeated interviews, and that while the use of open questions can decrease (e.g., Cederborg et al., 2008), the use closed or inappropriate questions can increase in follow up interviews (e.g., Petterson & Pipe, 2006).

The deviation from best-practice recommendations is particularly concerning and suggests that the interviewers and their interviewing style are likely to influence eyewitness reporting and undermine the accuracy of the information they elicit.

Conclusions

Research in psychology has shown that eyewitness memory is malleable and numerous variables affect the quality of the evidence that eyewitnesses report. While some of these variables are beyond the control of the Criminal Justice System, others derive from poor interviewing techniques. Thus, in order to standardise the interviewing practice and ensure that the information elicited from eyewitnesses is as accurate as possible, the PEACE framework was developed as a joint effort between practitioners and researchers in psychology. The PEACE model represents the authorised professional practice in investigative interviewing the UK and includes recommendations related to all aspects of the interviewing process, including the information gathering stage. A large body of research has identified the most effective way to extrapolate information from witnesses and this chapter has outlined the ideal structure of the interview and the most effective types of question used. While recognising that deviation from recommended best practice is still largely observed in the field, the goal of this PhD is to understand the impact that inappropriate interview structure, and inappropriate questions, can have on eyewitness accounts.

Chapter 2: Eyewitness Confidence

Introduction

Research on investigative interviewing has largely focused on understanding the impact of interviewing techniques, such as the types of question, on the quality of the information reported in response. In a typical study, participants are exposed to a to-be-remembered stimulus and then are asked to report what they can remember. Researchers often manipulate the types of question asked, or the retrieval facilitation techniques used. The effectiveness of the retrieval technique in eliciting accurate and complete accounts is then quantified by measuring several indexes of memory quality such as the amount and accuracy of the details reported. This paradigm allows researchers to understand the effect of the retrieval technique on memory reporting. However, it provides limited insight into how memory reporting occurs, which cognitive processes it involves, and how they operate. In a different yet relevant body of work, research on metamemory has focused more closely on the mechanisms that underpin memory reporting and has proposed a pivotal role for metacognitive processes such as memory confidence.

A framework for memory regulation

Research on metacognition has highlighted the relevance of many metacognitive processes that underpin memory retrieval and reporting. A fundamental assumption of the model is that people can self-direct their retrieval process and regulate the information they report. Under this perspective, people do not automatically retrieve and report information they have in memory, rather they engage in several decisions, including whether to start the retrieval process and when to interrupt it, and whether to report or withhold any of the information retrieved. For example, Barnes et al. (1999) proposed that in answer to an external input such as a question, a Feeling of Knowing (FOK) is evoked, and only when it exceeds a certain threshold is a decision made to start the memory search for the response. If the search is successful and an answer is retrieved, a confidence judgement is generated to assess the accuracy of the answer. If the magnitude of the confidence judgement exceeds a certain criterion, then the answer is likely to be reported, otherwise it will be withheld. If the candidate answer does not pass the satisfactory criterion for reporting, people might revaluate the strength on their FOK and any personal

motivation to re-start the memory task. If they still feel they know the answer to the question and are willing to continue searching, then they will engage in further effort to retrieve another candidate answer, otherwise they will end their memory search.

While Barnes et al.'s (1999) model encompasses metacognitive processes involved in memory retrieval and memory reporting, Koriat and Goldsmith (1996) focused primarily on the regulation of the information reported. The model proposes two core processes involved in the regulation of memory reporting: the monitoring and control process. The monitoring process refers to the array of judgements (e.g., FOK, Judgements of Learning (JOL), confidence judgements) that people have regarding their memory and the information they retrieve. These judgements are used to monitor or assess different qualities of the memories retrieved, such as their accuracy. The control process in turn uses these judgements to decide whether or not an answer should be reported. If the assessed probability of accuracy passes a certain criterion the answer is likely to be reported, otherwise it will be withheld.

In order to test the model, Koriat and Goldsmiths (1996) proposed a two-phase methodology – often referred to as Quantity - Accuracy Profile (QAP). In the first 'forced-report phase', participants are asked a series of memory questions to which they have to provide an answer and a confidence judgement. In a following 'free-report phase' participants are given the same test again and are invited to decide which answers they would like to volunteer and which they would prefer to withdraw. Three core findings showed that (1) participants reported less but more accurate information in the free-report phase, hence they were able to gain accuracy at cost of the amount of information reported, (2) participants were able to effectively (although not perfectly) discriminate between correct and incorrect answers via their confidence judgements, and finally (3) participants' tendency to report an answer depended strongly on the confidence judgement, as shown by the fact that volunteered answers had a relatively high mean confidence rate (.89). Thus, participants were able to increase the accuracy of their memory report by selecting the details to volunteer and withhold based on an evaluation of their probability of accuracy.

Memory regulation can be achieved also by other means, for example by regulating the specificity, or 'grain-size' of the information reported in order to achieve high accuracy (Goldsmith et al., 2002). In their study Goldsmith et al. (2002) gave participants a set of memory questions and asked them to provide both a fine-

grained answer and a coarse-grained answer, with associated confidence judgements for each. In a subsequent phase, participants were presented with their answers and were asked to volunteer only one out of the two answers given. The results of interest showed that (1) participants preferred to report fine-grained answers but only when these were confidently held. However, when unsure about their fine-grained answers, participants were successful in increasing accuracy by withdrawing the fine-grained answer and volunteering the coarse-grained answers instead, (2) participants were able to effectively (but not perfectly) discriminate between correct and incorrect answers at both level of grain-size, and finally (3) participants' confidence was a strong predictor of volunteering rate, hence participants tended to report answers where the assessed probability of correctness was high. Thus, confidence judgements - together with other available information (e.g., the level of informativeness of the answer, potential risks of error, external incentives) underpin the decision to volunteer an answer at different levels of grain-size (Ackerman & Goldsmith, 2008; Brewer et al., 2018; Evans & Fisher, 2011; Goldsmith et al., 2005). In summary, the framework of memory regulation depicts eyewitness memory reporting as a self – directed process whereby decisions to report or withhold information, and the level of granularity to report, depend strongly on how confident a person is that their memories are correct.

The basis of confidence judgements

Considering the relevance of confidence judgements in the regulation of memory output, it is important to understand how confidence is generated. A dual-process theory, referred to as cue-utilisation theory (Koriat, 2000; Koriat et al., 2008) proposes that the accuracy of a memory can rarely be accessed directly, and in most occasions it has to be inferred from the evaluation of available cues. Typically, two types of cues are evaluated when generating confidence judgements: information-based (theory-based) and experienced-based cues (affect-based). The former - information-based cues, refer to beliefs and thoughts that people hold regarding their own memory. For example, when people attempt to assess the quality of their memory they might evaluate pre-existing information regarding their own competencies and knowledge in that specific domain (e.g., 'I tend to perform well on memory tests'). While the latter - experienced-based cues, refer to cues generated during the retrieval process. For example, in order to assess the accuracy of a

memory, people might evaluate the retrieval experience such as how easily the memory came to mind, or how vivid or complete the memory appeared.

Information-based and experienced-based judgements

Within the eyewitness memory literature, information-based judgements have been investigated primarily in relation to social influence and its effects on confidence. For example, feedback on performance given during a lineup task or interview can lead an eyewitness to feel more or less confident in their memories. In eyewitness identification research, studies have shown that participants who are led to believe they have chosen the correct culprit, show higher confidence in their identification decision compared with those who had not received feedback. Similarly, confidence decreases when eyewitnesses are led to think they have picked the wrong person (e.g., Bradfield et al., 2002; Wells & Bradfield, 1998; Wells & Seelau, 1995). Similar results can be found in studies investigating social influence on confidence for recollection, for example leading eyewitness to think their memory is poor can reduce confidence (e.g., Leippe et al., 2006).

Further evidence that metacognitive judgements are likely to be influenced by the beliefs that people hold, is demonstrated in a study by Costermans et al. (1992) who found both FOK and confidence judgements to be correlated with beliefs that the answer to the question should be known (e.g., the question is in a familiar domain, and other people would know the answer). Finally, some evidence that confidence judgements might be reliant at least in part on information-based cues comes from the neuroimaging research. For example, Chua et al. (2006) observed greater activation in the brain regions associated with self-reflection when participants were making confidence judgements about a recognition task compared to when they were performing the task. This result suggests that confidence judgements are not solely reliant on the retrieval process but might involve addition self-reflective processes (see also Chua, 2012).

However, substantial evidence exists that experienced-based cues also influence confidence judgements. For example, Robinson et al. (2000) investigated the role of vividness as a basis for confidence judgements. In their study participants were asked a series of questions about a crime video. For each answer participants provided a confidence judgement, and three further estimates of subjective effort, vividness, and time taken to answer the questions. Vividness was found to be the best

predictor of confidence, showing that people strongly rely on this mnemonic cue when generating their confidence judgements. Similarly, Brewer et al. (2005) showed that confidence judgements are also based on the completeness of a memory. In this study participants were asked to recall sentences and to report a confidence judgement assessing whether the sentences they recalled were as originally encoded. Participants were found to associate higher confidence to the sentences they judged to be recollected completely as, rather than partially or not at all as those they originally encoded. A further experience-based cue used as indicator of memory accuracy is the ease with which the memory is retrieved. For example, Lindholm et al. (2018) interviewed participants about a video and asked them to provide a confidence rating for each statement reported. They analysed linguistic indicators of retrieval effort such as: hedges ("I guess" or "maybe"), and filler words ("uhm" or "you know") and found that these were negatively correlated with accuracy. Thus, when the number of hedges or pauses increased, confidence judgements decreased (see also Gustafsson et al., 2019; Kelley & Lindsay, 1993).

Further evidence that confidence judgements are based at least in part on the ease of retrieval can be found in studies that have manipulated the difficulty of the retrieval task. Winkielman et al. (1998) asked participants to recall events from their childhood with either a difficult task, whereby, they were asked to recall 12 events of their childhood, or an easy task where they were asked to recall four events. Participants that were given the difficult task judged their childhood memory as less complete than those that were given the easy task, presumably because the former (but not the latter) experienced retrieval difficulty (see also Belli et al., 1998; Gregg et al., 2019; Merckelbach et al., 2001; Schwarz et al., 1991). Similar results were found by Kebbell et al. (1996) who manipulated the difficulty of the questions asked and found that participants were significantly more confident in their correct answers to the easy questions compared to difficult questions – presumably because the former were more easily retrieved than the latter. Confidence for incorrect answers to easy questions was also higher than confidence for incorrect answers to difficult questions, however, crucially, the difference in mean confidence between correct and incorrect answers was far greater for easy than difficult questions. This indicates that when information come to mind easily confidence judgements better discriminate between correct and incorrect answers (see also Wheatcroft et al., 2015).

A further theoretical insight into how the ease of retrieval is used to inform metacognitive judgements can be found in Raghubir and Menon's (2005) model. The authors speculated that the subjective difficulty experienced during a retrieval task triggers a process whereby the participant attempts to assess the causes of such difficulty. Crucially, the model predicts that when the experienced retrieval difficulty is not expected, participants can (after discarding other possibilities) attribute this difficulty to the quality of their own memory. Raghubir and Menon's (2005) model is particularly important because it underlines the relevance of pre-existing expectations in the evaluation of retrieval cues, and potentially suggests that both information-based and experience-based cues concur in the formation of metacognitive judgements.

Confidence during the interview, what do we know?

If confidence judgements are dependent upon metacognitive experience during memory retrieval, then it is of interest to investigate the role of the interviewing techniques, and different ways in which they cue memory. In particular, different interviewing techniques are likely to prompt and direct eyewitness retrieval in different ways, and therefore differentially affect memory confidence. An attempt to understand how eyewitness confidence is influenced by interviewing techniques can be found in a study by Gwyer and Clifford (1997). They interviewed participants with either a Cognitive Interview (CI) or a Standard Interview (SI), either 48 or 96 hours after they had viewed a live staged event. Confidence was measured both before and after the interview. A pre-post confidence difference was calculated by subtracting confidence reported before the interview from the confidence reported after the interview. The results showed that confidence increased after the interview, and that this increase was larger for participants who had received the CI. The authors argued that the CI mnemonics could be responsible for the increase in confidence via the effect they had on participants' retrieval experience. When given a CI the participants reported many details and a complete narrative; this high performance in turn had likely promoted the impression that their memory was good. This interpretation is in line with the notion that the experience of memory completeness was used as a cue to build the confidence judgements.

While Gwyer and Clifford (1997) investigated how confidence changed depending on the type of interviewing received, Granhag et al. (2004) investigated

the effect of different types of interview on confidence and realism. They showed participants a video and interviewed them with either a CI or a SI, while a third group did not recall the event. After the interview, all participants were given a set of 45 forced-choice questions to which participants had to provide an answer and an associated confidence rating. Confidence for participants who initially had been interviewed with either the CI or SI was found to be very similar (68.8%, and 68.1%, respectively), and significantly higher than controls (63.6%). Furthermore, participants in the CI group (but not those in the SI) were overconfident in their memory compared with controls. However, in this study it is difficult to clearly isolate the direct effect of the types of interview on confidence considering that both confidence and overconfidence were measured on the information recognised during the follow-up forced-choice questions and not on the information recalled during the interview. As the authors also explain, a reiteration effect might have inflated confidence by mean of repeating statements already recalled. To overcome these limitations, Allwood et al. (2005) interviewed participants with either a CI or a SI, then transcribed and coded the interviews and two weeks later asked participants to rate their confidence in each unit of information (each statement) that they had provided during the interview. Confidence for participants in both groups was high (91% for the CI, and 91.2% for the SI), and statistically similar. Furthermore, participants in the two groups were similarly well calibrated when highly confident (i.e., 80-100% confidence). Contrary to Granhag et al. (2004), in this study participants in the CI showed under-confidence, but only on the information associated with lower confidence (0-20%, 20-40%).

Thus, the effect of CI and SI on confidence appear to be relatively similar; both types of interview are likely to promote realistic (although not perfect) confidence judgements, and increase eyewitness confidence in their own memory, although this increase is slightly larger for the CI. While the CI notoriously yields more information than the SI, its effect on confidence seems to be largely similar to that of the SI, perhaps due to the relatively similar structure of the two interviews, whereby a free report phase is followed by a questioning phase. An attempt to compare the effect that free report and follow-up focused questions have on confidence can be found in a study by Allwood et al. (2008). Here, participants were shown a video, and a week later they were asked to recall what they had seen in response to a free report prompt followed by a set of focused questions. In a

subsequent session, the authors coded the response to the free recall and focused questions in small units of information (or statements), and asked participants to report a confidence rating in each statement. Adult participants (and children) showed better realism and lower overconfidence in response to the free report invitation compared to focused questions. The authors argued that this difference was likely to be due to the higher level of control over memory reporting that participants have when they respond to a free recall invitation. Under free report conditions participants have control to report information that they were reasonably sure about and withhold information that they were less confident about. However, when answering focused questions, they exert less control because they are not able to control the questions, and thus were more likely to report answers they were less sure about (for similar results see also Yarmey & Yarmey, 1997). Somewhat similar results were found by Knutsson et al. (2011) who investigated whether memory confidence is influenced by requiring participants to expand on their free reported narrative with repeated recall requests and probing questions (i.e., TED questions). The relevant results for this discussion are those related to the different types of question asked. Here, confidence for information reported in response to the probing questions was lower than confidence for information reported in response to the free recall invitation. This suggests that in response to the probing questions, participants were likely reporting details that did not pass the report threshold during the free recall, i.e., details that had been withheld during the previous free report phase (see also Kontogianni et al., 2020).

The literature on eyewitness realism highlights two important considerations, (i) confidence judgements relating to information elicited by free recall and probing questions are more realistic than confidence judgements relating to information elicited by the focused questions and (ii) confidence judgements are higher for information reported in response to free report invitations and tend to decrease for information reported in answers to follow-up questions (probing and focused). The first consideration suggests that confidence judgements might be better predictors of accuracy under certain circumstances *only* (i.e., free recall). A second important point of note suggests that confidence might not be constantly held during the interview, and that it can change as a function of the types of question asked at different stages of the interview, presumably via the effect they have on the retrieval

cues (i.e., experience-based cues). This can be problematic because ideally an interview should not increase or decrease eyewitness confidence.

This PhD will explore these two considerations further. In particular, Study 1 will be focused on confidence shifts promoted by different types of interview, while in Study 2 and Studies 3a and 3b changes in confidence will be investigated in relation to the type of question asked. Finally, Study 4 will attempt to identify the conditions or circumstances related to the interview in which eyewitness confidence remains stable and a realistic predictor of accuracy.

On the confidence - accuracy relationship

The notion that confidence is a better predictor of accuracy under certain conditions is not new, and it is conceptually similar to the argument - developed in the field of eyewitness identification - that confidence is a good predictor of memory accuracy when memory is elicited under ideal conditions. For example, researchers have suggested that confidence in an identification decision is a relatively reliable predictor of accuracy if (and only if) it is elicited under pristine conditions (Brewer & Wells, 2006; Palmer et al., 2013; Sauer et al., 2010; Wixted & Wells, 2017; see also Wixted et al., 2015). As such, if a confidence judgement is collected as soon as possible, before any memory contamination occurs, and by using unbiased lineup procedures, then it can be highly reliable and extremely useful for the Criminal Justice System. However, while this general proposal is not directly disputed, many researchers have questioned whether the pristine conditions can really be met in real life (e.g., Berkowitz & Freda, 2018; Loftus & Greespan, 2017; Wade et al., 2018). Recently Sauer et al. (2019) underlined that even under controlled, unbiased experimental conditions almost 40% of the identification made with very high confidence are wrong, this variation in the confidence and accuracy relation is too high to safely conclude that confidence is a strong indicator of accuracy.

The concept that the relationship between confidence and accuracy is more reliable under certain conditions was initially proposed by Deffenbacher (1980) with the *optimality hypothesis*. Deffenbacher argued that when the conditions are not ideal for memory processing (e.g., high stress level, suggestive instruction or under longer retention intervals) the correlation between confidence and accuracy is more likely to be weak, however the correlation between the two variables improves the closer the conditions are to ideal (see also Leippe, 1980). The optimality hypothesis is based on

the assumption that confidence is a better predictor of accuracy when it is based on the direct access to the memory trace. Under ideal conditions the access to the memory trace is stronger and therefore confidence judgements are more likely to be reliably correlated with memory accuracy (Leippe & Eisenstadt, 2007). Conversely, under less ideal conditions confidence judgements are more reliant on the evaluation of the retrieval heuristics (e.g., fluency, ease of retrieval, completeness) which are often – but not always - valid indicators of accuracy.

In light of this, it becomes important to understand under which conditions eyewitness confidence judgements are correlated with memory accuracy, or, in metacognitive terms, it is important to understand when eyewitnesses are monitoring their memory effectively. Under conditions of effective memory monitoring eyewitnesses will also be more likely to effectively control the information they report by filtering out details that they believe (and are more likely) to be incorrect (Koriat & Goldsmiths, 1996; see also Goldsmith, 2015), or by regulating their specificity (grain-size). The literature has highlighted several examples of circumstances in which monitoring is likely to be poor. For example, when participants are asked deceptive questions, designed to trigger answers that are incorrect and associated with high confidence (e.g., What is the capital of Australia?), then memory monitoring is less effective (see Brewer et al., 2005; Brewer & Sampaio, 2011; Koriat & Goldsmiths, 1996). Further evidence of poor monitoring can be found in the literature on misinformation and memory suggestibility. For example, Loftus et al. (1989) found that participants tended to uptake and confidently hold post-event misinformation. Similarly, Howie and Roebers (2007) found evidence of poorer calibration for answers to misleading questions compared to answers to unbiased questions (see also Bohnam & Gonzalez-Vallejo, 2009; Tomes & Katz, 2000). Poor monitoring can be induced by encouraging memory elaboration, as shown by the literature on the imagination-inflation effect (e.g., Garry et al., 1996; Thomas & Loftus, 2002), and by repeatedly questioning eyewitness about their memories (e.g., Shaw, 1996; Shaw & McClure, 1996; however, for contrasting findings see Odinot & Wolters, 2006; Odinot et al., 2012). Finally, memory monitoring and in particular, discrimination - the ability of the confidence judgements to distinguish between correct and incorrect answers - has been found to decrease over time (e.g., Goldsmith et al., 2005; Shapira & Pansky, 2019), and in response to difficult questions (Kebbell et al., 1996).

To summarise, evidence has shown that monitoring is prone to error, however under conditions in which this process is not impaired, people are able to effectively (albeit not perfectly) assess the accuracy of their memory (e.g., Luna & Martin-Luengo, 2012; Pansky et al., 2005).

Conclusions

Eyewitness confidence has a very important role to play within an investigative interviewing context because it underpins eyewitness's memory regulation, and thus influences decisions regarding whether to report retrieved details, and if so, how to report them. That said, the literature on the basis of confidence has shown that confidence judgements are generated within an error-prone evaluation process, that takes into account both pre-existing knowledge and beliefs about one's own memory abilities, as well as cues generated during the retrieval process itself. The assessment of these cues is not always perfect and therefore confidence is not always a realistic indicator of accuracy. However, when attempting to determine the strength of the relation between confidence and accuracy it is important to consider the contextual factors relating to the retrieval attempt.

Researchers have investigated conditions under which confidence is less likely to be correlated with accuracy, some of which are extremely relevant for the investigative interviewing setting. For example, when time elapses between the encoding and the interview or when eyewitnesses are presented with misinformation, their ability to effectively monitor their memory accuracy is negatively impacted. Finally, evidence also suggests that the types of question asked might influence both confidence, and the confidence and accuracy relationship, and that eyewitness might have both higher confidence and better monitoring regarding the information they provide in free report and open questions as compared to focused questions. Based on these findings this PhD explores the interviewing conditions that are more likely to influence eyewitness confidence, and eyewitness's control over the memory reporting.

Chapter 3: Changes in Eyewitness Confidence

Introduction

Although researchers are still debating whether and how eyewitness confidence should be used in Court, both researchers and practitioners recognise that confidence is an important factor within the legal context. The relevance of eyewitness confidence is clearly captured by the cases of miscarriage of justice due to eyewitnesses whom confidently made a wrong identification. A compelling example can be found in the Thompson vs Cotton case. In 1985 Jennifer Thomson identified Ronald Cotton as the person who raped her. Despite at the time of the identification she was reported to "think" Cotton was the culprit, during the trail she become increasingly confident about her identification. Cotton was convicted largely on the base of her testimony and spent over ten years in prison for a crime he did not commit. It was only in 1995 – when the DNA testing become available, that Cotton was cleared of all charges and released from prison (Thomson-Cannino et al., 2010). This case clearly shows that confidence is not a perfect proxy of accuracy and confident eyewitnesses can be wrong. Yet, confident eyewitnesses are perceived as extremely persuasive, as shown by the surveys and empirical research conducted on police, attorneys and jurors (e.g., Brewer & Burke, 2002; Potter & Brewer, 1999). However, despite its relevance, the extent to which confidence is influenced by many variables at play during an interview is yet to be fully understood. In order to address this topic, a systematic review of the literature was conducted with the aim of identifying and understanding the factors that can lead to shifts in confidence within a recall task. The findings are discussed in relation to which of these factors have implications for the interviewing context.

Confidence malleability

Changes in confidence have been documented in the field of eyewitness identification investigating confidence malleability – the tendency that eyewitnesses become more or less confident regarding an identification judgement as a consequence of certain factors encountered either before or after the identification (Wells & Seelau, 1995). For example, Wells and Bradfield (1998) showed that post identification confirmatory feedback led to inflated certainty in the accuracy of the identification made. Furthermore, it led to the inflated perception of several qualities

related to the identification such as, the quality of the view of the culprit, the attention paid during encoding, how easily the identification was made, and the willingness to testify (for similar results see also Bradfield et al., 2002; Greenspan & Loftus, 2020). Similar, although somewhat smaller, effects have been found for disconfirming feedback which tends to deflate confidence (Luus & Wells, 1994; Steblay et al., 2014; Wells & Bradfield, 1998). Factors related to the administration of the lineup can also influence confidence. For example, leading eyewitness to think that the culprit is in the lineup when they are not, can inflate confidence in a misidentification, because the eyewitness will be more likely to make the identification, and to choose the person in the lineup that looks most familiar, albeit innocent (Leippe et al., 2009; Semmler et al., 2004; Steblay, 1997).

The effect of feedback on eyewitness confidence has also been found to influence memory recollection. For example, Leippe et al. (2006) found that eyewitnesses who received negative feedback about their memory report expressed lowered accuracy in their recollection compared to controls who did not receive the feedback (see also Lida et al., 2020). More recently, Gurney et al. (2014) investigated whether more subtle non-verbal feedback can lead to shifts in confidence. After watching a video of a crime, participants answered questions during which the interviewer either nodded (positive feedback) or shook their head (negative feedback). Participants who received the positive feedback were more confident in their answers than those who received the negative feedback, and the difference between the groups were more pronounced when participants reported to have noticed the head movement.

Eyewitness confidence is clearly strongly affected by social factors such as the administrator's feedback, and by factors related to information gathering procedures, such as instructions administered in identification lineups. However, confidence can shift even without external influence. For example, Granhag et al. (2000) found evidence that eyewitness confidence can change in relation to repeated attempts to (1) judge the confidence in a memory previously reported, and (2) reiterate confidence judgements previously reported. In their study, participants answered questions and reported accompanying confidence judgements relating to the accuracy of each answer given. When asked to provide a new confidence judgement regarding the answers they had given a week before, participants reported deflated confidence on correct answers. Participants who were asked to remember

the confidence judgments given a week before reported inflated confidence for both correct and incorrect answers (for similar results see also Granhag, 1997). In a previous study, Shaw and McClure (1996) found that when participants attempt to answer the same questions repeatedly, confidence increased (see also Shaw, 1996; for contrasting findings see Odinot & Wolters, 2006; Odinot et al., 2009). Thus, despite differing findings regarding the direction of the shift in confidence, these studies collectively show that confidence judgements regarding the accuracy of one's own memory are not stable.

Generalisability of findings

Clearly, a relatively large body of evidence has shown that confidence is malleable. The majority of this research has focused on the relationship between confidence and accuracy in recognition tasks, such as eyewitness identification. There are grounds to suggest that the findings from these studies might not be generalisable to other eyewitness tasks, such as memory recall in an investigative interview. First, memory for faces is often considered unique and, at least to some extent, dissociated from memory for non-face stimuli (e.g., Kanwisher & Yovel, 2006). Furthermore, the retrieval process involved in recognition tasks compared to that involved in recollection tasks can yield slightly different cues for confidence, because the former is likely to be driven by familiarity while the latter requires an intentional, more elaborated memory search. Here, several cues such as the retrieval effort or ease of retrieval, are more likely to be available during recollection tasks than recognition tasks (Robinson & Johnson, 1996; Robinson et al., 1997; see also Robinson et al., 2000). If the cues that underpin confidence judgements in recollection and recognition memory are different, then it is also possible that confidence for these two types of memory is affected to a different extent by any variable of influence.

Evidence that internal cues mediate confidence shifts caused by feedback is reported by Semmler and Brewer (2006) who found that when participants had good internal evidence that their identification was correct (quicker and more fluent identification), then 'confirming feedback' led to smaller confidence alteration compared with disconfirming feedback. In contrast, when the internal evidence for the identification decision was poor (slower and less fluent identification), then 'confirming feedback' caused a larger confidence alteration than 'disconfirming

feedback'. Thus, in this study fluency might have mediated the extent to which confidence was altered by the external feedback.

A second pitfall within the available literature is that the influence of different factors on confidence shifts is often investigated using between-subjects design, and therefore overlooks potential individual differences. Evidence of individual differences in relation to confidence judgements has been documented in the literature (e.g., Pallier et al., 2002; Soll, 1996; see also Thompson & Mason, 1996). More broadly, the large variability between confidence and accuracy relationship often found in the literature suggests that individuals vary in the way they monitor and control their memory (e.g., Deffenbacher et al., 1981; Thompson & Mason, 1996) and so these should also be considered.

The present review of the literature on confidence shifts

The goal of this systematic review is to identify and synthesise findings from studies that have examined confidence shifts within an episodic recall task, so that implications can be drawn for understanding the role of confidence on performance within an investigative interview. Thus, studies examining shifts in confidence were excluded if they featured a recognition memory task such as multiple - choice questions or yes/no questions, or a facial identification task. Studies were included if (a) they used to-be-remembered stimuli such as events or episodes, which are of a similar complexity to crimes or wrongdoings that eyewitnesses are asked to recall during an interview, and (b) they featured a recall test. Furthermore, we only included studies that analysed confidence within-subjects. Finally, we embrace a broad perspective in the analysis of confidence shifts compared with the literature on confidence malleability. While the latter often aims to identify the factors that lead to inflated or deflated retrospective confidence and *do not* affect accuracy, we analysed shifts in confidence independently from changes in memory accuracy. This is because we do not aim to analyse the reliability of confidence judgements (i.e., the confidence-accuracy relation) but rather to understand if, and under which conditions, the impression of accuracy changes. We are therefore able to include in our analysis prospective judgements and beliefs about the quality of one's own memory.

Method

Search strategy

The search was conducted on two electronic databases: Web of Science (Web of Knowledge) and PsycInfo. The first is an interdisciplinary database and the second is primarily focused on psychology and related disciplines, therefore a combination of both was likely necessary to access the literature of interest. A year range was introduced which identified studies published between 1990 to 2019 (June), whereby the former date refers to the publication year of the framework of metamemory developed by Nelson and Narens (1990), while the latter is the year I first conducted the search. The search strategy constituted of three keywords strings which identified studies focused on (i) a metacognitive judgement, (ii) which was aimed to assess the accuracy of a memory, (iii) and either changed or remained stable. In order to maximise the number of included studies at this stage, the strings were all searched *In Topic*.

The first string constituted the following keywords: (confidence NEAR/4 memor*) OR “confidence/accuracy” OR "retrospective confidence" OR (confidence NEAR/4 judg$ment*) OR (metacognit* NEAR/4 judg$ment*) OR (memor* NEAR/4 judg$ment*) OR (metamemor* NEAR/4 judg$ment*) OR "judg$ment* of learning". The Wildcard used were the asterisk (*), which includes the truncation, for example the keyword “memor*” finds both “memory” and “memorial”. The second Wildcard used was the dollar ($) which finds words that can be spelled in different ways, for example “judg$ment” finds studies including both the word “judgement” and “judgment”. I also used the Proximity operators such as “confidence NEAR/4 memor*”, in this example NEAR/4 finds the studies where the words “confidence” and “memory” are separated by a maximum of 4 words. Quotation marks were used to search for exact words sequence, for example "retrospective confidence" looks for studies containing this exact words-string. The two databases adopt similar Wildcard and Proximity operators but use different syntaxes, therefore this was adapted accordingly to the database guidelines.

In the first string the keywords “(confidence NEAR/4 memor*)”, “confidence/accuracy”, "retrospective confidence", and ‘(confidence NEAR/4 judg$ment*)’ were included to target studies on memory confidence. I predicted that relevant studies might have used more general labels, hence I also included the keywords “(metacognit* NEAR/4 judg$ment*)”, “(memor* NEAR/4 judg$ment*)”,

“(metamemor* NEAR/4 judg$ment*)”. Finally, I assumed that relevant studies might have compared judgements of learning (i.e., a judgment on the likelihood to remember a target memory) to retrospective memory, therefore the keyword "judg$ment* of learning" was also added.

The second string included the keywords (accura* OR correct*) AND memor*, these narrowed down the studies yielded by the first string and included only those focused on the judgements in memory accuracy. For example, this string excluded articles focused (exclusively) on judgements on vividness, or frequency judgements (i.e., a judgement on the number of correct answers given on a test).

Finally, the third string narrowed down the studies yielded by the two previous strings to those including either a reference to a change, or variation, or a lack of change (i.e., stability). The keywords used were: (stab* OR unchang* OR invariat* OR chang* OR malleab* OR inflat* OR retrospect* OR prospect* OR altera* OR variat* OR increas* OR decreas* OR shift* OR (repeat* recall*) OR (repeat* retriev*) OR (repeat* test*) or retell* OR (pre test* AND post test*) OR (pre recall* AND post recall*) OR (pre retriev* AND post retriev*) OR (before recall* AND after recall*) OR (before retriev* AND after retriev*) OR (before test* AND after test*) OR “repeated measure*” OR “within subject*”). We were particularly interested in studies that had measured confidence on two occasions, therefore we added keywords such as “repeat* recall*”, “repeat* retriev*”, and “pre recall* AND post recall*”. The last two keywords included were “repeated measure*” and “within subject*” which we expected to target studies that manipulated confidence within participants.

Inclusion and exclusion criteria

In order for a study to be included, five criteria had to be met: (1) the keyword confidence, or judgement of learning, or metacognitive judgement or a synonym had to be mentioned within the abstract; (2) the judgement measured had to relate to memory accuracy; (3) the judgement on the accuracy of the same memory had to be measured at least twice using a within participants design; (4) the memory had to be recalled and not recognised, and finally (5) the to-be-remembered memory had to be an episode or event.

The aim of the review was to quantify the malleability of confidence, therefore only experimental articles were considered. Non-experimental papers,

qualitative studies, book chapters and review papers were excluded. Furthermore, considering that this review is not focused on atypical cognitive functioning, I also excluded studies on mental health population. Last, studies were excluded if they were not published in English.

Screening and selection

The search yielded 827 results, of which 480 were from Web of Science, while the remaining studies were from PsycInfo. The selection strategy that followed was conducted in four stages. At each stage, studies that were not clearly excluded were eligible for the following selection stage.

The initial sifting stage was carried out in Zotero, which identified the duplicates, which were then excluded. Following this, the remaining studies were copied into an Excel file for the title screening stage. Studies that were clearly off-topic, for example studies on non-human animals, or studies on non-human memory or computer memory, were excluded.

Following this, the abstracts of the remaining studies were analysed. Here the reasons for exclusion depended on the information available in the abstract, which varied largely. Indicatively at this stage I excluded: (i) studies that did not mention confidence or a metacognitive judgement, (ii) those that clearly did not measure confidence on memory accuracy, (iii) studies featuring a clinical sample or mental health population, (iii) studies that clearly stated to have tested memory with a recognition test, such as an identification task or using multiple - choice questions, (iv) or to have used study material that did not meet Criterion 5, such as: word pairs, words lists, geometric or simple shapes, and (v) studies that had clearly measured confidence only on one occasion.

The remaining studies were then progressed to the full text reading stage. At this stage I scrutinised the method and result sections, and excluded studies that did not fully meet the inclusion criteria. In the last stage, I analysed the reference lists and citations of the included articles. Figure 3.1 shows a visual illustration of the selection process.

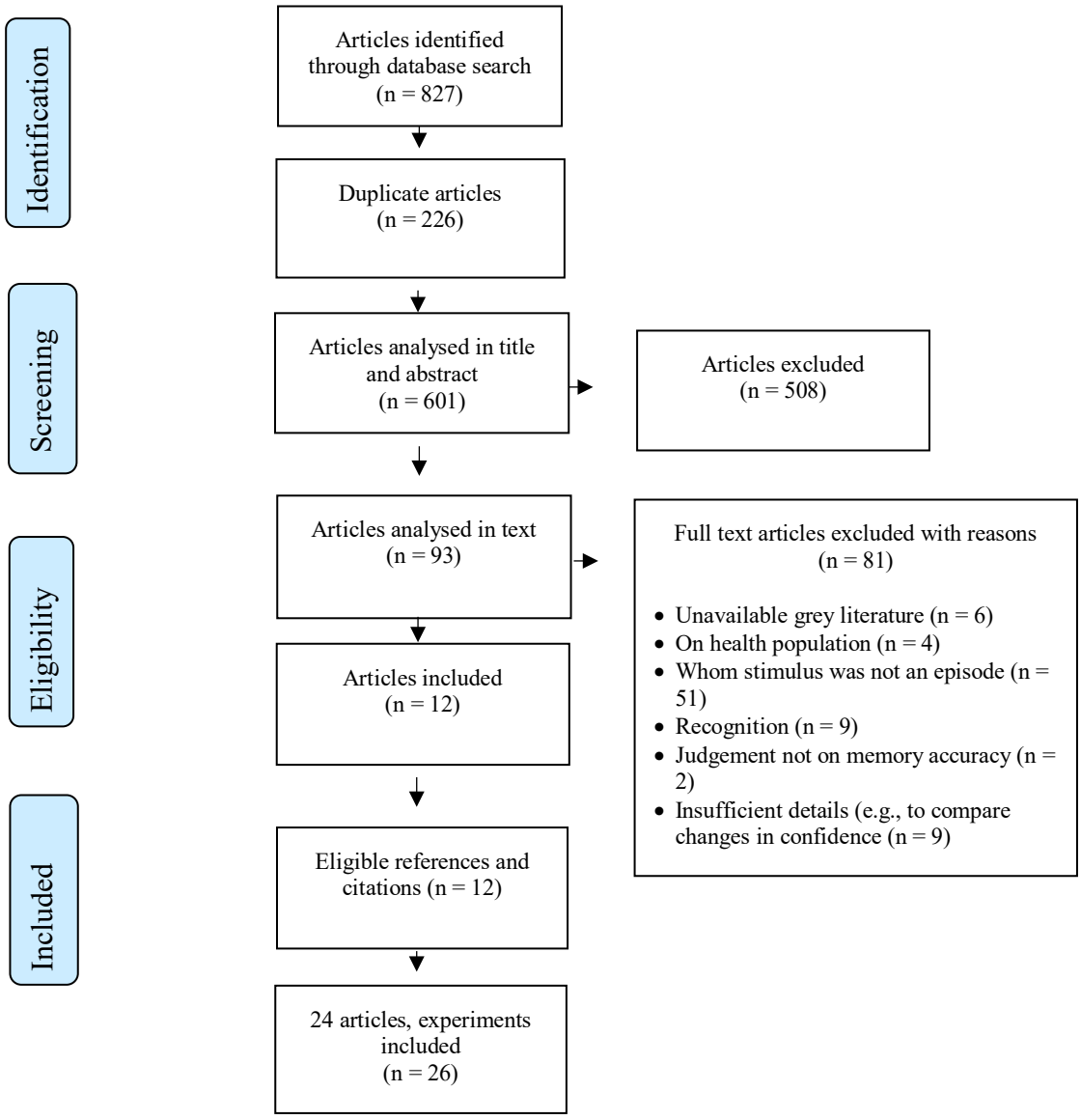


Figure 3.1. PRISMA chart illustrating the four stages.

Results

The search yielded 24 articles, comprising a total of 26 eligible studies. About 50% (n = 13) were in the field of on eyewitness memory, 42% (n = 11) were in the field of Flashbulb Memory (FBM), and 8% (n = 2) in the field of autobiographical memory. Although there is substantial overlap between the FBM and autobiographical memory, the former investigates personal memories related to a relevant public event, while the latter focuses on personal memories not necessarily linked to any public event.

All the studies tested adult participants; of these, 80% (n = 21) reported to have tested university/college students, two studies included students as young as 16

(Kraha et al., 2014; Gwyer & Clifford, 1997), while two studies included adults older than 60 (Holland & Kensinger, 2012; Wolters & Goudsmit, 2005). The latter were both in the field of FBM.

Memory related variables

Type of to-be-remembered event

The majority of the studies on eyewitness memory used short videos of crime related events in the encoding phase. The content (not always specified) included kidnapping, car accidents, and theft. Exceptions to the use of video as the to-be-remembered material included live staged events (Gwyer & Clifford, 1997; Winningham & Weaver, 2000), a photo of a crime (Sharps et al., 2012), and a description of a co-participant (Ebbesen & Rienick, 1998).

The studies on FBM asked participants to remember details of the circumstances in which they first heard about an extraordinary real-life event (e.g., the September 11th terrorist attack, Osama Bin Laden's assassination, Karol Wojtyła's death, or an important sports event). These details are often referred to as canonical or autobiographical. Beside the canonical FBM, in some studies participants are also asked to remember factual memories about the event, often referred to as event memories (Coluccia et al., 2010, studies 1 and 2; Hirst et al., 2015; Talarico & Moore, 2012; Wolters & Goudsmit, 2005). For example, Wolters and Goudsmit (2005) asked participants how many planes were involved in the September 11th terrorist attack. Four studies compared FBM to everyday memory (Kraha et al., 2014; Talarico & Rubin, 2003; 2007; Weaver, 1993).

Similar to the studies on FBM, the studies on autobiographical memory did not present to-be-remembered material, rather, participants were asked to recall personal events. An exception was Holland and Kensinger's (2012) study, where participants were asked to report their memory for a presidential election.

Type of memory test

As specified in the inclusion criteria (criterion 5), all studies included a retrieval test. Across studies the full lists of questions asked were rarely available, therefore the following descriptions are based on the information provided in the method sections. Five studies on eyewitness memory included closed questions requiring short answers (i.e., 5Wh) (Buratti & Allwood 2012; 2013; Michael &

Garry 2016, study 1; 2019, study 2; Turtle & Yuille 1994, study 1). Four studies used a mixture of 5Wh and TED questions or open prompts (Ebbesen & Rienick, 1998; Pezdek et al., 2007, study 1 and 2; Winningham & Weaver, 2000). The remaining studies adopted a more complex, interview-like memory test. For example, Odinot and Wolters (2006), and Odinot et al. (2012) used a general invitation followed by open questions (i.e., TED questions), while Sharps et al. (2012) used a free invitation followed by three requests to recall additional information. Gwyer and Clifford (1997) adopted the full CI (including the four mnemonics).

Across the studies on FBM, the majority reported to have used surveys which included a predominance of 5Wh questions. However, some authors (i.e., Kraha et al., 2014; Talarico & Rubin, 2003; 2007) refer to these as open-ended questions, therefore for these studies I was unable to clearly identify the type of test used. Finally, Weaver (1993), and Weaver and Krug (2004) also asked a mixture of 5Wh and TED type of prompt.

Regarding the studies on autobiographical memory, Holland and Kensinger (2012) used 5Wh questions, while Stone et al. (2013) used a free recall invitation.

Confidence related variables

Type of question

As specified in the inclusion criteria (criterion 2), all studies must have elicited a confidence or metacognitive judgment relating to memory accuracy. The way in which confidence judgements were solicited varied largely. In the studies on eyewitness memory, accuracy or correctness was often mentioned in the question eliciting the judgement (e.g., Odinot & Wolters, 2006; Pezdek et al., 2007). While these questions vary in clarity, they all ask participants to report a retrospective judgement to assess an answer given. An exception is Gwyer and Clifford (1997) who asked participants to report both a retrospective confidence judgement and a prospective judgement (i.e., "how confident are you that you will be able to answer correctly the questions that I am about to ask you", p. 126).

Similarly, the studies on FBMs largely used non-specific questions where participants were asked to rate confidence or their level of certainty in their recollection. For example, Stone et al. (2013) used a specific retrospective judgment where participants were asked how confident they were that "they were remembering each event as it *originally occurred*" (p. 252). A slight variation of this can be found

in the study by Weaver and Krug (2004), where participants were asked how confident they were that they answered the question correctly. Some, studies on FBM used composite measures of confidence by pulling together confidence ratings elicited by two questions. For example, Kraha et al. (2014) collapsed the ratings from two questions asking about (1) the belief that the memory participants had, really occurred the way they remembered it, and (2) if they could be persuaded that their memory was wrong (see also Talarico & Rubin, 2003; 2007). While, Hirst et al. (2015) asked half of the participants how confident they were in their recollection (a retrospective judgement), and the other half how well they thought they would remember the event in ten years (a prospective judgement), however they only reported results for the sample that answered retrospective judgements. A prospective judgement was also used by Holland and Kensinger (2012) who asked participants how well they thought they would remember the event in six months.

Granularity of the judgement.

Considering the variety of memory test used across the studies, it is important to also analyse the granularity or level of the detail of the memory participants were required to assess with each judgement. This is because confidence judgement asked at a question level can be different from confidence judgement on the overall recollection of a complex event. Here, the former is a much more precise estimate of accuracy than the latter, because it assesses the correctness of a smaller unit.

All the studies using closed or 5Wh questions asked for a confidence judgment after each answer, with the exception of Kraha et al. (2014), Talarico and Moore (2012), and Talarico and Rubin (2003; 2007), who collected retrospective confidence judgements in the recalled memory as a whole, and Holland and Kensinger (2012) who collected a prospective judgement in the memory as a whole. Michael and Garry (2016, study 1; 2019, study 2) asked for confidence ratings at both levels; after each question and in the memory as a whole.

The studies that adopted TED questions or open invitations generally collected confidence at the question level without clarifying which part of the answer given was to be assessed by the judgement. For example, Weaver (1993) asked participants “describe in as much detail as possible how you heard about the news of the beginning of the bombing of Iraq” (p. 41), and then asked them to report a confidence judgement in the whole answer given. An attempt to collect precise

confidence judgements on answers to open questions can be found in studies by Odinot and Wolters (2006), and Odinot et al. (2012), where participants were asked to write their answers in small units of information and were then asked a confidence judgement for each statement (e.g., 'question; 'What did the dog do when it came out of the water?' answer; 'it climbed on the bank'; 'it shook off the water'; 'it ran to his boss'', Odinot & Wolters 2006, p. 977). A similar methodology is used by Turtle and Yuille (1994, study 1), whereby participants were asked to report confidence ratings in both statements given in the written CI, and in their answers to ten closed questions about forensically relevant elements (however insufficient information was reported on the former measure, therefore we only analysed the latter). Finally, Gwyer and Clifford (1997) asked for global prospective and retrospective confidence judgements.

Confidence scale

The majority of the studies collected confidence judgements using a five or seven point Likert scale. However, Weaver (1993), and Gwyer and Clifford (1997) used a three point scale and a ten point scale, respectively. The remaining studies used a confidence scale ranging from 0% to 100%.

Experimental manipulations and factors analysed

The descriptive characteristics of the target memory and the confidence judgment outlined above provide an insight into the similarities and differences between the studies. However, of central importance for this review is an examination of the variables that led to either shifts in confidence or, alternatively, to confidence stability. Due to the focus on confidence measured within-subject, in this section I will only consider factors for which the effect has been investigated across confidence measurement times.

Regardless of whether studies focus on FBMs, autobiographical memory, or eyewitness memory, there is an observable consistency in the methodology applied, whereby participants are given a largely similar memory test on several occasions and at different time intervals. Despite this similarity, these studies differ with regards to whether the topic of interest is on shifts in confidence in relation to *time lags* (i.e., the time elapsed between recall tests), or whether the focus is on shifts in confidence in relation to *repeated recall attempts*.

All the studies on FBM and one study on autobiographical memory (Holland & Kensinger, 2012) focused on confidence shifts across time lags. The length of time-lag varied between studies but, in general, the studies adopted relatively wide time intervals. The narrowest time interval was found in Wolter and Goudsmit's (2005) study where participants were asked to recall the FBM twice; two weeks after, and again two months after, the target event was encoded. The longest time lag was found in the study by Hirst et al. (2015) that requested participants to complete a survey about the FBM event on five occasions, up until 119 months after the event was encoded.

In the FBM studies, all the remaining manipulations or factors which effect was measured on confidence shifts were (i) *types of memory*, for example, Coluccia et al. (2010) investigated confidence shifts in autobiographical FBM versus event memory; while Kraha et al. (2014); Talarico and Rubin (2003; 2007) and Weaver (1993) investigated confidence shifts in relation to FBM versus everyday memories. In addition, Ferre Romeu (2006) measured changes in confidence in relation to core (or canonical, e.g., when they heard about the news, where were they at the time, what were they doing) versus peripheral details (e.g., the clothing they were wearing when they heard about the news, what was their first thought, and what they did immediately afterward hearing the news). Other factors include (ii) *number of recall attempts* in Kraha et al. (2014) who reported results for participants that completed the survey two or three times; (iii) *emotional valence* analysed in Kraha et al. (2014) who studied the impact of a positive event on confidence shifts, and in Stone et al. (2013) who analysed confidence shifts for positive versus negative events. While Holland and Kensinger (2012) investigated shifts in confidence for an event perceived by participants as either positive, negative or neutral; (iv) *residency* was investigated by Hirst et al. (2015) that reported mean confidence for memories about the September 11th terrorist attack for participants who resided in New York versus those who resided elsewhere in North America (although no statistical comparison was reported on the means); (v) *social group* was manipulated in Talarico and Moore (2012) which investigated shifts in confidence in the memory for a sport event in two rival fan groups; (vi) *age,* was investigated in Holland and Kensinger (2012) how compared young, middle aged, and elderly, and (vii) *expectation*, investigated by Coluccia et al. (2010) who measured shifts in confidence for both surprising and expected FBMs. A final variable manipulated was (viii) *selective retrieval* in Stone et

al. (2013), here participants were asked to recall some memories but not others and confidence was measured for both practised and unpractised memories before and after the practice session.

Where confidence shifts across *repeated retrieval attempts* was of interest, most studies were in the field of eyewitness memory, presumably because witnesses are often interviewed on multiple occasions. For example, Winningham and Weaver (2000) asked participants to recall a staged event on five separate occasions, each time with a one-week interval between recall attempts, (see also Turtle & Yuille, 1994, study 1). The method used by Sharps et al. (2012) differed slightly in that there was no time delay between tests, instead participants first recalled the encoded event in response to a free-recall prompt, and then immediately after responded to three follow-up invitations to report additional information. While, Pezdek et al. (2007) manipulated the *number of presentations* of questions within the same recall attempt, here all participants were presented with three unanswerable questions once versus on three occasions (study 2).

Three further studies investigated confidence shifts as a function of repeated recall and (i) *retention interval*, i.e., the time elapsed between the encoding phase and the initial recall attempt. For example, Odinot and Wolters (2006) and Odinot et al. (2012), interviewed groups of participants at different retention intervals (one week or three weeks after encoding) and on different occasions (on two or three occasions). Similarly, Ebbesen and Rienick (1998) asked groups of participants to recall attributes of an unfamiliar co-participant (e.g., ethnicity, eye colour, hair colour) at different retention intervals (immediately, one day, one week or 28 days after encoding) and on different occasions (on two or four occasions).

The effect of (iii) *type of test* on changes in confidence was investigated in several eyewitness memory studies. In particular, Pezdek et al. (2007) asked participants both answerable and unanswerable questions on two separate occasions, while Michael and Garry (2016, study 1; 2019, study 2) investigated shifts in confidence in relation to either difficult or easy questions. The type of test was also manipulated in Gwyer and Clifford's (1997) study, whereby participants were interviewed with either a CI or a SI.

The remaining factors and manipulations measured in the eyewitness memory studies related to the type of test instruction given to participants. For example, Winningham and Weaver (2000) manipulated the (iv) *retrieval instruction*, whereby

participants were either pressured to report everything they could remember (pressure group), or not pressured and instead instructed to report only the information they were absolutely sure about (control group). While, Pezdek et al. (2007) asked participants in the initial recall test to either answer all the questions (forced guess group) or to volunteer their answers if they chose to do so (volunteer guess group). Buratti and Allwood (2012; 2013) investigated the instruction to improve realism. In both studies participants completed a confidence and an adjustment phase (a further extra adjustment phase was added in Buratti and Allwood, 2013). In the confidence phase, participants answered all the questions about the event and provided confidence judgements in the accuracy of their answers. In the adjustment (and extra adjustment) phase they were given the chance to modify their judgements in order to achieve maximum realism. Furthermore, additional metacognitive prompts and advice on how to improve the realism of the confidence judgements were manipulated. In the former participants were asked to either rate the ease with which they recalled their answers or to state if their remembered or knew the answer given (Buratti & Allwood, 2012). While in the latter participants were advised to improve realism by either decreasing confidence in the answers they thought were incorrect and increasing confidence in the answers they thought were correct, or by taking into account that "remembered" answers are generally accompanied by higher realism than "known" answers (Buratti & Allwood, 2013).

Changes in confidence and discussion

Shifts and stability in Flashbulb and everyday memory

All the studies that investigated confidence on FBMs report that these are generally associated with very high confidence. Out of the studies that investigated *time lags*, those that analysed the average confidence ratings associated with the canonical questions (e.g., what was the source of the news, where were they when they heard the news, whether other people were there and what they were doing) often reported confidence stability (Hirst et al., 2015; Talarico & Rubin, 2003; 2007; Wolters & Goudsmit, 2005). Partial evidence of confidence stability was reported by Weaver and Krug (2004) who analysed mean confidence ratings at the question level and found that confidence was generally stable. However, confidence was found to decrease in relation to two questions asking respectively about the clothing worn and

the first thought participants had when they first heard about the event. Ferre Romeu (2006) found decreased confidence across all questions. However, the magnitude of the decrease was different depending on the *type of memory;* in particular confidence decreased more for peripheral details than core details (for somewhat different results about similar questions see Weaver, 1993). Decreased confidence for FBMs as a function of time lags was also reported by Talarico and Moore (2012) who asked participants canonical questions and measured confidence about the overall memory rather than at questions level. Similar results were found by Kruha et al. (2014) who investigated an event not found to lead to the formation of a prototypical FBM, hence showing that confidence for non FBMs is likely to decrease over time. In line with the latter result, the studies that manipulated the *type of memory* reported that confidence for everyday memories decreased more over time than confidence for FBMs (Talarico & Rubin, 2003; 2007; Weaver, 1993). An exception was found by Coluccia et al. (2010) who reported evidence that confidence in both FBM and event memory followed a quadratic trend, whereby it decreased after five months and increased after one year.

Further factors analysed in the literature of FBM were the *number of recall attempts, social groups,* and *expectation*, of which none were found to have a statistically significant effect on shifts in confidence (Coluccia et al., 2010; Kruha et al., 2014; Talarico & Moore, 2012). Less clear conclusions can be made regarding *emotional valence, age* and *residency.* An examination of the mean confidence reported by Holland and Kensinger (2012) shows a decrease in confidence for participants that reported positive, negative and neutral emotions regarding the event, and this decrease was slightly more pronounced in the middle age group compared with the young and older group. However, clear conclusions cannot be made because the authors did not report information about the statistical comparisons between these means. While in Hirst et al. (2015) participants that lived in New York reported similar shits in confidence that those that lived in other cities in North America; however, no clear conclusion can be made because statistical comparisons between the means were not reported. Finally, Stone et al. (2013) found a significant interaction of *emotional valence*, *retrieval practice* and recall time, whereby non-practiced positive memories (but not non-practiced negative memories) were held less confidently after the practice session than before the practice session.

In summary, the research on FBM seems to report that long time intervals are likely to lead to decreased confidence for everyday or personal memories. However, the evidence regarding shifts in confidence for FBMs is less consistent; while some studies show that confidence remains stable over long periods of time, other find that confidence tends to decrease. The studies that have measured confidence at the 'question level' provide some insight into the specific aspects of the FBMs that are more likely to be associated with decreased confidence. In particular, confidence for peripheral details, such as the clothing worn and/or the first thought at the time, was more consistently found to decrease. On the contrary, confidence for the core (canonical) details related to the circumstances in which the participants heard about the news was more consistently (albeit not always) found to remain stable. An examination of memory performance showed that core details are more consistently remembered than peripheral details, which could explain why confidence for the former type of details tends to decrease less (Ferre Romeu, 2006). However, the difference between the decrease in confidence for core and peripheral details could also be explained in terms of the retrieval cues that underpin confidence, such as the ease of retrieval (e.g., Kelley & Lindsay, 1993). It is possible that peripheral details become more difficult to retrieve over time compared to core details, perhaps because the former are rehearsed less frequently than the latter (for a discussion on the relation between rehearsal and confidence for FBMs see Ferre Romeu, 2006; Talarico & Moore, 2012). The higher increase in retrieval difficulty over time for peripheral details compared with core details could in turn lead to a larger decrease in confidence for the former compared with the latter type of details.

The cues associated with the retrieval of FBMs can also explain the consistent finding that confidence for canonical details tends to remain very high after a long delay. FBMs are by definition vivid memories (Brown & Kulik, 1977), and vividness is an important cue for the formation of confidence judgements (Robinson et al., 2000). Research has showed that the vividness with which FBMs are retrieved tends to remain high even after a long delay (e.g., Talarico & Moore, 2012; Ferre Romeu, 2006), which could in turn explain why FBMs remain associated with high confidence. The high confidence associated with FBMs could also be explained in terms of ease of retrieval. FBMs are related to shocking, salient, and unique events that are more likely to stand out in the array of memories that people

hold, and it is possible that they remain particularly easy to retrieve even after very long period of time.

Shifts and stability in eyewitness memory

As with the literature on shifts and stability in FBM and everyday memory, the literature that investigated the effect of *repeated recall attempts* and *retention interval* on confidence in eyewitness memory presents some contrasting findings. For example, Turtle and Yuille (1994, study 1) stated that repeated recollections of the same information did not detect potential changes in confidence (however, in the latter study no statistical tests are reported to support this comparison). Furthermore, Ebbesen and Rienick (1998) found that confidence (and accuracy) did not change significantly between the initial and last recall attempt and this was independent of the retention interval. More recently, Winningham and Weaver (2000) also found confidence stability in relation to repeated recall attempts. Similar findings are reported by Odinot and Wolters (2006) and Odinot et al. (2012), who found confidence (and accuracy) to remain stable over repeated recall tests, independently of the retention interval. Confidence for correct answers was also found to remain stable across recall attempts, but inflated confidence for repeatedly reported errors was found in the longer retention interval group (3 weeks) in Odinot and Wolters (2006). However, the latter result was not replicated in Odinot et al. (2012).

In stark contrast are findings from Sharps et al.'s (2012) study, who found a decrease in confidence at each attempt to recall further information. While, Pezdek et al. (2007) found a non-significant main effect of recall time on confidence (this was significant in study 2), a main effect of type of question (*type of test*) and a significant interaction between recall time and type of question. Here, the interaction revealed that the direction of the change in confidence depended on the type of test. In particular, confidence for repeatedly reported correct answers to answerable questions decreased over time, while confidence for repeatedly reported guessed answers to unanswerable questions increased over time (similar findings were found in study 2).

Evidence that the *type of test* can lead to changes in confidence is also reported by Gwyer and Clifford (1997) who found a slight increase in confidence after the interviews, with the participants in the CI group reporting higher confidence increase than those in the SI (although the statistical significance of these shifts was

not reported). Another example is provided by Michael and Garry (2016, study 1; 2019, study 2), who showed that the difficulty of the test influenced changes in confidence. Here, a comparison between mean confidence in answer to the initial question of the test and confidence in the whole memory reported after the test showed that confidence decreased when the initial question was easy and increased when the initial question was difficult.

Some of the studies focusing on confidence in eyewitness memory examined *retrieval instruction*. Here, the type of retrieval instruction was not found to affect confidence changes, as showed by the lack of interaction between recall time and (i) the 'pressure instruction' in the study by Winningham and Weaver (2000), or (ii) the 'forced/voluntary guess instruction' in the study by Pezdek et al. (2007). Furthermore, in the latter study, the manipulation relating to the *number of presentations* did not interact with recall time, showing no effect on shifts in confidence. While Buratti and Allwood (2012; 2013) showed that the *instruction to improve realism* strongly influenced confidence shifts. In both studies participants tended to express lower confidence in the incorrect answers and not to change their confidence for correct answers, in the adjustment phase (but not in the extra adjustment phase in Buratti and Allwood, 2013), and were therefore able to improve their realism. Whereas, the additional manipulations (i.e., metacognitive prompts and the advice on how to improve realism) did not affect this adjustment significantly.

Overall, the studies on eyewitness memory seem to provide relatively convergent evidence that repeated recall attempts at different retention intervals and at time delays of at least one day, are not likely to lead to statistically significant changes in confidence. Whereas, repeated attempts to recall further information during the same test is likely to lead to decreased confidence. This pattern of results suggests that repeatedly asking participants to report additional information compared to asking them to recall the target event on multiple occasions might trigger different retrieval experiences that in turn could lead to the observed results. It is possible that when participants are asked to report details in addition to those they volunteered in the initial free recall phase, they are more likely to report details associated with higher retrieval difficulty compared to when they are asked to recall their memory on multiple occasions.

For example, Sharps et al. (2012) presented participants with three invitations to report any additional information they could remember. Each request might have

increasingly pressured participants to either report details that initially did not pass the report threshold and were selected to be withheld, or to search for details that were not initially accessed. In the first instance the initially withheld details were likely to be associated with low retrieval fluency and therefore low confidence, while in the second instance memory search might have triggered details that were increasingly more difficult to access and therefore likely to be associated with low confidence. On the contrary, when participants were presented with a similar test and reported their memory on several occasions, (e.g., Ebbesen & Rienick, 1998; Odinot & Wolters, 2006; Odinot et al., 2012), they were less likely to report details associated with low retrieval fluency and less likely to search for details difficult to access. Thus, in these studies participants did not show decreased confidence. This difference is important and suggests that pressuring participants for further information might lead to confidence decline, while repeatedly interviewing participants about their memories might only have a trivial impact on memory confidence.

The systematic review also suggests that confidence seems to change in both directions in relation to the type of test. Complex interview protocols, such as the CI and the SI, seem to lead to increased confidence, however the magnitude of such increase in unclear (Gwyer & Clifford, 1997). Furthermore, while answerable questions can seemingly decrease confidence when the same correct answer is repeatedly reported, unanswerable questions are likely to cause confidence inflation when incorrect answers are repeatedly reported (Pezdek et al., 2007). While the latter result implies that suggestible questions are likely to lead to confidence inflation, the former result is not in line with the general findings reported above; i.e., confidence remains relatively stable across repeated recall attempts. One possibility, although speculative in nature, is that the decrease in confidence for repeatedly reported correct answers found by Pezdek et al. (2007) was due to the interposition of the suggestive questions in the retrieval test. It is possible that suggesting details that participants had never encoded might have promoted the impression that their memory was not so good, hence the decision to decrease confidence for correct answers. However, this explanation does not account for the inflated confidence in incorrect answers to unanswerable and suggestive questions.

One final finding is related to the shifts in confidence made while attempting to improve realism. This result shows that confidence judgements are built within an

error-prone heuristic process, and it suggests that participants might be aware of potential unrealistic judgements they initially reported. Thus, when instructed they are able to report more appropriate and reliable confidence judgements.

Conclusions

The aim of this review was to (1) gather the available evidence relating to changes in confidence in a recalled memory, (2) identify the factors that are likely to lead to such changes, and (3) understand whether any of these factors are likely to influence eyewitness confidence within the context of interest – i.e., an investigative interview. We selected strict criteria to identify relevant studies and gather this evidence, therefore the literature that met our criteria was not large. However, in line with the research on confidence malleability, the eligible literature clearly supports the idea that confidence for episodic memory can change. Among the variables that are likely to promote such changes, some are directly relevant for the investigative setting.

In particular, the results related to the effect of repeated recall attempt and types of question are useful to pinpoint some of the circumstances in which eyewitness confidence is more likely to change and some in which confidence is more likely to remain stable. For example, the literature highlighted that repeated recall attempts do not appear to cause changes in confidence in circumstances in which the type of test (1) is similar across retrieval attempts, (2) is not misleading or suggestive, (3) does not repeatedly pressure for addition information, and (4) the time delay between interviews is relatively short (i.e., up to 7 days). Some of these results are not new, and underpin the already available guidelines on best-practice eyewitness interviewing. For example, it is well established that suggestive and misleading questions should not be asked within an interview because they are likely to elicit inaccurate memories from witnesses (e.g., ABE, Home Office, 2011). Similarly, research has shown that the access to a memory trace is likely to decay over time and it suggests that eyewitnesses should be interviewed as soon as possible after the incident (e.g., Gabbert et al., 2009). It appears that these suggestions are compatible to eyewitness confidence as they are to eyewitness memory. As such, repeatedly interviewing an eyewitness with similar, non-misleading interviews, conducted with relatively short time delays is unlikely to cause a significant shift in confidence.

Perhaps a less clear conclusion can be made regarding the encouragement to provide additional information. It appears that confidence is likely to decrease during the probing phase of an investigative interview, presumably because in answer to probing questions eyewitnesses are more likely to lower their report criterion and therefore report information they are less sure about, or because they engage in a more effortful memory search for the requested details. This explanation is in line with the literature on realism of confidence discussed in Chapter 2. In particular, with the finding that confidence for information recalled via follow-up probing questions is more likely to be lower than confidence for information recalled via a free recall invitation. Following this pattern of results, one aim of this PhD is to understand whether eyewitness confidence is likely to change during an interview, and in particular whether confidence is likely to remain stable during the free recall phase and decrease during the follow-up phase of the interview.

Chapter 4: Study 1. Investigating the effect of different quality of an initial interview on memory performance and memory confidence.

Introduction

Reporting a memory about a witnessed event in the context of an investigative interview is a very complex task. While, people retrieve event memories from the past on an everyday basis, the demands of an investigative interview are unique and often exceed the demands of everyday memory search. Normally, whether we remember all the details of an event or not is not vital, and often remembering the gist of a memory is sufficient to satisfy the purpose of the memory search. In comparison, the task for an eyewitness is to report everything they have seen, as details can make a difference for the outcome of the investigation. Remembering all the details of an event is not easy, neither a task we are accustomed to, therefore the support that the interviewers can offer during an interview is pivotal to ensure that the eyewitness's account is as accurate and detailed as it can be.

Furthermore, during the course of an investigation eyewitnesses are often asked to report their memory for the event seen more than once (Odinot et al., 2013), and it is important that the information gathered on each occasion is as accurate and detailed as possible. For example, if the eyewitness is the person reporting the crime, they are likely to be asked to recall what seen in the emergency call. Following this, they might be asked to report the crime to a frontline police officer on the crime scene (Gabbert et al., 2015). Finally, further into the investigative process investigators might ask an eyewitness to recall what seen once again in a more in-depth interview. Additional reasons for multiple interviews being required include if new evidence has been found - especially if this contrasts with the eyewitness's initial account (Gabbert et al., 2015), or if witnesses have not reported sufficient information in the first interview (e.g., La Rooy et al., 2009).

Rationale

As highlighted in Chapter 1, research in psychology has proposed several techniques and interviewing protocols aimed to enhance eyewitnesses' memory performance, while ensuring their accounts remain accurate. In general, such techniques have been found to be beneficial also across repeated interviews. In particular, research has shown that memory activation is likely to strengthen the

memory trace and increase the likelihood of accessing the trace in future (e.g., Roediger & Payne, 1982). Thus, recalling a memory in an initial interview increases the likelihood of remembering the memory in a subsequence interview; however, this is contingent to the quality of the initial interview (Hope et al., 2014; Marsh et al., 2005). A thorough memory activation following a good practice interview compared to a shallow memory activation, is more likely to improve memory performance in a subsequent interview.

Clearly, whether an interview can enhance eyewitnesses' performance is important; however, it is also important to understand whether an interview can impact other aspects of the eyewitness's cognition. As such, the research discussed in Chapter 2 showed also that eyewitness confidence can be affected during an interview. For example, the interviewer's feedback (e.g., Leippe et al., 2006; Lida et al., 2020), the type of interview (e.g., Gwyer & Clifford, 1997), or the retrieval prompts can influence confidence (e.g., Granhag et al., 2004; Knutsson et al., 2011). Investigating whether and why confidence changes is pivotal for the legal context, because confidence in turn underpins whether and how the information is reported (Koriat & Goldsmith, 1996). Furthermore, evidence suggests that confidence might also affect eyewitness's willingness to testify (Hafstad et al., 2004; Wells & Bradfield, 1998). Importantly, the literature on metamemory has shown that confidence depends on the retrieval process and its products, such as how complete the recalled memory seems (Brewer et al., 2005).

Thus, the primary aim of Study 1 is to investigate how the quality of the initial interview affects memory performance and in turn confidence in the quality of one's own memory, as well as confidence in engaging with the Criminal Justice System. Participants recalled their memory for the event seen on two occasions. Initially (Time 1) they received either a Best Practice Interview – designed by following the PEACE recommendations, or a Poor Practice Interview – which deviates from the best practice guidelines. At Time 2 all the participants received the same Free Recall test. Confidence in the accuracy of own memory and confidence in engaging with the CJS was measured before and after the initial Interview. Hypotheses and predictions are as follows:

Hypotheses

H_1: A Best Practice Interview is more likely to promote memory activation and to strengthen memory trace, which in turn lead to more complete, more accurate and more consistent accounts in both recall attempts. On the contrary a Poor Practice Interview is more likely to disrupt memory activation, which is turn is likely to lead to fewer and less accurate information being reported, and less consistent accounts.
Predictions: Participants in the Best Practice Interview group (compared to those in the Poor Practice Interview group) are expected to report more and more accurate accounts at Time 1 and at Time 2, and more consistent information at Time 2.

H_2: A Poor Practice Interview (but not a Best Practice Interview) is likely to disrupt memory performance and to yield poor accounts, which in turn is likely to promote (i) lowered confidence in own memory, and (ii) lowered confidence in engaging with the Criminal Justice System.
Predictions: after the initial interview participants in the Poor Practice Interview group are expected to show decreased confidence in the quality of their own memory and decreased confidence in engaging with the CJS. While the opposite pattern is expected for participants in the Best Practice Interview group.

Method

Design

A 2 (Initial Interview Type: Best Practice Interview, Poor Practice Interview) x 2 (Confidence measurement time: Time 1, Time 2) mixed design was used, with Initial Interview Type manipulated between subjects, and confidence manipulated within subjects. The DVs for memory recall at Time 1 and Time 2 were: (a) amount of correct details, (b) amount of incorrect details, (c) overall accuracy, and (d) consistent information. The DVs for confidence were: (e) confidence before the Initial Interview (Time 1), and (f) confidence after the Initial Interview (Time 2).

Participants

A total of 40 participants took part in the study (31 females; Mean age = 21.89, SD = 7.09). All participants were recruited in the Goldsmiths University, 50% were recruited from the first-year cohort of Psychology students, the remaining were students or staff members of different departments. All participants received either

course credits or £10 Amazon Vouchers as compensation for their participation in the experiment.

Materials

Video stimulus

A mock crime video (1 minute, 40 seconds) was used as stimulus material. The video was presented on a 17 inch HD screen. The event was non-violent robbery filmed in a Blockbuster video-rental store and a total of four males were involved in the scene. In the video a customer enters the shop and asks the shopkeeper where the new release can be found. He obtains direction and goes in a different section of the store. Shortly after, two males enter the shop and proceed straight to the section were the customer is browsing for videos. The two males confer briefly, then cover their faces with disguises and proceed to the front area to confront the shopkeeper. One of them walks to the door and guards the entrance, while the other asks for the money. After a few seconds, the customer who first entered the shop comes back to the front area and one of the robbers pushes him to the floor. Finally, the robbers take their disguises off and run out of the store.

Time 1: Initial Interview Type

Best Practice Interview. The Best Practice Interview followed the PEACE guidelines (College of Policing, 2020). However, we excluded the Planning and Preparation and Evaluation phase because they are not relevant for the purpose of this experimental study.

Free report phase*. Introduction and preferred name*. The researcher thanked the participant for taking part in the study, introduced herself, and asked for the interviewee's preferred name.

Instruction. Once established the preferred name the researcher acknowledged that she knew nothing about the event seen and to be able to understand what happened, she needed the interviewee to tell all they could possibly remember.

Retrieval aid. Three memory prompts were provided during the introduction, these were focused on the three categories of details: actions, location, and person description.

Initiating the Free Recall. In order to encourage completeness, participants were asked to report as many details as they could remember, including those that might appear irrelevant. The instruction was as follows: *"Thank you for coming and taking part in this study, my name is Alessandra, how would you like to be called? We are now going to talk about the video you have just seen. Please think that I know nothing about the event you have seen and to be able to reconstruct the event I need you to tell me all you know about it. I would like you to take the time you need and think about the event, focus your attention on the location in which the event happened. Think about the people involved, what they looked like, and what they did or said. When you are ready tell me everything you remember please do so in as many details as possible. All the information you have are equally important for me so please try not to skip any detail."*

Questioning phase. The questioning phase initiated when the participant had completed the Free Recall phase.

Follow up questions. The questioning phase included three TED follow up questions (tell, explain, describe); one was related to the main action (e.g., the robbery), one to the location in which the event happened (e.g., the shop), and one to the robbers' description (e.g., robber's clothing). These were built upon the participant's narrative. Each question was phrased by using the interviewee's own words and language style, (e.g., '*you told me that the robbers went back to the till where the guy was and shouted at him, tell me more about this scene'*). Hence, all the questions were focused on topics mentioned by the interviewee (interviewee-led probing of topic).

Closure. The participant was asked if there was any other detail they could remember, including details they were not asked about. Finally, the researcher thanked the participant about the information provided and explained that every detail reported was very important for the investigation.

Behavioural components. Two behavioural components were used to further support the recollection task and enhance cooperation; (1) open body language, and (2) active listening. These were kept constant throughout the free recall and questioning phase.

Open body language. The researcher's non-verbal behaviour appeared relaxed and the gestures confident but also polite and supportive. In order to establish and maintain a positive relationship, subtle similarities and mirroring techniques

were also used (e.g., if the participant was directed towards the researcher, the latter responded with a similar posture).

Active Listening and Facilitations. Active listening was expressed by maintaining eye contact without being invasive and keeping the posture directed towards the participant. Nods and other facilitation such as: "yes, go on", "okay, I understand" were also presented in the interview and used sensibly when needed (e.g., if the participant was struggling to recall information the researcher waited showing interest, and if after a few seconds the interviewee remembered that information, the researcher nodded a few times).

Poor Practice Interview. The Poor Practice Interview is composed of the most common deviations from the PEACE models observed in police interviews. In order to identify these components we randomly selected 10 transcripts of interviews conducted by police officers and found five common mistakes: lack of instructions for the interviewing process, lack of rapport building, use of closed questions, interviewer-led style of questioning, and absence of closure (e.g., Fisher et al., 1987).

Basic instructions about the interview process were found in two interviews out of ten. In both interviews the police officers clarified they intended to talk with the witnesses about the incident in order to gather information that could potentially be important for the investigation. In one interview the police officer also mentioned further possible stages the witness might be asked to take part in, such as testifying in Court.

In none of the ten interviews did the police officer attempt to build rapport through establishing the interviewee's preferred name, however in two interviews the police officers introduced themselves and specified their rank. Some attempt to connect or demonstrate interest in the interviewee's wellbeing was found in three interviews. Here, the police officers checked the interviewees were comfortable and did not feel intimidated by anybody.

5Wh and Yes/No type of questions were coded as closed questions, and were found in all the interviews, accounting for more than 85% of the questions asked in each interview, while the remaining 15% of the questions were phrased in the form of TED questions.

Instances of an interviewer-led style of questioning were found in all of the interviews. This style consisted of questions that shifted the interviewee's attention

constantly from topic to topic. For example, in one transcription the police officer asked if the robbers interacted with the interviewee during the robbery, followed by a question on where the interviewee was before entering the shop, followed by a question on the robber's description. Although these questions are useful for the interviewer and can elicit important information, topic-hopping such as this might not be compatible with the interviewee's recollection process. In line with Anderson's (1983) theoretical model of memory, by shifting interviewee's attention from topic to topic, the interviewee is less likely to engage in a thorough memory activation and might have less opportunities to remember each topic in detail.

Closure was found in two interviews out of ten. The police officers summarised the event and asked if there was any other information the interviewee wanted to add. In one interview the police officer further explained that the interviewee could be asked to testify in Court and invited the interviewee to ask any questions about this.

From the transcriptions we were unable to access information about non-verbal behaviour such as active listening expressed by nodding, or body posture and eye contact. Facilitation such as "okay" or "yes" were coded in all the interviews, however we did not find any sophisticated forms of encouragement for the recollection task. Because the facilitations observed are broadly used in everyday conversations and are not skillful techniques, we decide not to include them in the Poor Practice Interview.

A final component added was interruptions. We were unable to assess the amount of interruptions in the transcriptions; however, evidence exists that less skilled police officers do interrupt eyewitnesses' account (e.g., McLean, 1995; Fisher et al., 1987; Wright & Alison, 2004), therefore we include this in the Poor Practice Interview.

In sum, the Poor Practice Interview used in Study 1 comprised a lack of instructions for the interviewing process, a lack of social support, heavy use of closed questions, an interviewer-led style of questioning, and the absence of closure. Operationalisation of this was as follows;

Free recall phase. *Introduction.* In the introduction the interviewer thanked the participants and introduced herself but did not asked for the participant's preferred name.

Instructions. A general instruction about the recollection task was given, in particular the research informed the participants they were going to talk about the video they had just seen.

Retrieval aids. The retrieval prompts were identical to that used in the Best Practice Interview. Participants were asked to focus on the location of the event, the people involved and their actions.

Initiating the Free Recall. Finally, participants were asked to report what happened. The instruction for the free recall phase in the Poor Practice Interview condition was as follows: *"Thank you for coming and taking part in this study, my name is Alessandra. We are now going to talk about the event you have seen. Focus your attention on the location in which the event happened, the people involved, what they looked like, what they did or say and then tell me what happened".*

Interruptions. Between two and three interruptions were used in the Poor Practice Interview. An interruption was defined as an intrusion of the interviewer into the interviewee's narrative and it could occur when the participants had just concluded a sentence, but they could still potentially talk or when they were thinking in silence. Note that participants were never interrupted when they were actually speaking. The interruptions were used to introduce the follow up questions.

Questioning phase. *Follow up questions*. Three 5Wh follow up questions were asked in the Poor Practice Interview. As for the Best Practice Interview one question was focused on the main action, one on the location, and one on the robbers' description. The questions did not match the participants' language style as they were not phrased by using the participants' own words. Each question introduced a new topic area, and this was different from the topic the participants were talking or thinking about when the question was asked. For example, if the participant was talking about the robbers coming towards the till and shouting at the shopkeeper, the research interrupted and asked: "ok, how were these robbers dressed?" Finally, no Closure for the interview was presented in the Poor Practice Interview.

Behavioural components. In the Poor Practice Interview non-verbal support was not provided. The researcher sat on the chair and did not change her position throughout the interview. Also, no nodding or facilitations were not used in this interview.

Time 2: Free Recall Test

The recall test at Time 2 was a written Free Recall test, and it was the same for all participants. The instruction was as follows: '*Please focus the attention on the setting in which the event happened, the people involved, what they did or looked like. When you are ready please write down what happened'*. No time limit was given for this task.

Confidence questionnaire

The confidence questionnaire measured how confident participants were in their memory for the event seen. Furthermore, a section of the questionnaire investigated participants' confidence in engaging with the Criminal Justice System (Appendix A). Here, we identified three of the most significant stages a witness is likely to be involved in (Home Office, 2013): (i) report the crime, (ii) provide a statement, and (iii) provide evidence in Court. To these we added another possible stage that an eyewitness might be asked to participate in, (iv) an identification task. In the questionnaire participants were asked to imagine they were real witnesses of the robbery, and to rate how confident they were in their memory to be able to talk to the police, sign a statement, participate in an identification task and present their evidence in Court. All the confidence scales ranged from 10% to 100%.

Procedure

The experiment was conducted in a laboratory at Goldsmiths University, the room was equipped with a desk, a computer with headphones and two chairs facing each other. During the testing session no one apart from the participant and the researcher was present in the room. The experiment took about 30 minutes and was conducted in one session. Participants were informed through the consent form (Appendix B) that the study investigated the impact of different questioning styles on memory confidence, no other information about the hypotheses was given. The study procedure was briefly illustrated, in particular participants were told they were going to watch a crime-video, take part in a face-to-face interview and answer questions about how confident they felt in their memory. Before the experiment started participants were randomly assigned to one of the two experimental conditions (i.e., Best Practice Interview, or Poor Practice Interview). After that they were presented with the crime-video and reminded to pay attention to it. Then, participants were

interviewed accordingly to the condition they were assigned to. Participants were asked permission to audio record the interview and all participants agreed, therefore all the interviews were recorded. The confidence questionnaire was administered after the video but before the initial interview (Time 1), and again after the initial interview (Time 2). Once participants had completed the confidence questionnaire at Time 2, they were given the written Free Recall test. Finally, participants were debriefed (Appendix C) and compensated.

Coding

Accuracy coding

Information reported in the interview (Time 1) was transcribed, and the recall data provided at Time 1 and Time 2 (Free Recall test) was coded by using the same coding scheme (Appendix D). Each detail was coded as Action (a) whenever it referred to an action, Location (l) if it referred to the setting, and Person Descriptor (pd) if the detail referred to the people involved, their clothing and personal objects. As an example, the string *"the robber (pd) was blond (pd) and was standing (a) at the door (l)"* was coded as 1 x Action, 1 x Location and 2 x Person Descriptors[1]. Plurals were coded as two details. The details were coded as Correct if present in the video, or as Incorrect if not present in the video. Non-specific information was not coded; for example, in the string *"he said something",* the detail 'something' was not coded. Also, personal opinions such as *"I think they were accomplices",* and ambiguous information, such as *"somewhere in the shop"* were not coded. To ensure consistency across coding, ten (12.5%) randomly selected transcripts were coded by two research assistants. A percentage agreement between coders was calculated by dividing the number of details agreed by the number of total details (agreed and disagreed) and was found to be high (100%). Inter-rater reliability - assessed with Cohen's Kappa was 0.9, reflecting almost perfect agreement (McHugh, 2012).

[1] We did not have a hypothesis regarding the type of details reported, therefore we collapsed action, location and person descriptors details correct into number of correct; and action, location and person descriptors details incorrect into number of incorrect.

Consistency coding

Each detail reported at Time 2 was coded as Consistent if the same detail was recalled also at Time 1, while details reported at Time 2 but not at Time 1 were coded as new information.

Results

All statistical tests were performed at alpha level of .05 unless otherwise specified. Homogeneity of variance was assessed with Levene's test (1960).

Memory recall at Time 1 and Time 2

A series of 2 (Initial Interview Type: Best Practice Interview, Poor Practice Interview) x 2 (Confidence Measurement Time: Time 1, Time 2) mixed design ANOVAs were conducted on the DVs: correct, incorrect information, and accuracy rate.

For the correct information reported we found a significant main effect of Initial Interview Type $F(1, 38) = 5.04$, $p = .03$, $\eta^2_p = .12$; meaning that overall participants in the Best Practice Interview group reported more correct information than those in the Poor Practice Interview group ($M_{diff.} = 10.65$, $p = .03$, 95% CI [1.04, 20.21]). The main effect of Time was non-significant $F(1, 38) = 1.69$, $p = .20$, $\eta^2_p = .04$. However, we found a significant interaction $F(1, 38) = 14.97$, $p < .001$, $\eta^2_p = .28$. A post-hoc power analysis conducted in G*power (Erdfelder et al., 1996), with $\alpha = .05$, $n = 40$, and the average of the observed effect sizes, yielded a power $(1 - \beta)$ of .67 (ranging from .23 to .96). This analysis suggests the study might be underpowered and unable to detect small to medium size effects. Follow up comparisons as a function of Initial Interview Type showed that participants in the Best Practice Interview group reported more correct information than those in the Poor Practice Interview group, but this difference was only significant at Time 1 ($M_{diff.} = 16.05$, $p = .002$, 95% CI [6.53, 25.56]). Furthermore, comparisons as a function of Time showed that participants in the Best Practice Interview group reported significantly more correct information at Time 1 compared to Time 2 ($M_{diff.} = 7.25$, $p = .001$, 95% CI [3.23, 11.26]). While participants in the Poor Practice group reported more correct information at Time 2 compared to Time 1, but this difference did not reach significance (see Table 4.1 for Means and Standard Deviations at Time 1 and Time 2).

For the number of incorrect details we found a significant main effect of Initial Interview Type $F(1, 38) = 8.17$, $p = .007$, $\eta^2_p = .17$, meaning that across recall attempts participants in the Poor Practice Interview group reported more errors ($M_{diff.} = 2.02$, $p = .007$, 95% CI [.59, 3.45]). The main effect of Time was found non-significant $F(1, 38) = 3.41$, $p = .07$, $\eta^2_p = .08$. However, we found a significant interaction $F(1, 38) = 6.17$, $p = .017$, $\eta^2_p = .14$. Follow-up comparisons showed that at Time 2 (but not at Time 1) participants in the Poor Practice Interview reported more errors than those in the Best Practice Interview ($M_{diff.} = 3.00$, $p = .002$, 95% CI [1.15, 4.84]). Furthermore, participants in the Poor Practice Interview reported more errors at Time 2 compared to Time 1 ($M_{diff.} = 1.70$, $p = .004$, 95% CI [.57, 2.82]).

For the accuracy rate of the information reported we found a significant main effect of Initial Interview Type $F(1, 38) = 9.35$, $p = .004$, $\eta^2_p = .19$; a non-significant main effect of Time $F(1, 38) = 3.09$, $p = .08$, $\eta^2_p = .07$; and a non-significant interaction $F(1, 38) = 2.32$, $p = .13$, $\eta^2_p = .05$. Follow up comparisons showed that across the two time-points participants in the Best Practice Interview group were more accurate than those in the Poor Practice Interview group ($M_{diff.} = 4.13$, $p = .004$, 95% CI [1.39, 6.87]).

To summarise, when collapsing the results across recall attempts the quality of the initial interview was found to have a significant impact on the three DVs, as shown by the significant main effect of Initial Interview Type on the correct and incorrect information reported, and on the accuracy of the information reported. This impact was positive; overall participants in the Best Practice Interview reported more correct details, fewer errors, and more accurate information than those in the Poor Practice Interview. The non-significant main effect of Time on the DVs, showed that time in isolation did not affect the quality of information reported. However, the significant interaction for correct and incorrect showed that only at Time 1 the quality of the interview had a significant impact on the correct information reported. At this time point only participants in the Best Practice Interview group reported more correct details than those in the Poor Practice Interview group. While at Time 2 this was not the case as participants in the Best Practice Interview group decreased the amount of correct details reported, thus the difference in the amount of correct information reported between the two groups disappeared. On the contrary, the effect of the quality of the initial interview on the amount of incorrect details reported was only significant at Time 2. Only in this time point participants in the Poor Practice

Interview reported more errors than participants in the Best Practice Interview. At Time 2 participants in the Poor Practice interview showed an increase in the number of incorrect details reported compared to Time 1.

Table 4.1. Means (and Standard Deviations) of correct and incorrect details, and accuracy of information reported at Time 1 and Time 2 for the Best Practice Interview and Poor Practice Interview group.

	Time 1		Time 2	
	Best Practice	Poor Practice	Best Practice	Poor Practice
Correct	62.80 (15.66)	47.75 (14.01)	55.55 (16.23)	50.35 (16.32)
Incorrect	2.60 (1.35)	3.65 (2.77)	2.35 (2.11)	5.35 (3.49)
% Accuracy rate	95.68 (2.68)	92.56 (5.80)	95.52 (4.16)	90.36 (5.72)

Consistency

Each unit of information reported at Time 2 was coded as consistent if it was reported also at Time 1. We calculated the proportion of consistent information as a function of the amount of details (correct and incorrect) reported at Time 2. Overall at Time 2, approximately 88% of the information reported by the participants in the Best Practice Interview group was consistent with the information that had been reported at Time 1, compared to 75% for participants in the Poor Practice group; this difference was significant, $t(38) = 4.21$, $p < .001$, $d = 1.33$, ($M_{diff.} = 12.27$, 95% CI [6.38, 18.17]; see Table 4.2 for Means and Standard Deviations). We also explored the accuracy of the consistent details, and found a significant difference between groups $t(38) = 2.79$, $p = .01$, $d = .88$, ($M_{diff.} = 3.37$, 95% CI [.92, 5.81]), with participants in the Best Practice Interview group (M = 97.63; SD = 2.21) reporting more accurate consistent information than those in the Poor Practice Interview group (M = 94.26; SD = 4.92).

Conversely, in the Best Practice Interview group approximately 12% of the total information reported at Time 2 were new, against the 24% for the Poor Practice Interview group, and this difference was significant $t(38) = -4.17$, $p < .001$, $d = -1.32$, ($M_{diff.} = -12.07$, 95% CI [-17.92, -6.22]; see Table 3.2 for Means and Standard Deviations). No differences between groups was found on the accuracy of the new

information reported *t(38)* = -.41, *p* =.68, *d* = -.13, ($M_{diff.}$ = -2.95, 95% CI [-17.66, 11.75]).

In summary, as predicted participants in the Best Practice Interview were more consistent across interviews that those in the Poor Practice Interview group.

Table 4.2. Means (and Standard Deviations) of proportion and accuracy rate of consistent and new information in the Best Practice Interview and Poor Practice Interview group.

		Best Practice	Poor Practice
Consistent	Proportion	87.72 (7.46)	75.44 (10.66)
	Accuracy	97.63 (2.21)	94.26 (4.92)
New information	Proportion	12.27 (7.46)	24.34 (10.54)
	Accuracy	75.10 (25.19)	78.10 (19.96)

Changes in memory confidence

In order to investigate the effect of the quality of the initial interview on memory confidence, we collected subjective confidence ratings before and after the interviews. Our hypothesis was that confidence after the initial interview (Time 2) would decrease for participants in the Poor Practice Interview group and increase for participants in the Best Practice Interview group.

First, we looked at confidence reported before the interview (Time 1), and found no significant difference between groups *t(38)* = .65 *p* = .51, *d* = .20, ($M_{diff.}$ = 2.50, 95% CI [-5.27, 10.27]; see Table 4.3 for Means and Standard Deviations). Following this, a 2 (Initial Interview Type: Best Practice Interview, Poor Practice Interview) X 2 (Confidence Measurement Time: Time 1, Time 2) mixed design ANOVA revealed a non-significant main effect of Initial Interview Type $F(1, 38) = .37$, $p = .54$, $\eta^2_p = .01$; a significant main effect of Time $F(1, 38) = 5.82$, $p = .021$, $\eta^2_p = .13$, indicating that confidence decreased over time independently on the Initial Interview Type group ($M_{diff.}$ = -4.48, $p = .021$, 95% CI [-8.16, -.71]); and a non-significant interaction $F(1, 38) = 0.8$, $p = .97$, $\eta^2_p = .01$.

Thus, the results only partially support Hypothesis 2, in particular as predicted confidence in memory decreased after the initial interview, however contrary to our expectation this was independent on the type of initial interview.

Table 4.3. Means (and Standard Deviations) of confidence ratings reported at Time 1, and Time 2 for the Best Practice Interview and Poor Practice Interview group.

		Best Practice	Poor Practice
Confidence	Time 1	67.38 (13.46)	64.88 (10.65)
	Time 2	62.88 (13.57)	60.50 (17.08)

Changes in confidence in engaging with the Criminal Justice System

In order to investigate the effect of the quality of the initial interview on confidence in engaging with the Criminal Justice System, we collected four confidence measures. Participants were asked to rate how confident they were in their memory in relation to (i) talking with the police, (ii) signing a statement about what they had witnessed, (iii) providing identification evidence, and (iv) presenting evidence in Court. We first examined confidence ratings reported at Time 1 and found no difference between groups in the confidence measures reported: talking to the police $t(38)$ = -.59 p = .53, d = -.19, ($M_{diff.}$ = -4.00, 95% CI [-17.53, 9.53]); signing a statement $t(38)$ = -1.23, p = .22, d = -.39, ($M_{diff.}$ = -7.50, 95% CI [-19.77, 4.77]); participating in an ID task $t(38)$ = .32 p = .74, d = .10, ($M_{diff.}$ = 2.50, 95% CI [-12.94, 17.94]), and presenting evidence in Court $t(38)$ = -.44 p = .65, d = -.14, ($M_{diff.}$ = -3.50, 95% CI [-19.33, 12.37]; see table 4.4. for Means and Standard Deviations).

A series of 2 (Interview Type: Best Practice Interview, Poor Practice Interview) X 2 (Confidence Measurement Time: Time 1, Time 2) mixed design ANOVAs conducted on the four confidence measures revealed only a significant main effect of Time on confidence in presenting evidence in Court $F(1, 38) = 5.17$, p = .03, η^2_p =.12. No other significant main effect or interaction was found (see Table 4.5).

In conclusion, we found no support for Hypothesis 2. Confidence in relation to presenting evidence in Court decreased similarly for participants in Best Practice Interview and Poor Practice Interview group.

Table 4.4. Means (and Standard Deviations) of confidence in memory related to talking to the police, signing a statement, participating in an ID task, and presenting evidence in Court reported at Time 1 and Time 2 for the Best Practice Interview and Poor Practice interview group.

		Best practice	Poor practice
Talking to the police	Time 1	65.50 (21.63)	69.50 (20.97)
	Time 2	62.50 (22.44)	66.00 (25.01)
Signing a statement	Time 1	60.00 (18.04)	67.50 (20.22)
	Time 2	54.50 (20.82)	64.00 (24.79)
Participating in ID	Time 1	56.50 (21.58)	54.00 (25.83)
task	Time 2	53.00 (21.05)	50.50 (26.84)
Presenting evidence in	Time 1	56.50 (24.49)	60.00 (24.49)
Court	Time 2	44.50 (23.05)	58.50 (28.52)

Table 4.5. Results of the 2 x 2 ANOVA

Result	*variables*	df	test value	*p* value	Effect size (n_p^2)
		Talking to the police			
Main effect	Interview	(1, 38)	$F = 1.81$	= .18	$\eta^2_p = .04$
Main effect	Time	(1, 38)	$F = .05$	= .81	$\eta^2_p = .01$
Interaction	Interview x Time	(1, 38)	$F = 2.26$	= .14	$\eta^2_p = .05$
		Signing a statement			
Main effect	Interview	(1, 38)	$F = 1.78$	= .19	$\eta^2_p = .05$
Main effect	Time	(1, 38)	$F = 3.22$	= .08	$\eta^2_p = .07$
Interaction	Interview x Time	(1, 38)	$F = .15$	= .69	$\eta^2_p = .01$
		Participating in an ID task			
Main effect	Interview	(1, 38)	$F = .12$	= .72	$\eta^2_p = .01$
Main effect	Time	(1, 38)	$F = 1.44$	= .23	$\eta^2_p = .04$
Interaction	Interview x Time	(1, 38)	$F = .00$	= 1.00	$\eta^2_p = .00$
		Presenting evidence in Court			
Main effect	Interview	(1, 38)	$F = 3.12$	= .08	$\eta^2_p = .07$
Main effect	Time	(1, 38)	$F = 5.17$	= .03*	$\eta^2_p = .12$
Interaction	Interview x Time	(1, 38)	$F = 3.13$	= .08	$\eta^2_p = .07$

* indicates significance at $p < .05$

Discussion

Study 1 investigated the effect of the quality of the initial interview on (i) memory performance, (ii) confidence in the quality of own memory, and (iii) confidence in engaging with the Criminal Justice System. We found that participants in the Best Practice Interview reported more correct information, less errors and more accurate and consistent accounts over the two recall attempts. However, contrary to our expectation participants in the Best Practice Interview reported more correct information at Time 1 *only*, while participants in the Poor Practice Interview reported more errors at Time 2 *only*. Despite participants in the Poor Practice Interview reported less complete accounts than those in the Best Practice Interview,

confidence in their own memory and confidence in engaging with the CJS decreased similarly for participants in the two groups.

Overall, the quality of initial interview strongly affected the quality of the information reported. As such, the Best Practice Interview elicited more correct and more accurate information than the Poor Practice Interview. Thus, the Best Practice Interview compared with the Poor Practice Interview was successful in promoting a more in-depth memory activation and eliciting more correct details. The former interview was designed to support participants in their recollection task, it provided instructions to clarify the interviewers' expectation relating to the desired level of details. In addition, it ensured the opportunity to engage in an uninterrupted free recall that encouraged participants to think and activate their memory, and facilitated the access to large amount of target details. Furthermore, the use of open questions and compatible questioning style further facilitated the access to details by respectively (i) encouraging participants to concentrate their attention to more specific aspect of the event, and (ii) proving the cues that are more likely to reach the target details stored in memory. On the contrary, in the Poor Practice Interview, the lack of instruction was less likely to prepare participants for the interview, while the interruptions and topic-hopping were more likely to force eyewitness to shift their attention to different aspects of the memory, and overall to disrupt their recollection task. Furthermore, the use of closed questions and incompatible questioning style was less like to encourage memory activation in the probing phase of the interview.

The results of Study 1 showed that at Time 1 participants in the Best Practice Interview group benefitted from the format of their interview more than those in the Poor Practice Interview group, hence they were able to retrieve and report more correct information, however this effect was not carried over to the second recall attempt. At Time 2 the difference between groups in the amount of correct information reported disappeared, primarily because at Time 2 participants in the Best Practice Interview group reported significantly fewer correct details. The finding that participants can down-report in a subsequent recall test was unexpected but it is not new, and research investigating the carryover enhancing effect of a good quality initial interview have found similar results (Kraus et al., 2017, Krix et al., 2014; see also Memon et al., 1997). For example, in Krix et al.'s (2014) study participants recalled a memory with either a SAI or a Free Recall at Time 1, while at Time 2 participants in both groups recalled their memory with the same Free Recall

test. At Time 2 participants who initial received a SAI (but not those that received a Free Recall) reported significantly less correct information compared with Time 1. The authors explained that this result might be due to a report issue; in particular, the participants that received the initial SAI reported less correct details at Time 2, because the instruction of the Free Recall did not clarify the expectation that they had to report highly complete accounts. Similarly, in Study 1 participants in the Best Practice Interview group were specified the desired level of completeness during the initial interview, but not in the second Free Recall test. Consequently, at Time 2 they might have decided to withhold more details compared to Time 1. Interestingly, this harsher control over memory reporting did not lead to increased accuracy or to fewer incorrect information being reported, meaning that in the second recall attempt participants in the Best Practice Interview were withholding potentially valuable correct details.

The other unexpected result is that the Poor Practice Interview yielded more errors compared with the Best Practice Interview at Time 2 but not at Time 1. The lack of statistical difference between the groups in the amount of errors reported at Time 1 is surprising considering that the Poor Practice Interview included interruptions, closed questions, and an overall questioning style incompatible with the interviewee's memory reporting. It appears that at Time 1 our operationalisation of the quality of the initial interview had a larger effect on the correct rather than the incorrect information reported. In other words, at Time 1 the disruptive effect of the Poor Practice Interview had a larger effect in reducing the amount of correct information reported rather than in increasing the amount of errors. Certainly, the interruptions, the closed questions and the topic-hopping were likely reducing the access to the target details stored in memory. However, it remains the question as to why they did not lead to more errors. A possibility is that the short retrieval interval might have buffered the negative effect of the poor retrieval techniques on the amount of errors reported. For example, it is possible that at Time 1 participants' memory representation was still strong and so was the trace of the target details (i.e., details pertaining to the target event). Thus, despite the poor interviewing techniques the target details remained more likely to be reported than any competing and thus incorrect detail (i.e., a detail not pertaining to the target event). However, with longer delays and weaker memory representation, the combination of weaker traces and poor interviewing techniques could lead to more competing details being mistakenly

recalled (and reported) while searching for target details. Therefore, despite in Study 1 we found no evidence of a negative impact of the quality of the initial interview on the number of errors reported at Time 1, it is worth clarifying that this might be largely a consequence of the short retrieval interval we used. Furthermore, the results that overall participants in the Poor Practice Interview group reported less correct details and were overall less accurate than those in the Best Practice Interview group show that the quality of the initial interview remains key in eliciting good quality information from eyewitnesses.

Beside the short retrieval interval, a further limitation of Study 1 prevents us from making further conclusions relating to the impact of the quality of the initial interview on memory performance. In particular the study does not include a control group, thus when evaluating the effect of the quality of the initial interview on memory performance we are unable to determine whether the observed differences on the amount of correct and incorrect details and accuracy rate are due to the enhancing effect of the Best Practice Interview, or the detrimental effects of the Poor Practice Interview, or to a combinations of both.

In addition to investigating the impact of the type of the initial interview on memory performance, a further aim of Study 1 was to investigate how the quality of the initial interview affects confidence (i.e., confidence in the quality of the overall memory and confidence in engaging with the CJS). We expected participants in the Poor Practice Interview group to provide less complete accounts and therefore to decrease their confidence after the initial interview. On the contrary we expected participants in the Best Practice Interview group to report more complete and detailed accounts and therefore to increase their confidence after the initial interview. Results showed that the quality of the initial interview influenced the completeness of the accounts, but not confidence in the expected direction. Thus, despite the accounts reported by participants in the Poor Practice Interview were less complete than those reported by the participants in the Best Practice Interview, confidence decreased similarly in both groups. It is surprising that while the two initial interviews had a very different effect on the quality of the memory reported, their impact on confidence appeared to be largely similar.

It is worth mentioning that research investigating confidence in the information elicited by different quality interviews reported similar results. For example, Allwood et al. (2005) interviewed participants with either a CI or a SI

interview and found no difference between confidence ratings reported by participants in the two groups (see also Granhag et al., 2004). However, in Allwood et al.'s study neither of the interviews included elements of poor interviewing practice. In particular, the CI and SI were identical apart from the mnemonics "mental reinstatement of context" and "report everything instruction", that were included in the CI only. Both interviews presented a rapport phase, clear instructions, and probing invitations for additional information. Furthermore, in neither of the two interviews were participants asked specific questions. In sharp contrast, in Study 1 the two initial interviews were largely different. Specifically, the Poor Practice Interview included elements that previous research has found to be disruptive and detrimental for eyewitness reporting (Fisher et al., 1987; George; 1991), such as interruptions, lack of social support, and the use of interviewer-led questions and non-compatible questions. Based on these differences we predicted that the Poor Practice Interview would have a larger detrimental impact on memory performance and memory confidence compared to previous research.

Thus, it is unclear why the Best Practice Interview and Poor Practice Interview caused a similar decrease in memory confidence. Certainly, despite these two interviews include different interviewing techniques, they share some similarities that might explain the statistically comparable impact they had on confidence. For example, both the interviews include a questioning phase in which the participants are asked to report additional details they had not reported in the free recall phase. Research has showed that confidence in the information reported in response to forced choice questions (Allwood et al., 2008) or follow-up open invitations (Knutsson et al., 2011; Sharps et al., 2012) is lower than confidence on information reported in response to a free recall prompt. It is possible that in Study 1 participants' confidence decreased in the probing phase of the Best Practice Interview and Poor Practice Interview alike.

However, the question remains as to why confidence decreased in the probing phase of the interviews. Clearly, participants did not base their confidence judgements only on the completeness of their accounts, and it is likely that when asked about their confidence they evaluated other products of their own retrieval process. It is possible that the supposed difference between confidence for information reported in the free report and probing phase are due to the different retrieval products generated during the free recall phase compared to the probing

phase. In line with this, Allwood et al. (2008) suggests that in response to a free-report invitation, participants are more likely to exert higher control over their memory, thus they tend to report only information they are confident about. However, when asked specific questions they have less control and are more likely to report information they are less confident about. Thus, in Study 1 participants' confidence might have decreased in the probing phase, because during this stage of the interview in answer to probing questions the retrieval experience is more likely to yield lower confidence. However, Study 1 does not allow us to clearly conclude that confidence for participants in both groups decreased in the probing phase, not that decreased confidence was due to the products of the retrieval process generated during the probing phase. In particular, we are unsure whether the decrease in confidence depended on the type of retrieval prompt asked in the free and probing phase of the interviews. In Study 1, we aimed to design realistic interviews, that resemble real-life interviews. They are composed of serval different components and techniques; thus, it is difficult to disentangle the effects each element might have had on memory confidence. In other words, the operationalisation of quality of the initial interview we used does not allow us to isolate the potential elements of the interviews that might have caused the observed decrease in confidence. As a consequence, Study 1 does not allow us to fully understand the aspects of the two initial interviews that lead to the observed shifts in confidence.

A further limitation is that the complexity of the operationalisation we adopted raises issues related to consistency and replicability. The interviews conducted were not identical, as the interviewer's utterances dependent largely on the information reported by each participant. Similarly, it would be relatively difficult for other researchers to reproduce identical interviews and replicate the results of Study 1. In order to overcome these limitations, in Study 2 we adopted a more controlled operationalisation of the quality of initial interview and further investigate the impact of an initial interview on memory confidence.

In conclusion, Study 1 shows that the quality of the initial interview is likely to affect memory reporting. In line with existing literature an interview that follows the PEACE guidelines compared with an interview that diverts from such recommendations, is more likely to elicit good quality information in both the initial and the subsequent recall attempt. Furthermore, we found evidence that eyewitness confidence can change during an interview, and contrary to our expectation,

confidence appears to decrease independently on the quality of the initial interview received. However, due to the limitations of Study 1 we are unable to understand why confidence decreases and what elements of the Best Practice Interview and Poor Practice Interviews are likely to cause such change. Therefore, Study 2 aims to address the two remaining questions: (i) why does eyewitness confidence change during an interview, and (ii) what elements of the interview are likely to lead to such change?

Chapter 5: Study 2. Investigating the effect of free recall and closed questions on memory performance and memory confidence.

Introduction

In Study 1 we investigated; (i) the impact of the quality of the initial interview (Best Practice Interview and Poor Practice Interview) on memory performance and memory confidence. We speculated that participants in the Best Practice Interview would report more complete accounts and therefore increase their confidence after the initial interview. On the contrary we expected that participants in the Poor Practice Interview would report less complete accounts and therefore decreased confidence after the initial interview. The results only partially supported our hypotheses. In particular, we found that after the initial interview participants in the Best Practice Interview group reported more complete accounts compared with those in the Poor Practice Interview group (as measured by the amount of correct information reported), however, this difference did not lead to the expected difference in confidence shifts. Despite participants in the Best Practice Interview group reporting more complete accounts than those in the Poor Practice Interview group, confidence for participants in the two groups decreased to a similar extend after the initial interview.

Despite not finding support for our hypothesis, Study 1 shows that eyewitness confidence is malleable and can change during an interview. However, due to the complexity of the initial interviews adopted in Study 1, we were unable to understand why confidence changed and what factors led to such change. In order to address these remaining questions, and build on the limitations of Study 1, we adopted a more controlled manipulation of the quality of the initial interviews in Study 2, as well as investigating why confidence changes after an interview and what might cause such change.

Manipulation of the quality of the initial interview

In Study 1 we operationalised the quality of the two initial interviews by manipulating several interviewing techniques, such as verbal and non-verbal rapport, types of question used, and use of compatible questioning style. Similarly, in Study 2 we aimed to develop a 'best practice' and 'poor practice' initial interview that (a)

represented commonly accepted good and poor investigative interview practice, and (b) had replicable features within each interview to ensure consistency and experimental control. In order to do so the decision was made to focus on the differences in the types of question used in each initial interview. As discussed in Chapter 1, early and more recent research investigating the quality of real-life interviews highlighted that a common deviation from best-practice in interviewing witnesses relates to the use of questions (Carson & La Rooy, 2015; La Rooy et al., 2013; La Rooy et al., 2011; Loftus & Greenspan, 2017; Otgaar et al., 2018; Oxburgh et al., 2012; Wade et al., 2018). In particular, while best practice guidelines advocate for the use of open questions, in real life interviewers still strongly rely on the use of closed questions. Therefore, in this study we use the Free Report invitation as proxy of the Best Practice Interview and a set of Cued Recall questions as proxy of the Poor Practice Interview.

Study 2 was similar to Study 1. Here, participants were asked to recall their memory on two occasions. At Time 1 participants were given either a Free Recall Interview or Cued Recall Interview. While at Time 2 all participants received the same written Free Recall test. Confidence in overall memory and confidence in engaging with the Criminal Justice System was measured both before and after this initial interview.

Rationale: gap in memory

Research has showed that confidence in the information reported in response to forced-choice questions (Allwood et al., 2008), TED questions (Knutsson et al., 2011), and a probing invitations (Sharps et al., 2012) is lower than confidence in the information reported in response to a free recall invitation. A potential explanation is that during a free recall, participants can exert higher control over their own memory, and therefore they are more likely to report information they are sure about. On the contrary, when answering specific questions participants have less control over their memory, therefore they are more likely to report information they are not confidence about (Allwood et al., 2008). In line with this explanation, we speculate that when participants recall a memory via a free recall invitation, and therefore have higher control over their own retrieval process, they only think about and report the information they know, and do not necessarily focus on information they don't

know, thus their confidence remains unchanged. On the contrary, when answering closed questions participants' retrieval is largely dependent on the interviewer's questions, and their attention is more likely to be directed towards information they might not know. Thus, when participants answer specific questions, they are more likely to think about potential gaps in memory and therefore their confidence is more likely to decrease.

Confidence can influence the willingness to testify (Hafstad et al., 2004; Wells & Bradfield, 1998), thus confidence in engaging with the Criminal Justice System is expected to decrease in the Cued Recall Interview group and to remain stable in the Free Recall Interview group.

Research on metamemory suggests that memory confidence underpins the decision to regulate the information reported (Koriat & Goldsmith, 1996). For example, when a detail retrieved is associated with low confidence participants are less likely to volunteer it and more likely to withhold it. Thus, decreased confidence after the initial interview is expected to influence the quality of information subsequently reported. To test this hypothesis, we investigated whether decreased confidence (expected in the Cued Recall Interview group only) (i) leads to fewer and less accurate information being reported at Time 2, and (ii) is correlated with the amount and accuracy of information reported at Time 2. The hypotheses are as follows;

Hypotheses

H_1 : Under conditions of free recall, participants are likely to only think about the information they know, and their attention is unlikely to be directed towards unknown information, thus their confidence is likely to remain stable. On the contrary, under condition of directed retrieval participants' attention is more likely to be directed towards unknow information and therefore their confidence is more likely to decrease.

Predictions: Confidence is expected to remain stable for participants in the Free Recall Interview group, and to decrease for participants in the Cued Recall Interview group.

H_2: Memory confidence is likely to influence participants' confidence in engaging with the Criminal Justice System.
Predictions: Confidence in engaging with the CJS is expected to follow a similar pattern of that of memory confidence. Thus, we predict it to decrease in the Cued Recall Interview group and to remain stable in the Free Recall Interview group.

H_3: Decreased confidence following an initial interview is likely to influence the quality of information subsequently reported.
Prediction 1: At Time 2 participants in the Cued Recall Interview group are expected to report fewer correct and less accurate information than participants in the Free Recall Interview group.
Prediction 2: The decreased confidence after the initial interview is expected to be correlated with lower amount of information and lower accuracy of the information reported at Time 2.

Method

Design

A 2 (Initial Interview Type: Cued Recall vs Free Recall) x 2 (Confidence Measurement Time: Time 1, Time 2) mixed design was used, with Initial Interview Type manipulated between subjects, and Time manipulated within subjects. The DVs for memory recall at Time 2 were: (a) amount of correct details, and (b) accuracy of information reported. The DVs for confidence were: (c) confidence reported before the initial interview (Time 1), and (d) confidence reported after the initial interview (Time 2).

Participants

A total of 40 participants took part in the study (33 females; Mean age = 22.16, SD = 7.41). All participants were recruited in the Goldsmiths University, about half of the sample was recruited among the first-year cohort of Psychology students, the remaining were students of different departments. All participants received either course credits or £10 Amazon Vouchers as compensation for their time.

Materials

Video stimulus

The crime-video was the same as described in Study 1, and depicted a non-violent robbery filmed in a Blockbuster video-rental store.

Time 1: Initial Interview Type

Both Interviews were administered face-to-face.

Free Recall Interview. In the Free Recall interview the researcher started by introducing herself, and then gave a general instruction of the task by informing the participants that they were going to talk about the video they had just seen. Participants were asked to recall what they had seen in response to the general open-ended question: "What happened?" Three memory prompts were provided whereby the participant was asked to focus their attention of the setting of the robbery, the people involved, and their actions. No interruptions, facilitations, or follow up questions followed this recall. The instruction was as follows: *'My name is Alessandra, thank you for taking part in this experiment. We are now going to talk about the video you saw. Please focus your attention on the setting in which the event happened, the people involved, what they did, and what they looked like. When you are ready, please tell me what happened'*. No time limit was given for this task.

Cued Recall Interview. In the Cued Recall Interview the researcher started by introducing herself, and then informed the participants that they were going to talk about the video they had just seen, and that questions would be asked about the setting, the people involved, and their actions; thus, the participant was provided with the three memory prompts. The interview consisted of 24 5Wh- questions (e.g., what? where? who?) regarding actions (8 questions), location (8 questions), and person descriptions (8 questions) (Appendix E). The instruction was as follow *"My name is Alessandra, thank you for taking part in this experiment. We are now going to talk about the video you saw. I will ask you questions about the setting in which the event happened, the people involved, what they did, and what they looked like"*. After this the researcher asked the questions in the same order for all participants. Participants were told that they could say "I don't know" if they were not sure about their answer. No interruptions, facilitations, or further questions followed this recall.

Time 2: Free Recall Test

The second Free Recall test was a written memory test, the instruction was similar to that described for the Free Recall Interview, the instruction was as follows; *'Please focus your attention on the setting in which the event happened, the people involved, what they did, and what they looked like. When you are ready, please write down what happened'.* No time limit was given for this task.

Confidence questionnaire

The questionnaire was identical to that described for Study 1. Here we investigated both (a) confidence in the memory for the event seen and, (b) confidence in memory as to able to engage with the Criminal Justice System. The scale ranged from 10% to 100%.

Procedure

The basic procedure was the same as described for Study 1. Data collection was conducted in a laboratory at Goldsmiths University, and the interviews at Time 1 were conducted face-to-face. The experiment took about 20 minutes. After reading and signing the consent form (this was identical to that used for Study 1, see Appendix B), participants watched the crime video which was displayed on a 17-inche computer screen. After that they were randomly assigned to one of the two conditions and interviewed accordingly. The confidence questionnaire was administered after the video but before the interview (Time 1), and again after the interview (Time 2). When participants had completed the confidence questionnaire at Time 2, they were given the written Free Recall test to complete. Finally, all participants received the debrief form (this was similar to that used in Study 1, see Appendix C) and were compensated for their time.

Coding

Only the details reported in the Free Recall test at Time 2 were coded. It would not have been meaningful to compare recall at Time 1 due to the format of the interviews being too different. Each detail was coded as an Action (a) if it referred to an action, Location (l) if it referred to the setting of the event, and Person Descriptor (pd) if the detail referred to the people involved, their clothing, or personal objects.

The details were coded as Correct if present in the video, or as Incorrect if not present in the video. The coding scheme used was identical to that used for Study 1. Two researchers coded ten (25%) randomly selected transcripts. The percentage agreement between coders was calculated by following the same procedure as that described in Study 1 and it was found to be high (99%). Cohen's Kappa was 0.9, reflecting almost perfect agreement (McHugh, 2012).

Results

All statistical tests were performed at alpha level of .05 unless otherwise specified. Homogeneity of variance was assessed with Levene's test (1960).

Change in memory confidence

As in Study 1, confidence ratings were collected both before and after the initial interview. We predicted that confidence after the initial interview (Time 2) would decrease for participants in the Cued Recall Interview group, while confidence for participants in the Free Recall group would remain stable.

First, we looked at confidence rating reported before the interview, and found no difference between groups $t(38) = .48$ $p = .63$, $d = .15$, ($M_{diff.} = 1.87$, CI [-6.02, 9.77]; see Table 5.1 for Means and Standard Deviations). Following this, we investigated changes in memory confidence with a 2 (Initial Interview Type: Free Recall, Cued Recall) X 2 (Confidence Measurement Time: Time 1, Time 2) mixed ANOVA. We found a non-significant main effect of Initial Interview Type $F(1, 38) = 3.25$, $p = .08$, $n_p^2 = .08$, a significant main effect of Time $F(1, 38) = 12.92$, $p = .001$, $\eta^2_p = .25$, and a significant interaction $F(1, 38) = 16.51$, $p < .001$, $\eta^2_p = .30$. A post-hoc power analysis conducted in G*power (Erdfelder et al., 1996), with $\alpha = .05$, $n = 40$, and the average of the observed effect sizes, yielded a power $(1 - \beta)$ of .87 (ranging from .43 to .98). This analysis suggests the study might be underpowered and unable to detect small to medium size effects. Follow up comparisons showed that confidence for participants in the Cued Recall Interview was significantly lower than confidence for participants in the Free Recall Interview group at Time 2 only ($M_{diff.} = -15.50$, $p = .002$, 95% CI [-25.08, -5.92]). Comparisons as a function of Time confirmed that *only* participants in the Cued Recall Interview group decreased their confidence

significantly after the initial interview ($M_{diff.}$ = -16.37, p < .001, 95% CI [-22.49, -10.25]).

In summary we found support for our Hypothesis 1 in that confidence for participants in the Cued Recall Interview group decreased significantly after the interview, while confidence for participants in the Free Recall group remained stable.

Table 5.1. Means (and Standard Deviations) of confidence reported at Time 1, and at Time 2 for the Cued Recall Interview and Free Recall Interview group.

		Cued Recall Interview	Free Recall Interview
Confidence	Time 1	72.25 (14.53)	70.38 (9.67)
	Time 2	55.88 (15.92)	71.38 (16.72)

Changes in confidence in engaging with the Criminal Justice System

As in Study 1, in Study 2 we used four measures to investigate changes in participants' confidence in engaging with the different stages of the Criminal Justice System (i.e., talking to the police about their recollection for the event, signing a statement, taking part in an identification task, and presenting evidence in Court). We first examined confidence ratings reported at Time 1 and found no difference between groups in the confidence measures reported: talking to the police $t(38)$ = -.38 p = .70, d = -.12, ($M_{diff.}$ = -2.50, 95% CI [-15.69, 10.69]); signing a statement $t(38)$ = -.23, p = .81, d = -.07, ($M_{diff.}$ = -1.50, 95% CI [-14.62, 11.62]); participating in an identification task $t(38)$ = -.38 p = .70, d = -.12, ($M_{diff.}$ = 2.50, 95% CI [-15.64, 10.64]), and providing evidence in Court $t(38)$ = -.54 p = .65, d = -.17, ($M_{diff.}$ = -3.50, 95% CI [-15.04, 8.04]). See Table 5.2. for Means and Standard Deviations.

Following this, we performed a series of 2 (Initial Interview Type: Free Recall, Cued Recall) x 2 (Confidence Measurement Time: Time 1, Time 2) mixed ANOVAs on the DVs: (i) talking to the police, (ii) signing a statement, (iii) participating in an ID task, and (iv) presenting evidence in Court (see table 5.3). We found a significant main effect of Initial Interview Type on confidence in presenting evidence in Court *only* $F(1, 38) = 5.74$, p = .022, η^2_p = .13, meaning that across the two time points participants in the Cued Recall Interview group compared with those

in the Free Recall Interview group were significantly less confidence in providing evidence in Court ($M_{diff.}$ = -11.75, p = .02, 95% CI [1.82, 21.67]). The main effect of Time on the four DVs was found non-significant, meaning that confidence in engaging with the CJS did not change exclusively as function of time.

However, we found a significant interaction for (i) signing a statement $F(38)$ = 8.95, p = .005, n_p^2 = .19; (ii) participating in an ID task $F(1, 38)$ = 7.71, p = .008, η^2_p = .17; and (iii) presenting evidence in Court $F(1, 38)$ = 9.69, p = .004, n_p^2 = .20. Follow up comparisons as a function of Time confirmed that at Time 2 only participants in the Cued Recall Interview group significantly decreased their confidence in relation to (i) signing a statement ($M_{diff.}$ = -12.50, p = .004; 95% CI [-20.63, -4.36]); (ii) participating in an ID task ($M_{diff.}$ = -13.50, p = .005; 95% CI [-22.77, -4.22]); and (iii) providing evidence in Court ($M_{diff.}$ = -13.00, p = .001; 95% CI [-20.58, -5.41]). Consequently, at Time 2 participants in the Cued Recall Interview group compared with those in the Free Recall Interview group were significant less confidence in (i) signing a statement ($M_{diff.}$ = -18.50, p = .003; 95% CI [-30.36, -6.63]), (ii) participating in an ID task and ($M_{diff.}$ = -20.50, p = .014; 95% CI [-36.63, -4.36]); (iii) presenting evidence in Court ($M_{diff.}$ = -20.00, p = .001; 95% CI [-31.01, -8.99]).

In summary, we found support for Hypothesis 2, in that after the initial interview participants in the Cued Recall group decreased their confidence in engaging with the Criminal Justice System. On the contrary, confidence in engaging with the CJS for participants in the Free Recall group did not change significantly after the initial interview.

Table 5.2. Means (and Standard Deviations) of confidence in memory related to talking to the police, signing a statement, participating in an ID task, and presenting evidence in Court reported at Time 1 and Time 2 in the Cued Recall Interview and Free Recall Interview group.

		Cued Recall Interview	Free Recall Interview
Talking to the police	Time 1	64.00 (24.58)	66.50 (15.65)
	Time 2	60.00 (21.52)	72.00 (15.76)
Signing a statement	Time 1	66.50 (22.07)	68.00 (18.80)
	Time 2	54.00 (20.36)	72.50 (16.50)
Attending ID task	Time 1	59.50 (24.59)	62.00 (15.42)
	Time 2	46.00 (30.15)	66.50 (18.99)
Presenting evidence in Court	Time 1	55.00 (19.86)	58.50 (15.98)
	Time 2	42.00 (18.23)	62.00 (16.09)

Table 5.3. Results of the 2 X 2 ANOVAs

Result	*variables*	df	test value	*p* value	Effect size (n_p^2)
		Talking to the police			
Main effect	Interview	(1, 38)	$F = .31$	= .58	$\eta^2_p = .01$
Main effect	Time	(1, 38)	$F = 1.85$	= .18	$\eta^2_p = .05$
Interaction	Interview x Time	(1, 38)	$F = .01$	= .92	$\eta^2_p = .01$
		Signing a statement			
Main effect	Interview	(1, 38)	$F = 3.32$	= .07	$\eta^2_p = .08$
Main effect	Time	(1, 38)	$F = 1.98$	= .16	$\eta^2_p = .05$
Interaction	Interview x Time	(1, 38)	$F = 8.95$	= .005**	$\eta^2_p = .19$
		Participating in an ID task			
Main effect	Interview	(1, 38)	$F = 3.12$	= .08	$\eta^2_p = .07$
Main effect	Time	(1, 38)	$F = 1.73$	= .17	$\eta^2_p = .04$
Interaction	Interview x Time	(1, 38)	$F = 7.71$	= .008**	$\eta^2_p = .17$
		Presenting evidence in Court			
Main effect	Interview	(1, 38)	$F = 5.74$	= .022	$\eta^2_p = .13$
Main effect	Time	(1, 38)	$F = 3.21$	= .08	$\eta^2_p = .07$
Interaction	Interview x Time	(1, 38)	$F = 9.69$	= .004**	$\eta^2_p = .20$

** indicates significance at $p < .01$

Memory recall at Time 2

At Time 2, all participants were given the same Free Recall test. Here we predicted that participants in the Cued Recall Interview group – which reported lowered confidence following the interview, would report poorer quality of information (as measured by: amount of correct details, and accuracy of information) compared to those in the Free Recall Interview group. However, a series of Independent t-tests showed no difference in either the amount of correct details

reported, *t(38)* = -1.07, *p* = .28, *d* = -.33, ($M_{diff.}$ = -5.45, 95% CI [-15.73, 4.81]), or in the accuracy of information reported, *t(38)* = -.78, *p* = .43, *d* = -.25, ($M_{diff.}$ = -1.32, 95% CI [-4.71, 2.07]). See Table 5.4 for Means and Standard Deviations.

In summary, we found no support for our Hypothesis 3; in particular, participants in the Cued Recall group, despite showing a drop in confidence after the initial interview, did not report poorer quality accounts in a subsequent recall attempt, as measured by the amount of correct details and the overall accuracy of the details reported at Time 2.

Table 5.4. Means (and Standard Deviations) of amount of correct details, amount of incorrect details, and accuracy of information at Time 2 for the Cued Recall Interview and Free Recall Interview group.

	Cued Recall Interview	Free Recall Interview
Correct details	46.45 (16.56)	51.90 (15.48)
% Accuracy rate	92.17 (6.24)	93.49 (4.14)

Relation between confidence and memory recall at Time 2

Our Hypothesis 3 predicted that the decrease in confidence reported after the initial interview would lead to poorer quality accounts reported at Time 2. Thus, we expected that decreased confidence would be correlated with the amount of correct details and the accuracy of the information reported at Time 2. For each participant we calculated a decrease in confidence measure by subtracting the confidence ratings reported at Time 1 from the confidence ratings reported at Time 2. Following this, Pearson Correlations were conducted between decreased confidence and (i) amount of correct details, and (ii) accuracy of the details reported at Time 2. We found a non-significant relationship between the decreased confidence and both the amount of correct details reported at Time 2, *r(*38) = .06, *p* = .71; and the accuracy of the information reported, *r(*38) = -.06, *p* = .69.

In summary, contrary to our expectations, we found no support for the hypotheses that decreased confidence after the initial interview is associated with the amount or accuracy of the information reported at Time 2.

Discussion

Study 2 further investigated how and why the quality of an initial interview can affect eyewitness confidence. The primary aim was to understand when eyewitness confidence is likely to change and what are the factors of an interview likely to cause such change. We predicted that when the initial interview triggers a free or undirected retrieval, participants are less likely to think about the information they do not know and therefore their confidence is unlikely to change. On the contrary when the interview directs participants' retrieval, their attention is more likely to focus on the information they do not know and therefore their confidence is more likely to decrease. We found support for this hypothesis in that participants that received a Cued Recall Interview decreased their confidence after the initial interview, while participants in the Free Recall Interview group showed stable confidence after the initial interview. Furthermore, the results of Study 2 confirm the predictions that confidence in Engaging with the Criminal Justice System changed in a similar direction. Thus, after the initial interview participants in the Cued Recall Interview were less confident is signing a statement about what they had witnessed, providing identification evidence, and presenting evidence in Court. Finally, we speculated that that decreased confidence following the initial interview would impact the quality of information subsequently reported. However, we found no support for this hypothesis, as shown by the trivial difference in the quality of information reported at Time 2 by participants in the two interview groups, and by the non-significant correlation between decreased confidence and amount of correct details and accuracy of the information reported at Time 2.

Researchers have shown that the information reported in response to a free report invitation is more likely to be associated with higher confidence compared with information reported in response to specific or follow up questions (Allwood et al., 2008; Knutsson, et al., 2011; Sharps et al., 2012), and they have argued that this is due to the higher control that participants have over their own memory search in response to free recall invitations compared to more specific questions. However, we did not know whether these differences lead to potential changes in confidence. In Study 2, we adopted a within-subjects methodology and investigated the conditions of the interview that are more likely to cause confidence changes and those that are more likely to leave confidence stable. In line with our hypothesis, Study 2 confirmed that eyewitness confidence can change after an interview, however such

change is contingent on the quality of the initial interview. A good practice initial interview – that promotes a free and undirected retrieval, is more likely to leave confidence unchanged. On the contrary, a poor practice interview - that directs participants' retrieval on specific details of the memory is more likely to decrease confidence.

In Study 2 we found evidence that the quality of the initial interview can influence confidence in engaging with the Criminal Justice System. Therefore, practitioners should be aware that when the interviewing process exposes potential gaps in memory, and eyewitnesses experience their memory as less good than they initially thought, they might lose confidence in their ability to engage with the investigation and its stages. This is concerning for the legal context, because decreased confidence following a poor interview might lead eyewitness to become reluctant in taking part in future interviews and cooperate further with the police. Furthermore, the loss of confidence in the quality of own memory, could lead eyewitnesses to perceive themselves as unable to help the investigation, with detrimental consequence for their wellbeing – especially if the eyewitness is also the victim of the crime being investigated (Diesen, 2012).

On the other hand, Study 2 shows that decreased confidence following a poor initial interview might not necessarily lead participants to report poorer accounts in a subsequent recall attempt. Research on metamemory has showed that confidence is pivotal in the control the information reported (Koriat & Goldsmith, 1996), and it is surprising that the decreased confidence in the quality of own memory did not show the expected effects on subsequent memory performance. However, our methodology is unlikely to capture the effect that decreased confidence might have had on eyewitness's cognitive processes. For example, decreased confidence after the initial interview could have affected participants' monitoring process. It is possible that decreased confidence led participants to systematically associate lowered confidence to the details they retrieve independently on their objective accuracy. However, in Study 2 the methodology does not allow us to investigate participants' monitoring process.

Similarly, the statistical tests used in Study 2 are not appropriate to investigate the relation between confidence and accuracy of information reported. Previous research suggests that correlation analysis might not be ideal to investigate the relation between confidence and accuracy. In particular, researchers suggest that

this analysis can provide useful information relating to the shared variance between subjective confidence and objective accuracy, but it does not provide information about over or under-confidence (i.e., Brewer & Wells, 2006; Brewer et al., 2002; Juslin et al., 1996; Weber & Brewer, 2004; Wixted et al., 2015). Thus, the analysis adopted in Study 2 might be unable to fully capture the effect that decreased confidence had on how participants were monitoring the accuracy of their own memory.

In conclusion Study 2 provide evidence that eyewitness confidence is not stable, and that the quality of an initial interview can impact memory confidence. An interview that promotes free retrieval is unlike to cause shifts in confidence, presumably because eyewitness do not focus their attention on information unknown, and they are likely to only think about information they do know about. On the contrary, an interview that directs eyewitness' attention towards specific details is more likely to cause shifts in confidence because attention is more likely to be directed towards unknown information. Furthermore, we found evidence that a poor practice interview, compared with a good practice interview, is more likely to decrease confidence in one's own memory and confidence in engaging with the Criminal Justice System. Finally, we found no evidence that decreased confidence following a poor practice interview leads participants to report poorer accounts in a subsequent recall attempt.

In Study 3, we aim to better understand the cognitive processes that underpin our results. In particular we investigate further the specific aspects of the directed retrieval process (i.e., retrieval products) likely to cause decreased confidence following a poor practice interview. Furthermore, by adopting more appropriate methodologies and statistical tests, we examine whether decreased confidence impacts eyewitnesses' monitoring (Study 3a) and control (Study 3b) of the information subsequently reported.

Chapter 6: Study 3. Investigating the effect of ease of retrieval on memory confidence and subsequent memory output and memory monitoring

Introduction

Study 2 tested the hypothesis that when participants engage in a free and undirected retrieval task (i.e., Free Recall Interview) they only recall the information they are sure about and do not necessarily focus their attention on the information they do not know, thus their confidence is unlikely to change. On the contrary, when participants answer questions and their retrieval is directed (i.e., Cued Recall Interview), their attention is more likely to be led towards unknown information, therefore confidence is more likely to decrease. We also expected that the confidence in engaging with the CJS would change in the same direction as confidence in the own memory. Finally, we speculated that decreased confidence following a poor practice initial interview would lead participants to report poorer accounts in a subsequent recall attempt.

As expected, confidence decreased only after a Cued Recall Interview, while it remained stable following a Free Recall Interview. Furthermore, confidence in engaging with the CJS followed a similar pattern; as such, confidence in signing a statement, providing identification evidence, and providing evidence in Court decreased following a Cued Recall Interview, while it remained stable following a Free Recall Interview. However, contrary to our predictions decreased confidence did not lead participants to report fewer or less accurate information and was not associated with the amount of correct details and the accuracy of the details reported in a subsequent recall attempt.

In Study 3a we investigated the cognitive processes that underpin these results. In particular, we focus our attention on (i) the specific cognitive processes that are likely to lead to decreased confidence, and (ii) the effects that decreased confidence following an interview has on subsequent memory monitoring.

Rationale: expanding the gaps in memory hypothesis

In Study 2 we speculated that eyewitness confidence might decrease only when during the search in memory, participants become aware of details they do not know. This is likely to promote the impression that the memory for the event seen is

less good than they had previously thought. During a Cued Recall Interview, in answer to each question participants attempt to retrieve the information triggered by the question; if this is not accessible or not available, participants are likely to experience a degree of retrieval difficulty. The retrieval difficulty that derives from attempting to access unavailable or un-accessible information can in turn lead to reduced memory confidence. However, during a Free Recall Interview, participants are unlikely to experience retrieval difficulty, presumably because they are more likely to search for the details they know and that are easy to access. In Study 3a we investigate whether retrieval difficulty experienced when answering different initial recall tests might underpin the decrease in confidence following an initial recall test.

As discussed in Chapter 2, Raghubir and Menon (2005) suggested that unexpected retrieval difficulty can be attributed to the quality of own memory. Furthermore, the metamemory literature has provided evidence that the ease of retrieval influences subsequently made confidence judgements relating to the quality of own memory (e.g., Kelley & Lindsay, 1993; Gustafsson et al., 2019; Lindholm et al., 2018). Further evidence that the ease of retrieval is used to build confidence judgements regarding the quality of own memory comes from the literature manipulating the difficulty of the retrieval task (Belli et al., 1998; Gregg et al., 2019; Merckelbach et al., 2001; Schwarz et al., 1991; Winkielman et al., 1998). Building on this literature we extend the gap in memory hypothesis and speculate that confidence is likely to change as a function of the ease of retrieval experienced during the initial recall attempt. In particular, confidence is more likely to decrease when the recall test is likely to trigger details that are difficult to retrieve. On the contrary when the recall test is unlikely to trigger information difficult to retrieve, confidence is more likely to remain stable.

Confidence is pivotal in the regulation of information reported (Koriat & Goldsmith, 1996), but in Study 2 we found no evidence that decreased confidence leads participants to report poorer accounts. However, we argued that due to our methodology we were unable to fully understand the impact of decreased confidence on eyewitnesses' monitoring of information subsequently reported. For example, it is possible that decreased confidence in the quality of own memory led participants to systematically associate lowered confidence to the details they subsequently recalled, independently on their objective accuracy. Thus, a second aim of Study 3a is to investigate whether decreased confidence following an initial recall test influences

eyewitnesses' monitoring process. In order to measure the relationship between the confidence judgements and the accuracy of the information subsequently reported we used a different statistical test: the calibration analysis.

From correlation to calibration

It has been suggested that correlation analysis might not be the ideal test to investigate the relationship between confidence and accuracy, and researchers have proposed the use of *calibration analysis* instead (Brewer et al., 2002; Juslin et al, 1996; Olsson, 2000). One major advantage of the calibration analysis is that it provides clear information about how participants diverge from perfect calibration, and whether and to what extent they over or underestimate the accuracy of their own memory. On the contrary, a correlation can only provide information about the shared variance between confidence and accuracy and the direction and strength of their relationship. As such, researchers have argued that calibration might be more informative for the CJS, and Juslin et al. (1996) provided an example to explain this advantage. If an eyewitness is 80% confident about a positive identification made, the Court might want to know how reliable an identification made with 80% confidence really is. In this circumstance information about how over or underconfident eyewitnesses can be in an identification made is more useful than knowing about the shared variance between confidence and accuracy.

Calibration analysis measures the extent to which subjective confidence matches objective accuracy. It is calculated by plotting the proportion of correct answers given in each confidence category (i.e., 0-20%, 30-40%, 50-60%, 70-80%, 90-100%). The resulting calibration curve provides information about perfect calibration and divergence from it. Here, the diagonal line represents perfect calibration; on the contrary a curve that fall above the diagonal line shows under-confidence, and a curve that falls below the diagonal line shows overconfidence. A participant is perfectly calibrated if as an example, 80% of the correct answers given is associated with 80% confidence. However, if 80% of the correct answers is given with 40% confidence, the participant is under-confident, i.e., their confidence judgements are underestimating the probability of correctness of their own answers. Similarly, if 80% of the correct answers is given with 100% confidence, the participant is overestimating the probability of correctness of their own answers, and

therefore they are showing overconfidence in the accuracy of their memory. Two further indexes can be calculated to describe the curve: the calibration index (C-index), and the over/under-confidence index (O/U confidence index). C-index is calculated with the equation 1 by finding the differences between mean confidence *(c_j)* and proportion correct answers *(a_j)* for each confidence category *(j)*, then multiplying the squared difference by the number of total answers *(n_j)* given in the category, and finally by summing those across confidence categories and dividing them by the total number of answers reported *(n)*.

$$\frac{1}{n}\sum_{j=1}^{J} n_j\left(c_j - a_j\right)^2 \qquad (1)$$

C-index ranges between 0 – representing perfect calibration, and 1 - representing poor calibration. In order to calculate the O/U confidence index the same formula is used, however the differences between mean confidence and proportion correct are not squared. O/U confidence index ranges between -1 and +1, whereby negative scores represent under-confidence and positive scores represent over-confidence (for a guide on how to calculate C-index and O/U confidence index see Brewer, et al. 2002; Brewer & Wells, 2006). Finally, a further index - the Adjusted-Normalised Discrimination Index (ANDI) can be calculated to measure whether participants are using their confidence judgements to successfully discriminate between correct and incorrect answers.

$$NDI = \frac{\frac{1}{N}\sum_{t=1}^{N} n_t\,(c_t - c)^2}{c(1-c)} \qquad (2) \qquad\qquad ANDI = \frac{N*NDI - T + 1}{N - T + 1} \qquad (3)$$

In formula 2, c_t represents the mean proportion of correct responses in each confidence category *(t)*, and *c* is the overall proportion of correct answers given. The NDI is adjusted for number of confidence categories *(T)*, and number of total answers *(N)* (see formula 3), (for a detailed guide on how to calculate ANDI see Yaniv et al., 1991).

In Study 3a we adopted the calibration analysis to investigate how effectively participants monitor the accuracy of the information they report after the initial recall test.

Study 3a

Study 3a aims to investigate (i) the specific cognitive processes that are likely to lead to decreased confidence, and (ii) the effects that decreased confidence following an initial recall test has on subsequent memory monitoring. We asked participants to recall the event seen on two occasions. At Time 1, they answered either a Difficult Cued Recall, and Easy Cued Recall, or a Free Recall initial test, while a fourth group did not recall the event at this stage (Control group). At Time 2 half of the sample was given a second Free Recall test, while the other half answered a Neutral Cued Recall test. Confidence was measured before and after the initial interview and after the second Free Recall test. Furthermore, a confidence rating was required after each question of the initial recall test for participants in Difficult Cued Recall and Easy Cued Recall group, and for each question of the Neutral Cued Recall test. The hypotheses and predictions are as follows:

H_1: Confidence will decrease only when participants are more likely to experience retrieval difficulty in accessing details in their memory. On the contrary when participants are less likely to experience retrieval difficulty confidence is less likely to change.
Prediction: Confidence is expected to decrease only in the Difficult Cued Recall group. While confidence is expected to remain stable in the Easy Cued Recall, Free Recall, and Control group.

H_2: Decreased confidence reported by participants in the Difficult Cued Recall group will have an impact on subsequent memory monitoring.
Prediction: We predict that participants in this group will not report poorer accounts (in the second Free Recall test) but will show under-confidence in the information subsequently reported (in the second Neutral Free Recall test).

Method

Design

We used 4 (Initial Recall Test: Difficult Cued Recall, Easy Cued Recall, Free Recall, and Control) x 3 (Confidence Measurement Time: Time 1 - before Initial Recall, Time 2 - after Initial Recall, and Time 3 - after Second Recall) mixed design,

with confidence measured within subjects. The dependent variables related to confidence were: (a) confidence at Time 1 (before the Initial Recall test), (b) confidence at Time 2 (after the Initial Recall test), (c) and confidence at Time 3 (after the Second Recall test).

The dependent variables for memory report in the second Free Recall test included; (d) amount of correct information and (e) accuracy of information reported in the second Free Recall test. The dependent variables for confidence-accuracy calibration after the second Neutral Cued Recall test were: (f) C-index, (g) ANDI, and (h) OU (over/under-confidence) index.

Participants

A total of 134 participants took part in the study (99 females, age range 18 to 51 years, M = 23.61, SD = 6.56). They were recruited among students and staff members of all departments at Goldsmiths University. Participants took part in the experiment in exchange for either course credits, or £5 compensation.

Materials

Video stimulus

The video was the same as described for Study 1, it involved a robbery carried out in a Blockbuster video-rental store.

Time 1: Initial Recall Test

Difficult Cued Recall and Easy Cued Recall Test. The Difficult Cued Recall and Easy Cued Recall questions included 16 closed questions (Appendix F). To build the set of questions we followed a two stages process. First, we randomly selected 10 Free Recall transcripts from Study 2 and calculated the mean correct details reported, for each character in each scene. This allowed us to identify the scenes and the characters in the video that participants remembered the most. We then formulated 38 potentially easy questions on details relating to the scenes and the characters with numerically higher means; and 39 potentially difficult questions on details relating to scenes and characters with numerically lower means. Second, we conducted a pilot test where participants (n = 23) were shown the target video followed by the 77 questions. Participants were asked to respond only if they felt

they were able to answer the question, and to write 'I don't know' where appropriate rather than guessing. Finally, we selected 16 difficult questions with a mean correct percentage response rate lower than 40 (overall Mean = 21.51; range = 8 to 33), and 16 easy questions with a mean correct response rate higher than 60 (overall Mean = 81.75; range = 63 to 100).

Free Recall. In the Free Recall group participants were asked to report any information that they could remember about the target event. In order to allow for free control on memory reporting, no further instructions were given. The instruction was as follows: *"Please write down in your own words everything you remember about the video. Use the space below. There is no time limit for this task"*.

Control. The control task participants were presented with four geometric images, and were required to select the image that matched a single rotated image. The task included nine different trials. The overall duration to complete the task was similar to that of the two Cued Recall groups.

Time 2: Second Test

Second Free Recall Test. The instruction for this test was as described for the Free Recall test described above; participants were asked to *"Write down in your own words everything you remember about the video. Use the space below. There is no time limit for this task"*.

Second Neutral Cued Recall Test. This test constituted eight closed questions (Appendix G). In order to identify the neutral questions, we used data from the pilot study described above. From this, we chose eight questions associated with a mean correct response rate of between 40 and 60 (overall Mean = 48.37; range = 45 to 54). In order to discourage guessing, participants were instructed to write "I don't know" if they were not sure about their response. The instruction was as follows: *"Please answer the questions and provide a confidence rating to indicate how confident you are that your answer is correct. It is very important that you don't guess. If you don't know the answer, please write down "I don't know"*.

Confidence

Confidence was measured immediately after the video and before the Initial Recall test, immediately after the Initial Recall, and immediately after the second recall. Participants were asked to think about their memory for the event, and rate how confident they were that the information they remembered could help the police to accurately reconstruct the event. The confidence scale ranged from 0% to 100%, where 0 indicated 'not confident at all', and 100 indicated 'completely confident'. The confidence question was the same across the three-points time.

Procedure

Participants were tested either individually or in small groups (up to 10) in a computer laboratory in the university campus. The study was administered as a Qualtrics survey (https://goldpsych.eu.qualtrics.com/jfe/form/SV_50hXwUsDGDzEKtD) and took between 20 and 30 minutes to complete. The survey was designed to randomly allocate participants to each condition[2]. As soon as the participant(s) arrived in the laboratory, they were invited to take a seat in front of the computer and were given the headphones. During group testing we ensured participants could only see their monitor by placing them far apart from each other. After reading the consent form participants were shown the non-violent video clip and then completed a three-minute word-search filler task. Following this, participants recalled the event accordingly to the Initial Recall test they were assigned to (i.e., Difficult Cued Recall, Easy Cued Recall, or Free Recall), the Control group did not recall the event at this stage, rather they completed the rotation task as described above. In the Difficult Cued Recall and Easy Cued Recall test participants were asked to provide a confidence rating in relation to each answer given. After this, participants were presented with a three-minute word-search filler task, which was different from the first they had completed before. Finally, participants in all groups were given the second Recall test. Here, sixty-nine participants were allocated to the second Free

[2] However, the function "evenly present elements" was not activated, thus groups have unequal number of participants. There were 35 participants in Difficult Cued Recall, 32 in the Easy Cued Recall, 33 in the Free Recall, and 34 in the Control group. In the Second Recall test, 69 participants answered the second Free Recall test and 65 answered the second Neutral Cued Recall test.

Recall test, while sixty-five participants were assigned to the second Neutral Cued Recall test. All participants reported a confidence rating immediately after watching the video but before the Initial Recall, immediately after the Initial Recall test, and immediately after the second Recall test (see Figure 6.1). Furthermore, in the second Neutral Cued Recall test, participants provided a confidence rating in relation to each answer given. Finally, participants read the debrief form and were compensated for their time.

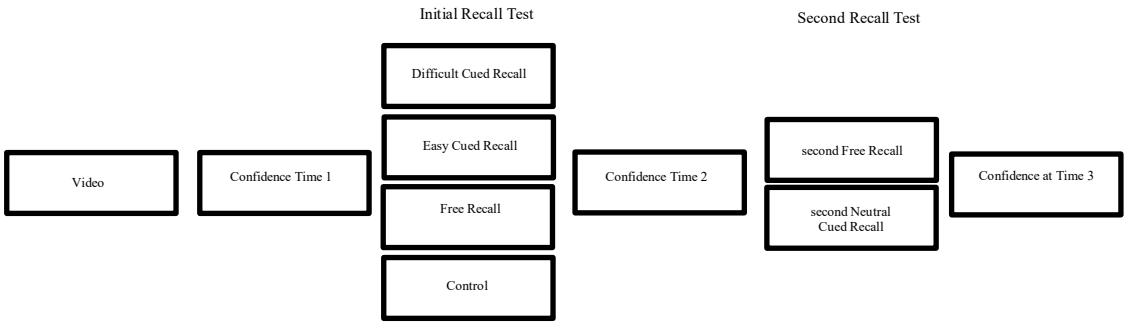


Figure 6.1. Procedure of Study 3a

Coding of the second Free Recall Test

Information reported in the second Free Recall test was transcribed and coded for amount and accuracy. Coding was conducted accordingly to the same principles illustrated for Study 1. In order to ensure inter-coder reliability an independent researcher coded 14 (20%) randomly selected second Free Recall tests. The percentage agreement between coders was found to be high (95%). Cohen's Kappa was .84, reflecting almost perfect agreement (McHugh, 2012).

Results

A conventional *alpha* level of .05 was used, where relevant for non-significant results we present Bayesian analysis (conducted with JASP software) in order to quantify evidence in support of the null hypothesis. We interpret Bayes factors by using Jarosz and Wiley's (2014) guidelines whereby a BF_{01} of 1-3 is interpreted as anecdotal evidence in support of the null, 3-10 as moderate, 10-30 as strong, and 30-100 as very strong (see also Wagenmakers et al., 2018).

Manipulation check

First, we tested whether during the initial recall participants in the Difficult Cued Recall group were experiencing lower ease of retrieval compared with those in the Easy Cued Recall group. In order to do so, we looked at the number of "I don't know" responses reported. These responses suggest that participants might have experienced a degree of difficulty when attempting to access a suitable answer to the question but failing to do so. As expected, participants in the Difficult Cued Recall reported more "I don't know" responses (M = 5.14, SD = 3.04) than those in the Easy Cued Recall group (M = 1.31, SD = 1.49) and this difference was significant $t(65) = 6.45$, $p < .001$, $d = 1.59$, ($M_{diff} = 3.83$, 95% CI [2.64, 5.01]). Following this, we tested whether participants in the Difficult Cued Recall group were overall less confident in the answers given to the initial recall test compared with those in the Easy Cued Recall test. As expected, confidence in answers to the Difficult Cued Recall was lower (M = 41.58, SD = 17.68) than confidence in the answers to the Easy Cued Recall questions (M = 77.19, SD = 15.81), and this difference was significant $t(65) = -8.65$, $p < .001$, $d = -2.12$, ($M_{diff} = -35.60$, 95% CI [-44.82, -27.39]).

Confidence stability

We performed a 2 x 4 mixed Analysis of Variance with confidence measurement time (Time 1 - before Initial Recall, and Time 2 - after Initial Recall) as a within-subjects factor, and Initial Recall test (Difficult Cued Recall, Easy Cued Recall, Free Recall, and Control) as a between-subjects factor. We found a significant main effect of Time on confidence, $F(1, 130) = 41.93$, $p < .001$, $\eta^2_p = .24$, a significant main effect of Initial Recall test on confidence $F(3, 130) = 4.58$, $p = .004$, $\eta^2_p = .09$, and a significant interaction $F(3, 130) = 34.73$, $p < .001$, $\eta^2_p = .44$. A post-hoc power analysis conducted in G*power (Erdfelder et al., 1996), with $\alpha = .05$, $n = 134$, and the average of the observed effect sizes, yielded a power $(1 - \beta)$ of .99 (ranging from .70 to 1). This analysis suggests the study might be slightly underpowered and unable to detect a small size effect. Bonferroni post hoc tests showed that confidence before the Initial Recall test was not different between groups (see Table 6.1 for Means and Standard Deviations). At Time 2, participants in the Difficult Cued Recall group were significantly less confident than those in the Easy Cued Recall ($M_{diff} = -18.75$, $p < .001$, 95% CI [-30.57, -6.94]), the Free Recall

(M_{diff} = -27.78, p < .001, 95% CI [-39.50, -16.06]), and Control group (M_{diff} = -24.51, p < .001, 95% CI [-36.14, -12.88]). No other significant differences between groups were found at this time point.

Furthermore, at Time 2, confidence for participants in the Difficult Cued Recall group was significantly lower compared with confidence reported at Time 1 (M_{diff} = -28.28, p < .001, 95% CI [-33.00, -23.57]). On the contrary confidence for participants in the Easy Cued Recall, Free Recall and Control group did not show a significant change compared with confidence reported at Time 1. In sum, as expected, only the Difficult Cued Recall questions caused a drop in subjective confidence in the memory of the target event.

An exploratory analysis was performed on confidence ratings reported at Time 3 - after the second Recall test for participants that completed the Free Recall test only. Here we were interested in examining if confidence for participants in the Difficult Cued Recall group would increase after the second Free Recall test. A 3 x 4 mixed Analysis of Variance with confidence measurement time (Time 1 - before Initial Recall, Time 2 - after Initial Recall, and Time 3 - after the second Free Recall test) as a within-subjects factor, and Initial Recall test (Difficult Cued Recall, Easy Cued Recall, Free Recall, and Control) as a between-subjects factor was used. The assumption of sphericity was violated therefore we looked at Greenhouse-Geisser estimates. There was a significant main effect of Time on confidence, F(1.83, 118.91) = 11.21, p < .001, η^2_p = .15, a significant main effect of Initial Recall test on confidence F(3,65) = 3.28, p = .03, η^2_p = .13, and a significant interaction F(5.49, 118.91) = 7.30, p < .001, η^2_p = .25. Bonferroni post hoc comparisons showed that at Time 3 participants in the Difficult Cued Recall group did not increase their confidence significantly (M_{diff} = 7.22, p = .09, 95% CI [-.89, 15.34]). On the contrary participants in the Easy Cued Recall group and Control group showed a significant increase in confidence, respectively (M_{diff} = 13.52, p = .001, 95% CI [5.17, 21.88]), and (M_{diff} = 11.77, p = .003, 95% CI [3.41, 20.12]). At Time 3, participants in the Difficult Cued Recall group were significantly less confidence than those in the Free Recall group (M_{diff} = -18.49, p = .04, 95% CI [-36.47, -.51]), and Control group (M_{diff} = -23.20, p = .005, 95% CI [-41.18, -5.22]), but not less confidence than participants in the Easy Cued Recall group (M_{diff} = -17.90, p = .05, 95% CI [-35.89, .07]), although this last result approached statistical significance.

Finally, participants in the Free Recall group did not show any significant change in confidence at Time 3.

In sum, after the second Free Recall test, participants in the Difficult Cued Recall group did not increase their confidence and therefore remained significantly less confident than participants in the Free Recall and Control group.

Table 6.1. Means (and Standard Deviations) of confidence provided at Time 1 and Time 2 in the four groups.

	Time 1	Time 2
Difficult Cued Recall	61.71 (17.57)	33.43 (18.30)
Easy Cued Recall	55.31 (17.41)	52.19 (17.36)
Free recall	56.67 (14.50)	62.42 (18.88)
Control	62.65 (16.72)	56.47 (16.67)

Performance in the second Free Recall Test

We did not expect decreased confidence following an initial recall test to influence the quality of the information reported in the second Free Recall test. A series of four one-way Analyses of Variance tests, revealed no significant difference between groups in amount of correct information $F(3, 65) = .76$, $p = .51$, $\eta^2 = .03$ or accuracy of the information reported in the second recall test $F(3, 65) = .34$, $p = .80$, $\eta^2 = .02$ (see Table 6.2 for Means and Standard Deviations).

We also conducted a Bayesian analysis which compares the "null model" and "model with the expected effect" in order to assess the model that better predicts our data. We have moderate to strong evidence in support of the null for number of correct answers ($BF_{01} = 5.85$) (i.e. our data are approximately 6 times as likely to occur under the null model as they are under the alternative model), and accuracy ($BF_{01} = 8.99$). In line with the results of Study 2, these findings suggest that the decreased confidence after the initial recall test does not impact the number of correct and the overall accuracy of information freely reported in a second Free Recall test.

Table 6.2. Means (and Standard Deviations) of correct details and accuracy of the information reported in the second Free Recall test in the four groups.

	Difficult Cued Recall	Easy Cued Recall	Free Recall	Control
correct details	47.50 (23.71)	50.88 (19.88)	47.18 (15.28)	56.47 (21.78)
% accuracy rate	90.47 (5.63)	90.96 (4.68)	90.80 (5.39)	92.22 (6.09)

Monitoring in the second Neutral Cued Recall Test

We were interested in investigating the effect of decreased confidence following an initial recall test on confidence-accuracy calibration in the second Neutral Cued Recall test. Here, we expected participants in the Difficult Cued Recall group to show under-confidence in the information reported. In order to test this hypothesis each answer was coded as correct if present in the video or incorrect if not present in the video. The 'Don't know' answers were coded as incorrect and these were always associated with 0% confidence[3]. The confidence categories were reduced from 11 (0%, 10%, ... 100%) to five (0-20%, 30-40%, 50-60%, 70-80% and 90-100%). A series of one-way ANOVAs showed that there was no difference between groups in C-index $F(3, 64) = 1.21$, $p = .31$, $\eta^2_p = .05$; or under/over-confidence index $F(3, 64) = .17$, $p = .91$, $\eta^2_p = .01$, and no difference in the ANDI $F(3, 64) = .40$, $p = .75$, $\eta^2_p = .01$ (see Table 6.3 for Means and Standard Deviations and Figure 6.2 for the calibration curves).

Bayesian analysis showed we have moderate to strong evidence in support of the null for C-index ($BF_{01} = 3.62$), ANDI ($BF_{01} = 7.99$), and over/under-confidence index ($BF_{01} = 10.02$). These results suggest that the change in confidence after the initial recall test does not impact confidence-accuracy calibration in a subsequent memory test.

[3] There was no difference between groups in the number of don't know answers reported in the second Neutral Cued Recall test $F(3, 64) = 1.57$, $p = .21$, $\eta^2_p = .06$.

Table 6.3. Means (and Standard Deviations) of the C-index, OU index and ANDI in the four group.

	Difficult Cued Recall	Easy Cued Recall	Free Recall	Control
C- index	.16 (.13)	.09 (.07)	.15 (.09)	.14 (.09)
OU index	.08 (.25)	.09 (.13)	.06 (.22)	.11 (.19)
ANDI	.55 (.42)	.52 (.35)	.40 (.37)	.52 (.46)

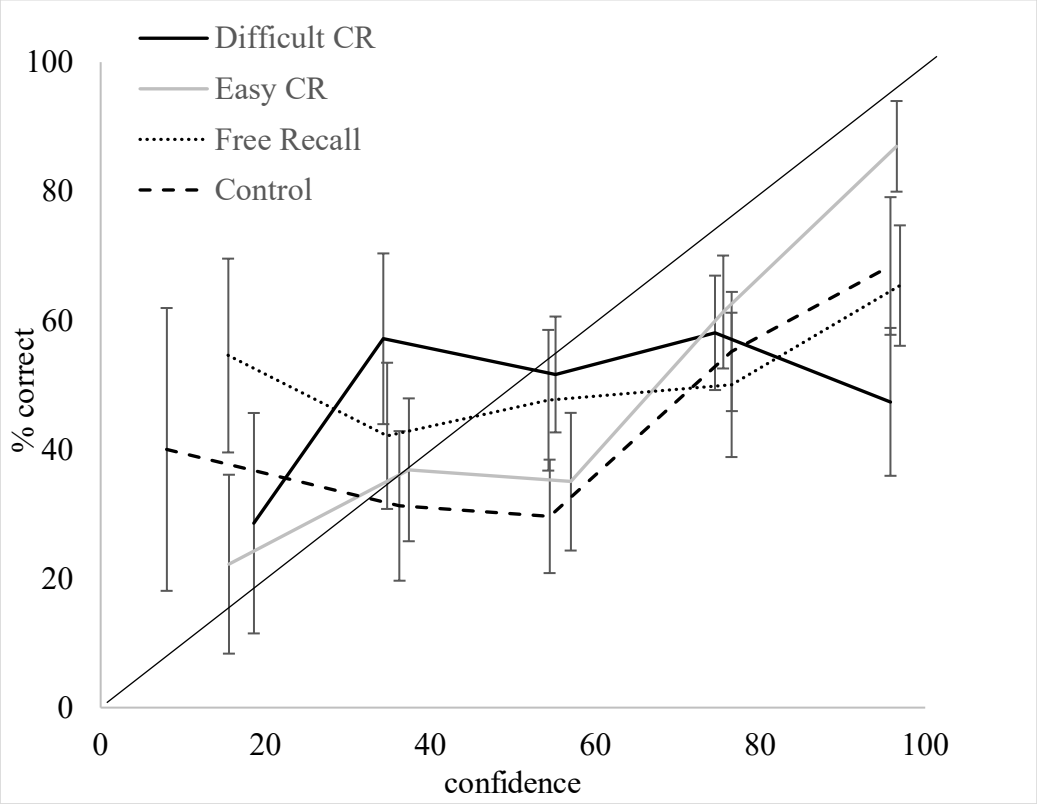


Figure 6.2. Calibration curves for the Difficult Cued Recall, Easy Cued Recall, Free Recall, and Control group in Study 3a.

Discussion

Study 3a investigated (i) the cognitive process that underpin decreased confidence following an initial recall attempt, and (ii) the impact of decreased

confidence following an initial recall attempt on subsequent memory monitoring. We speculated that confidence would *only* decrease when during the retrieval process participants are likely to experience retrieval difficulty. As such, we expected only participants in the Difficult Cued Recall group to report decreased confidence following the initial recall test. Furthermore, we expected decreased confidence to affect participants' calibration and to lead them to systematically associate lower confidence judgements to the information subsequently reported. Our results only partially support our hypotheses.

In particular, as predicted participants' confidence decreased only in the Difficult Cued Recall group, presumably because during the initial recall test participants in this group were consistently required to report information difficult to recall. This experienced difficulty in answering the questions of the initial recall test might have led participants to believe that their memory was poorer than they initially thought. However, contrary to participants in the Difficult Cued Recall group, participants in the Easy Cued Recall group were more likely to easily recall the details triggered by the questions and were unlikely to engage in an effortful retrieval search. Thus, their confidence in the quality of the target memory remained stable. Similarly, participants in the Free Recall group were unlikely to experience retrieval difficulty, presumably because they only focused their attention on the details they knew and that were easily accessible, and did not necessarily attempt to retrieve information they did not know or that were difficult to recall. Thus, they did not report decreased confidence after the initial recall test.

Overall, these results are in line with and extend the results of Study 2. Like in Study 2, in Study 3a we found evidence that a Free Recall test is less likely to cause significant shifts in confidence compared with a Cued Recall test. However, we found that the Cued Recall questions are likely to cause decreased confidence only when they trigger information that are difficult to remember. In other words, participants are likely to decrease their confidence when they experience a degree of retrieval difficulty during the initial test. As such, it is not necessarily the type of question that causes the decrease in confidence following an initial recall attempt, but rather the difficulty with which the to-be-remembered detail is retrieved. Consequently, when the Cued Recall question is easy or triggers details that are known and easily retrieved, participants' confidence is not likely to change. This result is important and provides initial evidence that directed retrieval can leave

confidence unaltered in so far as it directs participants' attention towards information that are not experienced as difficult to recall.

We found no support for our second hypothesis. As in Study 2 decreased confidence following an initial recall test did not lead participants to report poorer accounts in a second Free Recall test. However, contrary to our expectation decreased confidence (i) did not lead to under-confidence - as showed by the lack of difference between groups in Over/Under-confidence index, and (ii) did not affect how participants monitored the information they reported subsequently - as shown by the lack of difference between groups in the Calibration index and ANDI. Overall, the calibration indexes across groups ranged between .09 and .16 and were relatively close to 0 - representing perfect calibration. While the over/under-confidence indexes ranged from .06 to .11 showing that participants across groups were only slightly overconfident. These results are in line with previous research and overall, they suggest that participants are relatively well calibrated when judging the quality of their own memory (e.g., Luna & Martin-Luengo, 2012; Pansky et al., 2005).

In Study 3a we found no evidence that decreased confidence following an initial recall test influences subsequent calibration. However, we found initial evidence that decreased confidence might persist as an overall confidence judgement on the quality of one's own memory. In particular, we collected confidence ratings after the second Free Recall test. In this time point, despite participants across groups receiving the same Free Recall test, participants in the Difficult Cued Recall group remained significantly less confident than participants in the Free Recall and Control group, and substantially less confident than those in the Easy Cued Recall group (although this last comparison only approached statistical significance). Taken together, these results show that decreased confidence following a difficult initial recall test might not impact the monitoring and the quality of the information reported subsequently, however decreased confidence can persist as an overall judgement on the quality of own memory. This result is important for the legal context, because participants who report lowered confidence are also likely to report lowered confidence in engaging with the CJS and its stages, as shown in Study 2. Furthermore, evidence exists that lowered confidence in the quality of one's own memory might increase susceptibility to post-event suggestions (Jaeger et al., 2012; Leippe et al., 2006), thus undermining the reliability of the information that the eyewitness might report in future recall attempts.

Despite these results Study 3a is not without limitations. In particular, the calibration analysis in this study was performed on a relatively small data set and therefore our results might not be sufficiently reliable. Researchers have argued that the calibration analysis requires a large set of data. For example, Juslin et al. (1996) suggest analysing at least 200 data points for each confidence category. This can be achieved by either increasing the number of participants or by increasing the number of confidence judgements collected (via increasing the number of questions asked). In Study 3a, sixty-five participants answered eight questions in the second Cued Recall test, and thus our data set did not reach the sufficient amount of data points[4]. We therefore believe that the results yielded by the calibration analysis in this study are indicative but not sufficiently reliable. Furthermore, the limited sample adopted in Study 3a raises the issue of the reliability of the remaining findings. In recent years, the replication crisis has highlighted the need to adopt more rigorous research practice (Earp & Trafimow, 2015), such as increasing the power of the studies by testing larger samples. In an attempt to address these limitations in Study 3a we reported a Bayesian analysis for the non-significant results in order to better quantify the evidence in favour of the null hypothesis. We found moderate to strong evidence for our non-significant results and believe that these can be considered an additional evidence that our results are due to a true null rather than a lack of statistical power. Nevertheless, in order to adhere with good practice in research in the following studies larger samples are adopted.

In conclusion, in Study 3a we investigated (i) the cognitive processes that underpin the decrease in confidence following an initial recall test, and (ii) the effects of decreased confidence on participants' monitoring process. We found evidence that confidence decreases only when participants are likely to experience retrieval difficulty during an initial recall test, on the contrary when participants are unlikely to experience retrieval difficulty confidence is likely to remain stable. However, decreased confidence following an initial recall test is unlikely to lead participants to report poorer accounts in a subsequent recall test, and to affect their monitoring process by leading to under-confidence. However, an exploratory analysis provided initial evidence that decreased confidence following an initial recall test might persists as an overall confidence judgement in the quality of own memory. We have

[4] In Study 3a the data points for each confidence category ranged from 19 to 45.

also highlighted that Study 3a presents a series of important limitations that undermine the reliability of these findings, thus Study 3b was designed to replicate and expand the findings of Study 3a.

Study 3b

Study 3b was conducted primarily to address the limitations of Study 3a, and to increase the size of the sample, and the number of questions asked in the second Neutral Cued Recall test. Furthermore, we intended to address the H_2 of Study 3a with a more appropriate methodology. In particular, in Study 3a we speculated that decreased confidence following an initial recall test (i) would not lead participants to report poorer quality information in a subsequent recall test, and (ii) it would affect their monitoring process, leading them to associate lowered confidence judgements to the information reported subsequently. While the second part of the hypothesis investigated how participants monitor the information subsequently reported, the first part relates to how participants control this information.

In Study 3a we investigated participants' control over memory reporting with a Free Recall test. While this methodology enabled us to gauge the impact of decreased confidence on a memory freely reported subsequently, it might not be ideal to measure the impact that decreased confidence has on the control process. As discussed in Chapter 2, previous research investigating participants' control over the information reported has often adopted the Quantity - Accuracy Profile (QAP) methodology (Koriat & Goldsmiths, 1996). This method allows researchers to better understand how the confidence judgements associated with the details retrieved is used to decide the details selected to be reported and those selected to be withheld.

Thus, in Study 3b we aimed to replicate the results relating to the impact of retrieval difficulty experienced during the initial recall test on memory confidence, and to investigate the impact of decreased confidence on the monitoring and *control* of the information subsequently reported. The procedure was similar to that of Study 3a. Participants were asked to recall the event seen on two occasions. At Time 1 they were given either a Difficult Cued Recall, and Easy Cued Recall, a Free Recall initial recall test, a fourth group did not recall the event at this stage (Control group). After a short delay, all participants received a second Neutral Cued Recall test. Confidence ratings were collected before and after the initial recall test, and after each question of the Neutral Cued Recall test, the Difficult Cued Recall, and the Easy Cued Recall initial recall test. We maintained the hypotheses identical to those of Study 3a:

H_1: Confidence after an initial recall attempt decreases only when participants experience retrieval difficulty.

Prediction: We expect confidence to decrease only for participants in the Difficult Cued Recall group.

H_2: Decreased confidence reported by participants in the Difficult Cued Recall group will have an impact on subsequent memory monitoring.
Prediction: We predict that participants in the Difficult Cued Recall group will show under-confidence in the information subsequently reported.

H_3: Decreased confidence does no impact participants' control process.
Predictions: We expect participants in the Difficult Cued Recall group not to exert harsher control over their memory reporting.

Method

Design

In Study 3b, we used a 4 (Initial Recall test: Difficult Cued Recall, Easy Cued Recall, Free Recall, and Control) x 2 (Confidence measurement time: Time 1 - before Initial Recall, Time 2 - after Initial Recall) mixed design, with confidence measured within subjects. Dependent variables for memory monitoring were; (a) C-index, (b) ANDI, and (c) over/under-confidence measured in the Second Recall test. The dependent variable for memory control was (d) volunteering rate.

Participants

Participants were recruited online by using the platform Testable; the final sample included 169 participants (76 females, aged ranged from 18 to 56 years, M = 33.71 SD = 8.70).

Materials

Video Stimulus

The crime-video was as described for Study 1. While the Difficult Cued Recall, Easy Cued Recall, Free Recall, and Control tests and their instructions were exactly as described for Study 3a.

Time 2: Second Neutral Cued Recall Test

In the second recall test, all participants answered 42 Neutral Cued Recall questions (Appendix H). In order to identify the neutral questions, we ran a pilot study where participants (n = 10) watched the stimulus video and answered 111 questions about it. We instructed them to only answer the questions they were sure about, and to answer "I don't know" to the questions they did not know the answer to. The neutral questions selected (e.g., What colour was the bag used for the robbery?) were those with a mean correct response percentage rate of between 30 and 70 (total M = 44.29, SD = 12.71).

Procedure

The procedure was identical to that of Study 3a, with only three differences. First, data collection was conducted online rather than in the lab. Second, as in Study 3a participants completed the neutral Cued Recall test, however in Study 3b participants were instructed to provide an answer to all questions even if they were not sure. After each answer participants were asked to rate how confidence they were that the answer given was correct; we refer to this as the Forced-Report phase. Third, after the Forced-Report phase, participants were asked to imagine that they were real witnesses of the robbery they had seen in the video, and were reminded that the information they were going to provide was extremely important for the investigation. After this, all the participants were presented with the second Neutral Cued Recall test and their answers and were asked to (a) volunteer the answers that they were sure about, and (b) withdraw any answer that they were not sure about (confidence ratings were not displayed at this stage); we refer to this as the Free-Report phase[5]. As for Study 3a, confidence was measured before and after the initial recall test. Instructions and confidence scale were identical to those used in Study 3a.

Results

Manipulation check

[5] On occasion, and despite instructions, some participants had reported 'I don't know' in response to questions in the Neutral Cued Recall test. Thus, a third response option ('Not Applicable') was provided in the Free-Report phase for participants to select instead of 'volunteer' or 'withdrawn'.

A manipulation check was performed to verify that participants in the Difficult Cued Recall group were more likely to experience retrieval difficulty during the initial recall test than participants in the Easy Cued Recall group. As in Study 3a, we looked at the amount of "I don't know" responses reported, and found that participants in the Difficult Cued Recall group reported more "I don't know" responses (M = 5.60, SD = 3.32) than those in the Easy Cued Recall group (M = 2.70, SD = 2.43) and this difference was significant $t(83) = 4.59$, $p < .001$, $d = .50$, ($M_{diff} = 2.89$, 95% CI [1.63, 4.145). In addition, and consistently with Study 3a, confidence in the answers to Difficult Cued Recall questions was lower (M = 41.90, SD = 16.63) than confidence in the answers given in the Easy Cued Recall questions (M = 64.68, SD = 19.58); this difference was also statistically significant $t(83) = -5.72$, $p < .001$, $d = -.64$, ($M_{diff} = -22.77$, 95% CI [-30.62, -14.92]).

Confidence stability

To examine if we replicated the decrease in confidence after the initial recall test in the Difficult Cued Recall group, we conducted a 2 (Confidence measurement time: Time 1 - before Initial Recall, Time 2 - after Initial Recall, after Second Recall) x 4 (Initial Recall test: Difficult Cued Recall, Easy Cued Recall, Free Recall, and Control) mixed ANOVA. We found a significant main effect of Time on confidence, $F(1, 165) = 21.02$, $p < .001$, $\eta^2_p = .13$, a significant main effect of Initial Recall test on confidence, $F(1, 165) = 6.47$, $p < .001$, $\eta^2_p = .08$, and a significant interaction $F(3, 165) = 17.53$, $p < .001$, $\eta^2_p = .24$ (see Table 6.4 for Means and Standard Deviations in each group). A post-hoc power analysis conducted with G*power (Erdfelder et al., 1996), with an alpha level of .05, $n = 169$, and observed effect sizes, yielded a power of .98 to 1. Bonferroni post-hoc tests showed that confidence was not significantly different between groups before at Time 1. However, at Time 2, participants in the Difficult Cued Recall group were significantly less confident than those in the Easy Cued Recall ($M_{diff} = -20.78$, $p < .001$, 95% CI [-33.79, -7.76]), Free Recall ($M_{diff} = -30.71$, $p < .001$, 95% CI [-43.80, -17.62]), and Control group ($M_{diff} = -14.04$, $p = .028$, 95% CI [-27.14, -.97]). Participants in the Difficult Cued Recall group also reported a significant decrease in confidence at Time 2 compared to Time 1 ($M_{diff} = -22.14$, $p < .001$, 95% CI [-28.54, -15.74]). Whereas, confidence at Time 2 compared to Time 1 did not change

significantly for participants in the Easy Cued Recall, Free Recall, and Control group. Thus, consistently with the result of Study 3a, only participants in the Difficult Cued Recall group reported decreased confidence after the initial recall attempt.

Table 6.4. Means (and Standard Deviations) of confidence reported at Time 1 and Time 2 in the four groups.

	Time 1	Time 2
Difficult Cued Recall	58.57 (20.19)	36.43 (23.14)
Easy Cued Recall	60.70 (21.53)	57.21 (21.30)
Free recall	64.05 (23.48)	67.14 (22.44)
Control	52.14 (22.91)	50.48 (24.94)

Monitoring of memory in the second Neutral Cued Recall Test

We were interested in understanding whether decreased confidence following an initial recall attempt affected subsequent memory monitoring. In order to increase our data in each confidence category we reduced the number of categories from 11 to five (0-20%, 30-40%, 50-60%, 70-80% and 90-100%)[6]. The confidence category 0% registered the least data points and was therefore merged with the category 10-20%. Three one-way ANOVAs were conducted on the C-index, ANDI, and over/under-confidence index. We found no significant differences between groups in the C-index $F(3, 165) = 1.05$, $p = .37$, $\eta^2 = .01$, or in relation to over/under-confidence $F(3, 165) = 1.73$, $p = .16$, $\eta^2 = .03$. Furthermore, no difference were found between groups in the ANDI $F(3, 165) = .56$, $p = .64$, $\eta^2 = .01$ (see Table 6.5 for Means and Standard Deviations, and Figure 6.3 for the confidence-accuracy calibration curves). Bayesian analysis showed moderate to strong evidence in support for the null for C-index ($BF_{01} = 9.34$), ANDI ($BF_{01} = 16.78$), and OU index ($BF_{01} = 4.18$).

Finally, in order to investigate how well calibrated participants were, we compared the C-index and OU index to 0 (perfect calibration), and ANDI to 1

[6] Data points per confidence categories ranged from 131 to 659, the category 30-40% in the Difficult Cued Recall, Easy Cued Recall, and Free Recall group registering less data points than 200.

(perfect discrimination). We found that in all groups C-index was significantly different from 0, [$t(42) = 6.58$, $p < .001$, $d = 1.05$ for the Difficult Cued Recall; $t(43) = 12.20$, $p < .001$, $d = 1.84$ for the Easy Cued Recall; $t(42) = 5.50$, $p < .001$, $d = .84$ for the Free Recall; and $t(42) = 10.58$, $p < .001$, $d = 1.63$ for the Control group], but OU measures were not [$t(42) = 1.68$, $p = .10$, $d = .03$ for the Difficult Cued Recall, $t(43) = .91$, $p = .36$, $d = .13$ for the Easy Cued Recall, $t(42) = .41$, $p = 68$, $d = .06$ for the Free Recall, and $t(42) = -1.47$, $p < .15$, $d = -.23$ for the Control group]. Furthermore ANDI in all groups was significantly different from 1 [$t(42) = -22.01$, $p < .001$, $d = -3.3$ for the Difficult Cued Recall; $t(43) = -22.19$, $p < .001$, $d = -3.3$ for the Easy Cued Recall; $t(42) = -23.06$, $p < .001$, $d = -3.5$ for the Free Recall; and $t(42) = -18.02$, $p < .001$, $d = -2.81$ for the Control group].

In sum, confirming the results of Study 3a, participants in the Difficult Cued Recall group compared with participants in the other groups: (i) were not less calibrated (as measured by the C-index), (ii) were not systematically associating lower confidence judgements to the information reported (as measured by the OU index), and (iii) were not less effective in using their confidence judgements to discriminate between correct and incorrect information (as measured by the ANDI). Furthermore, overall calibration and discrimination for participants across groups were significantly different from perfect.

Table 6.5. Means (and Standard Deviations) of C- index, ANDI, and over/under-confidence measure in the four groups.

	Difficult Cued Recall	Easy Cued Recall	Free Recall	Control
C-index	.09 (.09)	.09 (.05)	.12 (.05)	.09 (.09)
ANDI	.25 (.22)	.22 (.23)	.24 (.21)	.28 (.25)
OU measure	.05 (.19)	.03 (.21)	.01 (.18)	-.04 (.18)

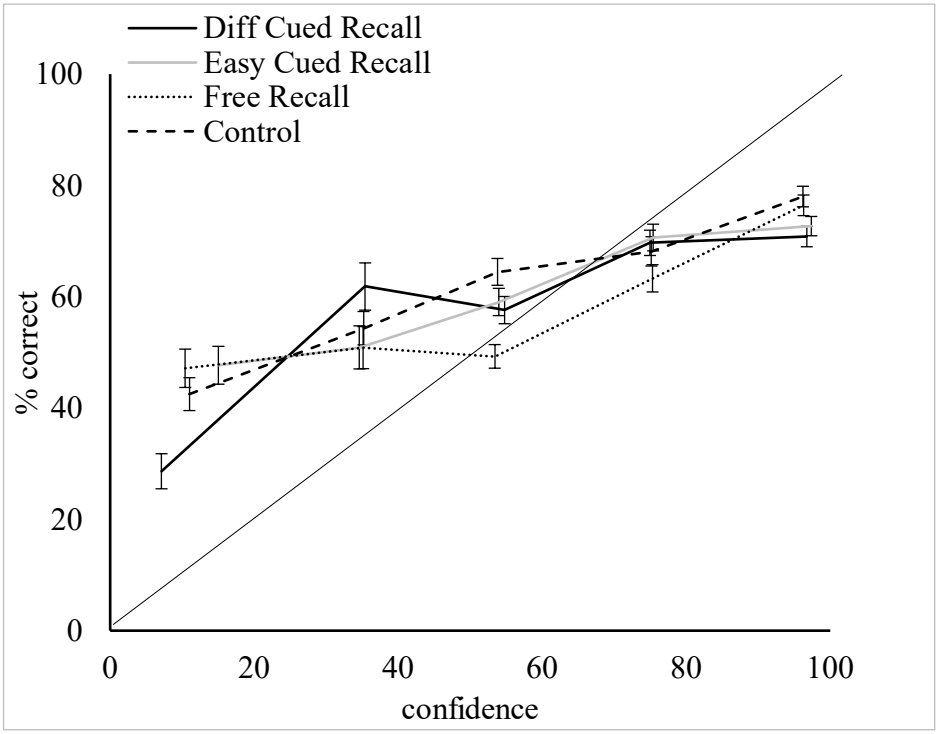


Figure 6.3. Calibration curves for the Difficult Cued Recall, Easy Cued Recall, Free Recall, and Control group in Study 3b

Control of memory reporting during the second Neutral Cued Recall Test

Four participants were excluded from this analysis because they answered "I don't know" to over half of the 42 questions. First, we assessed memory performance during the second Neutral Cued Recall test. A one-way ANOVA found that there was no significant difference between groups on overall accuracy $F(3, 161) = .55$, $p = .64$, $\eta^2 = .01$ in the Forced-Report phase. To test whether participants in the Difficult Cued Recall group were exerting stricter control over the information reported by volunteering fewer items in the Free-Report phase, we coded each answer as "volunteered" (V) or "withdrawn" (W), and excluded the "I don't know" answers[7] and a further 244 answers (3.5% of the total) across the whole sample, because they were associated with the option "non-applicable". At this point there was no difference between groups in relation to the number of answers included $F(3, 161) = 2.33$, $p = .08$, $\eta^2 = .04$. A volunteering rate was calculated for each participant

[7] A one-way ANOVA found no significant difference between groups on the number of "I don't know" answers in the Forced-Report phase, $F(3,161) = 2.27$, $p = .08$, $\eta^2 = .04$.

by dividing the number of volunteered answers by the total (correct and incorrect) number of answers given. A one-way ANOVA showed that there was no difference between groups on the volunteering rate $F(3, 161) = .65$, $p = .59$, $\eta^2 = .01$. A Bayesian analysis confirmed we have moderate evidence for the null regarding volunteering rate ($BF_{01} = 8.425$). In sum, confirming Study 3a, participants in the Difficult Cued Recall group were not exerting harsher control over their memory.

Discussion

Study 3b was conducted to further test the hypotheses of Study 3a and it featured: (i) a larger sample size, a (ii) larger data set to conduct a reliable confidence-accuracy calibration analysis, and (iii) a more controlled methodology to investigate the effect of decreased confidence following an initial recall test on participants' control over the information subsequently reported. The results of Study 3b replicate those of Study 3a. In particular, participants' confidence is more likely to decrease after an initial recall test when this triggers details that are difficult to access. On the contrary when the initial test is unlikely to trigger details difficult to access, participants' confidence is likely to remain stable. Furthermore, decreased confidence following an initial recall test does not impact subsequent memory monitoring and control over the information subsequently reported.

While we evaluated the first set of results in the discussion of Study 3a, in this section we focus our attention on the results that decreased confidence does not appear to affect monitoring and control over the information subsequently reported. These results are not necessarily in contrast with the largely accepted theoretical framework of memory regulation (Koriat & Goldsmith, 199b). First, our results are consistent; in particular, the lack of difference between groups in how the information was subsequently freely reported (Study 2, and Study 3a) and controlled (Study 3b), is due to the lack of significant difference between groups on how the memory was monitored (Study 3b). Koriat and Goldsmith (1996) suggest that in order to observe a difference in the way in which a memory is controlled, it is necessary to also observe a difference in how the memory is monitored. Thus, in our studies participants in the Difficult Cued Recall group compared with participants in the other groups did not show a difference in how they controlled their memory

reporting, because they were not monitoring their memory any differently compared to participants in the other groups.

Despite this consistency, it remains a question as to why we find evidence that decreased confidence after an initial recall attempt does not seem to impact subsequent memory monitoring and regulation. We consider two possible explanation for these results. First it is possible that confidence judgements used to monitor and control the information reported are primarily based on the retrieval experience generated during the *ongoing* test, and largely disregard any previous retrieval attempt. For example, it is possible that during the second recall test participants in the Difficult Cued Recall group based their confidence judgements on the ongoing retrieval experience, and disregarded the initial recall attempt, and their retrieval experience and the overall performance at that stage. This interpretation would suggest that confidence judgements might be more dependent on the ongoing retrieval process (i.e., experienced-based) rather than the previous belief that the own memory for the target event might be poor (i.e., information-based). Overall, this interpretation is in line with the existing evidence that confidence judgements are largely depend on the retrieval experience and its products (e.g., Kelley & Lindsay, 1993; Lindholm et al., 2018; Pansky & Goldsmith, 2014)

However, a second possible explanation is that our methodology is not capturing the impact of decreased confidence on the monitoring and control of the information subsequently reported. For example, in both Study 3a and 3b only three minutes elapsed between the initial and the second recall attempt, and it is possible that this time was insufficient to fully integrate the new belief as a new heuristic. Thus, although we found no evidence that decreased confidence impacted the subsequent monitoring and control over the information reported it is worth mentioning that this result is contingent to the short delay used in our studies.

In Study 3b we replicated the result that confidence is likely to decrease when participants are asked to recall details that are difficult to retrieve. On the contrary, when participants are not required to recall information that are difficult to retrieve their confidence is more likely to remain stable. This result is important and highlights a previously undocumented negative effect that Cued Recall questions can have on eyewitness cognition; in particular, this type of question is more likely to dent eyewitness' confidence. However, our results also suggest that closed questions do not always lead to shifts in confidence. Specifically, if they are easy or likely to

trigger details that are easy to remember, they are unlikely to cause any change in confidence. Closed questions can be a useful tool for the interviewer, and they can help eyewitnesses to focus on and report specific details of the event. As discussed in Chapter 1, researchers do not discourage the use of closed questions, rather they suggest using them at the right time during an interview (e.g., Gabbert et al., 2015a; 2015b; Griffiths & Milne, 2006). Therefore, it is important to understand whether and how Cued Recall questions can be posed without causing any change in memory confidence. However, the results gathered so far cannot be directly applicable to a real-life interview, because in real-life interviewers cannot foresee how easy or difficult a question is going to be for the eyewitness. Thus, in Study 4 we attempt to fill this gap by investigating how these findings can be implemented in a more realistic interview.

Chapter 7: Study 4. Investigating shifts in confidence in a good practice and a poor practice interview

Introduction

The results so far have highlighted two core findings: (i) confidence in the quality of one's own memory can decrease following a retrieval attempt, and (ii) decreased confidence following an initial recall attempt does not affect the quality of the details subsequently reported, nor does it impact participants' monitoring and control of the information reported subsequently. Here, we address the first result and examine whether it holds within more realistic types of investigative interview.

Study 2 showed that confidence is likely to change as a function of the quality of the initial interview. In particular, an interview that poses Cued Recall questions is more likely to lead to decreased confidence than an interview that presents a Free Recall invitation. However, a more in-depth examination of the cognitive processes that underpin this result suggests that it is not the type of question that lead to decreased confidence, rather the retrieval difficulty experienced in answering the different questions. As such, Study 3a and Study 3b confirmed that Cued Recall questions only cause decreased confidence in so far as they trigger details that are difficult to recall. However, when Cued Recall questions prompt details experienced as easy to recall, they are more likely to leave confidence unaltered. In the discussion of Study 3b, we argued that this result is important because it suggests that there might be circumstances in which Cued Recall questions can be asked without causing significant shifts in eyewitness confidence. However, in real-life interviews it is not easy to foresee when a Cued Recall question is likely to be experienced as easy or difficult. Thus, in Study 4, we address this limitation and examine how our previous results can be used to predict shifts in confidence within more realistic investigative interviews.

Rationale

The finding that confidence is more likely to decrease in response to Cued Recall questions compared with Free Recall questions suggests that during an interview confidence is likely to remain stable after the Free Recall phase and is likely to decrease after the probing phase (e.g., when Cued Recall questions are

asked). However, as shown by the results of Study 3a and Study 3b, Cued Recall questions only lead to decreased confidence in so far as they trigger details that are experienced as difficult to recall. Thus, it is possible that the probing phase promotes decreased confidence only when the probing Cued Recall questions are experienced as difficult to answer.

We speculated that confidence following a Free Recall invitation is unlikely to decrease, because during a free and undirected retrieval experience, participants tend to only recall details relatively easy to access. This result suggests that the freely reported details are more likely to be experienced as easy to retrieve, and conversely, the unreported details – should they be elicited via further questioning – are more likely be experienced as relatively difficult to retrieve. Hence, we speculate that following a free-recall phase, the probing Cued Recall questions are likely to be experienced as easy or difficult depending on whether the questions focus on topics that have already been reported by the witness during their Free Recall account. We therefore hypothesise that during an interview, confidence will decrease when the probing Cued Recall questions are *unrelated* to the information reported in the Free Recall phase, and will remain stable when the probing Cued Recall questions are *related* to the information freely reported. As such, when an interview follows the best practice guidelines (as outlined in Chapter 1) relating to; (i) the interview structure and hierarchical use of different types of question, whereby the Free Recall phase precedes the probing phase, and; (ii) the use of compatible questions, whereby the probing questions are compatible with the interviewee's free account, then we expect confidence to remain stable. On the contrary when the interview does not follow the best practice guidelines relating to interview structure and compatible use of questions, then we predict confidence to decrease significantly after the probing phase.

As discussed in Chapter 2, focused questions (i.e., Cued Recall questions) tend to promote poorer realism (i.e., poorer calibration and discrimination) compared with Free Report invitations. Researchers have proposed that this might depend on the difficulty of the task (Allwood et al., 2008). In particular, answering focused questions is more difficult than responding to a Free Recall invitation, because participants are required to answer questions formulated by others, and that might request information that had been poorly encoded. Based on this line of argument, we predict a difference in realism for easy Cued Recall questions (i.e., Cued Recall

questions related to the information freely reported), and difficult Cued Recall questions (i.e., Cued Recall questions unrelated to the information freely reported). Specifically, we predicted the related Cued Recall questions to promote better realism (i.e., better calibration and discrimination) compared with the unrelated Cued Recall questions.

To test these hypotheses, we manipulated the order of the interview phases, and the types of probing Cued Recall questions posed. Participants first watched a short crime video and were then interviewed about what they had seen. The probing phase either preceded or followed the Free Recall phase. Of the participants that received the Free Recall phase before the probing phase, half answered the Cued Recall questions related to the information they had reported in the Free Recall phase (referred to subsequently as Related Cued Recall questions) followed by the Cued Recall questions unrelated to the information reported in the Free Recall phase (referred to subsequently as Unrelated Cued Recall questions), while the remaining half answered the Unrelated Cued Recall questions *prior to* the Related Cued Recall questions. Confidence was measured after the Free Recall phase and after each section of the probing phase; following the Unrelated Cued Recall questions and the Related Cued Recall questions. Last, participants reported a confidence rating for each probing Cued Recall questions (the procedure is illustrated in Figure 7.1).

Hypotheses

H_1 and prediction: During an investigative interview, initial confidence will remain stable following the Free Recall phase and will decrease after the probing phase.

H_2: Cued Recall questions *related* to the information freely reported are more likely to trigger details that are easily retrieved, therefore they are less likely to lead to decreased confidence; on the contrary Cued Recall questions *unrelated* to the information freely reported are more likely to trigger details experienced as difficult to retrieve and therefore more likely to decrease confidence.
Predictions: confidence will decrease after answering unrelated Cued Recall questions and remain stable after answering related Cued Recall questions.

H_3: Cued Recall questions unrelated to the information freely reported are expected to be more difficult and consequently to promote poorer realism compared with Cued Recall questions related to the information freely reported.
Predictions: Participants will be relatively poorly calibrated (indicated by C- index, OU index, and ANDI) when answering unrelated Cued Recall questions compared with related Cued Recall questions.

Method

Design

A 3 (Interview Type: mixed CR +FR; FR + related CR first, FR+ unrelated CR first) x 4 (Confidence Measurement Time: initial confidence (Time 1), confidence at Time 2, confidence at Time 3, and confidence at Time 4) mixed design was used, with Interview Type manipulated between-subjects, and confidence measured within-subjects. The dependent variables for memory reporting were: (a) number of correct details reported, and (b) accuracy of the information reported. The dependent variables for confidence in the memory for the target event were: (c) initial confidence at Time 1, (d) confidence at Time 2, (e) confidence at Time 3, and (f) confidence at Time 4. Finally, the dependent variables for calibration were: (g) C-index, (h) ANDI, and (e) OU measures.

Participants

An a-priori power analysis conducted in G*power (Erdfelder et al., 1996) revealed that n = 99 participants were required to detect a medium effect size η^2_p = .06 with a power (1 – β) set at .80, and α = .05. In order to gather sufficient data points for the calibration analysis the number of participants tested was increased slightly. A total of 125 participants took part in the study (females = 100, Mean age = 21.85, SD = 5.05). One participant was excluded because - due to a researcher's mistake they were not given one task of the procedure (they were not asked to provide confidence ratings for the details of the Free Recall test). All participants were recruited among the first-year cohort Psychology students at Goldsmiths University. They all received course credits as compensation for their time.

Material

Video stimulus

The video clip depicting a robbery in a Blockbuster video-rental store was the same as described in Study 1.

The Interview phases

Three interviews were used: the mixed CR+FR, the FR + Related CR first, and the FR + Unrelated CR first. They all constituted of two phases: a Free Recall, and Probing phase.

Free Recall phase. The Free Recall phase was administered as a written Free Recall test. Participants were instructed to imagine the researcher knew nothing about the video and to write down what they remembered about it. In order to allow for free control over their reporting, no restrictions were imposed in relation to requests for accuracy or detail, and no time limit was given for this task.

Probing phase. The probing phase was administered verbally and constituted of 89 predetermined 5Wh Cued Recall questions (Appendix I). The questions focused on the four characters in the video (i.e. dark-haired robber, light-haired robber, customer, and shopkeeper) including their person description and actions, the bag used for the robbery and the location in which the robbery took place. Participants in the three groups answered all the questions. The instruction was identical for the three groups, all participants were asked to always answer the question even when not sure about their answer. All the participants consented to be audio recorded, therefore the predetermined Cued Recall questions were recorded on a smartphone.

The order of the Interview phases

The order of the interview phases was different depending on condition;

The mixed CR + FR. In this interview, the probing phase preceded the Free Recall phase. In the probing phase of this interview, the predetermined Cued Recall questions were administered in random order (we refer to this as mixed Cued Recall).

The FR + Related CR first. In this interview the Free Recall phase preceded the probing phase. Here, the predetermined Cued Recall questions were administered as follows: participants received first the questions related to the information they

reported in the Free Recall (we refer to these as Related Cued Recall) followed by the remaining questions (we refer to these as Unrelated Cued Recall).

The FR + Unrelated CR first. In this interview the Free Recall phase preceded the probing phase. The predetermined Cued Recall questions were administered in the opposite order compared to that described above for the FR + Related CR first. In particular, participants received first the questions unrelated to the information they reported in the Free Recall (i.e., Unrelated Cued Recall), followed by the remaining questions (i.e., Related Cued Recall).

Confidence

Confidence in the memory for the event seen was measured on four occasions throughout the interview: after the video (initial confidence at Time 1), after each interview phases, and after each Cued Recall type. The confidence ratings were collected through a written questionnaire and the instruction was the same on each occasion. Here, participants were asked to think about their memory for the event seen, and to report how confident they were that their memory and the information they were able to report would help the police to accurately reconstruct the event and solve the crime.

A confidence judgement was also required after each predetermined Cued Recall questions, and for each detail reported in the Free Recall test. All the confidence scales ranged from 0% to 100%, whereby 0% represented "not confident at all", and 100% represented "completely confident".

Procedure

Participants were tested individually in a quiet laboratory in the university campus. The experiment took between 60 and 75 minutes and was completed in one session. As soon as the participant arrived in the lab, they were invited to take a seat in front of a computer and were provided with the consent form (Appendix J). If the participant agreed to participate the researcher checked that the procedure was clear, and invited questions. After this, the researcher placed the confidence scale on the participant's desk (the scale was left on the desk and was available to the participant during the entire procedure) and explained what it was, and how to use it. They were told that 0% on the scale represented *"not confident at all that the answer given is*

correct", while 100% meant *"complete confidence that the answer is correct".* The following example was used to clarify the use of the scale: *"If I was asked what is the capital of the UK, and my answer was London – I would be 100% confident that my answer is correct. However, if I was asked, what is the capital of Suriname, and my answer was London – then I would be 0% confident that my answer is correct".* After this, participants were given a set of headphones and asked to watch the video clip, followed by a three minutes word-search filler task. Immediately after the filler task participants were asked to provide a confidence rating (initial confidence at Time 1). At this stage participants were reminded that the confidence rating they were asked to report was related to their confidence in their memory of the video not about their performance on the word-search task. Up until this stage the procedure was identical for all participants; afterwards the procedure was different for each group (see Figure 6.1)

Participants in the mixed CR+FR group received the probing phase followed by the Free Recall phase. Immediately after they had answered all the predetermined Cued Recall questions, they provided a second confidence rating (confidence at Time 2). After this task, participants were asked to complete the written Free Recall test, followed by the seven - minute word-search filler task. In the time participants took to complete the filler task, the researcher coded the written Free Recall test for amount and accuracy. Here each coded detail was underlined with a blue pen. Once the participant had completed the filler task, they were asked to provide a confidence rating for each detail underlined in blue on the Free Recall account they had previously provided. The instruction for this task was as follows: "*I would need you to report a confidence rating for each piece on information you provided in the Free Recall test. I am going to indicate pieces of information highlighted in blue and ask you the question "How confident are you that this information is correct?" You should answer by giving a confidence rating – please do this using the scale you have in front of you."* In the mixed CR + FR group confidence was measured on three occasions: after the video (initial confidence at Time 1), after the mixed Cued Recall questions (confidence at Time 3) and after the Free Recall test (confidence at Time 4).

Participants in the FR + Related CR first, were given the Free Recall phase followed by the probing phase. The written Free Recall test was administered on a paper sheet placed on top of transfer paper which copied the participant's free report

on a second paper sheet. Once the participants had completed the Free Recall test, they were given the seven minutes filler task. While the participant completed the filler task the researcher coded the Free Recall test for amount of information (here details were highlighted in blue), and two research assistants worked on the second copy of the Free Recall to identify the Related Cued Recall questions out of the pool of predetermined Cued Recall questions. Here, a question was considered *related* if it was associated to a detail reported in the Free Recall test. For example, if in the Free Recall the participant mentioned the "dark-haired robber", all the predetermined Cued Recall questions related to the dark-haired robber's person description were considered as Related Cued Recall questions (e.g., What was the dark-haired robber ethnicity? How old was the dark-haired robber?). Similarly, if in the Free Recall phase the participant mentioned the dark-haired robber's 'disguise', then all the predetermined Cued Recall questions related to the dark-haired robber's disguise were considered Related Cued Recall (e.g., What type of disguise was the dark-haired robber wearing?). At the end of this procedure all the remaining predetermined Cued Recall questions were considered *Unrelated* Cued Recall questions. Once participants had completed the filler task, they were asked to provide a confidence rating on the details reported in the Free Report (this procedure was identical as that described for the mixed CR + FR group above). Next, participants answered the predetermined Cued Recall question in this order: the Related Cued Recall questions were asked *before* the Unrelated Cued Recall questions. Confidence in the memory for the event seen was measured on four occasions for participants in the FR + Related CR first group: after the video (initial confidence at Time 1), after the Free Recall test (confidence at Time 2), after the Related Cued Recall questions (confidence at Time 3), and after the Unrelated Cued Recall questions (confidence at Time 4).

The procedure of the FR + Unrelated CR first was identical to that described for the FR + Related CR first with one difference. The predetermined Cued Recall questions were given in the opposite order. Participants in this group answered the *Unrelated* CR questions first and the *Related* CR question after. Confidence in the memory for the event seen was measured on four occasions: after the video (initial confidence at Time 1), after the Free Recall test (confidence at Time 2), after the unrelated CR questions (confidence at Time 3), and after the related CR questions (confidence at Time 4).

Finally, participants in all groups received the written debrief form (Appendix K) and were invited to ask any question they had.

Coding

The Free Recall tests were coded for amount and accuracy of information by following the same coding scheme described for Study 1.

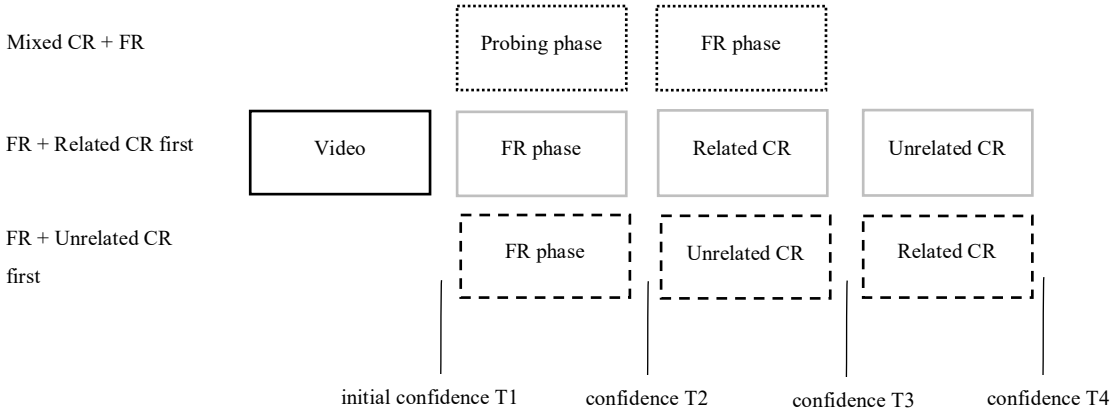


Figure 7.1. Procedure of Study 4

The Pilot Study

An initial Pilot Study (n = 5) highlighted one issue with the procedure; when answering the predetermined Cued Recall questions, participants tended to also elaborate and explain their answer. This led to lengthy answers that included many additional details, some of which were correct and some that were incorrect. We believed this was problematic because participants were then asked to report a single confidence rating in the accuracy of an answer that now contained multiple details. In order to overcome this issue, we modified the instruction of the predetermined Cued Recall questions and encouraged participants to (a) provide short answers, and (b) avoid elaboration. The revised instruction was as follows: "*The questions I am going to ask you are focused on specific details, and require a one or two word answer, i.e., you should be able to answer with one or two words. For this task it is important that you always answer the question, even if you are not sure about it. After each question I will ask you how confident you are in your answer, you should only answer by giving a confidence rating – please do this by using the scale that you*

have in front of you. It is important that you do not justify your answer or make any further comment during this task. Later I will be available to discuss any comment or thought you might have."

A second Pilot Study with the new instruction was conducted (n = 5) and no further issues were observed, therefore data collection commenced. The data from the second Pilot Study was included within the Study 4 sample.

Results

We used a conventional *alpha* level of .05 unless otherwise specified. Homogeneity of variance was assessed with Levene's test (1960).

Manipulation check

We examined mean confidence associated with the information reported in each of the three interview stages (Free Report, Related Cued Recall questions, and Unrelated Cued Recall questions). As predicted, overall confidence in the details reported in the Free Recall (M = 92.09; SD = 6.95) was higher than confidence in the information reported in response to the Unrelated Cued Recall questions (M = 41.19; SD = 17.51), and this difference was statistically significant $t(82) = 23.03$, $p < .001$, $d = 2.52$, $M_{diff} = 42.90$, 95% CI [39.20, 46.60]. Similarly, confidence in answers reported in response to the Related Cued Recall questions (M = 76.38; SD = 12.41) was significantly higher than confidence in answers to the Unrelated Cued Recall questions $t(82) = 18.51$, $p < .001$, $d = 2.03$, $M_{diff} = 27.19$, 95% CI [24.27, 30.11]. Overall, these results show that retrieval difficulty is higher in answers reported in response to Unrelated Cued Recall questions, compared to Related Cued Recall questions and Free Recall invitations.

Confidence stability

In order to investigate confidence stability across the interview phases we performed an initial 3 (Interview Type: mixed CR+FR; FR + Related CR first; and FR+ Unrelated CR first) x 3 (Confidence Measurement Time: initial confidence at Time 1, confidence at Time 2, and confidence at Time 4) mixed ANOVA, with Interview Type manipulated between subjects, and confidence measured within-subjects. In this analysis we did not include confidence at Time 3 because this was

only measured for participants in the FR + Related CR first and FR+ Unrelated CR first, but not for those in the mixed CR + FR group. The assumption of Sphericity was violated therefore we report the Greenhouse-Geisser estimates. There was a significant main effect of Time on confidence $F(1.87, 226.83) = 22.83$, $p < .001$, $\eta^2_p = .16$, a significant main effect of Interview Type $F(2, 121) = 4.05$, $p = .02$, $\eta^2_p = .06$, and a significant interaction $F(4, 226.83) = 11.51$, $p < .001$, $\eta^2_p = .16$. See Figure 7.2 for Means and Standard Deviations. Bonferroni post-hoc comparisons showed that there was no difference between groups in confidence reported at Time 1. However, at Time 2 confidence for participants in the mixed CR + FR was significantly lower than that of participants in the FR+ Related CR first ($M_{diff} = -22.48$, $p < .001$, 95% CI [-32.79, -12.18]), and FR+ Unrelated CR first ($M_{diff} = -18.29$, $p < .001$, 95% CI [-28.66, -7.92]). As expected, at Time 2 only participants in the mixed CR+FR reported significantly lower confidence compared to their baseline confidence ($M_{diff} = -17.31$, $p < .001$, 95% CI [-24.46, -10.16]). On the contrary, participants in the other two groups did not report significant changes in confidence at Time 2 compared to their confidence reported at Time 1. At Time 4 we found no difference between groups in confidence ratings. As expected participants in the FR+ Related CR first, and FR+ Unrelated CR first groups reported significantly decreased confidence compared to Time 2 (FR+ Related CR first $M_{diff} = -13.33$, $p < .001$, 95% CI [-19.88, -6.78], and FR+ Unrelated CR first $M_{diff} = -16.58$, $p < .001$, 95% CI [-23.21, -9.95]). Furthermore, participants in the mixed CR + FR group reported significantly increased confidence compared to Time 2 ($M_{diff} = 8.53$, $p = .007$, 95% CI [1.91, 15.16]), however, as expected for this group, confidence at Time 4 was not significantly different than initial confidence reported at Time 1.

In summary, in the mixed CR+FR group, confidence reported after the probing phase (Time 2) (but not confidence reported after the Free Recall phase - Time 4) was significantly lower than initial confidence (Time 1). Similarly, in the other two groups, confidence reported after the probing phase (Time 4) (but not confidence reported after the Free Recall phase – Time 2) was significantly lower than initial confidence reported at Time 1.

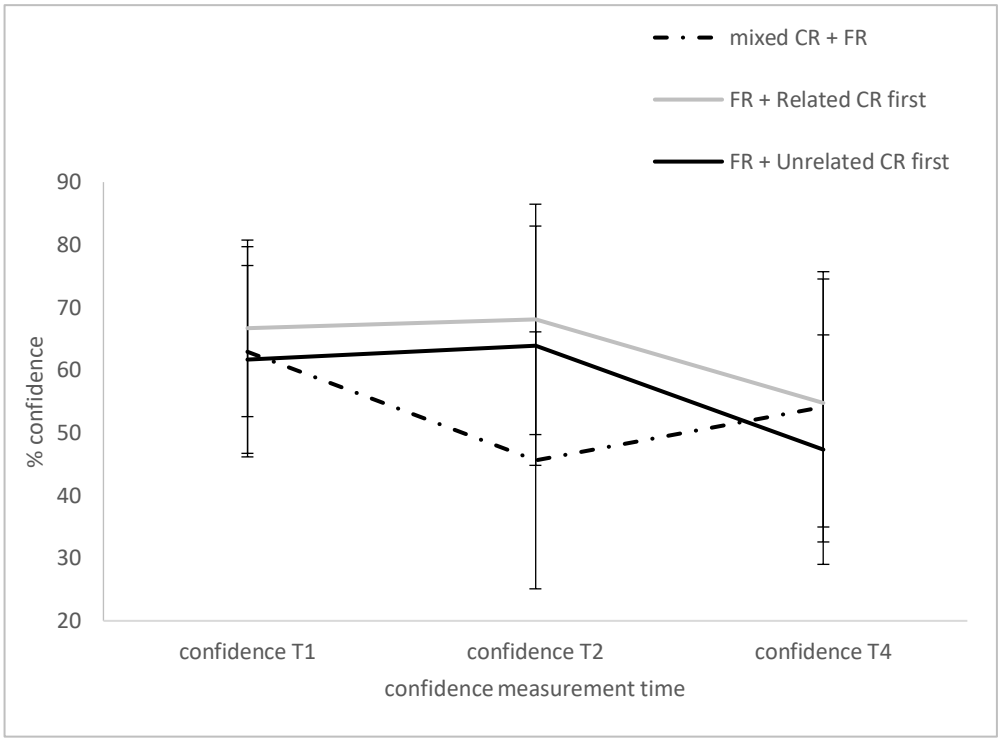


Figure 7.2. Means (and Standard Deviations) of confidence at Time 1, confidence at Time 2, and confidence at Time 4 for participants in the mixed CR +Fr, FR + Related CR first, and the FR + Unrelated CR first.

Confidence stability across Cued Recall type

In order to investigate the impact of Related Cued Recall and Unrelated Cued Recall questions on confidence in memory for the event seen. A 2 (Interview Type: FR + Related CR first; and FR+ Unrelated CR first) x 4 (Confidence Measurement Time: initial confidence Time 1, confidence at Time 2, confidence at Time 3, and confidence at Time 4) mixed design ANOVA was conducted. The assumption of Sphericity was violated therefore we report the Greenhouse-Geisser estimates. We found a significant main effect of Interview Type, $F(1, 81) = 13.21$, $p < .001$, $\eta^2_p = .14$, a significant main effect of Time on confidence $F(2.50, 203.09) = 26.22$, $p < .001$, $\eta^2_p = .25$, and a significant interaction $F(2.50, 203.09) = 14.03$, $p < .001$, $\eta^2_p = .15$ (see Figure 7.3 for Means and Standard Deviations). Bonferroni post-hoc comparisons showed no differences between the two groups on confidence reported at Time 1 ($M_{diff} = -4.95$, $p = .12$, 95% CI [-11.30, 1.39]), and confidence reported at Time 2 ($M_{diff} = -4.19$, $p = .31$, 95% CI [-12.39, 3.99]). However, a significant difference was found on confidence reported at Time 3; here confidence for participants answering the Unrelated Cued Recall questions was significantly lower

than confidence for those answering the Related Cued Recall questions (M_{diff} = -26.13, p < .001, 95% CI [-33.31, -18.95]). At Time 3 participants answering the Unrelated Cued Recall questions decreased their confidence significantly compared to confidence reported at Time 2 (M_{diff} = -21.46, p < .001, 95% CI [-29.66, -13.26]), while participants answering the Related CR questions did not report significant changes in confidence compared to Time 2 (M_{diff} = .47, p = 1, 95% CI [-7.62, 8.57]). Finally, at Time 4 no difference was found between groups on confidence ratings reported (M_{diff} = -7.44, p = .07, 95% CI [-15.77, .88]). Here, participants that answered the Unrelated Cued Recall questions reported significantly lowered confidence compared to Time 3 (M_{diff} = -13.81, p < .001, 95% CI [-19.89, -7.72], while those answering the Related Cued Recall questions increased their confidence compared to Time 3, however this increase was not significant (M_{diff} = 4.87, p = .21, 95% CI [-1.28, 11.03]).

In summary, confidence in both groups decreased following the Unrelated Cued Recall questions and did not change significantly following Related Cued Recall questions. We found support for our hypothesis that Related Cued Recall questions are less likely to decrease in confidence compared with Unrelated Cued Recall questions.

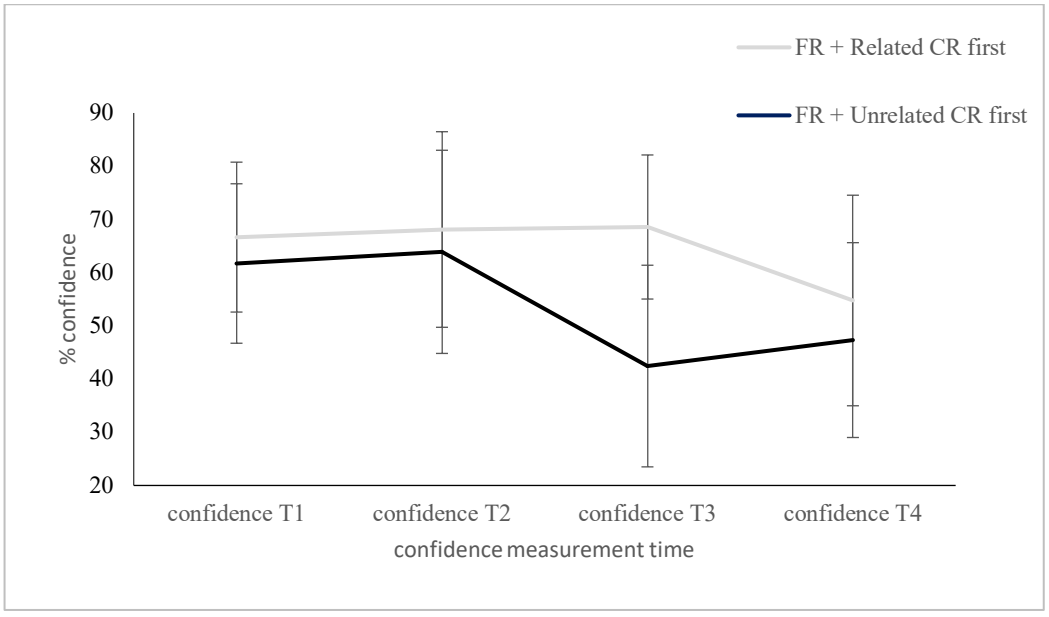


Figure 7.3. Means (and Standard Deviations) of confidence at Time 1, confidence at Time 2, confidence at Time 3, and confidence at Time 4 for participants in the FR + Related CR first, and the FR + Unrelated CR first

Calibration on Related CR and Unrelated CR questions

We were interested in examining if participants' confidence and accuracy were more calibrated in answers to the Related Cued Recall questions compared to the Unrelated Cued Recall questions. We performed a series of 2 (Cued Recall type: Related Cued Recall, and Unrelated Cued Recall) x 2 (Cued Recall order: Related CR first, Unrelated CR first) mixed design ANOVAs on (a) C-index, (b) ANDI, and (c) OU measures. We did not expect order to have a significant effect on the calibration and discrimination measured.

On the C-index we found a significant main effect of Cued Recall type $F(1, 81) = 20.96$, $p < .001$, $\eta^2_p = .20$, here independently of the order in which the questions were asked C-index was lower for the Related Cued Recall compared with the Unrelated Cued Recall questions ($M_{diff} = -.04$, $p < .001$, 95% CI [-.06, -.02]). We found non-significant main effect of Cued Recall order $F(1, 81) = .7$, $p = .37$, $\eta^2_p = .01$. However, we found a significant interaction $F(1, 81) = 6.41$, $p = .01$, $\eta^2_p = .07$. Bonferroni follow-up comparisons showed that the C-index for the Related Cued Recall questions was lower than that for the Unrelated Cued Recall questions ($M_{diff} = -.07$, $p < .001$, 95% CI [-.09, -.04]) in the Related CR first group *only*; meaning that calibration on the Related Cued Recall questions is better than calibration for the Unrelated Cued Recall questions *only* when the former are not preceded by the latter. Furthermore, the C-index for the Related Cued Recall questions was lower in the FR + Related CR first group compared with the Unrelated CR first group ($M_{diff} = -.03$, $p = .02$, 95% CI [-.06, -.01]), (see Table 7.1 for Means and Standard Deviations).

On ANDI, we found a significant main effect of Cued Recall type $F(1, 81) = 22.06$, $p < .001$, $\eta^2_p = .21$, meaning that independent of the order in which the questions were asked, discrimination was higher for the Related Cued Recall questions compared to the Unrelated Cued Recall questions ($M_{diff} = .17$, $p < .001$, 95% CI [.10, .24]). The main effect of Cued Recall order was not significant $F(1, 81) = 1.19$, $p = .27$, $\eta^2_p = .01$, nor was the interaction $F(1, 81) = .94$, $p = .33$, $\eta^2_p = .01$.

On the OU indexes we found a non-significant main effect of Cued Recall type $F(1, 81) = .11$, $p = .74$, $\eta^2_p = .00$, a significant main effect of Cued Recall order $F(1, 81) = 4.88$, $p = .03$ $\eta^2_p = .06$, meaning that across all the Cued Recall questions (Related Cued Recall and Unrelated Cued Recall) participants in the FR + Related CR first group compared with those in the FR + Unrelated CR first were less

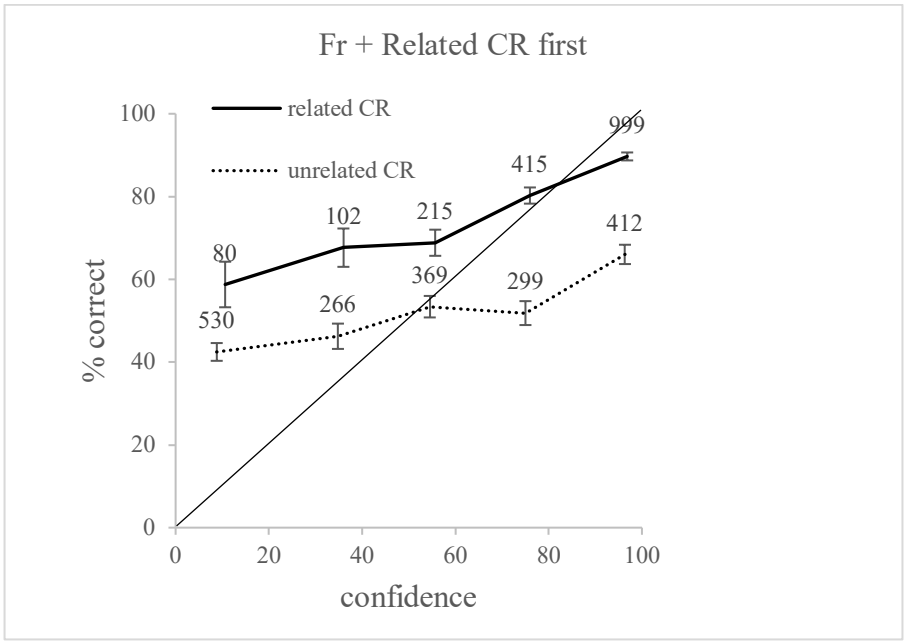
underconfident (M_{diff} = .07, p = .03, 95% CI [.01, .13]). Finally, the interaction was found to be non-significant $F(1, 81) = .14$, $p = .71$, $\eta^2_p = .00$.

Table 7.1. Means (and Standard Deviations) of C-index, ANDI, and OU measures for the answers to Related Cued Recall and Unrelated Cued Recall in the two groups.

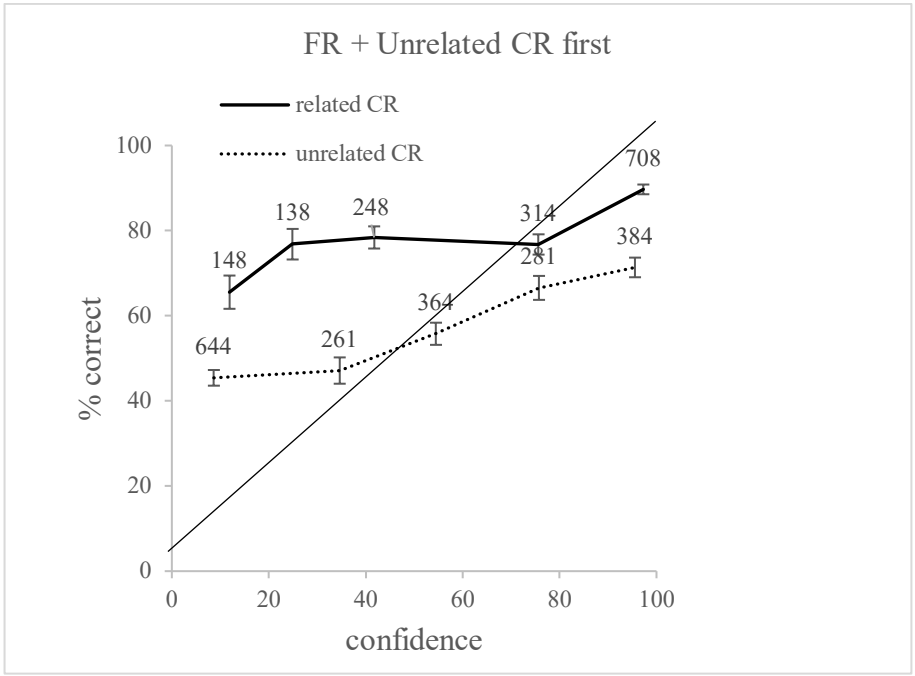
		FR + Related CR first	FR + Unrelated CR first
C-index	Related CR	.06 (.04)	.09 (.07)
	Unrelated CR	.13 (.07)	.11 (.04)
OU measure	Related CR	-.02 (.11)	-.09 (.17)
	Unrelated CR	-.01 (.19)	-.09 (.19)
ANDI	Related CR	.41 (.22)	.34 (.24)
	Unrelated CR	.20 (.25)	.20 (.21)

In summary, we found support for our Hypothesis 3, whereby regardless of the order in which the (Related and Unrelated) Cued Recall questions are asked, calibration and discrimination were more reliable for the Related Cued Recall questions compared with the Unrelated Cued Recall questions, as measured by (a) a lower C-index, and (b) a higher ANDI on the former compared with the latter. However, calibration for the Related Cued Recall is significantly more reliable than that for the Unrelated Cued Recall questions *only* if the former questions precede the latter, as shown by (a) the lower C-index for Related (compared to the Unrelated) Cued Recall questions in the FR + Related CR first group, and (b) the overall lower under-confidence for participants in the FR+ Related CR first compared with the FR+ Unrelated CR first group. On the contrary, if the Related Cued Recall questions are preceded by the Unrelated Cued Recall questions, calibration on the Related Cued Recall and Unrelated Cued Recall questions become statistically similar. This can also be observed in the calibration curves. In particular, the solid lines – representing calibration for the Related Cued Recall questions, are closer to the diagonal line in the FR + Related CR first group only (Panel a). Furthermore, only in the FR + Related CR first group (Panel a) is the solid line closer to the diagonal line than the dashed line – representing calibration on the unrelated CR questions. These patterns in the curves show that calibration for Related Cued Recall questions in

better than calibration for the Unrelated Cued Recall questions only when the formers precede the latter.



Panel a



Panel b

Figure 7.4. Calibration curves (and data points for each confidence interval) for the related CR and unrelated CR questions for participants in the FR + related CR &

unrelated CR (panel a), and for participants in the FR + unrelated CR & related CR (panel b).

Memory report

Although the impact on memory reporting was not the primary focus of this study, in order to report a complete picture of the effectiveness of the Related Cued Recall questions compared to the Unrelated Cued Recall questions, we investigated their impact on memory reporting. First we tested whether there was any difference in the quality of information reported in the Free Recall test and, as expected, we found no difference between groups in the amount of correct information reported $t(81) = 1.94$, $p = .06$, $d = .42$, or the accuracy of the information reported $t(81) = -.82$, $p = .41$, $d = -.18$[8]. In order to investigate the effect of Cued Recall type of memory reporting, a 2 (Cued Recall type: Related Cued Recall, and Unrelated Cued Recall) x 2 (Cued Recall order: Related Cued Recall first, Unrelated CR first) mixed design ANOVA was performed on (a) number of correct details, and (b) accuracy rate in response to the Related Cued Recall and Unrelated Cued Recall questions[9].

On the number of correct details reported we found a significant main effect of Cued Recall type $F(1, 81) = 31.34$, $p < .001$, $\eta^2_p = .28$, meaning that independently of the order in which the questions were asked participants reported more correct information in response to the Related Cued Recall compared with the Unrelated Cued Recall questions ($M_{diff} = 8.47$, $p < .001$, 95% CI [5.39, 11.57]). We found a non-significant main effect of Cued Recall order $F(1, 81) = .26$, $p = .61$, $\eta^2_p = .00$, meaning that the order in which the questions were asked did not impact on the number of correct information reported. Finally we found a significant interaction $F(1, 81) = 5.63$, $p = .02$, $\eta_p^2 = .06$. Bonferroni follow-up comparisons showed that participants in the FR + Related CR first group reported significantly more correct details in answer to the Related Cued Recall compared to the Unrelated Cued Recall questions ($M_{diff} = 12.00$, $p < .001$, 95% CI [7.79, 16.21]), as did those in the FR +

[8] A research assistant coded 25 (20%) randomly selected transcripts. The percentage agreement was found to be high (94%). Cohen's Kappa was .81, reflecting almost perfect agreement (McHugh, 2012).

[9] Overall, participants answered on average 41 Related Cued Recall questions, there was no difference in the number of Related Cued Recall questions answered between groups ($M_{diff} = 5.41$, $t(81) = 5.08$, $p = .05$, $d = .55$, 95% CI [.26, 8.26]).

Unrelated CR first group (M_{diff} = 4.85, p = .02, 95% CI [5.93, 9.11]). However, participants in the FR + Related CR first, compared to those in the FR + Unrelated CR first group, reported significantly more correct information in response to the Related Cued Recall questions (M_{diff} = 3.98, p = .03, 95% CI [.382, 7.57]) but not in response of Unrelated Cued Recall questions. This means that participants are likely to report more correct information in response to the Related Cued Recall questions (but not in response to Unrelated Cued Recall questions) only when the Related Cued Recall questions are answered *before* the Unrelated Cued Recall (see Table 7.2 for Means and Standard Deviations).

On the accuracy rate, we found a significant main effect of Cued Recall type $F(1, 81) = 440.96$, $p < .001$, $\eta^2_p = .85$, meaning that regardless of the order in which the questions were presented, participants were significantly more accurate in response to the Related Cued Recall compared to the Unrelated CR questions (M_{diff} = 28.29, $p < .001$, 95% CI [25.61, 30.98]). Finally, we found a non-significant main effect of Cued Recall order $F(1, 81) = .98$ $p = .86$, $\eta^2_p = .00$, and a non-significant interaction $F(1, 81) = .72$ $p = .39$, $\eta^2_p = .01$.

In summary the Related Cued Recall questions compared to the Unrelated Cued Recall questions yielded more correct and more accurate information. The order in which the questions were asked had only an impact on the number of correct information reported. In particular, we found initial evidence that if the Related Cued Recall questions are not preceded by the Unrelated Cued Recall questions, then participants are more likely to report more correct information.

Table 7.2. Means (and Standard Deviations) of number of correct, and accuracy rate of information reported in response of Cued Recall type (Related Cued Recall vs Unrelated Cued Recall) in both groups

.

		FR + Related CR first	FR + Unrelated CR first
Related CR	Correct	35.83 (8.90)	31.85 (7.48)
	Accuracy rate	83.31 (7.39)	82.31 (8.89)
Unrelated CR	Correct	23.83 (6.64)	27.00 (7.77)
	Accuracy rate	53.71 (11.20)	55.15 (10.85)

Discussion

The aim of Study 4 was to investigate shifts in confidence within a realistic interview, while still maintaining a high level of experimental control. Based on the results of Study 2 we hypothesised that, during an interview, confidence is unlikely to change after the Free Recall phase, but it is likely to decrease as a result of the probing phase. Furthermore, based on the results of Study 3a and 3b we predicted that, during the probing phase, confidence would only decrease in response to Cued Recall questions that were unrelated to the information freely reported in the Free Recall phase, presumably because these are likely to be experienced as difficult to answer. On the contrary, we did not expect confidence to change after answering Cued Recall questions that were related to the information freely reported in the Free Recall phase, presumably because these are more likely to be experienced as easy to answer. Our last hypothesis related to the realism of the confidence judgements reported in answers to the Cued Recall questions. In particular, we predicted that participants' confidence and accuracy would be better calibrated for answers given to related vs. unrelated Cued Recall questions.

We found supporting evidence for all the hypotheses. In particular, initial confidence was found to be statistically similar to confidence reported after the Free Recall phase, but significantly lower than confidence reported after the probing phase. However, the results of Study 4 show that if the questions of the probing phase are compatible to the information that had been freely reported previously, then confidence is unlikely to change. Therefore, as in Study 3a and Study 3b, the results of Study 4 show that the format of probing question is not responsible for shifts in confidence, rather it is the difficulty experienced in answering the question that leads to decreased confidence. We found that participants are more likely to easily retrieve details that were related to units of information freely reported, than they were to details that were related to units of information that had not been mentioned in the Free Recall phase. Thus, when the probing Cued Recall questions trigger details relating to information already recalled, confidence is less likely to change. On the contrary, when the probing Cued Recall questions seek details relating to information not yet recalled, confidence is more likely to decrease.

Overall, this result shows that confidence strongly depends on the cues generated while retrieving a memory (i.e., ease of retrieval), and is in line with the literature (e.g., Gustafsson et al., 2019; Kelley & Lindsay, 1993; Lindholm, 2018).

However, we found initial evidence that the retrieval process and its cues are not exclusively dependent on internal factors (such as the quality of a memory), rather they can be influenced by external variables, such as the interview type used to trigger the memory. Thus, if the interview is designed to maximise the ease of retrieval, confidence is unlikely to change. In Study 4 we found evidence that an interview featuring open questions followed by closed questions and, specifically, where the closed questions are compatible with the interviewee's own account, might promote ease of retrieval and thus confidence stability. This result is in line with the broader literature on investigative interviewing, and it suggests that the best practice guidelines related to the interview structure, the hierarchical organisation of different types of question asked and the use of compatible questions, might all be beneficial for confidence as well as memory output. In other words, these techniques are not only likely to yield complete and accurate accounts, and but also confidence stability.

Finally, the results of Study 4 support the hypothesis that participants are better calibrated in response to Cued Recall questions when these are related to the information freely reported in the Free Recall phase. In particular, we found both calibration and discrimination to be higher for the related compared to the unrelated Cued Recall questions, as measured by the lower C-index, and higher ANDI for the former compared to the latter. Overall, this result is important, and it suggests that the extent to which confidence is a reliable predictor of accuracy might depend, at least in part, on the way the memory is cued. As such, confidence appears to be a better predictor of accuracy when participants are asked compatible rather than incompatible questions.

However, Study 4 also showed an unexpected impact of Cued Recall question order. In particular, while discrimination was more reliable for related compared to the unrelated Cued Recall questions, regardless of the order in which these were asked, calibration was more reliable for related compared to unrelated Cued Recall questions *only* when the former preceded by the latter. The impact of Cued Recall order was also evident on the Over/Under-confidence indexes. In particular, participants were more confident in response to all the Cued Recall questions when the related Cued Recall questions were asked before the unrelated CR questions. We did not expect the order of Cued Recall questions to affect calibration, therefore it is difficult to explain why calibration for related Cued Recall questions diminished when the unrelated Cued Recall question were asked first. It is

possible that participants found answering unrelated Cued Recall (i.e., experienced as difficult) particularly tiresome, and when they were subsequently asked to answer related Cued Recall questions, they were also less motivated, which in turn affected their ability to reliably monitor the accuracy of their memory. However, the order in which the Cued Recall questions were asked was found also to influence memory reporting, thus taken together, these results show that when the compatible questions are asked after non – compatible questions, both the ability to monitor one's own memory, and memory reporting, are impoverished.

Study 4 is not without limitations. In particular, in order to collect a large dataset for the calibration analysis, participants were asked a large set of Cued Recall questions. This was clearly a demanding task that potentially had an impact on participants' attention and concentration throughout the task. As such, it is recommended that future studies increase the sample size and decrease the number of Cued Recall questions asked, should this be possible. Another issue relates to the ecological validity of the Study; although one important aim of Study 4 was to build upon the findings of the prior studies within a relatively realistic investigative interview, we acknowledge that in real-life eyewitnesses are unlikely to be asked such a large amount of Cued Recall questions, or to be asked compatible and incompatible Cued Recall questions in the spilt order adopted in this Study. In real-life interviews incompatible questions are more likely to be interposed among more compatible questions and this might affect confidence in different ways.

However, the results of Study 4 are important and shed light on the impact of different aspects of an interview on eyewitness confidence. In particular, the structure of the interview, the sequential organisation of the probing questions asked, and their compatibility with the free account, all contribute to shape confidence as well as memory output. Following best practice guidelines appears to promote confidence stability in a memory reported during an interview, and to increase the extent to which eyewitness confidence predicts accuracy. These results are important and contribute to our understanding of the impact that different interviewing techniques can have on eyewitnesses.

Chapter 8: General Discussion

Introduction and overview

The information eyewitnesses report is vital at all stages of the investigative process, and it is often the only source of evidence in a case. The errors eyewitnesses report can result in delays, a waste of police resources, and - at worst - in the apprehension of innocent people while guilty perpetrators remain at large. Therefore, it is essential that eyewitnesses remain engaged with the Criminal Justice System and report accurate accounts at all the stages of the investigative process. However, the information eyewitness report is not always correct, and a wealth of empirical evidence shows that they can and do make mistakes. In the past forty years, researchers have investigated the most effective way to extrapolate accurate and detailed accounts from memory. Based on such work, interviewing techniques and best practice guidelines have been made available to ensure that investigators are successful in eliciting accurate and complete accounts from witnesses. In the UK, best practice guidelines in eyewitness interviewing are summarised in the PEACE model, currently recognised as the official framework of professional practice in the field on eyewitness interviewing. Despite the large body of literature in support of the PEACE model and its techniques, very little is known about how these techniques impact eyewitnesses' confidence.

Generally, the effectiveness of eyewitness interviewing is measured in terms of the quality of the information the techniques are able to elicit, such as the amount, the accuracy and the granularity or specificity of the information reported. Clearly, whether interviewing techniques can extrapolate complete, accurate, and detailed accounts from witnesses is pivotal. However, this focus overlooks the processes that underlie eyewitnesses' memory reporting, and it does not fully unable an understanding of the impact that these techniques might have on the eyewitnesses.

Research on metamemory suggests that memory reporting is not a simple automatic process. As such, when people are asked a question, they don't automatically initiate their memory search and report the information they retrieved. Rather they engage in several cognitive operations that underpin decisions about whether and how report a recalled memory. It has been suggested that confidence is central when making decisions related to the regulation of memory output (e.g.,

Ackerman & Goldsmith, 2008; Brewer et al., 2018; Koriat & Goldsmith, 1996; Goldsmith et al., 2002). Thus, in order to fully understand the effectiveness of interviewing techniques in eliciting high quality information, it is important to investigate how the techniques influence memory confidence.

In line with this, the aim of this PhD is to enhance our understanding of the impact that interviewing techniques have on eyewitness confidence. Study 1 investigated whether a Best Practice compared to a Poor Practice interview would affect confidence differently. However, contrary to predictions, a Best Practice interview and a Poor Practice interview led to a statistically similar decrease in (i) confidence in the quality of a memory, and (ii) confidence in engaging with the Criminal Justice System. However, a more controlled methodology whereby a Free Recall and Cued Recall interview were used as proxies of a Good Practice and a Poor Practice interview (Study 2) showed that confidence in the quality of one's own memory and confidence in engaging with the CJS, are unlikely to change in response to a Free Recall invitation, and likely to decrease in response to Cued Recall questions. A further investigation of the impact of different types of question on the cognitive processes that underlie confidence judgements (Study 3a and 3b) showed that Cued Recall questions only lead to decreased confidence in so far as they trigger details that are difficult to remember, whereas when the Cued Recall questions elicit easily accessible information, confidence is unlikely to change.

A second important aim of this PhD was to investigate whether decreased confidence following an initial recall attempt would impact the amount and accuracy of the information subsequently reported. Contrary to expectations, Study 2 showed that the decrease in confidence following a Cued Recall interview was not correlated with a decrease in the quality of information reported in a subsequent recall test. Furthermore, Study 3a and 3b showed that a self-reported decrease in confidence does not impact on the ability to monitor and control the information in a subsequent recall test. In other words, while global confidence judgements differed between conditions, participants' confidence-accuracy calibration, and their volunteering rates, did not. Finally, Study 4 - designed to build on the previous results - investigated the interviewing conditions most likely to promote (i) confidence stability and (ii) effective monitoring. As predicted, the results showed that when the structure of the interview and the type of probing questions asked follow best

practice guidelines, confidence is less likely to change, and the monitoring of the information reported in the probing phase is more effective.

The malleability of confidence in the interviewing context: a gap in the literature

A large body of literature has investigated the effectiveness of memory confidence and underlined its malleable nature, although we do not know if eyewitness confidence changes during an interview. Research suggests that people are effective in monitoring their memories and controlling the information they report. In other words, people are able to assess the likely accuracy of their memories via their confidence judgements, and then use these judgements to regulate the amount and type of information that is reported (Koriat & Goldsmith,1996; Goldsmith et al., 2002; Pansky et al., 2005). This confidence-driven process is generally effective in supporting the accurate reporting of information from memory. It is therefore important that confidence is not artificially inflated or deflated during an interview. In other words, the extent to which witnesses believe that their memories are correct should not be altered within the interviewing setting. However, confidence is malleable, and research has shown that it can be influenced by external variables. For example, the literature investigating the effect of social influence on confidence has shown that the interviewer's feedback can shift confidence, leading eyewitnesses to either increase or decrease confidence in their own memory (e.g., Luus & Walls, 1994; Steblay et al., 2014; Wells & Bradfield, 1998; Wells & Seelau, 1995).

Despite the evidence that confidence is malleable and can be easily altered, the literature lacks a systematic examination of the impact that variables at play during an interview can have on eyewitness confidence. In order to fill this gap, this PhD investigated targeted factors of an interview that are likely to cause changes in confidence. In particular, this research focuses on the impact that the interview structure and the types of question asked during an interview have on eyewitness confidence and on the cognitive processes that underpin it.

The effect of interview structure and different types of questions on confidence malleability

Research has investigated the organisation of different types of question within an interview that are most likely to yield accurate and detailed account from witnesses. Current best practice guidelines suggest following a hierarchy of the question types, whereby different questions are asked at different stages of the interview (Gabbert et al., 2015a; 2015b; Griffiths & Milne, 2006). This interview structure is designed to maximise memory recollection and minimise errors. Ideally an interviewee should always be given a free and uninterrupted recall opportunity, and the Free Recall invitation should always precede the probing phase whereby more specific types of question can be asked. This hierarchical organisation of the question types ensure the opportunity for eyewitnesses to initiate their memory search in a way that better suit their own memory representation, and at a pace that is appropriate for them (Gabbert et al., 2015a; 2015b).

After a Free Recall opportunity has been given, the interviewers can engage in a probing phase. Here, the freely reported narrative can be further explored via the use of different types of probe, such as Cued Recall questions. These questions direct the eyewitnesses' attention towards single aspects or specific details of a memory, and they can help the eyewitnesses to reach details of the memory overlooked during the Free Recall attempt. In this phase of the interview, the interviewer has a more active role, thus the opportunity to contaminate eyewitnesses' memory can increase and so the probability of eliciting errors (e.g., Dando, 2013; Dando et al., 2009; Kontogianni et al., 2020). However, several techniques can be used to minimise the interviewer's contamination and researchers suggest probing eyewitnesses' memory with questions that are compatible to the eyewitnesses' own free report (Fisher et al., 2011; Gabbert et al., 2015a; 2015b; Kebbell et al., 2001). For example, the questions asked at this stage of the interview should match the content of the information reported in the Free Recall phase.

While research in investigative interviewing has focused on the impact that the hierarchical use of question types has on the quality of information reported, research on metamemory has investigated whether different types of question elicit information associated with different degree of confidence. Evidence shows that information elicited via Free Recall invitations compared to Cued Recall questions is generally associated with higher confidence (Allwood et al., 2008; Knutsson et al., 2011). The suggested explanation is that when answering Free Recall invitations eyewitnesses have a high control over the output of their memory. Here, they can

decide the details to report based on the evaluation of their likely accuracy, and presumably they tend to select only the information they are highly confident about. On the contrary, when answering Cued Recall questions eyewitnesses have less control over their memory, because they do not control the questions, and they are more likely to report information they are less sure about. Based on this literature this PhD investigated whether eyewitness' confidence in their memory for the whole event would be influenced by different types of question asked during the different phases of an interview.

Overall, the findings of this research consistently show that different types of question, asked at different stages of an interview are likely to trigger eyewitnesses' retrieval processes in different ways and therefore to differently affect memory confidence. In particular, during a free and undirected recall attempt, participants are likely to pay attention and retrieve the information they know, and as a consequence their confidence in the memory for the event seen is more likely to remain stable. However, when answering Cued Recall questions, eyewitnesses' retrieval process is directed towards information that is more likely to be unknown, thus gaps in memory are more likely to be highlighted and confidence is more likely to decrease.

Furthermore, the investigation of the processes upon which confidence is based (i.e., retrieval difficulty), provide a more comprehensive understanding of the effect that different types of question might have on confidence. In particular, the results showed that the questions are likely to cause a decrease in confidence *only* in so far as they trigger information that are difficult to access. As such it is not the type of question that is likely to cause a shift in confidence but rather the difficulty in retrieving the details triggered by the questions. When freely recalling a memory, eyewitnesses are likely to engage in the retrieval of highly memorable information and they are unlikely to attempt to retrieve information that is difficult to access. On the contrary, when answering Cued Recall questions – where the eyewitnesses cannot control the information to be recalled, their retrieval process is more likely to be directed towards information that are difficult to access. Here, Cued Recall questions are *only* likely to cause a decrease in confidence when they trigger information that is difficult to access. On the contrary, when Cued Recall questions are easy, i.e., more likely to trigger information that are easily accessible, confidence is likely to remain stable.

Overall these results support an understanding of the impact of different types of question on eyewitness confidence. In particular, a consistent finding within this PhD is that confidence can change during an interview and the way in which memory is triggered is pivotal in determining such change. Confidence is only likely to decrease in response to questions that trigger information that is difficult to access. The likelihood of eliciting difficult information increases under directed retrieval, i.e., when Cued Recall questions are asked. As such, we present a new, previously undocumented, effect that Cued Recall questions can have on eyewitnesses. Research shows that, in real-life, interviewers still rely strongly on Cued Recall questions to elicit information from witnesses (e.g., Carson & La Rooy, 2015; La Rooy et al., 2013; La Rooy et al., 2011; Loftus & Greenspan, 2017; Otgaar et al., 2018; Oxburgh et al., 2012; Wade et al., 2018). Therefore, it is of interest to underline that an interview dominated by the use of Cued Recall questions is likely to lead eyewitnesses to feel unsure about the extent to which their memory for an event seen is accurate.

The effect of best-practice and poor-practice interviews on confidence

Beside extending our knowledge of the effects that Free Recall and Cued Recall questions have on confidence in memory for events, the results of this PhD also provide the opportunity to predict the conditions of a realistic interview in which confidence is likely to remain stable and the conditions in which confidence in more likely to change. Here, the result of interest is that Cued Recall questions compared with Free Recall invitations are more likely to lead to decreased confidence, and that the extent to which Cued Recall questions lead to decreased confidence is contingent to the difficulty experienced in answering them.

Based on these findings, Study 4 featured two realistic interviews in which the Cued Recall questions asked in the probing phase were either easy or difficult. Here, the prediction was that confidence would only decrease in response to the latter type of Cued Recall question, but not in response to the former. We speculated that any Cued Recall question focused on the information freely reported by the eyewitness in the Free Recall phase would likely be experienced as easy. Conversely, any Cued Recall questions focused on information not freely reported in the Free Recall phase would likely be experienced as difficult. As predicted, confidence remained stable when the Free Recall was followed by Cued Recall questions

probing information freely reported in the Free Recall phase. Conversely, it decreased when the Cued Recall questions focused on information not reported in the previous Free Recall phase. In other words, following best-practice hierarchy of asking questions based on witness-led topic areas is unlikely to significantly alter eyewitness confidence.

Taken together, these results add to the existing literature on investigative interviewing and to the body of findings in support of best practice in interviewing witnesses. In sum, across four studies this PhD showed that eyewitness's confidence is likely to be unaffected by the interview, should this follow best-practice recommendations relating to: (i) the structure and hierarchical use of questions - whereby a Free Recall invitation precedes the probing phase, and (ii) the use of compatible questions in the probing phase - whereby these questions target *exclusively* information freely reported in the Free Recall phase.

Potential risks of shifts in confidence in the interviewing setting

Research has shown that memory confidence can influence the willingness to testify (Hafstad et al., 2004; Wells & Bradfield, 1998); and in this PhD initial evidence suggests that should eyewitness confidence decrease following a poor practice interview, their confidence in engaging with the Criminal Justice System is also likely to decrease. In particular, it was found that following a Cued Recall interview (but not following a Free Recall interview), not only did confidence in the accuracy of a memory decrease, but so did also confidence in the reported likelihood of (i) signing a witness statement, (ii) participating in an identification task, and (iii) reporting evidence in Court.

These results are important because they broaden our understanding of the potential risks related to confidence shifts during a poor practice interview. These findings suggest that if the interview leads eyewitnesses to lose confidence in their memory, then they are less likely to engage in other aspects of the investigation (for similar findings see Wells & Bradfield, 1998). This can be particularly problematic in cases whereby the eyewitness is the primary source of information. Researchers have argued that eyewitness testimony is pivotal in many investigations, and in order to ensure justice and apprehend the perpetrator investigators often rely on the information eyewitnesses report (Fisher & Geiselman, 1992; Wells & Olson, 2003). Ensuring eyewitnesses remain cooperative and engaged with the different stages of

the investigative process can increase the chances of identifying potential leads and solving the crime. Best practice guidelines encourage interviewers to leave the eyewitnesses with a positive state of mind, should they be asked to continue collaborating with the investigation (Bull, 2010). In addition to current recommendation, this PhD reports initial evidence that conducting a good quality interview might also contribute to ensuring eyewitnesses' engagement.

The relation between shifts in confidence and subsequent confidence-accuracy calibration

Given the main findings, discussed above, a second important aim of this PhD was to investigate whether the consistently observed decrease in confidence following an initial recall attempt would impact people's ability to effectively monitor and control the information in a subsequent recall test. The results gathered across this PhD did not support this hypothesis. In particular, decreased confidence following an initial recall attempt (reported by participants in the Difficult Cued Recall group only) did not lead participants to freely report poorer accounts in a subsequent recall test. This result was confirmed by the findings that decreased confidence following an initial recall attempt did not lead participants (i) to report under-confidence in the information subsequently recalled, or (ii) to exert harsher control over the information reported.

Although this set of findings were not expected, they consistently confirm only a trivial effect of decreased confidence in the memory for the whole event on the monitoring and control of information subsequently reported, which is ultimately a positive finding. We interpreted these results as evidence that the monitoring and control over the information reported is largely based on the task in hand rather than being influenced by previously declared beliefs regarding the accuracy of memory. In other words, these results suggest that participants appeared to disregard their previously declared belief that confidence in one's memory had declined, and instead, based their confidence judgements relating to the subsequent memory task on current perceptions of the retrieval process and its cues (e.g., the ease of retrieval). As such, these results are in line with the notion that confidence is primarily experienced-based and heavily reliant upon the qualitative experience of retrieval (e.g., Brewer et al., 2005; Gustafsson et al., 2019; Kelley & Lindsay, 1993; Lindholm et al., 2018; Robinson et al., 2000).

Although these results are consistent and in line with the literature, it is worth mentioning that in real-life settings the delay between two interviews is likely to be more significant than that adopted in this research (Gabbert et al., 2015; La Rooy et al, 2008; Odinot et al., 2013). As such, the methodology adopted in this PhD studies does not allow to draw confident conclusions or to make practical recommendations for the applied context. In particular, the lack of a statistically significant effect of decreased confidence on subsequent monitoring and control of information, is contingent to the short delay used in the studies. Consequently, I am unable to conclude whether decreased confidence impacts upon subsequent memory monitoring and regulation, should a longer delay elapse between the initial and subsequent recall attempt. As such, this particular finding should be considered carefully, and further research should investigate whether decreased confidence in the memory for an event might affect eyewitnesses' monitoring and control after a longer and more realistic delay.

A contribution to the debate on the confidence- accuracy relationship

A further contribution of this PhD relates to the evidence that eyewitnesses' monitoring, i.e., the extent to which their confidence judgements predict memory accuracy, might depend (at least in part) on the quality of the interview given. It was found that eyewitnesses are more effective in monitoring the accuracy of their memories when the interview follows best practice recommendations relating to the type and organisation of the questions used. In particular, Study 4 tested the hypothesis that participants' confidence and accuracy would be better calibrated if the Cued Recall questions asked in the probing phase related to the topics freely-reported by the participant in the Free Recall phase (but not if the follow-up questions were *unrelated*). As expected, participants' calibration and discrimination were found to be more reliable when the interview followed recommended best-practice relating to the use of compatible questions in the probing phase. This result is important; providing an initial insight into the interviewing conditions in which eyewitnesses confidence is a more reliable predictor of accuracy, and contributing to the evidence-base supporting the practice of witness-led interviewing.

Although the extent to which confidence is a good predictor of accuracy is still debated in the eyewitness literature, research on identification memory has shown that confidence can be a good predictor of accuracy, under conditions

whereby confidence has not been influenced, e.g., short retention interval, unbiased line-ups, and the absence of social feedback (see Brewer & Wells, 2006; Palmer et al., 2013; Sauer et al., 2010; Wixted & Wells, 2017; also Wixted et al., 2015). The studies in this PhD confirm this finding in the context of memory recall rather than recognition. Specifically, confidence and accuracy in memory recall appears to be better calibrated when the interview is witness-led, rather than featuring interviewer-led questions. In particular, our results show that participants' monitoring (as measured by calibration analysis) during the probing phase is more reliable when the Cued Recall questions are compatible with the eyewitness's freely reported account. However, this is only the case when non-compatible Cued Recall questions do not precede compatible Cued Recall questions.

Taken together, these results support the literature on identification and they underline that the characteristics of an interview not only influence eyewitnesses' confidence in the memory for the whole event, but also how effectively eyewitnesses are at monitoring their memories. In particular, the types of question asked during the probing phase, and whether these questions are compatible with the information freely reported, contribute to shape the extent to which eyewitnesses' confidence is a reliable predictor of accuracy.

Limitations of the PhD

Despite the contributions to the field of investigative interviewing, and to the current debate on the reliability of memory confidence, this PhD is not without limitations. In addition to the short delay between the initial and subsequent recall attempts discussed above in this Chapter, it is important to mention also some of the methodological issues relating to (i) the sample size, and (ii) the ecological validity of the studies.

Researchers have shown that statistical significance is strongly influenced by the size of the sample. Studies adopting small samples are more likely to incur false negative (Type II errors), occurring when the researcher rejects the experimental hypothesis, when this is true (Goodwin, 2010). It is therefore important to ensure that studies are run with enough participants to be able to detect the predicted effect, should this exist. A sound experiment can be achieved by using a Power Analysis, that can assist in determining the most appropriate sample size, depending on

established parameters (e.g., alpha level, effect size, type of inferential test, and desired statistical power) (Faul et al., 2007). The issue of incurring a Type II error due a small sample size is particularly problematic when the results do not reach statistical significance and the hypotheses are rejected. In these circumstances the researcher is unable to establish whether the hypothesis is rejected because the effect does not exist or because it has not been detected (i.e., due to a low powered study). In order to mitigate these limitations, in this PhD I have adopted alternative statistical tests (i.e., Bayesian analysis) to establish whether the non-significant results were more likely to be reliable (i.e., due to a true null) or to be due to a lack of power. Despite allowing me to draw conclusion more confidently, the importance of adopting more appropriate size samples in order to build more robust experiments is acknowledged.

A further limitation of this PhD relates to the ecological validity of the studies, which on occasions has prevented me from making strong conclusions (e.g., the issue of the short delay discussed above in this Chapter). It is well established that the artificial nature of empirical studies reduces the extent to which the results can be applied to real-life contexts (Schmuckler, 2001). This can be particularly problematic for applied research, that ultimately aims to be informative for practitioners. In particular, throughout the studies and in order to achieve experimental control, participants were shown a crime video, and this experience differs greatly from witnessing a crime in real life. Similarly, in this PhD the interviews are either delivered by myself, or presented via an online survey. While in real life eyewitnesses are interviewed by an authoritative police officer, and this can result in a completely different experience. These differences are important, and they can influence people and their cognitive functioning in ways experimental studies cannot fully capture.

Considering these limitations and the applied nature of the results of this PhD, it is essential to replicate these findings in order to enhance the validity and reliability of the results found.

Suggestions for future research

The findings of this PhD have highlighted several important research questions that are currently unanswered. Despite this body of research demonstrating

that confidence can change during an interview, and that poor practice interviews are likely to lead to decreased confidence, the extent to which this is problematic is not yet understood. As such, further research should investigate the effects that decreased confidence following an interview can have on eyewitnesses' performance.

Research has shown that the belief regarding the strength of one's own memory can influence the decision to take on board post-event information suggested by others (Gabbert et al., 2012). For example, when deciding whether the information suggested by a co-witness should be believed, participants can compare the quality of their own memory to the perceived quality of a co-witness's memory. Research has found that when participants believe the co-witness's memory is stronger than their own memory, they are more likely to report suggested details (especially when their memory is objectively poor) (Allan et al., 2012). In line with this research, it is not unreasonable to think that decreased confidence following a poor practice interview might increase reliance on the co-witness's memory, and thus increase the extent to which an eyewitness conforms to the co-witness's account.

Evidence exists that decreased confidence following negative feedback exacerbates suggestibility. For example, Leippe et al. (2006) found that participants who believed to have poor memory following receiving negative feedback were more likely to report suggested details compared with those that did not receive the negative feedback (see also Leippe et al., 2009; Lida et al., 2019). Based on these results, it is possible to formulate the hypothesis that decreased confidence following a poor practice interview might affect suggestibility similarly, and future research should investigate this further.

Overall, further research should not only focus on replicating the findings of this PhD, but also on investigating further the effects that shifts in confidence can have on eyewitness and on the quality of the information they report.

Conclusions

To conclude, this PhD addresses a gap in the literature by investigating the effect of targeted interviewing techniques on eyewitness confidence. Good practice interviewing techniques are unlikely to alter confidence, while poor practice techniques are more likely to lead to a drop in confidence. These results add to the wealth of evidence in favour of good practice interviewing techniques relating to the

type and order of questions used. As such, the use of (i) Free Recall invitations, (ii) hierarchical organisation of different types of question (from Free Recall to Cued Recall questions), and (iii) compatible witness-led questions appear to promote confidence stability and more effective monitoring (i.e., calibration). Furthermore, this PhD reports evidence of a novel and undocumented detrimental effect of poor practice interviewing techniques on eyewitness confidence. As such, it increases our understanding of the risks relating to interviewing styles that are interviewer-led and dominated by the use of closed questions. Cumulatively, the studies in this PhD have shown that confidence - like memory - is malleable and can be easily influenced by the quality of the interview given. Ensuring that eyewitnesses are interviewed by following best practice guidelines is pivotal in preserving confidence and memory alike.

References

Ackerman, R., & Goldsmith, M. (2008). Control over grain size in memory reporting:With and without satisficing knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(5), 1224–1245. https://doi.org/10.1037/a0012938.

Aldridge, J., & Cameron, S. (1999). Interviewing child witnesses: questioning techniques and the role of training. *Applied Developmental Science 3*(2), 136–147. https://doi.org/10.1207/s1532480xads0302_7.

Allan, K., Midjord, J. P., Martin, D., & Gabbert, F. (2012). Memory conformity and the perceived accuracy of self versus other. *Memory & cognition*, *40*(2), 280–286. https://doi.org/10.3758/s13421-011-0141-9.

Allwood, C. M., Ask, K., & Granhag, P. A. (2005). The Cognitive Interview: Effects on the realism in witnesses' confidence in their free recall. *Psychology, Crime & Law, 11*(2), 183–198. https://doi.org/10.1080/10683160512331329943.

Allwood, C. M., Innes-Ker, A. C., Homgren, J., & Fredin, G. (2008). Children's and adults' realism in their event-recall confidence in responses to free recall and focused questions. *Psychology, Crime & Law, 14*(6), 529-547. https://doi.org/10.1080/10683160801961231.

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning & Verbal Behavior, 22*(3), 261–295. https://doi.org/10.1016/S0022-5371(83)90201-3.

Anderson, R. C., & Pichert, J. W. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior, 17*(1), 1-12. https://doi.org/10.1016/S0022-5371(78)90485-1.

Barnes, A. E., Nelson, T. O., Dunlosky, J., Mazzoni, G., & Narens, L. (1999). An integrative system of metamemory components involved in retrieval. In D. Gopher & A. Koriat (Eds.), *Attention and Performance XVII – Cognitive regulation of performance: Interaction of theory and application* (pp. 287-313). The MIT Press.

Belli, R. F., Winkielman, P., Read, J. D., Schwarz, N., & Lynn, S. J. (1998). Recalling more childhood events leads to judgments of poorer memory: Implications for the recovered/false memory debate. *Psychonomic Bulletin & Review, 5*(2), 318–323. https://doi.org/10.3758/BF03212958.

Berkowitz, S. R., & Frenda, S. J. (2018). Rethinking the confident eyewitness: A reply to Wixted, Mickes, and Fisher. *Perspectives on psychological science: A journal of the Association for Psychological Science*, *13*(3), 336–338. https://doi.org/10.1177/1745691617751883.

Bohnam, A. J., & Gonzalez-Vallejo, C. (2009). Assessment of calibration for reconstructed eye-witness memories. *Acta Psychologica, 131*(1), 34-52. https://doi.org/10.1016/j.actpsy.2009.02.008.

Bradfield, A. L., Wells, G. L., & Olson, E. A. (2002). The damaging effect of confirming feedback on the relation between eyewitness certainty and identification accuracy. *Journal of Applied Psychology, 87*(1), 112–120. https://doi.org/10.1037/0021-9010.87.1.112.

Brewer, N., & Burke, A. (2002). Effects of testimonial inconsistencies and eyewitness confidence on mock-juror judgments. *Law and Human Behavior, 26*(3), 353–364. https://doi.org/10.1023/A:1015380522722.

Brewer, N., Keast, A., & Rishworth, A. (2002). The confidence-accuracy relationship in eyewitness identification: The effects of reflection and disconfirmation on correlation and calibration. *Journal of Experimental Psychology: Applied, 8*(1), 44–56. https://doi.org/10.1037/1076-898X.8.1.44.

Brewer, W. F., Sampaio, C., & Barlow, M. R. (2005). Confidence and accuracy in the recall of deceptive and nondeceptive sentences. *Journal of Memory and Language, 52*(4), 618-627. http://dx.doi.org/10.1016/j.jml.2005.01.017.

Brewer, N., Vagadia, A. N., Hope, L., & Gabbert, F. (2018). Interviewing witnesses: Eliciting coarse-grain information. *Law and human behavior*, *42*(5), 458–471. https://doi.org/10.1037/lhb0000294.

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied, 12*(1), 11–30. https://doi.org/10.1037/1076-898X.12.1.11.

Bull, R. (2010). The investigative interviewing of children and other vulnerable witnesses: Psychological research and working/professional practice. *Legal and Criminological Psychology, 15*(1), 5-23. https://doi.org/10.1348/014466509X440160.

Buratti, S., & Allwood, C. M. (2012). Improved realism of confidence for an episodic memory event. *Judgment and Decision Making, 7*(5), 590–601.

Buratti, S., & Allwood, C. M. (2013). The effects of advice and "try more" instructions on improving realism of confidence. *Acta Psychologica*, *144*(1), 136–144. https://doi.org/10.1016/j.actpsy.2013.05.008.

Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition*, 5(1), 73–99. https://doi.org/10.1016/0010-0277(77)90018-X.

Carson, L., & Rooy, D. J. (2015). 'Commonsense psychology' is a barrier to the implementation of best practice child interviewing guidelines: A qualitative analysis of police officers' beliefs in Scotland. *Journal of Police and Criminal Psychology, 30*(1), 50–62. https://doi.org/10.1007/s11896-013-9139-5.

Carter, C. A., Bottoms, B. L., & Levine, M. (1996). Linguistic and socioemotional influences on the accuracy of children's reports. Law and Human Behavior, 20(3), 335–358. https://doi.org/10.1007/BF01499027.

Cederborg, A. C., Orbach, Y., Sternberg, K. J., & Lamb, M. E. (2000). Investigative interviews of child witnesses in Sweden. Child Abuse and Neglect 24(10), 1355–1361. https://doi.org/10.1016/S0145-2134(00)00183-6.

Cederborg, A. C., La Rooy, D. J., & Lamb, M. E. (2008). Repeated interviews with children who have intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities*, *21*(2), 103-113. https://doi.org/10.1111/j.1468-3148.2007.00372.x.

Cherryman, J., & Bull, R. (2001). Police officers' perceptions of specialist investigative interviewing skills. *International Journal of Police Science & Management*, *3*(3), 199-212. https://doi.org/10.1177/146135570100300302.

Chua, E. F. (2012). Evaluating confidence in our memories: Results and implications from neuroimaging and eye movement monitoring studies of metamemory. In L. Nadel & W. P. Sinnott-Armstrong (Eds.), *Oxford series in neuroscience, law and philosophy. Memory and law* (pp. 119–141). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199920754.003.0005.

Chua, E. F., Schacter, D. L., Rand-Giovannetti, E., & Sperling, R. A. (2006). Understanding metamemory: neural correlates of the cognitive process and subjective level of confidence in recognition memory. *NeuroImage*, *29*(4), 1150–1160. https://doi.org/10.1016/j.neuroimage.2005.09.058.

Clarke, C., & Milne, R. (2001). National evaluation of the PEACE investigative interviewing course. Report no: PRAS/149. London: The Home Office.

College of Policing (2020). *Investigative Interviewing.* College of Policing, Authorised Professional Practice. https://www.app.college.police.uk/app-content/investigations/investigative-interviewing/#peace-framework.

Collins, R., Lincoln, R., & Frank, M. G. (2002). The effect of rapport in forensic interviewing. *Psychiatry, Psychology and Law, 9*(1), 69-78. https://doi.org/10.1375/132187102760196916

Coluccia, E., Bianco, C., & Brandimonte, M. (2010). Autobiographical and event memories for surprising and unsurprising events. *Applied Cognitive Psychology, 24*(2), 177-199. https://doi.org/10.1002/acp.1549.

Costermans, J., Lories, G., & Ansay, C. (1992). Confidence level and feeling of knowing in question answering: The weight of inferential processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(1), 142–150. https://doi.org/10.1037/0278-7393.18.1.142.

Dando, C. (2013). Drawing to remember: External support of older adults' eyewitness performance. *PloS One, 8*, e69937. https://doi.org/10.1371/journal.pone.0069937.

Dando, C., Wilcock, R., & Milne, R. (2009). The cognitive interview: The efficacy of a modified mental reinstatement of context procedure for frontline police investigators. *Applied Cognitive Psychology, 23*(1), 138–147. https://doi.org/10.1002/acp.1451.

Davies, G. M., Westcott, H. L., & Horan, N. (2000). The impact of questioning style on the content of investigative interviews with suspected child sexual abuse victims. *Psychology, Crime and Law 6*(2), 81–97. https://doi.org/10.1080/10683160008410834.

Davis, S. L., & Bottoms, B. L. (2002). Effects of social support on children's eyewitness reports: A test of the underlying mechanism. *Law and Human Behavior, 26*(2), 185–215. https://doi.org/10.1023/A:1014692009941.

Deffenbacher, K. A. (1980). Eyewitness accuracy and confidence: Can we infer anything about their relationship? *Law and Human Behavior, 4*(4), 243–260.

https://doi.org/10.1007/BF01040617.

Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law & Human Behavior, 28*(6), 687–706. https://doi.org/10.1007/s10979-004-0565-x.

Diesen, C. (2012). Therapeutic jurisprudence and the victim of crime. In T.I. Oei, & M.S. Groenhuijsen (Eds.), *Progression in forensic psychiatry: About boundaries* (pp. 579-598). Dordrecht, The Netherlands: Kluwer.

Dysart, J. E., & Lindsay, R. C. L. (2007). The effects of delay on eyewitness identification accuracy: Should we be concerned? In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology, Vol. 2. Memory for people* (pp. 361–376). Lawrence Erlbaum Associates Publishers.

Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in psychology, 6*(621), 1-11. https://doi.org/10.3389/fpsyg.2015.00621.

Ebbesen, E. B., & Rienick, C. B. (1998). Retention interval and eyewitness memory for events and personal identifying attributes. *Journal of Applied Psychology, 83*(5), 745–762. https://doi.org/10.1037/0021-9010.83.5.745.

Evans, J. R., & Fisher, R. P. (2011). Eyewitness memory: Balancing the accuracy, precision and quantity of information through metacognitive monitoring and control. *Applied Cognitive Psychology, 25*(3), 501-508. https://doi.org/10.1002/acp.1722.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. https://doi.org/10.3758/bf03193146.

Ferré Romeu P. (2006). Memories of the terrorist attacks of September 11, 2001: a study of the consistency and phenomenal characteristics of flashbulb memories. *The Spanish journal of psychology*, *9*(1), 52–60. https://doi.org/10.1017/s1138741600005977.

Fisher, R. P., Geiselman, R. E., & Amador, M. (1989). Field test of the cognitive interview: Enhancing the recollection of actual victims and witnesses of crime. *Journal of Applied Psychology, 74*(5), 722–727. https://doi.org/10.1037/0021-9010.74.5.722.

Fisher, R. P., Geiselman, R. E., & Raymond, D. S. (1987). Critical analysis of police interview techniques. *Journal of Police Science and Administration, 15*, 177-185.

Fisher, R. P., McCauley, M. R., & Geiselman, R. E. (1994*). Improving eyewitness testimony with the Cognitive Interview.* In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 245–269). Cambridge University Press.

Fisher, R. P., Milne, R., & Bull, R. (2011). Interviewing Cooperative Witnesses. *Current Directions in Psychological Science, 20*(1), 16–19. https://doi.org/10.1177/0963721410396826

Gabbert, F., Memon, A., & Allan, K. (2003). Memory conformity: Can eyewitnesses influence each other's memories for an event? *Applied Cognitive Psychology, 17*(5), 533–543. https://doi.org/10.1002/acp.885.

Gabbert, F., Hope, L., Carter, E., Boon, R., & Fisher, R. P. (2015). The role of initial witness accounts within the investigative process. In G. Oxburgh, T. Myklebust, T. Grant, & R. Milne (Eds.), *Communication in investigative and legal contexts: integrated approaches from forensic psychology, linguistics and law enforcement* (pp. 107-132). Wiley-Blackwell.

Gabbert, F., Hope L., & Fisher, R. P. (2009). Protecting eyewitness evidence: examining the efficacy of a self-administered interview tool. *Law and Human Behavior*, *33*(4), 289–307. https://doi.org/10.1007/s10979-008-9146-8.

Gabbert, F., Hope, L., Fisher, R. P., & Jamieson, K. (2012). Protecting against misleading post-event information with a self-administered interview. *Applied Cognitive Psychology, 26*(4), 568–575. https://doi.org/10.1002/acp.2828.

Gabbert, F., Hope L., La Rooy, D. J., McGregor, A., Milne, R., & Ellis, T. (2015b). Introducing a PEACE-compliant 'Structured Interview Protocol' to enhance the quality of investigative interviews. [iIIRG. https://www.iiirg.org/assets/Gabbert-et-al.-SIP-talk-for-IIRG-pdf].

Gabbert, F., Hope, L., Luther, K., Wright, G., Ng, M., & Oxburgh, G. (in press). Exploring the use of rapport in professional information-gathering contexts by systematically mapping the evidence base. *Cognitive Psychology, Applied.*

Gabbert, F., Hope, L., McGregor, A., Ellis, T., La Rooy, D. J., & Milne, R. (2015a). SIP Structured Interview Protocol. [Pamphlet].

Gabbert, F., Wright, D. B., Memon, A., Skagerberg, E. M., & Jamieson, K. (2012). Memory conformity between eyewitnesses. *Court Review: The Journal of the American Judges Association, 48*(1-2) 36– 43.

Garry, M., Manning, C. G., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin & Review, 3*(2), 208–214. https://doi.org/10.3758/BF03212420.

Geiselman, R. E., & Fisher, R. P. (1989). The cognitive interview technique for victims and witnesses of crime. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 191–215). Springer Publishing Company.

George, R. C. (1991). *A field und experimental evaluation of the cognitive interview.* Unpublished Master's Thesis, Polytechnic of East London.

Goldsmith, M. (2016). Metacognitive quality-control processes in memory retrieval and reporting. In J. Dunlosky & S. K. Tauber (Eds.), *Oxford library of psychology. The Oxford handbook of metamemory* (pp. 357–385). Oxford University Press.

Goldsmith, M., Koriat, A., & Weinberg-Eliezer, A. (2002). Strategic regulation of grain size memory reporting. *Journal of Experimental Psychology: General, 131*(1), 73–95. https://doi.org/10.1037/0096-3445.131.1.73.

Goldsmith, M., Koriat, A., & Pansky, A. (2005). Strategic regulation of grain size in memory reporting over time. *Journal of Memory and Language, 52*(4), 505-525. http://dx.doi.org/10.1016/j.jml.2005.01.010.

Goodwin, C. J. (2010). *Research in psychology: Methods and design (6th ed.).* John Wiley & Sons Inc.

Granhag, P. A. (1997). Realism in eyewitness confidence as a function of type of event witnessed and repeated recall. *Journal of Applied Psychology, 82*(4), 599–613. https://doi.org/10.1037/0021-9010.82.4.599.

Granhag, P. A., Ask, K., & Giolla, E. M. (2014). Eyewitness recall: An overview of estimator-based research. In D. S. Lindasy, & T. J. Perfect (Eds.), *The SAGE handbook of applied memory*, (pp. 541–558). SAGE Publications.

Granhag, P. A., Jonsson, A. C., & Allwood, C. M. (2004). The cognitive interview and its effect on witnesses' confidence. *Psychology, Crime & Law, 10*(1), 37–52. https://doi.org/10.1080/1068316021000030577.

Granhag, P. A., Strömwall, L. A., & Allwood, C. M. (2000). Effects of reiteration, hindsight bias, and memory on realism in eyewitness confidence. *Applied*

Cognitive Psychology, 14(5), 397–420. https://doi.org/10.1002/1099-0720(200009)14:5<397::AID-ACP658>3.0.CO;2-M.

Greenspan, R. L., & Loftus, E. F. (2020). Eyewitness confidence malleability: Misinformation as post-identification feedback. *Law and Human Behavior, 44*(3), 194–208. https://doi.org/10.1037/lhb0000369.

Gregg, J., Upadhyay, S. S. N., Kuntzelman, K., Sacchi, E., & Westerman, D. L. (2019). Parallel effects of retrieval ease on attributions about the past and the future. *Acta Psychologica, 193,* 96–104. https://doi.org/10.1016/j.actpsy.2018.12.007.

Griffiths, A., & Milne, R. (2006). Will it all end in tiers? Police interviews with suspects in Britain. In T. A. Williamson (Ed.) *Investigative Interviewing: rights, research, regulation* (pp.167–189). Willan.

Griffiths, A., & Milne, R. (2010). The application of Cognitive Interview techniques as part of an investigation. In A. Carol, & M. J. Fisher (Eds.) *Consultancy and advising in Forensic practice: Empirical and practical guidelines* (pp. 69-90). John Wiley & Sons.

Griffiths, A., & Walsh, D. (2018). Qualitative analysis of qualitative evaluation: an exploratory examination of investigative interviewers' reflections on their performance. *Psychology, Crime & Law, 24*(4), 433–450. https://doi.org/10.1080/1068316X.2017.1390115.

Gustafsson, P. U., Lindholm, T., & Jönsson, F. U. (2019). Predicting accuracy in eyewitness testimony with memory retrieval effort and confidence. *Frontiers in Psychology 10*. https://doi.org/10.3389/fpsyg.2019.00703.

Gurney, D. J., Vekaria, K. N., & Howlett, N. (2014). A nod in the wrong direction: Does non-verbal feedback affect eyewitness confidence in interviews? *Psychiatry, Psychology and Law, 21*(2), 241–250. https://doi.org/10.1080/13218719.2013.804388.

Gwyer, P., & Clifford, B. R. (1997). The effects of the cognitive interview on recall, identification, confidence and the confidence/accuracy relationship. *Applied Cognitive Psychology, 11*(2), 121–145. https://doi.org/10.1002/(SICI)1099-0720(199704)11:2<121::AID-ACP443>3.0.CO;2-L.

Hafstad, G. S., Memon, A., & Logie, R. (2004). Post-identification Feedback, Confidence and Recollections of Witnessing Conditions in Child Witnesses. *Applied Cognitive Psychology, 18*(7), 901–912. https://doi.org/10.1002/acp.1037.

Hallgren, K. A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutorials in quantitative methods for psychology*, *8*(1), 23–34. https://doi.org/10.20982/tqmp.08.1.p023.

Hamilton, G., Brubacher, S. P., & Powell, M. B. (2016). Investigative Interviewing of Aboriginal Children in Cases of Suspected Sexual Abuse. *Journal of child sexual abuse*, *25*(4), 363–381. https://doi.org/10.1080/10538712.2016.1158762.

Hershkowitz, I. (2001). Children's responses to open-ended utterances in investigative interviews. *Legal and Criminological Psychology, 6*(1), 49–63. https://doi.org/10.1348/135532501168190.

Hirst, W., Phelps, E. A., Meksin, R., Vaidya, C. J., Johnson, M. K., Mitchell, K. J., Buckner, R. L., Budson, A. E., Gabrieli, J. D. E., Lustig, C., Mather, M., Ochsner, K. N., Schacter, D., Simons, J. S., Lyle, K. B., Cuc, A. F., & Olsson, A. (2015). A ten-year follow-up of a study of memory for the attack of September 11, 2001: Flashbulb memories and memories for flashbulb events. *Journal of Experimental Psychology: General, 144*(3), 604–623. https://doi.org/10.1037/xge0000055.

Holland, A. C., & Kensinger, E. A. (2012). Younger, middle-aged, and older adults' memories for the 2008 U. S. Presidential Election. *Journal of Applied*

Research in Memory and Cognition, 1(3), 163–170. https://doi.org/10.1016/j.jarmac.2012.06.001.

Home Office (UK) (2011). *Achieving Best Evidence in Criminal Proceedings: guidance on interviewing victims and witnesses, and using special measures.* London: HMSO.

Home Office (UK) (2013). *The Witness Charter. Standards of care for witnesses in the criminal justice system*. London: HMSO https://www.cps.gov.uk/sites/default/files/documents/victims_witnesses/witness_charter.pdf.

Hope, L., Gabbert, F., Fisher, R. P., & Jamieson, K. (2014). Protecting and enhancing eyewitness memory: The impact of an initial recall attempt on performance in an investigative interview. *Applied Cognitive Psychology*, *28*(3), 304-313.https://doi.org/10.1002/acp.2984.

Howes, L.M. (2020). Interpreted investigative interviews under the PEACE interview model: Police interviewers' perceptions of challenges and suggested solutions. *Police Practice and Research, 21*(4), 333–350. https://doi.org/10.1080/15614263.2019.1617145.

Howie, P., & Roebers, C. M. (2007). Developmental progression in the confidence-accuracy relationship in event recall: Insights provided by a calibration perspective. *Applied Cognitive Psychology, 21*(7), 871–893. https://doi.org/10.1002/acp.1302.

Ibabe, I., & Sporer, S. L. (2004). How you ask is what you get: On the influence of question form on accuracy and confidence. *Applied Cognitive Psychology, 18*(6), 711-726. https://doi.org/10.1002/acp.1025.

Jaeger, A., Lauris, P., Selmeczy, D., & Dobbins, I. G. (2012). The costs and benefits of memory conformity. *Memory & Cognition, 40*(1), 101–112. https://doi.org/10.3758/s13421-011-0130-z.

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes Factors. *The Journal of Problem Solving, 7*(1). https://doi.org 10.7771/1932-6246.1167.

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

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *361*(1476), 2109–2128. https://doi.org/10.1098/rstb.2006.1934.

Kebbell, M. R., Milne, R., & Wagstaff, G. F. (2001). The Cognitive Interview in forensic investigation: A review. In G. B., Traverso & L. Bagnoli (Eds), *Psychology and Law in a changing world* (pp. 185-198). Harwood Academic.

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

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory & Language, 32*(1), 1-24. https://doi.org/10.1006/jmla.1993.1001.

Kieckhaefer, J. M., Vallano, J. P., & Schreiber Compo, N. (2014). Examining the positive effects of rapport building: When and why does rapport-building benefit adult eyewitness memory? *Memory, 22*(8), 1010–1023. https://doi.org/10.1080/09658211.2013.864313.

Knutsson, J., Allwood, C. M., & Johansson, M. (2011). Child and adult witnesses: The effect of repetition and invitation-probes on free recall and metamemory realism. *Metacognition and Learning, 6*(3), 213–228. https://doi.org/10.1007/s11409-011-9071-y.

Kontogianni, F., Hope, L., Taylor, P.J., Vrij, A., & Gabbert, F. (2020). "Tell me more about this…": An examination of the efficacy of follow-up open questions following an initial account. *Applied Cognitive Psychology, 34(*5), 972-983. https://doi.org/10.1002/acp.3675.

Koriat, A. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. *Consciousness and Cognition, 9*(2), 149-171. https://doi.org/10.1006/ccog.2000.0433.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*(3), 490-517. http://dx.doi.org/10.1146/annurev.psych.51.1.481.

Koriat, A., Nussinson, R., Bless, H., & Shaked, N. (2008). Information-based and experience-based metacognitive judgments: Evidence from subjective confidence. In J. Dunlosky, & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 117-135). US: Psychology Press.

Kraha, A., Talarico, J. M., & Boals, A. (2014). Unexpected positive events do not result in flashbulb memories. *Applied Cognitive Psychology, 28*(4), 579–589. https://doi.org/10.1002/acp.3039.

Kraus, U., Zeier, F., Wagner, W., Paelecke, M., & Hewig, J. S. (2017). Comparing the quality of memory reports in different initial eyewitness questioning approaches. *Cogent Psychology, 4*(1), 1403063. https://doi.org/10.1080/23311908.2017.1403063.

Krix, A. C., Sauerland, M., Gabbert, F., & Hope, L. (2014). Providing eyewitnesses with initial retrieval support: What works at immediate and subsequent recall? *Psychology, Crime & Law, 20*(10), 1005–1027. https://doi.org/10.1080/1068316X.2014.902456.

La Rooy, D. J., Brubacher, S.P., Aromäki-Stratos, A., Cyr, M., Hershkowitz, I., Korkman, J., & Stewart, H. (2015). The NICHD protocol: A review of an internationally-used evidence-based tool for training child forensic interviewers. *Journal of Criminological Research, Policy, and Practice, 1*(2), 76–89. https://doi.org/10.1108/JCRPP-01-2015-0001.

La Rooy, D. J., Earhart, B., & Nicol, A. (2013). Joint Investigative Interviews (JIIs) conducted with children in Scotland: a comparison of the quality of interviews conducted before and after the introduction of the Scottish Executive (2011) guidelines. *Scots Law Times*, *31*, 4-10.

La Rooy, D. J., Lamb, M. E., & Memon, A. (2011). Forensic Interviews with Children in Scotland: A Survey of Interview Practices Among Police. *Journal of Police and Criminal Psychology*, *26*(1), 26-34. https://doi.org/10.1007/s11896-010-9072-9.

La Rooy, D. J., Lamb, M. E., & Pipe, M. E. (2009). Repeated interviewing: A critical evaluation of the risks and potential benefits. In K. Kuehnle & M. Connell (Eds.), *The evaluation of child sexual abuse allegations: A comprehensive guide to assessment and testimony* (p. 327–361). John Wiley & Sons Inc.

Lamb, M. E., Malloy, L. C., & La Rooy, D. J. (2011). Setting realistic expectations: developmental characteristics, capacities and limitations. In M. E. Lamb, D. J. La Rooy, L. C. Malloy, & C. Katz (Eds.), *Children's testimony: a handbook of psychological research and forensic practice* (2nd ed., pp. 15-48). (Wiley Series in the Psychology of Crime, Policing and Law). Wiley-Blackwell. https://doi.org/10.1002/9781119998495.ch2.

Lamb, M. E., Orbach, Y., Hershkowitz, I., Esplin, P. W., & Horowitz, D. (2007). A structured forensic interview protocol improves the quality and informativeness of investigative interviews with children: a review of research using the NICHD Investigative Interview Protocol. *Child abuse & neglect*, *31*(11-12), 1201–1231. https://doi.org/10.1016/j.chiabu.2007.03.021.

Lamb, M. E., Sternberg, K. J., Orbach, Y., Esplin, P. W., Stewart, H., & Mitchell, S. (2003). Age differences in young children's responses to open-ended invitations in the course of forensic interviews. *Journal of Consulting and Clinical Psychology, 71*(5), 926–934. https://doi.org/10.1037/0022-006X.71.5.926.

Leippe, M. R. (1980). Effects of integrative memorial and cognitive processes on the correspondence of eyewitness accuracy and confidence. *Law and Human Behavior, 4*(4), 261–274. https://doi.org/10.1007/BF01040618.

Leippe, M. R., & Eisenstadt, D. (2007). Eyewitness confidence and the confidence-accuracy relationship in memory for people. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology, Vol. 2. Memory for people* (pp. 377–425). Lawrence Erlbaum Associates Publishers.

Leippe, M. R., Eisenstadt, D., & Rauch, S. M. (2009). Cueing confidence in eyewitness identifications: Influence of biased lineup instructions and pre-identification memory feedback under varying lineup conditions. *Law and Human Behavior, 33*(3), 194–212. https://doi.org/10.1007/s10979-008-9135-y.

Leippe, M. R., Eisenstadt, D., Rauch, S. M., & Stambush, M. A. (2006). Effects of social-comparative memory feedback on eyewitnesses' identification confidence, suggestibility, and retrospective memory reports. *Basic and Applied Social Psychology, 28*(3), 201–220. https://doi.org/10.1207/s15324834basp2803_1z.

Levene, H. (1960). Robust tests for equality of variances. In: Olkin I et al (Eds) *Contributions to probability and statistics. Essays in honor of Harold Hotelling* (pp. 278–292). Stanford University Press.

Lida, R., Itsukusima, Y., & Mah, E., Y. (2020). How do we judge our confidence? Differential effects of memory feedback on eyewitness accuracy and confidence. *Applied Cognitive Psychology, 34*(2), 397-498. https://doi.org/10.1002/acp.3625.

Lindholm, T., Jönsson, F. U., & Liuzza, M. T. (2018). Retrieval effort cues predict eyewitness accuracy. *Journal of Experimental Psychology: Applied, 24*(4), 534-542. http://dx.doi.org/10.1037/xap0000175.

Loftus, E. F. (1982). Interrogating eyewitnesses – good questions and bad. In R. Hogarth (Ed.) *Question Framing and Response Consistency* (pp. 51–63). Josey-Bass.

Loftus, E. F. (2005). Searching for the neurobiology of the misinformation effect. *Learning & memory, 12*(1), 1–2. https://doi.org/10.1101/lm.90805.

Loftus, E. F., Donders, K., Hoffman, H. G., & Schooler, J. W. (1989). Creating new memories that are quickly accessed and confidently held. *Memory & Cognition, 17*(5), 607–616. https://doi.org/10.3758/BF03197083.

Loftus, E. F., & Greenspan, R. L. (2017). If I'm certain, is it true? Accuracy and confidence in eyewitness memory. *Psychological Science in the Public Interest, 18*(1), 1-2. https://doi.org/10.1177/1529100617699241.

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

Loftus, E. F., & Zanni, G. (1975). Eyewitness testimony: The influence of the wording of a question. *Bulletin of the Psychonomic Society, 5*(1), 86–88. https://doi.org/10.3758/BF03336715.

Luna, K., & Martín-Luengo, B. (2012). Confidence-accuracy calibration with general knowledge and eyewitness memory cued recall questions. *Applied Cognitive Psychology, 26*(2), 289-295. https://doi.org/10.1002/acp.1822.

Luther, K., Snook, B., Barron, T., & Lamb, M. E. (2015). Child interviewing practices in Canada: A box score from field observations. *Journal of Police and Criminal Psychology, 30*(3), 204–212. https://doi.org/10.1007/s11896-014-9149-y.

Luus, C. A. E., & Wells, G. L. (1994). The malleability of eyewitness confidence: Co-witness and perseverance effects. *Journal of Applied Psychology, 79*(5), 714–723. https://doi.org/10.1037/0021-9010.79.5.714.

Marsh, E. J., Tversky, B., & Hutson, M. (2005). How eyewitnesses talk about events: Implications for memory. *Applied Cognitive Psychology, 19*(5), 531–544. https://doi.org/10.1002/acp.1095.

McHugh, M.L. (2012). Interrater reliability: The kappa statistic. *Biochemia medica*, *22*(3), 276–282.

McLean, M. (1995). 4. Quality Investigation? Police Interviewing of Witnesses. *Medicine, Science and the Law*, *35*(2), 116–122. https://doi.org/10.1177/002580249503500205

Meissner, C. A., Sporer, S. L., & Schooler, J. W. (2007). Person descriptions as eyewitness evidence. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology, Vol. 2. Memory for people* (pp. 3–34). Lawrence Erlbaum Associates Publishers.

Memon, A., Hope, L., & Bull, R. (2003). Exposure duration: effects on eyewitness accuracy and confidence. *British journal of psychology*, *94*(3), 339–354. https://doi.org/10.1348/000712603767876262.

Memon, M., Linsey, W., Bull, R., & Koehnken, G. (2011). Isolating the effects of the cognitive interview techniques. *British Journal of Psychology, 88*(2), 179-197. https://doi.org/10.1111/j.2044-8295.1997.tb02629.x.

Memon, A., Wark, L., Holley, A., Bull R., & Koehnken, G. (1997). Eyewitness Performance in Cognitive and Structured Interviews. *Memory, 5*(5), 639-656. https://doi.org/10.1080/741941481.

Merckelbach, H., Wiers, R., Horselenberg, R., & Wessel, I. (2001). Effects of retrieving childhood events on metamemory judgments depend on the questions you ask. *British Journal of Clinical Psychology, 40*(2), 215–220. https://doi.org/10.1348/014466501163553.

Michael, R. B., & Garry, M. (2016). Ordered questions bias eyewitnesses and jurors. *Psychonomic bulletin & review*, *23*(2), 601–608. https://doi.org/10.3758/s13423-015-0933-1.

Michael, R. B., & Garry, M. (2019). How do ordered questions bias eyewitnesses? *Memory, 27*(7), 904 - 915. https://doi.org/10.1080/09658211.2019.1607388.

Milne, R., & Bull, R. (1999). *Investigative interviewing: Psychology and practice.* Wiley.

Myklebust, T., & Alison, L. (2000). The current state of police interviews with children in Norway: How discrepant are they from models based on current issues in memory and communication? *Psychology, Crime & Law, 6*(4), 331–351. https://doi.org/10.1080/10683160008409810.

Myklebust, T., & Bjørklund, R. A. (2006). The effect of long-term training on police officers' use of open and closed questions in field investigative interviews of children (FIIC). *Journal of Investigative Psychology and Offender Profiling, 3*(3), 165-181. http://dx.doi.org/10.1002/jip.52.

Nash, R. A., Nash, A., Morris, A., & Smith, S. L. (2016). Does rapport-building boost the eyewitness eyeclosure effect in closed questioning? *Legal and Criminological Psychology, 21*(2), 305–318. https://doi.org/10.1111/lcrp.12073.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), The psychology of learning and motivation (pp. 1–45). Academic Press.

Odinot, G., Memo, A., La Rooy, D. J., & Millen, A. (2013). Are two interviews better than one? Eyewitness memory across repeated Cognitive Interviews. *PLoS ONE 8*(10): e76305. https://doi.org/10.1371/journal.pone.0076305.

Odinot, G., & Wolters, G. (2006). Repeated Recall, Retention Interval and the Accuracy-Confidence Relation in Eyewitness Memory. *Applied Cognitive Psychology, 20*(7), 973–985. https://doi.org/10.1002/acp.1263.

Odinot, G., Wolters, G., & Lavender, T. (2009). Repeated partial eyewitness questioning causes confidence inflation but not retrieval-induced forgetting. *Applied Cognitive Psychology, 23*(1), 90–97. https://doi.org/10.1002/acp.1443.

Odinot, G., Wolters, G., & van Giezen, A. (2013). Accuracy, confidence and consistency in repeated recall of events. *Psychology, Crime & Law, 19*(7), 629-642. https://doi.org/10.1080/1068316X.2012.660152.

Olsson, N. (2000). A comparison of correlation, calibration, and diagnosticity as measures of the confidence–accuracy relationship in witness

identification. *Journal of Applied Psychology, 85(*4), 504–511. https://doi.org/10.1037/0021-9010.85.4.504.

Orbach, Y., Hershkowitz, I., Lamb, M. E., Sternberg, K. J., Esplin, P. W., & Horowitz, D. (2000). Assessing the value of structured protocols for forensic interviews of alleged child abuse victims. *Child abuse & neglect*, *24*(6), 733–752. https://doi.org/10.1016/s0145-2134(00)00137-x.

Orbach, Y., & Lamb, M. E. (2001). The relationship between within-interview contradictions and eliciting interviewer utterances. *Child Abuse & Neglect*, *25*(3), 323-333. https://doi.org/10.1016/S0145-2134(00)00254-4.

Otgaar, H., Howe, M. L., Merckelbach, H., & Muris, P. (2018). Who Is the Better Eyewitness? Sometimes Adults but at Other Times Children. *Current directions in psychological science*, *27*(5), 378–385. https://doi.org/10.1177/0963721418770998.

Oxburgh, G. E., Myklebust, T., & Grant, T. D. (2010). The question of question types in police interviews: a review of the literature from a psychological and linguistic perspective. *International Journal of Speech, Language and the Law*, *17*(1), 45-66. http://dx.doi.org/10.1558/ijsll.v17i1.45.

Oxburgh, G. E., Ost, J., & Cherryman, J. (2012). Police interviews with suspected child sex offenders: does question type, empathy, or interviewer training influence the amount of investigation relevant information obtained? *Psychology, Crime & Law*, *18*(3), 259-273. https://doi.org/10.1080/1068316X.2010.481624.

Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L., & Roberts R. D. (2002). The role of individual differences in the accuracy of confidence Judgments. *The Journal of General Psychology, 129*(3), 257-299. https://doi.org/10.1080/00221300209602099.

Palmer, M. A., Brewer, N., Weber, N., & Nagesh, A. (2013). The confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention. *Journal of Experimental Psychology: Applied, 19*(1), 55–71. https://doi.org/10.1037/a0031602.

Pansky, A., Koriat, A., & Goldsmith, M. (2005). Eyewitness Recall and Testimony. In N. Brewer & K. D. Williams (Eds.), *Psychology and law: An empirical perspective* (pp. 93–150). The Guilford Press.

Paterson, H. M., Kemp, R. I., & Forgas, J. P. (2009). Co-witnesses, confederates, and conformity: Effects of discussion and delay on eyewitness memory. *Psychiatry, Psychology and Law, 16*(1), 112–124. https://doi.org/10.1080/13218710802620380.

Paterson, H. M., Kemp, R. I., & McIntyre, S. (2011). Can a witness report hearsay evidence unintentionally? The effects of discussion on eyewitness memory. *Psychology, Crime & Law, 18*(6), 505-527. https://doi.org/10.1080/1068316X.2010.510117.

Patterson, T., & Pipe, M. E. (2009). Exploratory assessments of child abuse: Children's responses to interviewer's questions across multiple interview sessions. *Child abuse & neglect*, *33*(8), 490–504. https://doi.org/10.1016/j.chiabu.2008.12.012.

Pezdek, K., Sperry, K., & Owens, S. M. (2007). Interviewing witnesses: the effect of forced confabulation on event memory. *Law and Human Behavior*, *31*(5), 463–478. https://doi.org/10.1007/s10979-006-9081-5.

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

Potter, R., & Brewer, N. (1999). Perceptions of witness behaviour–accuracy relationships held by police, lawyers and mock-jurors. *Psychiatry, Psychology and Law, 6*(1), 97–103. https://doi.org/10.1080/13218719909524952.

Powell, M. B., & Snow, P. C. (2007). Guide to questioning children during the free narrative phase of an investigative interview. *Australian Psychologist 42*(1), 57–65. https://doi.org/10.1080/00050060600976032.

Price, H. L., & Roberts, K. P. (2011). The effects of an intensive training and feedback program on police and social workers' investigative interviews of children. *Canadian Journal of Behavioural Science, 43*(3), 235- 244. https://doi.org/10.1037/a0022541.

Quas, J. A., Wallin, A. R., Papini, S., Lench, H. C., & Scullin, M. H. (2005). Suggestibility, social support, and memory for a novel experience in young children. *Journal of Experimental Child Psychology, 91*(4), 315–341. https://doi.org/10.1016/j.jecp.2005.03.008.

Raghubir, P., & Menon, G. (2005). When and why is ease of retrieval informative? *Memory & Cognition, 33*(5), 821–832. https://doi.org/10.3758/BF03193077.

Robinson, M. D., & Johnson, J. T. (1996). Recall memory, recognition memory, and the eyewitness confidence–accuracy correlation. *Journal of Applied Psychology, 81*(5), 587–594. https://doi.org/10.1037/0021-9010.81.5.587.

Robinson, M. D., Johnson, J. T., & 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. https://doi.org/10.1037/0021-9010.82.3.416.

Robinson, M. D., Jonson, J. T., & Robertson, D. A. (2000). Process versus content in eyewitness memory monitoring. *Journal of Experimental Psychology, 6*(3), 207-22. http://dx.doi.org/10.1037/1076-898X.6.3.207.

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

Roediger, H. L., & Payne, D. G. (1982). Hypermnesia: The role of repeated testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*(1), 66–72. https://doi.org/10.1037/0278-7393.8.1.66.

Sanders, G. S. (1986). On increasing the usefulness of eyewitness research. *Law and Human Behavior,10*, 333-336.

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

Schmuckler, M. A. (2001). What is ecological validity? A dimensional analysis. *Infancy, 2*(4), 419–436. https://doi.org/10.1207/S15327078IN0204_02.

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simmons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology, 61*(2), 195-202. https://doi.org/10.1037/0022-3514.61.2.195.

Scott, A. J., Tudor-Owen, J., Pedretti, P., & Bull, R. (2015). How intuitive is PEACE? Newly recruited police officers' plans, interviews and self-evaluations. *Psychiatry, Psychology and Law, 22*(3), 355–367.https://doi.org/10.1080/13218719.2014.949397.

Semmler, C., & Brewer, N. (2006). Postidentification Feedback Effects on Face Recognition Confidence: Evidence for Metacognitive Influences. *Applied Cognitive Psychology, 20*(7), 895-916. https://doi.org/10.1002/acp.1238

Semmler, C., Brewer, N., & Wells, G. L. (2004). Effects of Post-identification feedback on eyewitness identification and nonidentification confidence. *Journal of Applied Psychology, 89*(2), 334–346. https://doi.org/10.1037/0021-9010.89.2.334.

Shapira, A. A., & Pansky, A. (2019). Cognitive and metacognitive determinants of eyewitness memory accuracy over time. *Metacognition and Learning, 14*, 437-461. https://doi.org/10.1007/s11409-019-09206-7.

Shapiro, L. R. (2006). The effects of question type and eyewitness temperament on accuracy and quantity of recall for a simulated misdemeanor crime. *Emporia State Research Study, 43*(1), 1-7.

Shapiro, L. R., Blackford, C., & Chen, C. F. (2005). Eyewitness Memory for a Simulated Misdemeanor Crime: The Role of Age and Temperament in Suggestibility. *Applied Cognitive Psychology, 19*(3), 267–289. https://doi.org/10.1002/acp.1089.

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

Sharps, M. J., Herrera, M., Dunn, L., & Alcala, E. (2012). Repetition and reconfiguration: Demand-based confabulation in initial eyewitness memory. *Journal of Investigative Psychology and Offender Profiling, 9*(2), 149–160. https://doi.org/10.1002/jip.1361.

Shaw, J. S. III (1996). Increases in eyewitness confidence resulting from postevent questioning. *Journal of Experimental Psychology: Applied, 2*(2) 126–146. https://doi.org/10.1037/1076-898X.2.2.126.

Shaw, J. S. III, & McClure, K. A. (1996). Repeated postevent questioning can lead to elevated levels of eyewitness confidence. *Law and Human Behavior, 20*(6), 629–653. https://doi.org/10.1007/BF01499235.

Shepherd, E. (1991). Ethical interviewing. *Policing, 7*, 42–60.

Shepherd, E. (1986). Conversational core of policing. *Policing, 2*, 294–303.

Soll, J. B. (1996). Determinants of overconfidence and miscalibration: The roles of random error and ecological structure. *Organizational Behavior and Human Decision Processes, 65*(2), 117–137. https://doi.org/10.1006/obhd.1996.0011.

Snook, B., Luther, K., Quinlan, H., & Milne, R. (2012). Let 'em Talk!: A Field Study of Police Questioning Practices of Suspects and Accused Persons. *Criminal Justice and Behavior*, *39*(10), 1328–1339. https://doi.org/10.1177/0093854812449216.

St-Yves, M. (2006). *The psychology of rapport: Five basic rules.* In T. Williamson (Ed.), *Investigative interviewing: Rights, research, regulation* (pp. 87–106). Willan Publishing.

Steblay, N. M. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. *Law and Human Behavior, 21*(3), 283–297. https://doi.org/10.1023/A:1024890732059.

Steblay, N. K., Wells, G. L., & Douglass, A. B. (2014). The eyewitness post identification feedback effect 15 years later: Theoretical and policy implications. *Psychology, Public Policy, and Law, 20*(1), 1–18. https://doi.org/10.1037/law0000001.

Sternberg, K. E., Lamb, M. E., Hershkowitz, I., Esplin, P., Redlich, A., & Sunshine, N. (1996). The relationship between investigative utterance types and the informativeness of child witnesses. *Journal of Applied Developmental Psychology 17*(3), 439 – 451. https://doi.org/10.1016/S0193-3973(96)90036-2.

Stone, C. B., Luminet, O., & Hirst, W. (2013). Induced forgetting and reduced confidence in our personal past? The consequences of selectively retrieving emotional autobiographical memories. *Acta Psychologica*, *144*(2), 250–257. https://doi.org/10.1016/j.actpsy.2013.06.019.

Sutherland, R., & Hayne, H. (2001). The effect of postevent information on adults' eyewitness reports. *Applied Cognitive Psychology, 15*(3), 249–263. https://doi.org/10.1002/acp.700.

Talarico, J. M., & Moore, K. M. (2012). Memories of 'the rivalry': Differences in how fans of the winning and losing teams remember the same game. *Applied Cognitive Psychology, 26*(5), 746–756. https://doi.org/10.1002/acp.2855.

Talarico, J. M., & Rubin, D. C. (2003). Confidence, not consistency, characterizes flashbulb memories. *Psychological Science*, *14*(5), 455–461. https://doi.org/10.1111/1467-9280.02453.

Talarico, J. M., & Rubin, D. C. (2007). Flashbulb memories are special after all; in phenomenology, not accuracy. *Applied Cognitive Psychology, 21*(5), 557–578. https://doi.org/10.1002/acp.1293.

Thomas, A. K., & Loftus, E. F. (2002). Creating bizarre false memories through imagination. *Memory & Cognition, 30*(3), 423–431. https://doi.org/10.3758/BF03194942.

Thompson-Cannino, J., Cotton, R., & Torneo, E. (2009). *Picking Cotton: Our memoir of injustice and redemption*. St. Martin's Press.

Thompson, W. B., & Mason, S. E. (1996). Instability of individual differences hi the association between confidence judgments and memory performance. *Memory & Cognition, 24*(2), 226-234. https://doi.org/10.3758/bf03200883.

Tomes, J. L., & Katz, A. N. (2000). Confidence–accuracy relations for real and suggested events. *Memory, 8*(5), 273-283. https://doi.org/10.1080/09658210050117708.

Turtle, J. W., & Yuille, J. C. (1994). Lost but not forgotten details: Repeated eyewitness recall leads to reminiscence but not hypermnesia. *Journal of Applied Psychology, 79*(2), 260–271. https://doi.org/10.1037/0021-9010.79.2.260.

Vallano, J. P., & Schreiber Compo, N. (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. https://doi.org/10.1002/acp.1789.

Verkampt, F., Dodier, O., Milne, B., & Ginet, M. (2019). An analysis of the quality of French investigative interviews with children: Age of the witness does matter. *Police Practice & Research*. https://doi.org/10.1080/15614263.2019.1658581.

Wade, K., Nash, R., & Lindsay, D. S. (2018). Reasons to doubt the reliability of eyewitness memory: Commentary on Wixted, Mickes, and Fisher (2018). *Perspectives on Psychological Science, 13*(3), 339–342. https://doi.org/10.1177/1745691618758261.

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau Q.F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R.D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review, 25*(1), 35-57. https://doi.org/10.3758/s13423-017-1343-3.

Walsh, D., & Milne, R. (2008). Keeping the PEACE? An Analysis of the taped interview performance of benefit fraud investigators within the DWP. Legal and Criminological Psychology, 13, 39-57. https://doi.org/10.1348/135532506X157179.

Waterhouse, G. F., Ridley, A. M., Bull, R., La Rooy, D. J., & Wilcock, R. (2016). Dynamics of repeated interviews with children. *Applied Cognitive Psychology, 30*(5), 713–721. https://doi.org/10.1002/acp.3246.

Weaver, C. A. III (1993). Do you need a "flash" to form a flashbulb memory? *Journal of Experimental Psychology: General, 122*(1), 39–46. https://doi.org/10.1037/0096-3445.122.1.39.

Weaver, C. A. III, & Krug, K. S. (2004). Consolidation-like effects in flashbulb memories: Evidence from September 11, 2001. *The American Journal of Psychology, 117*(4), 517-530. ttps://doi.org/10.2307/4148989.

Weber, N., & Brewer, N. (2004). Confidence-accuracy calibration in absolute and relative face recognition judgments. *Journal of Experimental Psychology: Applied, 10*(3), 156–172. https://doi.org/10.1037/1076-898X.10.3.156.

Wells, G. L., & Bradfield, A. L. (1998). "Good, you identified the suspect": Feedback to eyewitnesses distorts their reports of the witnessing experience. *Journal of Applied Psychology, 83*(3), 360–376. https://doi.org/10.1037/0021-9010.83.3.360.

Wells, G. L., & Seelau, E. P. (1995). Eyewitness identification: Psychological research and legal policy on lineups. *Psychology, Public Policy, and Law, 1*(4), 765–791. https://doi.org/10.1037/1076-8971.1.4.765.

Wheatcroft, J. M., Wagstaff, G., & Manarin, B. (2015b) The influence of delay and item difficulty on eyewitness confidence and accuracy. *International Journal of Humanities and Social Science Research 1*: 1–9.

Wheeler, R. L., & Gabbert, F. (2017). Using Self-Generated Cues to Facilitate Recall: A Narrative Review. *Frontiers in Psychology, section Theoretical and Philosophical Psychology.* https://doi.org/10.3389/fpsyg.2017.01830.

Winkielman, P., Schwarz, N., & Belli, R. F. (1998). The role of ease of retrieval and attribution in memory judgments: Judging your memory as worse despite recalling more events. *Psychological Science, 9*(2), 124–126. https://doi.org/10.1111/1467-9280.00022.

Winningham, G. R., & Weaver, C. A. III (2000). The effects of pressure to report more details on memories of an eyewitness event. *European Journal of Cognitive Psychology, 12*(2), 271-282. https://doi.org/10.1080/09541446.2000.10708707.

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

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

Wolfman, M., Brown, D., & Jose, P. (2016). Talking past each other: Interviewer and child verbal exchanges in forensic interviews. *Law and Human Behavior, 40*(2), 107-117. https://doi.org/10.1037/lhb0000171.

Wolters, G., & Goudsmit, J. J. (2005). Flashbulb and event memory of September 11, 2001: Consistency, confidence and age effects. *Psychological Reports, 96*(3 Pt 1), 605-619. https://doi.org/10.2466/pr0.96.3.605-619.

Wright, A. M., & Alison, L. (2004). Questioning sequences in Canadian police interviews: Constructing and confirming the course of events? *Psychology,*

Crime & Law, 10(2), 137-154. https://doi.org/10.1080/1068316031000099120.

Yaniv, I., Yates, J. F., & Smith, J. K. (1991). Measures of discrimination skill in probabilistic judgment. Psychological Bulletin, 110(3), 611–617. https://doi.org/10.1037/0033-2909.110.3.611.

Yarmey, A. D. (1993). Adult age and gender differences in eyewitness recall in field settings. *Journal of Applied Social Psychology, 23*(23), 1921-1932. https://doi.org/10.1111/j.1559-1816.1993.tb01073.x.

Yarmey, A. D., & Yarmey, M. J. (1997). Eyewitness recall and duration estimates in field settings. *Journal of Applied Social Psychology, 27*(4), 330–344. https://doi.org/10.1111/j.1559-1816.1997.tb00635.x.

Appendix A

Confidence in Engaging with the Criminal Justice System

Imagine that the event you've seen was a real robbery and you witnessed it while you were in the store looking for a DVD, based on your memory for this incident, how confident would you be to:

a. Talk with a police-officer?

10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

b. Provide and sign a witness statement? *A witness statement is a document recording the evidence of a person, which is signed by that person to confirm that the contents of the statement are true.*

10 % 20% 30% 40% 50% 60% 70% 80% 90% 100%

c. Go to an identification parade? *The identification parade involves watching a three 3 minutes video, made up from a number of volunteers who look similar to the suspect. Each clip plays for approximately 15 seconds. Each person starts off looking directly at the camera, then turns the head from left to right, finishing by facing the camera again.*

10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

d. Provide evidence in Court? *Sometimes witnesses can be requested to go to court and give evidence in a trial. This requires the witness to review the statement provided to the police officer after the incident, and then answer questions, under oath, asked by the prosecution and defence.*

10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Appendix B

Participant information sheet and consent form

The study you are taking part in investigate the impact of different styles of interviewing on people's confidence in their own memory quality. We are also interested in the effect of repeated questioning on people's memory performance.

As a participant you will be asked to watch a short video clip and afterward to recall what you saw in the video in a face to face interview. Before and after the interview you will be given a short questionnaire asking you to rate how confident you are in your memory for the event witnessed.
Finally, you'll be asked to write down what you can remember about the video clip.
Your participation is entirely voluntary. You may decide to stop being a part of the research study at any time without explanation or consequence. The data collected do not contain any personal information about you and your identity will remain anonymous. Please note that you are free to withdraw your data from analysis up until the point we have entered it into the dataset in an anonymous format. At this stage we will have no way to associate it with you, and it will no longer be possible to withdraw it.
The Departmental Ethics Committee has reviewed and approved this research.

If you have any questions about the study, please feel free to ask.
Alessandra Caso: acaso006@gold.ac.uk
Supervisor: Fiona Gabbert (f.gabbert@gold.ac.uk)

If you are happy to take part, please complete the following:

AGE:
GENDER:
STUDENT ID:
Signed Date:

_____ _____

Appendix C

Debrief form

The Criminal Justice System relies on people's memory to assert the truth and convict offenders for the crimes they are responsible for. Interviewing witnesses is a crucial part of an investigation. However human memory is a fragile process and it can easily be influenced by a number of external variables. People's confidence in their memory for an event can also be influenced. The present study therefore investigates the impact of different interviewing styles on memory confidence. Research in cognitive psychology has looked at the effect of different recall tests on memory confidence in an experimental setting.

In real life an eyewitness's memory is in first place tested in a face to face interview with a police officer. Research in applied psychology in a forensic field has developed advanced techniques for interviewers with the aim of both reducing the impact of bad interview practice on witnesses' memory and providing supportive techniques to help the interviewee in the recollection task.

In this study "Poor Practice interview" or a "Good Practice interview" was used to elicit information with the aim to explore their influence on memory confidence. A final memory test will enable an insight into changes in recall accuracy, and amount of information provided in relation to confidence levels expressed.

For further queries regarding the study please contact me:
Alessandra Caso Email: a.caso@gold.ac.uk
Supervisor: Fiona Gabbert Email: f.gabbert@gold.ac.uk

Hope you enjoyed it!
Thank you for taking part in my study.

Appendix D

Coding Scheme

General rules:

- Correct:
- info presented in the video
- correct attribution of subject to a specific action
- Incorrect:
- Info not presented in the video
- Incorrect attribution of subject to a specific action

NO NOT SCORE REPETITIONS

NO NOT SCORE NOT CLEAR INFORMATION

NO NOT SCORE OPINIONS OR ASSUMPTIONS

Actions (a):

- What characters do: (he stayed (a), he entered (a)).
- Movements from an area to another: (they went (a) towards the cashier, he came (a) back to the front).

  DO NOT SCORE unclear action or opinion: (they seemed to be looking for something *(do not score)*).
- Conversations:
- What people actually say: (he said: 'it's a robbery!' (a), he said it was a robbery (a)).

  DO NOT SCORE unclear information: (he said something *(do not score)*).
- People who are spoken to: (the robber said to the cashier (a); he said to his mate (a)).

Location (l):

- Details about the store:
- Specific part of the store related to an action: (he went into a different area (l)).

  DO NOT SCORE: description of the store if not related to an action: (the robbery happened in a block buster store *(do not score)*).

- Complex descriptions that help picturing the store or a part of it: (they went towards the area where the cashier can't see them (l)).
- Characters' position in the store: (he stayed by the door (l)).
- Characters' position used to indicate a part of the store: (he went toward the guy at the tills (l)).

Person description (pd):

- The character described: (one (pd) of the two robbers was wearing a jeans).
  DO NOT SCORE a re-mentioned person when it has already been coded: (they (2pd) were dressed in black, they *(do not score)* had jeans and jackets).
  SCORE attribution of descriptors to the right person: ((they (pd) were dressed in black, the blond one (pd) had a green hoodie).
- Clothes: (he was wearing a jeans (pd), t-shirt (pd)).
- Colours: (black (pd) jacket, blue (pd) jeans).
- Fabric: (wool (pd) hat, lather (pd) jacket).
- Objects carried: (the pull out their masks (pd)).
- Other descriptors: (light (pd) brown, long (pd) brown hair, up to the shoulder (pd))
- Objects mentioned: (he gave him a bag (pd)).

Appendix E

Cued Recall condition

My name is Alessandra, thank you for taking part in this experiment. We are now going to talk about the video you saw. I will ask you questions about the setting in which the event happened, the people involved, what they did, and what they looked like.

Setting:

1. Which colour was the store decorated with?
2. Where was the customer standing when the robbers were looking at the DVDs?
3. There was any window? If yes where was it located?
4. What brand was advertised on the vending machine?
5. By which brand was the bag given to the shopkeeper by the robber?
6. What price was advertised on the poster hanged above the counter?
7. How many tills did the shopkeeper take the money from?
8. What kind of light devices lightened the store?

Actions:

1. What was the shopkeeper looking at when the robbers came into the shop?

1. Which side did the robbers go to cover up their faces?
2. What did the shopkeeper do after giving the money to the robber?
3. What happened to the customer before the robbers left the store?
4. Where was the robber with the balaclava standing while the other one was talking with the shopkeeper?

5. Which side did the robbers go after leaving the store?

6. What did the robbers do after the customer got pushed and before they left the store?

7. In which pocket did the robber hide his hand when he said that he had a gun?

8. Which of the robbers did push the customer?

Person descriptions:

1. What was the customer, who first come into the shop wearing?

2. What colour was the shopkeeper's hair?

3. What was the light -haired robber wearing?

4. What was the brown - haired robber wearing?

5. What did the light - haired robber use to cover his face up?

6. What kind of accent did the robbers have?

7. What colour was the shopkeeper's t-shirt?

8. What did the brown - haired robber use to cover his face up?

Appendix F

Difficult Cued Recall questions

1. How many doors were in the store?
(One)
2. What kind of Dvds was the customer browsing when the robbers were walking through the aisles?
(New releases)
3. What did the light - haired robber touch during the robbery? (List all)
(Door), *(if they answered correctly and no error is reported score as correct)*
4. What pattern was on the dark - haired robber's jacket?
(Two white stripes)
5. Which of the people present in the store left the shop before? (If one of the robbers, please specify if dark or light - haired)
(Dark haired robber)
6. What colour was the clothing the light - haired robber was wearing underneath the hoodie?
(White)
7. How did the robbers communicate to each other they were ready to go to the till and rob the store?
(They nodded/look at each other)
8. What objects in the store did the dark - haired robber touch? (List all)
(Two DVDs, the till, the computers, and the door) *(if they answered correctly and no error is reported score as correct)*
9. What happened to the bag after the robbery?
(It was not displayed)
10. What kind of shoes was the light - haired robber wearing?
(Grey/brown trainers)
11. Who touched the customer? (List all), (if robber, please specify if dark or light - haired)
(Light - haired robber)
12. What kind of shoes was the dark - haired robber wearing?
(Black trainers)

13. Who did the dark - haired robber speak to? (List all)

(The shopkeeper and the light - haired robber)

14. What did the robber throw at the shopkeeper when he first said "this is a robbery"?

(Two DVDs)

15. What price was advertised on the posters hanging all over the store?

(£4.99)

16. Who did the light - haired robber speak to? (List all)

(All those present in the video)

Easy Cued Recall questions

1. Where did the shopkeeper direct the customer?

(To his left)

2. What kind of item was mainly sold in the store?

(DVDs)

3. From how many tills did the shopkeeper take the money from?

(Two)

4. Where did the shopkeeper take the bag from?

(The dark - haired robber handed it over)

5. Where was the shopkeeper standing?

(Behind the till)

6. Where did the dark - haired robber go to cover himself up?

(In the room where the DVDs are/ left/back of the store)

7. What colour was the customer's hat?

(Grey/black)

8. What side of the store did the light - haired robber go to cover himself up?

(In the room where the DVDs are/ left/back of the store)

9. What was the customer wearing on the upper part of the body?

(A hoodie/jumper)

10. What kind of bag did the shopkeeper use to put the money in?

(Tesco/carrier bag)

11. What colour was the dark - haired robber's trousers?

(Dark blue/black)

12. Where was the customer while the robbery was going on?

(In the separate/DVD room)

13. What did the light - haired robber use to cover himself up?

(Balaclava)

14. What did the shopkeeper do immediately after putting the money in the bag?

(He lays down on the floor)

15. What colour was the dark - haired robber's jacket?

(Dark blue/black)

16. From how many tills did the robber demand the money?

(Two)

Appendix G

Neutral Cued Recall questions (Study 3a)

1. What colour was the shopkeeper t-shirt?

(Blue)

2. What colour was the store decorated in? (List all)

(Blue, yellow, white) *(at least blue and yellow – if an incorrect colour is present score as incorrect)*

3. What was the light - haired wearing underneath his jacket?

(Hoddie/ jumper)

4. What kind of trousers was the dark - haired robber wearing?

(Baggy/ tracksuit/ sport trousers)

5. What colour was the shopkeeper hair?

(Blond, dark blond/ light brown)

6. Where was the dark - haired robber standing during the most part of the robbery?

(By the till)

7. What did the dark - haired robber do immediately before leaving the store?

(Took off the mask/ talk to themselves/ shouting)

8. What did the light - haired do immediately after before leaving the store?

(Took off the mask/ talk to themselves/ shouting)

Appendix H

Neutral Cued Recall questions (Study 3b)

1. What colour is the light - haired robber's jacket?
   (Black)
2. Indicate at least one individual in the shop who is wearing a hooded jumper (If robber specify which one)
   (Customer or light - haired robber)
3. Which person in the shop is wearing a short sleeves top?
   (Shopkeeper)
4. What colour is the shopkeeper's hair?
   (Brown/blond, blond, light brown, dark blond)
5. Which robber does step first in the DVDs section of the shop?
   (Light - haired robber)
6. 45. Where was the light - haired robber standing during the majority of the robbery?
   (Near the door)
7. Which robber does enter the shop first?
   (Light - haired robber)
8. Does the light - haired robber look at the shopkeeper when entering the store?
   (No)
9. Which hand does the robber hide in his pocket when he says that he has a gun?
   (Right)
10. With which hand does the shopkeeper empty the first register?
    (Right)
11. With which hand does the shopkeeper empty the second registers?
    (Right)
12. How many main rooms are there in the shop?
    (Two)
13. Which individuals in the shop are wearing a hat?
    (Customer and one robber)
14. Which robber is taller?
    (Dark - haired robber)

15. With which hand does the robber hand the bag over to the shopkeeper?
    (Right)
16. Where was the dark - haired robber standing during the majority of the robbery?
    (In front of the till)
17. Where do the robbers have their masks hidden?
    (In the pocket/in the coats)
18. Where exactly does the shopkeeper put his hands when laying down on the floor?
    (On the floor/ above his head/ in front of him)
19. Does the customer look at the robbers when they are in the DVDs section of the shop?
    (No)
20. Which individuals do leave the shop? (If robber specify which one)
    (Both robbers)
21. Which individuals do touch the bag? (List all. If robber, specify which one)
    (Dark - haired robber and shopkeeper)
22. What were the <u>main</u> colours of the store? (List at least two)
    (Blue, yellow, creamy yellow, grey)
23. Which person in the shop has the longest hair?
    (Shopkeeper)
24. How does the shopkeeper open the cash registers?
    (Typing/ inserting codes on keyboard)
25. Do the robbers make eye contact with the shopkeeper when they first enter the shop?
    (No)
26. How many people leave the shop by the end of the robbery?
    (Two)
27. Where does the customer fall off when he is pushed over?
    (Stairs, isles)
28. How was the light outside the shop?
    (Dark/no light)
29. What colours was the bag? (List at least two colours)
    (White, blue, red)
30. What is the shopkeeper's ethnicity?
    (White/ Caucasian)

31. What does the robber with the hat use to cover up the lower part of his face?
    (Scarf/ his hoody/collar/ neck of top)
32. Which individual does NOT step into the DVDs section of the store?
    (Shopkeeper)
33. Who did touch the counter?
    (Dark - haired robber or the shopkeeper)
34. How many customers are there in the shop?
    (One)
35. Which person in the shop has dyed hair?
    (Light - haired robber)
36. Which robber does put on his disguise first?
    (Light- haired robber)
37. Does the customer speak?
    (Yes)
38. Which individuals do speak at least ones?
    (All)
39. What part of the customer's body does the robber hit when he pushes him over?
    (Shoulder, chest, upper body)
40. How many men are there in the shop?
    (Four)
41. Which robber is wearing a full - face mask?
    (Light - haired robber)
42. How many cash registers are there in the shop?
    (Two)

Appendix I

Predetermined 5Wh Cued Recall questions

Light haired robber:

Person description

1. How old was the light- haired robber?

(Between 18-30)

2. What was the light- haired robber ethnicity?

(Caucasian/ white/ white British)

3. Where was the light- haired robber standing during the majority of the robbery?

(In front to the door/ behind the other robber/at the entrance)

4. How long was the light- haired robber's hair?

(Short not too short – longer than the other – up to his hears)

5. What colour exactly was the light- haired robber's hair?

(Blond/ strawberry blond/ dyed blond/ light blond)

Disguise

6. What type of disguise/mask is the light - haired robber wearing?

(Balaclava/ sky mask/ full face mask/ mask with holes)

7. What colour was the light - haired robber's disguise/mask?

(Black)

8. Where does the light – haired robbers have his disguise/mask hidden?

(Pocket)

Clothing

9. What colour was the light - haired robber's jacket?

(Black)

10. What type of materials was the light - haired robber's jacket?

(Leather)

11. What is the light - haired robber wearing underneath his jacket?

(Jumper – hoody)

12. What colour was the light - haired robber's jumper?

(Cream/ creamy green/ light brown)

13. What kind of trousers was the light - haired robber wearing?

(Jeans – baggy)

14. What colour were the light - haired robber's trousers?

(Blue)

15. What kind of shoes was the light - haired robber wearing?

(Trainers)

16. What colour were the light - haired robber's shoes?

(Brownish – green)

Actions

17. What did the light - haired robber say to the shopkeeper?

(Did not talk to him)

18. What did the light - haired robber do at the door?

(Watching out/ guarding)

19. Where did the light - haired robber take the disguise off?

(Near the door/ at the entrance/ in the main room)

20. When did the light - haired robber take the disguise off?

(Before leaving / after the robbery)

Dark haired robber:

Person description

22. How old was the dark - haired robber?

(Between 18-30)

22. What was the dark - haired robber's ethnicity?

(Caucasian/ white/ white British)

23. Where was the dark - haired robber standing during the majority of the robbery?

(Front of the counter/ near the counter)

24. How long was the dark - haired robber's hair?

(Short – very short)

25. What colour exactly was the dark - haired robber's hair?

(Brown – black)

Disguise

26. What type of disguise/mask was the dark - haired robber wearing?

(Hat and scarf/ turtleneck)

27. Where did the dark - haired robbers have his disguise/mask/hat hidden?

(Pocket)

28. What type of hat was the dark - haired robber wearing?

(Beanie/ woolly/ winter hat)

29. What colour was the dark - haired robber's hat?

(Black/ dark grey)

30. What part of the face did the dark - haired robber's scarf/turtleneck cover up?

(Up to his nose)

Clothing

31. What type of materials was the dark - haired robber's top?

(Waterproof/ cotton/ rain jacket)

32. What colour was the dark - haired robber's top?

(Black/ dark blue)

33. How many strips were there on the dark - haired robber's top?

(Two)

34. Where exactly on the haired robber's top were there the strips?

(His left)

35. What colour were the strips on the dark - haired robber's top?

(Grey/ white)

36. What type of trousers was the dark - haired robber wearing?

(Baggy/ tracksuit)

37. What colour were the dark - haired robber's trousers?

(Black/ dark blue)

38. What type of shoes was the dark - haired robber wearing?

(Trainers)

39. What colour were the dark - haired robber's shoes?

(Black)

Actions

40. With which hand did the dark - haired robber bang the computer monitor?

(Right)

41. How many DVD case(s) did the dark - haired robber throw at the shopkeeper?

(Two)

42. What did the dark - haired robber say to the shopkeeper?

(This is a robbery/ give me the money/ hurry up)

43. Where did the dark - haired robber take the disguise off?

(Near the door/ at the entrance/ in the main room)

44. When did the dark - haired robber take the disguise off?

(Before leaving/ after the robbery)

The customer

Person description

45. How old was the customer?

(Between 18-30)

46. What was the customer's ethnicity?

(Caucasian/ white/ white British)

47. Where was the customer standing during the majority of the robbery?

(Other room/ DVD room/ small room)

48. What colour was the customer's hair?

(Dark brawn/ black)

49. What type of beard did the customer have?

(Stubble/ short/ moustache)

Clothing

50. What type of top was the customer wearing?

(Tracksuit/ jumper)

51. What colour was the customer's top?

(Black)

52. How many strips were there on the customer's top?

(Two)

53. What colour were the strips on the customer's top?

(White)

54. How long was the customer's hair?

(Short)

55. What type of hat was the customer wearing?

(Beanie/ woolly hat/ winter hat)

56. What colour was the customer's hat?

(Black/ dark grey)

57. What type of trousers was the customer wearing?

(Tracksuits/ baggy/ joggers)

58. What colour were the customer's trousers?

(Black)

Actions

59. What did the customer ask the shopkeeper for?

(DVDs/ direction)

60. What exactly did the customer say to the robber?

(On the left)

The shopkeeper

Person description/position

61. How old was the shopkeeper?

(Between 18-30)

62. What was the shopkeeper's ethnicity?

(Caucasian/ white/ white British)

63. Where was the shopkeeper standing during the majority of the robbery?

(Behind the counter)

64. What colour was the shopkeeper's hair?

(Blond)

65. How long was the shopkeeper's hair?

(Long/ below his ears/ above the shoulders)

Clothing

66. What kind of top was the shopkeeper wearing?

(T- shirt/ polo shirt/ shirt/ blockbuster top)

67. What colour was the shopkeeper's top?

(Blue)

68. What kind of trousers was the shopkeeper wearing?

(Smart/ work trousers/ Chinos)

69. What colour were the shopkeeper's trousers?

(Grey)

Actions

70. In which direction did the shopkeeper direct the customer?

(To the left)

71. What did the shopkeeper say to the customer?

(To the left)

72. What did the shopkeeper say to the robber?

(Ok)

73. With which hand did the shopkeeper empty the registers?

(Right)

74. In what position did the shopkeeper lay on the floor?

(On his stomach/ flat)

75. Where did the shopkeeper put his hands when laying down?

(Above his head/ near his head/ on the floor)

Bag

76. What type of bag was used?

(Carrier bag/ shopping bag/ plastic bag/ supermarket bag)

77. What material was the bag used made of?

(Plastic)

78. What brand was the bag?

(Tesco)

79. What was main colour of the bag?

(White)

80. Where did the shopkeeper take the bag from?

(From the robber)

81. What colours was the pattern seen on the bag?

(Red/ Blue)

82. Who exactly did touch the bag (list them all, if robber - specify which one)?

(Dark - haired robber/ shopkeeper)

The confrontation

83. How was the customer confronted by the robber?

(Pushed)

84. What part of the customer's body did the robber touch when he pushed him?

(Chest/ shoulder)

General

85. How many people did leave the shop by the end of the robbery?

(Two)

86. How many people were left in the shop by the end of the robbery?

(Two)

87. How many computers were there in the shop?

(Two)

88. How many tills were there in the shop?

(Two)

89. How many aisles were there in the DVDs section of the store?

(Two)

Appendix J

Participant information sheet and consent form

This study aims to investigate the effect of different types of questions on eyewitness' confidence. Research in psychology has showed that open-ended questions are more effective than closed questions in eliciting detailed and reliable information for witnesses. In this study we explore this finding further by looking at how the aforementioned types of question contribute to shape confidence in the quality of own memory.

As a participant you will be asked to watch a short video clip and afterward to recall what you saw in the video with both open-ended and closed-questions. In different stages of the study you will be asked to rate how confident you are in your memory. A researcher will give you detailed instruction about each task you are asked to complete, and they will be available to answer any question you might have. After the study you will be fully debriefed about the study procedure, and the group you were assigned to.

Your participation is entirely voluntary. You may decide to stop being a part of the research study at any time without explanation or consequence. The data collected do not contain any personal information about you and your identity will remain anonymous. Please note that you are free to withdraw your data from analysis up until the point we have entered it into the dataset in an anonymous format. At this stage we will have no way to associate it with you, and it will no longer be possible to withdraw it.
The Departmental Ethics Committee has reviewed and approved this research.

If you have any questions about the study, please feel free to ask.
Alessandra Caso: acaso@gold.ac.uk
Supervisor, Prof. Fiona Gabbert; f.gabbert@gold.ac.uk

If you happy to take part, please complete the following:

AGE:

GENDER:

STUDENT ID:

Signature: _____ Date: _____

Appendix K

Debrief

The Justice System relies on witnesses' memory to assert the truth and convict offenders for the crimes they are responsible for. Interviewing witnesses is a crucial part of an investigation, however human memory is a fragile process and can easily be influenced by the type of questions asked. Research has showed that some types of question can have detrimental effects on memory, for example leading questions that introduce information that the witness did not mention can encourage the witness to assimilate that information in their original memory.

After thirty years of research in eyewitness memory, we have learned what types of question are less likely to change a witness's memory, and which are more effective in eliciting reliable information. For example, there is a general consensus that open-ended questions are more likely to elicit accurate information than closed-questions, or yes/no questions, or 5Wh. type of questions (why, what, who, where, when). However, the extent to which different types of question change people's confidence is yet to be explored.

As human beings we have the ability to think about our cognition. In particular, research on metamemory has studied the array of thoughts and beliefs that we hold regarding our memory, and how we generate them. In particular the experienced-based approach speculates that confidence depends on the retrieval process. When we are asked a question, we retrieve some possible answers, those are associate with a number of cues (i.e. vivideness, fluency, completeness). Confidence in the accuracy of a retrieved answer is built on these cues: i.e., we are likely to be more confident in an answer that has come to mind quickly and seems clear and detailed.

In this study we tested participants' memory for the event seen with different types of question: 'Free-Recall', and 'Cued-Recall'; furthermore we divided the Cued Recall test in questions that are related with information reported in the Free Recall, and questions that are not related to the information reported in the Free Recall. Across three conditions we manipulated the order in which these questions were asked: group a) answered Cued Recall questions before the Free Recall test, group b) answered the Free Recall test first, and subsequently answered Cued Recall questions (related

questions before and unrelated afterward), group c) answered a Free Recall test and then Cued Recall questions (unrelated questions first and subsequently related questions afterward). Confidence was measured throughout the study.

Our hypothesis is that when freely reporting a memory the participants' attention is not directed towards information not know, because they tend to only report what comes to mind, and confidence in expected to be higher. On the contrary when answering Cued Recall questions, attention is directed towards information that might not be known, and confidence is expected to be lower. We also predict that confidence for information elicited through related Cued Recall questions will be higher than confidence for information elicited through unrelated Cued Recall questions. This is because the former are focused on information initial associated with high confidence (i.e. freely reported in the Free Recall).

For further queries regarding the study please contact me:

Alessandra Caso, Email: a.caso@gold.ac.uk
Supervisor: Fiona Gabbert, Email: f.gabbert@gold.ac.uk

Hope you enjoyed it!

Thank you for taking part in my study.