THESIS EXAMINING THE ACCURACY OF WITNESS STATEMENTS FOR REAL

WORLD CRIMINAL EVENTS

By

ALEXANDRA LILLEY

A tbesis submitted to the University of Birmingham for the degree of

DOCTORATE IN FORENSIC PSYCHOLOGY PRACTICE (ForenPsyD)

School of Psychology

College of Life and Environmental Sciences

University of Birmingham

2023

UNIVERSITY OF BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

Table of Contents

List of Figures

List of Tables

Abstract

Eyewitness testimony is paramount to the solving of crimes by giving police forensically relevant evidence to apprehend suspects. However, eyewitness memory errors are also a leading cause of wrongful convictions (Innocence Project, 2020) which points to the importance of gaining accurate memory information from witnesses. Accuracy can be affected by both the witness being mistaken in their recollection or actively being dishonest. This thesis examines how the accuracy of memory reporting can be affected and discusses theory explaining memory reporting. This is achieved through examining the current field literature on memory reporting in a systematic literature review (Chapter Two). Next, a secondary data analysis on real world crimes examines the potential mechanisms that may underlie memory reporting in the field to enhance accuracy (Chapter Three). A psychometric critique of Statement Validity Analysis, with particular focus on the Criteria-Based Content Analysis component of this tool, is next discussed, which attempts to determine the credibility of witness statements (Chapter Four). Finally, the conclusions and the theoretical and practical implications of the thesis are discussed in Chapter Five. The thesis identifies future research areas needed to better understand memory reporting in the field and the possible theoretical mechanisms underlying witness memory reporting. It also suggests implications for practice, including for UK police interviewing practice, to improve the accuracy and completeness of witness statements. Ultimately, such implications should improve the accuracy of testimony, reduce wrongful convictions, and ensure that witnesses are not underestimated.

Acknowledgements

I would like to thank both my supervisors who have supported me throughout the last three years. Firstly, Dr Heather Flowe, who passed on all her wisdom to me for the first half of the doctorate. I will always be grateful for your support. Also, to Dr Melissa Colloff who took me on and always showed so much kindness as well as knowledge. Thank you for supporting me both personally and academically.

To my family who have all supported me throughout this process. My parents have been there to read through endless drafts of this thesis and were always there to emotionally support me through this process. Thank you. Also, to my brothers and grandparents who I cherish dearly.

I would also like to thank my fellow ForenPsyD course-mates. We have both struggled and thrived together and it has been such an honour to go through this journey with you all. It has been a pleasure to see the amazing psychologists (and researchers) you have turned into.

Lastly, I want to thank my amazing friends for always being there and providing a source of comfort in this process. Thank you for distracting me and giving me laughs when needed. Georgia C, you have been there for me throughout my whole life, and I can only thank you for standing by me throughout everything, it means the world. Georgia B, thank you for being there through all the stress of this process; I am so proud of the people we have become together. Megan, my fellow forensic psychology buddy, I am so grateful for you and know you will thrive at whatever path you go down.

Chapter One

It has been estimated that there were 1.6 million violent offences in England and Wales in the year ending March 2021, with 21.3% of the adult population having been a victim of a violent crime in their lifetime (ONS, 2021). However, convictions for these crimes have fallen. For example, although rape charges have increased by almost five per cent over the year ending March 2021, convictions have decreased (CPS, 2021). This may suggest that there is often not enough evidence to gain a legitimate conviction. Bystanders and victims who witness crimes (hereafter *witnesses*) are often interviewed by the police as part of the investigation and a witness may also be asked to testify in court. The quality of witness statements and testimony can play an important role in securing a conviction (Goldstein et al.,1989). It is imperative therefore that we identify factors that can influence the accuracy of witness memory.

Eyewitness testimony is a term that refers to the recollection of an event from memory which the individual witnessed and is frequently associated with criminal events (Junnarkar & Lakhani, 2021). Witness accounts may involve describing an individual who was thought to be involved in the crime to help police apprehend a suspect. Witness accounts are also used in court to help juries come to a decision in terms of innocence or guilt of an apprehended person. Numerous studies have shown eyewitness testimony to be a strong form of evidence which the juries consider when making a decision (Lindsay et al.,1981; Brewer & Burke, 2002; Lieberman et al., 2008). However, this can be problematic as jurors often over evaluate the reliability of eyewitness testimony (Brigham & Bothwell, 1983). Indeed, eyewitness error has been found to be one of the leading causes of wrongful convictions (Wise et al., 2014).

What might cause inaccurate eyewitness testimony? Loftus (1996) states that the memory process consists of three stages: the acquisition stage, the retention stage, and the retrieval stage, which all can alter what is ultimately remembered. In the acquisition stage, the witness may focus on certain

aspects of the crime to the exclusion of other aspects which determines what features they can remember. In the retention stage, some of the encoded information may undergo changes by the witness undertaking post event activities, such as interviews from the police or discussing the event with other witnesses. This stage is often unconscious. In the retrieval stage, the witness uses both the information they acquired in the encoding stage and any information they may have gained subsequently to recall the event (Loftus, 1996). Therefore, this affects the information that the witness can recall to the police.

Research psychologists conduct studies on eyewitness memory with the ultimate goal of improving procedures in the criminal justice system and enhancing witness memory accuracy. There are different types of studies. Laboratory studies are experiments that manipulate an independent variable (e.g., factors that can influence memory reporting) while controlling for extraneous variables to measure the dependent variable (e.g., memory reporting). In a typical laboratory study, participants watch a mock crime, and are then asked to recall the event. Researchers typically use this method because this environment allows the researcher to create and control the mock events (i.e., they can establish cause and effect relationships due to the controlled environment; Chae, 2010) and hence participant memory can be scored for accuracy and completeness when manipulating various factors (such as stress, distance, type of interview, etc). From laboratory studies, it has been concluded that eyewitness memory is often inaccurate under certain circumstances (Wells & Olson, 2003). Criminal justice experts also have beliefs that certain factors of a crime can affect the accuracy of recall, such as presence of a weapon, alcohol intoxication and the age of the witness (Kassin et al., 2001). However, laboratory studies do lack external validity, which is the extent to which you can generalise findings of a study to other situations. In contrast to laboratory studies, field studies do not take place in a controlled environment. A type of field study is a natural/ quasi experiment which does not manipulate an independent variable, but instead involves observing changes that are associated with a naturally occurring independent variable.

Lastly, archival studies involve studying already-collected data to test associations between variables. Both of which have the benefit of a high level of external validity.

Eyewitness memory research has been predominantly conducted in the laboratory (e.g., Loftus et al., 1987). It is therefore important to conduct field studies to see if results from laboratory studies hold in the real world. The current thesis focuses on studies of memory reporting in real-world crimes and hence these shall be referred to as field studies from hereafter.

Most of this literature, however, assumes that witnesses are intending to provide honest accounts. However, dishonesty also affects the accuracy of statements and hence this needs to be considered as well. For many crimes there is often no physical evidence and no other witnesses to the crime meaning that a victim account can be the sole evidence for the case (e.g., allegations of sexual abuse; Vrij, 2015). With the importance placed on testimonies in these cases in investigations and court proceedings, it is important that accounts are accurate and honest to prevent wrongful convictions of individuals.

Laboratory research has led to the development of procedures that police use in practice to maximise the accuracy of witness testimony. This includes the Achieving Best Evidence (ABE) guidance (Ministry of Justice, 2011) which sets out evidence-based recommendations for conducting police interviews in the UK. The guidance outlines that police interviewers should ask witnesses to provide a free recall account, and report everything, no matter how trivial, in their own words without interruption. A questioning phase takes place, in which witnesses are asked open-ended and closed questions until all the information relating to the crime event has been obtained (Ministry of Justice, 2011). Other procedures used in practice include the Statement Validity Assessment (SVA) (Undeutsch, 1984; 1989), which is used to determine the creditability of child and adult witness testimonies, particularly in sexual abuse cases and is accepted as evidence in

some North American and several West European court procedures (Goodman & Melinder, 2007). The goal of both procedures is to enable witnesses to provide accurate and complete testimony, through reducing either mistaken or dishonest information.

Whilst laboratory-based research has advantages in that it has allowed for the development of these procedures, the sole reliance on laboratory research to make inferences about the accuracy of real-world witnesses has been repeatedly criticised for lacking external validity (Holleman et al., 2020; see also chapter 2). Supporting this, research that has been conducted in the field has found that the variables shown in laboratory studies to affect memory reporting for crime events (e.g., retention interval, presence of a weapon) do not necessarily affect memory reporting in the real world (e.g., Yuille & Cutshall, 1986; Cooper et al., 2002). This is discussed further in Chapter Two.

One explanation for why memory may be less effected in real-world scenarios compared to in laboratory studies is provided by what is termed the quantity accuracy trade-off framework (Koriat & Goldsmith, 1996). This theory states that when individuals are questioned they use a monitoring mechanism that assesses the correctness of a potential answer, and a control mechanism that determines whether to volunteer it. The participant sets a response criterion, and they report the response if it exceeds this criterion. That is, an individual only provides an answer if the certainty in their response exceeds their threshold. The quantity accuracy trade-off also considers the gain for providing correct information relative to the cost of providing wrong information, which would further alter the criterion (Koriat & Goldsmith, 1996). For example, as there are potentially detrimental consequences of getting information wrong, such as a wrongful conviction, an eyewitness may only report information of which they are very sure.

Thesis Aim

Proper treatment and interviewing of witnesses is likely to provide the most accurate and complete memory evidence used to apprehend guilty perpetrators and prevent wrongful convictions of innocent individuals. The findings from laboratory studies that state that certain factors can affect accuracy of memory reporting for crimes has been well established and has led to the development of procedures that are used in the criminal justice system. However, there has been less consensus on the findings of field studies and how the findings of such studies can be utilised to help improve the accuracy of witness testimony. **Consequently, the aim of this thesis is to better understand the accuracy of witness statements in the real world.** This will be achieved by conducting a systematic literature review to explore the current field research on witness memory reporting, including which factors have been investigated and their effects on memory, and attempting to establish mechanisms that could underlie memory reporting in the field. Furthermore, there is an evaluation of Statement Validity Analysis (SVA), a tool that attempts to distinguish between truthful and deceptive memory reporting. Through gaining further understanding into what affects the accuracy of witness memory reporting, the findings of this thesis could allow for strategies to be recommended to police for interviewing witnesses when these factors are present to improve the accuracy of testimony. There is currently a lack of witness memory studies conducted outside of the laboratory, and therefore the finding of this thesis will be able to provide results that are generalisable to the real-life crime situations. If results are replicated, findings could be used to educate officers as well as providing recommendations for best practice when interviewing witnesses, ultimately reducing adverse consequences that are associated with poor testimony for both witnesses and wrongfully accused.

Chapter Summaries

To achieve the thesis aim, Chapter Two contains a systematic literature review which explores the existing field research conducted on real life crimes with an aim of understanding the current findings and the implications for knowledge of eyewitness memory. The review explores aspects such as the characteristics of the studies, how statements were analysed, what factors were

investigated and their association with the accuracy or completeness of witness memory reports. The review concludes that factors that were expected to affect memory (as identified by laboratory studies) did not affect memory for the crimes studied in real life crime situations. This may mean that laboratory studies may not be able to adequately instantiate and capture the extraneous factors that may influence memory processes that operate in real world scenarios. Given these considerations, we need to investigate the strategic regulation of memory reporting in witnesses not only in the laboratory but also in real world conditions. However, the studies reviewed did not consider a theoretical understanding for their results. This is a goal for future research.

Chapter Three outlines research that attempts to fill the gap in the field. The study is a secondary data analysis that analysed police crime incident reports for 509 crimes which included information such as crime details (e.g., weapon used, nature of any injury, lighting, duration of crime, distance from crime, etc.) and the suspect's physical appearance (including age, gender, race, height, weight, build, eye colour, hair colour, hair length, hair type, facial hair, complexion). An identical checklist was also completed by the police when a suspect was arrested allowing for correspondence between the witness' description of the perpetrator and the suspect's appearance to be determined (coded as either 'exact match' or 'no match'). The aim of the study was to see if the correspondence was affected when there were memory compromising factors (i.e., crime details that may affect memory) present. It aimed to investigate the quantity accuracy trade-off framework (Koriat & Goldsmith, 1996) and how the quantity accuracy trade-off may underlie witness recall of details for perpetrators in real crime events.

Chapter Four reviews Statement Validity Analysis with particular focus on its main component, Criteria-Based Content Analysis (CBCA). This is a tool that can also be used to determine accuracy but focuses on whether the witness was dishonest rather than inaccurate in their memory reporting.

Lastly, Chapter Five summarises the findings and conclusions across the thesis and presents the implications and recommendations for future research and practice in the criminal justice sector.

Chapter Two

A Systematic Literature Review of Research on Eyewitness Memory Reporting in Real-Life Forensic Contexts

Abstract

Background: Much laboratory research has focused on studying eyewitness memory reporting with the aim of improving its accuracy, but it is not clear to what extent findings from these studies are also replicated in real life crime events.

Objective: The present study systematically analysed the empirical literature on eyewitness memory reporting in real life forensic contexts to evaluate the methods and measures that have been used to examine eyewitness memory performance and the findings of these studies (i.e., how different factors are associated with memory accuracy and completeness).

Search Strategy: Automatic searches were conducted using Web of Science (Core collection), OVID (Psychinfo) and ProQuest (Social Sciences Premium Collection). A search strategy combining terms of (Eyewitness or synonyms) near to (Memory and synonyms) and (Field or synonyms) near to (Study and synonyms) was used. Additional references were identified through reference list searches.

Study Selection: All references obtained from the searches were screened using the inclusion and exclusion criteria before studies were quality assessed.

Main Results: 16 papers were included in the review. Types of crime, statement, and participant characteristics, witnessing conditions investigated, how memory reporting was collected and analysed are reviewed. Research has mostly studied memory reporting in robberies. Overall, the research compared memory reporting information to facts from the crime, as well as considering length of statement, and number and types of broad details reported by witnesses. Conclusions of the studies and which factors were associated with memory reporting is also discussed.

Conclusions: The findings of laboratory research may not generalise to real life crime situations. Furthermore, it was found that memory compromising factors can differentially affect the length

and accuracy of statements. Future research is needed into understanding the theories that underlie

witness memory reporting and evaluating the possible role of memory control mechanisms.

Introduction

Witnesses often report their memories to the police by giving a verbal account of the crime, which can provide pivotal information for criminal investigations (Reno et al., 1999). These statements help investigations to proceed and can impact trial outcomes (Farrington & Lambert, 1997; Coupe & Griffiths, 1996). However, while some archival research has focused on eyewitness identification from lineups (e.g., Horry et al., 2012) relatively few studies have investigated verbal reports given by witnesses in real world cases (Sporer, 1996). Given the important role testimony can play in investigations and the courtroom, a key issue for criminal justice practitioners and legal fact finders is understanding how witnessing conditions (e.g., degree of lighting, level of stress, duration of exposure to the culprit, among other factors) can influence the accuracy of witness memory reports. A key issue that influences accuracy is the extent to which eyewitnesses can appropriately regulate their memory reporting when they recall a criminal event to the police. According to strategic memory regulation accounts (Koriat & Goldsmith, 1996), to maintain accuracy witnesses should regulate their memory reporting such that they tend to report only the details that are likely to be accurate. This is achieved by controlling the amount and precision of the details that are recalled and reported to the police.

It has been shown in laboratory studies that people can successfully regulate their memory reporting (Brewer & Weber, 2008; also see Wixted & Wells, 2017). The results indicate that people tend to volunteer information when they are relatively confident that it is likely to be accurate (Koriat & Goldsmith, 1996). This is referred to as the quantity accuracy trade-off framework, which holds that when individuals are questioned by the police, they use a monitoring mechanism that assesses the likely correctness of a candidate memory response to an investigator's question, and a control mechanism that determines whether to volunteer this response. This control mechanism sets a criterion, or a probability value, that the candidate answer is likely to be correct. This criterion takes into consideration the gain for providing

correct information relative to the cost of providing wrong information. An individual would only answer if the candidate answer exceeded the criterion (Koriat & Goldsmith, 1996). For example, in high stakes scenarios (such as providing testimony), there are high costs for providing wrong information (misleading the investigation, wrongfully convicting an innocent individual, etc.) and therefore an individual would presumably set a higher criterion and only provide answers they are most confident in to prevent these consequences from occurring. This mechanism has been applied to the study of eyewitness memory (Goldsmith et al., 2014) to suggest that even if the encoding conditions are unfavourable and include extraneous memory compromising factors (e.g., poor lighting), as is expected in a real-life crime situation, individuals may volunteer less information with less specificity, but this does not necessarily mean that the volunteered information would be less accurate.

However, experts believe and state that they would be willing to testify in court, that unfavorable witnessing conditions can affect the quantity and accuracy of eyewitness memory recall (Kassin et al., 2001). These beliefs are reportedly based on the results of laboratory studies, which have investigated a host of factors thought to impact memory accuracy. In particular, Kassin et al. (2001) asked experts about their beliefs regarding 30 factors that could affect eyewitness memory. Examples of such factors include the weapon focus effect, which refers to the decreased memory accuracy associated with the presence of a weapon that impairs recall of an event (see Fawcett et al., 2013, for a meta-analysis) and stress, with a meta-analysis indicating that relatively high levels of stress can negatively impact on eyewitness memory accuracy (Deffenbacher et al., 2004). The length of the retention interval has also been found to be influential, with laboratory studies indicating that memory accuracy decreases with increasing retention interval length (e.g., Flin et al., 1992; Odinot & Wolters, 2006). Other factors include the effects of acute alcohol intoxication on memory performance, with early work suggesting that alcohol decreases accuracy (for example, Yuille & Tollestrup, 1990) and more recent work

showing that alcohol decreases the quantity but not the accuracy of information reported (Jores et al., 2019). Also, the own-race effect, wherein identification accuracy is higher for own race compared to other race faces (see Meissner & Brigham 2001 for a meta-analysis). In summary, laboratory research indicates that there are a large number of factors that can influence memory accuracy and based on this, experts are willing to testify about these effects in real life cases. However, questions remain about whether we can apply this laboratory research to real-life cases.

Laboratory-based research has advantages in that it can establish and measure memory accuracy due to the controlled environment, which is important to establish cause and effect relationships (Chae, 2010). However, the sole reliance on laboratory research to make inferences about the accuracy of real-world witnesses has been repeatedly criticized for lacking external validity (Holleman et al., 2020). For example, laboratory studies are restricted in the amount of stress that they can impose on their participants; therefore, in laboratory studies, participants will not experience intense emotions, such as fear and anxiety that accompany witnessing an actual crime (Penrod et al., 1995). These emotions trigger the flight or fight response and result in neuro-chemical effects on memory, which can be difficult to instantiate in a laboratory context (Schacter, 1996), but may influence memory. Several field studies have investigated eyewitness memory utilizing 'unpleasant' or staged scenarios as the to-be-remembered event. However, external validity of this research may also be relatively low when attempting to generalize to real world crimes that involve witnesses (Flowe et al., 2009; Flowe et al., 2018), for the same reasons as above. Furthermore, there is an increase in attentional demands on an individual who is experiencing an actual crime. For instance, there are competing objectives such as avoiding injury or planning an escape, which may decrease a witness' attentional focus on the crime. Whereas, in laboratory studies, participants have fewer attentional demands and may therefore pay greater attention to the scenario (Lane, 2006). Another consideration is that the statements

and testimony that witnesses provide in actual criminal cases may result in serious consequences (e.g., conviction of a defendant who may not be guilty) whereas this is not the case for participant witnesses in laboratory studies (Wagstaff et al., 2003). As such, in real-life situations witnesses may exercise greater control when reporting their memories about the events they witnessed. Laboratory compared to real world witnesses may set a lower criterion when they report their memories, resulting in a greater quantity of information reported and lower accuracy rates. Lastly, most laboratory studies are conducted with participants who are university students. University students may differentially regulate their memory reporting compared to other witnesses of crimes, and thus, theories based on data from laboratory studies may have low external validity (Pozzulo, 2006). Overall, some would argue that laboratory studies may not be able to adequately instantiate and capture the extraneous factors that may influence memory processes that operate in real world scenarios. Given these considerations, we need to investigate the strategic regulation of memory reporting in witnesses not only in the laboratory but also in real world conditions.

The study of witness memory reporting in actual cases is also important because the approach used by criminal investigators when questioning witnesses may substantively differ compared to the way in which researchers question participants. The interviewing approach might differentially affect memory reporting strategies and the accuracy of information reported. For example, in rape cases, victims frequently report that police who took their statements did not believe them, and they found the interview process extremely stressful (McMillan & Thomas, 2009). This could lead to a reluctance on the part of victims to disclose information about the crime which gives an incomplete account of events and thereby possibly preventing justice. Another consideration is that the approach used by criminal investigators in evaluating the likely accuracy of statements and testimony may also be problematic. For example, an analysis of 250 DNA exoneration cases indicated that the eyewitness testimony of a witness was a contributory

factor leading to wrongful convictions of the defendant (Garrett, 2011). The wrongful conviction cases demonstrate the profound effect inaccurate testimonies can have on criminal justice outcomes. Importantly, however, all the witnesses in those cases informed the police taking their statement that they had low confidence in the accuracy of the information they had reported (Garrett, 2011). This raises the question of whether eyewitness memory per se led to the wrongful conviction. Arguably, the police and prosecutors should have given less weight to the eyewitness evidence in those cases where the memory evidence was weak, and thus, it was faulty criminal investigation procedures that led to the wrongful conviction. Wixted (2018) discuss the questionable eyewitness-identification procedures that police commonly employ and how these have led to wrongful convictions.

As another example, in the UK, the Achieving Best Evidence (ABE) guidance sets out evidence-based recommendations for conducting police interviews to enable witnesses to provide accurate and complete testimony (Ministry of Justice, 2011). However, a closer examination of the ABE guidance raises questions about whether it may increase memory reporting errors. The guidance provides that police interviewers should ask witnesses to provide free recall account, and report everything, no matter how trivial, in their own words without interruption. A questioning phase takes place, in which witnesses are "systematically probed using open-ended and specific-closed questions until all the relevant material relating to it has been obtained" (Ministry of Justice, 2011, p.76). However, these questions lead witnesses to potentially report details about which they are unsure that they otherwise may not have reported, thereby reducing the accuracy of their memory report overall. Utilising the quantity accuracy trade-off framework, accuracy of testimony has been suggested to be higher during the free recall phase than the questioning phase (Flowe et al., 2019). It is therefore important that the results from memory reporting in real crimes is utilised to be able to inform how to question witnesses to produce the most accurate memory information.

There has been no previous review of research investigating the statements and testimony given by real world witnesses. There have been reviews of field studies on lineup performance (Wells et al., 2020); however, memory reporting mechanisms may differ for face identification compared to free recall tasks (Koriat & Goldsmith, 1996). Therefore, there is an outstanding need to collate the findings from field studies that have studied real world crimes to better understand what this body of work reveals about how witnesses may be strategically reporting their memories, and what further research may be needed to address knowledge gaps. Ultimately, this will provide education opportunities for criminal justice practitioners who may have beliefs that certain testimonies may be less 'accurate' under certain circumstances which may unconsciously affect the interviewing process and ultimately the memory information received. Furthermore, gaining an understanding of how witnesses of crimes monitor their memory would allow for strategies to be recommended for police and courtroom interview procedures which maximise the most effective memory information from the witness. These impacts would reduce the adverse effects that are associated with poor testimony on both witnesses and the wrongfully accused.

The specific objectives of the current review are to describe: 1) the range of demographic characteristics of participants, as well as the types of crimes, measures, and designs that have been used in field research; 2) the types of memory performance measures that have been used in field research; 3) the conclusions from the studies as to which factors affected memory reporting; and 4) the dominant theoretical perspectives that have been used to understand witness memory reporting performance.

Method

Scoping

A search to determine whether any pre-existing literature reviews on how researchers have studied memory for crime events in a naturalistic context was run and no reviews were found. Databases that were searched included Cochrane Database of Systematic Reviews (CDSR), The Centre for Reviews and Dissemination (DARE) and Campbell Collaboration. A search was also conducted on PROSPERO, a site where authors can prospectively register their systematic review, and there was no indication that a similar review was being conducted. Searches were also extended to electronic databases including Psychinfo, OVID and Web of Science, in which the search was limited to 'reviews'. Again, there was no indication that a previous review on this topic has been conducted.

Sources of Literature

The following databases were used in the search:

- Web of Science (Core collection)

- OVID (Psychinfo)

- ProQuest (Social Sciences Premium Collection)

These databases most frequently publish research on the psychology of remembering in forensic contexts, and hence are appropriate for the research topic at hand. The search was restricted to peer reviewed research papers only. The reference lists of the papers selected for review were also examined to locate any additional papers that met the inclusion criteria.

Search terms

The search was conducted using a combination of the following search terms (Appendix A):

1. (Eyewitness* or Bystander* or Witness* or Observer*) near/2 (Memor* or Recall or Remember* or Recollect* or Testimon* or evidence*)

2. (Field or Natural*) near/2 (stud* or experiment* or research)

3. 1 AND 2

Inclusion criteria

Studies were selected for inclusion if they analysed adult witness memory performance captured via written or verbal statements of an actual crime that they had witnessed. There were many studies that focused on child memory reporting and hence this topic would benefit from a separate, dedicated, review. Furthermore, many of the papers obtained from the search utilised lineup procedures to test memory; these were excluded to meet the aims of the review. Many studies returned from the initial search studied memory performance using fictitious scenarios as the to-be-remembered event, such as exposing participant witnesses to negative, traumatic, stressful events, live or staged. However, arousal and other factors can differ between crime and other negative events which could affect generalisability (Penrod et al., 1995). Therefore, only the studies that analysed memory reporting in adult witnesses following their exposure to an actual real world crime event were included in the final sample.

Assessment of Methodological Quality

22 studies that met the eligibility criteria were subjected to a quality assessment. There is no existing quality checklist to assess memory research. However, many elements of quality assessments are common to all types of study design. Therefore, the checklist used was designed by Protogerou and Hagger (2020) as it included items that were applicable to the aims of the current review. There are 20 items in this checklist; however, items 18 and 19 were omitted as

these related to ethical considerations of whether participants were provided with informed consent/ debrief. As most of the studies included in the current review gathered data from police files, these items are not relevant. The remaining 18 items are scored with the options: "yes", "no", "not stated clearly", or "not applicable" based on the information in the paper to be assessed. Studies need to receive a "yes" response to 12 out of 18 items to receive an overall "acceptable" quality score and be included in the present review (Protogerou & Hagger, 2020) (copy of checklist in Appendix B). Dual coding was also conducted on all papers deemed appropriate for quality assessment for reliability. Discrepancies were resolved through sourcing a third parties' opinion. Following the quality assessment, six studies in total were excluded due to not meeting the acceptable cut-off score (Quality Assessment shown in Appendix C). Included studies had an average quality assessment score of 13 from a maximum possible score of 18 (range from 12 to 16); see figure 1 below.
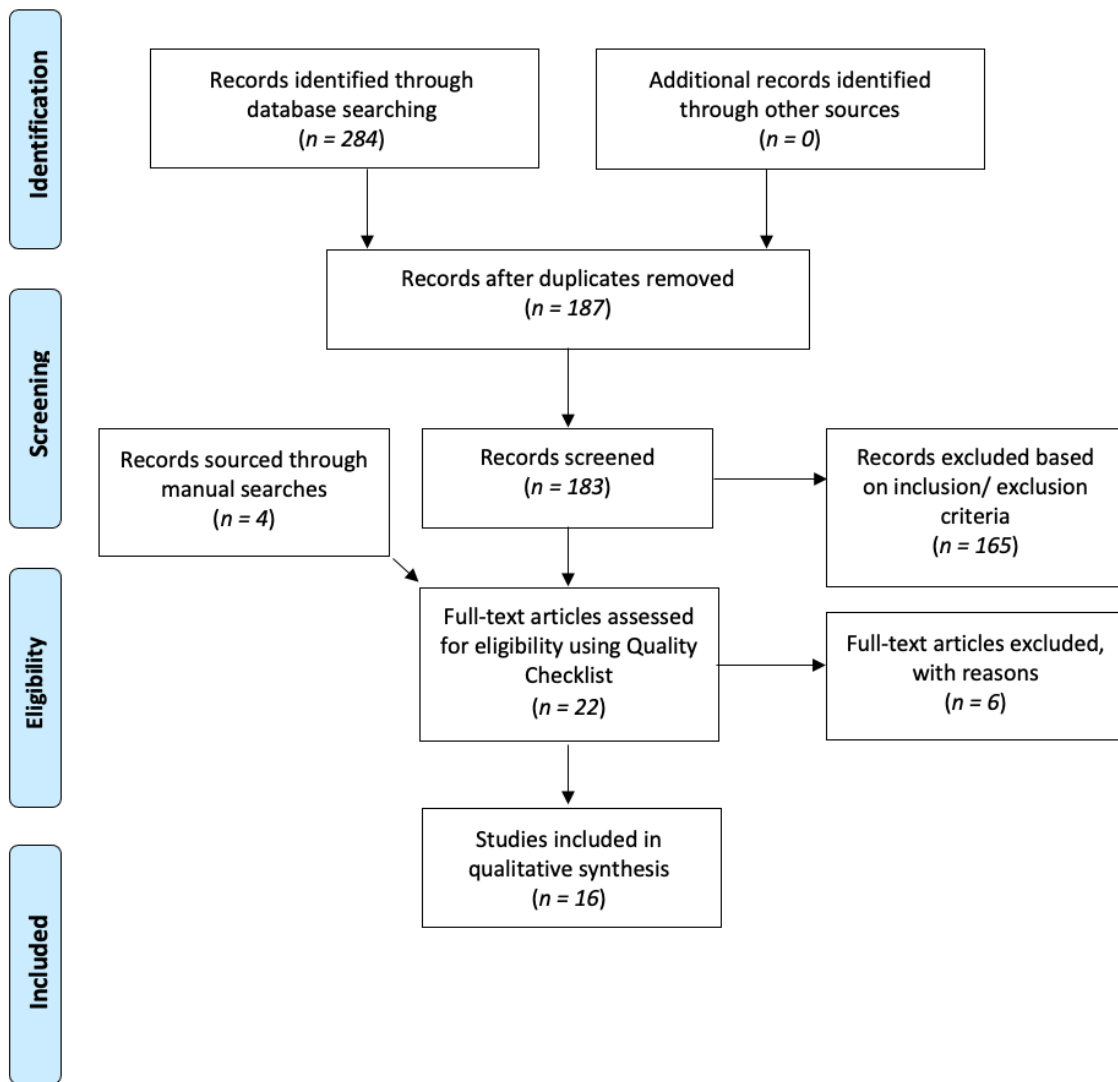
Figure 1 PRISMA flow diagram of article selection process

Results

The methodological characteristics for each study were extracted and are presented in Table 1. The studies included in the review were published between 1986 and 2022.

Table 1 Study characteristics.

Study	Year	N	Participants	Age (M unless stated)	Crime Event	Crime Information	Statement Type	Factors Investigated For Effect on Memory Reporting
Yuille & Cutshall	1986	13	Witnesses	-	Homicide and attempted murder case	A shooting incident in which one person was killed and another was seriously injured	Police statements and research interviews	Retention interval, stress, and misleading information
Macleod & Shepherd	1986	379	Witnesses	-	Assault cases	Cases occurring in Aberdeen during 1982	Police statements	Seriousness of crime and sex of witness
Yuille & Kim	1987	45	Victims, witnesses, and police officers	-	Homicides, attempted murders, fatal hit and run accidents, sexual assaults, robberies, assaults, special investigations	Cases covering the period 1978 and 1985 from the Vancouver Police files	Police statements	Hypnosis
Fisher et al.	1989	88	Victims	-	Commercial robbery or purse-snatching cases	Cases chosen on the basis that the participants had an interview and observed a suspect	Police statements	Cogntive interview
Christianson & Hübinette	1993	110	Victims and witnesses	40.5	Bank robbery cases	22 bank robberies committed in the city of Stockholm between January 1989 and May 1990	Research interviews	Retention interval and witness role
Tollestrup et al.	1994	81	Victims and witnesses	32.34	Robbery cases	119 robberies committed between 1987 and 1989 in Vancouver	Police statements	Presence of a weapon and witness role
van Koppen & Lochun	1997	1,313	Witnesses	-	Robbery cases	Nearly all the robbery cases that were committed in the Netherlands in 1992	Police statements	Distance, crime duration, retention interval, witness position, seriousness of crime, use of disguise, obstructions, sex of witness, and age of witness
Woolnough & MacLeod	2001	19	Victims and witnesses	26.4	Assault and breach of peace cases	Case selection based on those that were no longer under police investigation and had occurred in Stirling, Scotland between 1996 and 1997	Police statements	Seriousness of crime
Cooper et al.	2002	51	Victims	35.19	Sexual assault cases	Sexual assault of prostitutes	Research interviews	Presence of a weapon
Wagstaff et al.	2003	70	Victims and witnesses	Age range from below 20 to above 70	Robbery, rape, and assault cases	Cases from files of two British police forces	Police statements	Seriousness of crime, presence of a weapon, and age of witness
Fahsing et al.	2004	236	Witnesses	28.6	Robbery and fraud cases	Cases that targeted banks and post offices in Oslo between January 1999 and December 2001	Police statements	Presence of a weapon, witness role, crime duration and number of perpetrators
Odinot et al.	2009	14	Witnesses	27.5	Robbery case	A single robbery on 9th February 2007 in the Netherlands	Research interviews	Retention interval and confidence
Brónimann et al.	2013	47	Witnesses, expert witnesses, and civil parties	57	Crimes of the Democratic Kampuchea	Those responsible for the deaths of millions of Cambodians through forced labour and genocide	Trial testimonies	Different types of interviewers
Granhag et al.	2013	29	Witnesses	-	Homicide case	Stabbing of Swedish Foreign Minister Anna Lindh on 10th September 2003	Police statements	Accuracy of memory
Vredeveldt et al.	2015	95	Witnesses	34.72	Sexual assault, robbery, attempted murder, and homicide cases	Cases put to Facial Identification Unit in the South African Police Service	Police statements	Eye-closure interview
Ashkenazi & Fisher	2022	60	Victims, witnesses, civillians and soliders	31.38	Terrorist attacks	Cases that investigated individuals who were suspected or known to be involved to recognized terrorist organizations	Police statements	Cognitive interview

Note: '-' indicates that the information was unknown.

Participants

As shown in Table 1, participants in studies investigating memory reporting for crime events in the field involved a variety of participants, including victims, witnesses, police officers, expert witnesses, and civil parties with some studies using more than one type of participant. However, the vast majority examined either witnesses or victims, with ten studies examining witnesses, and seven examining victims. This could be because it is common for there to be more witnesses than victims and witnesses are more easily interviewed first due to potentially needing less attention from other professionals such as first aid. Furthermore, it is also important to add that the term 'witnesses' and 'victims' are used interchangeably in the literature, meaning that whilst some studies have separated these others may have not. There are also three papers that use other types of participants, such as police officers, civilians, and soldiers.

The mean ages, when reported, of the participants varied widely from 26.4 to above 40.5 (see Table 1). The ages shown contrast those of laboratory studies in which the majority are conducted on university students. Most participants who were examined for the current studies were older than 'typical' undergraduate university age.

Crime Event

There was also a wide range of crime events that were explored, with six papers reporting more than one crime type. As seen in Table 1, most of the studies considered robbery, with 11 (73.3%) examining this crime type. Other crime types included: sexual assaults (n=6), physical assault (n=6), murder (n=6), attempted murder (n=6), road traffic accidents (n=3), special investigations (n=1; defined as an 'investigation for a specific purpose arising outside the ambit of an ordinary investigation '(Law Insider (n.d.)), fraud (n=1), terrorist

attacks (n=1), and crimes within the Democratic Kampuchea (n=1). Notably, most of the crimes included in the review involve some degree of violence, indicating the focus of most of the literature focuses on violent crimes.

Six studies analysed more than one type of crime. For example, Yuille and Kim (1987) studied several homicides, attempted murders, fatal hit and run accidents, sexual assaults, robberies, assaults, and special investigations. There were a few studies (seven studies) that analysed a single crime type, such as Tollestrup et al. (1994) who studied 119 robbery cases. However, there were three studies that utilised a single case study in which there were numerous witnesses. For example, Granhag et al. (2013) studied a single homicide case in which there were 29 witnesses. Similarly, Yuille and Cutshall (1986) studied a shooting incident, in which one person was killed and another was seriously injured, that had 13 witnesses.

Types of Statement Analysed and Coding Approaches

This section describes the types of statements that were analysed across studies and the different coding methods that were applied to assess the statements. It is notable that the studies varied with respect to the amount of detail provided about the coding procedure used.

Police Statements

A total of 12 studies analysed the statements that victims and witnesses had given to the police after they had witnessed the crime. The statements were coded for analysis using different approaches across the studies. One approach involved scoring police statements for the amount of information reported by the witness, such as by assigning one point for

every unit of information reported by the witness (e.g., Woolnough & MacLeod, 2001; Fisher et al., 1989). As another example, Yuille and Kim (1987) coded the amount of information reported and whether it was descriptive (i.e., provided evidence about the appearance of a person or object) or about an action (i.e., provided evidence about the actions of an individual). The same method was also used by several other studies including Macleod and Shepherd (1986) and Yuille and Cutshall (1986). A similar method was used by Ashkenazi and Fisher (2022), but they defined a unit of information as a statement that composed of a noun and a predicate (an active verb or an adjective describing the noun).

Fahsing et al. (2004) also analysed police statements, extracting and coding offender attributes and then calculating the overall frequency score for each description given by an eyewitness. These units were grouped into categories of basic features (e.g., the offender's gender, height, build and age) and details (which consisted of all other attributes). Descriptions of the perpetrator were analysed in four other studies from witnesses' statements to the police (Tollestrup et al., 1994; Granhag et al., 2013; Wagstaff et al, 2003) and in court (van Koppen & Lochun, 1997). Another approach used by Vredeveldt et al. (2015) was to analyse units of information provided by the witness relating not only to the appearance of the perpetrator, but also the modus operandi, appearance of other witnesses, and other details. These details were coded as 'visual', 'auditory' or 'other' modality. They also rated perceived forensic relevance on a 7-point Likert scale.

Research Interviews

Four studies analysed interviews which were conducted by the researchers, rather than utilising statements that witnesses gave to law enforcement. Christianson and Hübinette

(1993) asked bank employees who were witnesses and victims of robberies to complete questionnaires to measure memory accuracy, measured by a comparison of information gathered from the police reports. Questions asked were concerned with the witness' memory for the robbery and their emotional reactions to experiencing the robbery. The questionnaire included 31 items; 16 questions were multiple-choice items that asked the witness to describe the actions that occurred, as well as to describe people and objects, and circumstances in which the robbery occurred (time, date, day of the week and number of customers). To assess emotional reactions, witnesses were asked to rate their emotions (e.g., unpleasantness, afraid, angry and threat) using an 11-point Likert-type scale. They were also asked to describe the physiological reactions they experienced, the vividness of their memory, and to provide demographic information about themselves such as their sex, age, years of bank employment, prior experiences of bank robberies and background information. The mean number of correct details reported was determined by comparing information gathered from the police reports.

Cooper et al. (2002) interviewed sexual assault victims. Victims were asked to recall a sexual assault using the Adult 'Stepwise' Assault Interview protocol (Yuille, 1990) which is used as an investigative tool for allegations of sexual assault and domestic violence. This technique involves eliciting a free narrative of the event from victims and witnesses. Open-ended questions are also asked to gain more information and specific questions are used only to resolve any uncertainties. The authors argue that this elicits unbiased accounts. They analysed the statements victims gave by partitioning the information across six types of categories (person, object, action, relational, subjective, and conversational details).

Odinot et al. (2009) also conducted research interviews with witnesses three months after the crime event. The witnesses were asked to rate how accurate they thought their memory for the crime was using a seven-point scale, with 1 indicating 'very uncertain' and 7 indicating 'absolutely certain'. To examine the memory, witnesses were first asked for a free recall of the event and to draw a floor plan. After this, the interviewer asked more specific questions to have the witness elaborate on the information they had provided, focusing, in particular, on forensically relevant details, which included full description of the robbers, the gun, the bag used, the position and acts of the robbers, and the position and acts of the witnesses and his/her colleagues. Lastly, they were asked if they had spoken to anyone else about the robbery if they ever thought back about the robbery and how much the incident affected them emotionally using a 7-point scale. These interviews were audio recorded and then transcribed before being scored about the number of units of information provided about person, object, and action details.

Lastly, Yuille and Cutshall (1986) interviewed witnesses to a gun store robbery that ended in the death of the robber, utilizing police statements to score accuracy. Witnesses were asked to rate the stress they had experienced at the time of the crime using a 7-point scale, with 1 indicating 'perfectly calm' and 7 'extreme anxiety or stress'. They were also asked about their emotional state prior to the incident and if they had experienced any negative effects since the incident. The researchers also included two misleading questions into the interview. For example, half of the witnesses were asked whether they saw *the* broken headlight and the other half were asked if they saw *a* broken headlight with their being no broken headlight at the scene. Like other studies, the statements were coded into single units of information, and then categorised into descriptive and action details. The authors

reconstructed the event based on available information. Accuracy was judged by calculating the proportion of details that were correct from the witness's account.

<u>Trial Testimonies</u>

Brónnimann et al. (2013) analysed the transcripts of testimony that witnesses gave at trial for crimes of the Democratic Kampuchea. A content analysis was conducted on the transcribed account using the Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al., 2007) which categorises and quantifies word use. It then counts a percentage of a text's sample words which fall into a predefined category. This means that text samples can be compared against each other even if the length varies.

Overall, most studies on memory for crimes in the field have analysed police statements. The coding systems varied between studies, but most often researchers counted the *number* of units of information reported, conditioning the results on different types of details (e.g., person, action), with some attempting to measure accuracy (e.g., Yuille and Cutshall, 1986). Researchers also employed interviews to gain further insight into witness memory retrieval processes which might not otherwise be inferred from witness statements in police reports or trial testimony.

Assessing Statements

There are several ways in which the studies have assessed the statements with respect to memory reporting. Some studies utilized more than one method.

<u>Verified by Other Information</u>

Ten studies utilised other information to determine the accuracy of the statements given. One study which utilised this procedure was Yuille and Kim (1987) which compared police statements with case evidence. Another study by Yuille and Cutshall (1986) compared both police and research interview information with the facts of a shooting incident. This included: (a) police reports of the forensic evidence, (b) reports that were taken from ambulance staff and other personnel who attended the scene, (c) any other witness statements other than those included in the analysis and (d) autopsy and medical reports. Similarly, Christianson and Hubinette (1993) verified a research interview statement through comparing it with the police reports. This included comparing it to the police reports taken at the time of the robbery and to photographs/ films at the scene of the crime. A mean accuracy score was then calculated for each participant.

Wagstaff et al. (2003) used a scoring system of a Likert type scale ranging from 'very accurate' (1) to 'very inaccurate' (2) to measure the accuracy of the statements. Witness descriptions of offenders were coded and compared with the National Intelligence Bureau Form (NIB74) in which the true description of the offender was detailed by the arresting officer and hence could be used to infer accuracy. A similar scale was used by Fahsing et al. (2004). They compared the witness reported attributes against crime scene video recordings and, depending on the degree of correspondence, assigned memories to different categories. The categories were: (a) correct, when there was a perfect match between the descriptor and the verification data; (b) partly correct, when a descriptor was correctly but incompletely described; (c) incorrect, when a descriptor did not match the verification data; and (d) unverifiable, when correctness could not be determined.

Similarly, van Koppen and Lochun (1997) used a similar method to assess their statements. They compared information from statements with data on the robbers' true appearance taken from the national database on offenders (Herkenningsdienst Systeem). They rated the witness accuracy by rating the details given on a three-point scale (wrong, partially correct, and correct). Tollestrup et al. (1994) also used descriptions of individuals eventually charged with a crime and logged by police officers which were compared to the statements given by the witnesses and victims.

Odinot et al. (2009) determined accuracy by comparing each unit of information from the witness with the information on security videos. The units were verified as either correct or incorrect. Another study that utilised CCTV evidence was Woolnough and MacLeod (2001), as well as Granhag et al. (2013). In Granhag et al.'s (2013) study, CCTV cameras took photographs which were clear enough to be able to identify features of the offender. The reported attributes by the witnesses were compared to these photographs and were coded as wither 'correct', 'partly correct', 'incorrect', or 'unverifiable'.

Length/ Number of Words in Statements

Ten studies analysed the length of the statements to assess memory reporting. This method of assessment involved quantifying each unit of information in the statement. The effectiveness of each statement was then determined, with more information equating to more effectiveness. This method was utilised by studies, including Fisher, Geiselman and Amador (1989), Tollestrup et al. (1994), van Koppen and Lochun (1997), Cooper et al. (2002), Fahsing et al. (2004), Granhag et al. (2013), and Vredevelt et al. (2015).

Some studies utilised more than one measure relating to length/ number of words in statements. For example, Macleod and Shepherd (1986) looked at the length of the statement as a measure (as described above with quantifying units of information) as well as measuring the amounts of different kinds of information (e.g., action vs descriptive units). Similarly, Brónnimann et al. (2013) used both the length of the statement (quantifying units of information) and the number of affective/ cognitive words used when investigating the influence of different types of interviewers on witness statements. Lastly, Ashkenazi and Fisher (2022) used the number of units of information and number of new units of information. They conducted two interviews on the same individuals to see whether the type of interview would influence the information given by the witness.

Detail and Comprehensiveness of Statements

Two studies examined the detail and comprehensiveness of statements to determine their effectiveness. Vredeveldt et al. (2015) looked at the perceived quality and detail of information provided as one measure to assess the statements. This was achieved by a senior police officer rating the information for perceived forensic relevance. This method was also utilized by Ashkenazi and Fisher (2022) who rated the utility of the information based on the intelligence contribution of the information and the extent to which this information could help to 'solve' the event. This was rated on an 11-point Likert scale ranging from 0 (not at all) to 10 (very much so).

Overall, several methods have been utilised to assess the statements. Most studies, when information is available, verified the statements with other sources of information to determine the accuracy of the statements. However, some studies (N = 10) utilised the length of the statement to assess the memory information given by participants. Other

methods such as comprehensiveness (i.e., the level of detail) of information have also been used to determine the effectiveness.

Possible Memory Compromising Factors Investigated

The studies were involved in researching the effect of factors on eyewitness recall. This section discusses whether the results are consistent with laboratory research.

<u>Interview Type</u>

Six studies examined the relationship between interview type and memory reporting for crime events. Brónnimann et al. (2013) studied the effects of different types of interviewers on the quality of the testimony given by witnesses at trial. They found that when witnesses were interviewed by judges, statements contained four times more words than when interviewed by legal professionals (prosecutors, civil party lawyers and defence lawyers). They also found that more affective and cognitive process words were used during examination by civil party lawyers than by judges and prosecutors, which may be because witnesses have an increased trust in their own lawyers allowing them to speak about more personal and emotional information. There were also more cognitive process words during interviews by defense lawyers which reflected the higher cognitive activation due to a more interrogative questioning style to create contradictions in the witness.

Odinot et al. (2009) studied the effect of free recall versus immediate subsequent questioning. These questions were more specific questions that followed-up on the general information provided during free recall. They found that details provided in initial free recall were more

accurate than details recalled in the questioning phase when statements were compared against CCTV footage to determine accuracy.

Fisher et al. (1989) investigated the effects of the cognitive interview (CI) on witness statements. The CI is an interview procedure which was developed based on accepted principles of memory. It provides a set of instructions to the interviewer which encourages the witness to (a) reinstate the context of the original event and (b) to retrieve memory by using a variety of different retrieval routes (Geiselman et al., 1985). Investigators were trained in the CI (through four 60-minute group sessions). They concluded that the CI allowed for an increase, compared to interviews conducted before training, in investigative-relevant information, defined as physical descriptions of the assailants and relevant actions, as well as clothing, weapons, vehicles, objects taken, and conversations. This provides strong support for the effectiveness of CI in field investigations. They suggested that the effectiveness could be due to the short delay in questioning, the use of witness-compatible questioning and the context reinstatement that is utilised as part of the CI procedure. However, this study did not assess statement accuracy.

A more recent study by Ashkenazi and Fisher (2022) also studied the effect of the CI on statements for terrorist attacks. Everyone was interviewed twice; once using a Standard Interview (SI) and again using either a SI or CI to see whether the latter would elicit additional or increased utility of information. They concluded that those who were interviewed a second time with a CI provided more information and more new information (not already in the first interview) and that this information was also rated as being more useful, emphasising the importance of the retrieval environment for witnesses. This supports the Fisher et al., (1989) study on the utility of the CI in investigations. However,

both studies (Ashkenazi & Fisher, 2022; Fisher et al., 1989) only examine the length of the statement, rather than the accuracy and hence further research is needed into this. It is also possible that the CI elicited more incorrect information, but this has not yet been tested.

The effect of the eye closure interview was researched by Vredeveldt et al. (2015). The background to this is that eye closure is proposed to have benefits in improving performance on cognitive tasks (Glenberg et al., 1998), improves recall on both visual and auditory aspects of events (Perfect et al., 2008) and mitigates cross-modal impairment caused by auditory distraction (Perfect et al., 2011). The procedure was compared to a control condition with no eye closure. The procedure for both groups involved: (1) obtaining a description of the event, (2) obtaining a description of the perpetrator and (3) constructing a facial composite. The eye-closure interview instructed witnesses to keep their eyes closed during the recall of the event and the perpetrators. The authors concluded that whilst the amount of information did not increase, the perceived quality (detail) of this information was enhanced with an eye closure interview. The authors suggested this could be because closing one's eyes could increase the intensity of emotions experienced (Lerner, et al., 2009) and hence increase focus towards central details relevant to the case. They used this to argue that this procedure could be useful to police interviews in obtaining valuable information. However, this study did not measure accuracy and hence it is unknown if the information recalled is accurate.

Hypnosis was studied by Yuille and Kim (1987). There has been research to suggest that hypnosis can improve memory gains in witnesses (Kroger & Doucé, 1980) and hence they wanted to investigate its effects in a field study. Results showed that hypnosis-aided interviews obtained twice as much accurate information from a witness and therefore

concluded that using hypnosis in police procedures could assist with the retrieval of information. However, the study did not measure the amount of inaccurate information and hence it cannot be known if this also increased with hypnosis.

Stress and Arousal (Seriousness of Crime)

Five studies investigated the relationship between stress and memory reporting. Physiological arousal is hypothesised to increase with increasing levels of violence, which is thought to deleteriously affect memory encoding and retrieval (Christianson, 1992a). However, three of the five studies found stress is associated with improvements in accuracy of memory. Wagstaff et al. (2003) operationally defined stress as the level of violence that occurred during the crime, with stress levels experienced in each case coded into one of five categories: (a) no violence, (b) low violence (including indirect verbal threats and nonviolent physical contact), (c) medium violence (including direct verbal threats and a single aggressive act involving contact), (d) high violence (including multiple threats and acts of violence) and (e) very high violence (including multiple threats and acts of violence, high levels of agitation and actual use of a weapon to injure the victim). They found that violence predicted increased accuracy on the recall of hair colour of the offender, but not the other details such as age, height, build, hairstyle, date, time, and other people. They suggested that hair colour was considered a more 'central detail' of the crime compared to the other details and hence these results could be considered supportive of the hypothesis that violence improves the accuracy of recall of central compared to peripheral information.

However, Woolnough and MacLeod (2001) defined stress as incident seriousness and emotional impact. They judged how violent they perceived the incident to be, the extent of

injuries, if they thought the incident would have an emotional effect on the victim and if they thought the event would have an emotional effect on the witnesses. Each rating was made on a 6-point scale (1=not at all to 6= extremely). The incidents were then ranked from most to least serious. From this they concluded that when there was enhancement of arousal (measured through incident seriousness), this resulted in more accurate memory. The authors proposed that this may be due to more personal involvement and a need for survival; the potential detrimental effect of high arousal may be counteracted by the novelty of the situation and the need to maintain a high state of awareness.

Compared to the other studies, Yuille and Cutshall (1986), measured stress by the self-reports from witnesses. From this they found that their stress level at the time of the event had no negative effects on subsequent memory. In fact, those who self-reported higher stress levels (scores of 5-8) were more accurate in their police and research interview compared to those that self-reported lower stress levels (scores of 1-4), supporting the other studies (Woolnough and MacLeod, 2001; Wagstaff et al., 2003).

However, two of the five studies have suggested that stress might impair memory reporting. For example, van Koppen and Lochun (1997) investigated the effected of 'estimated threat' and concluded that an increase in threat resulted in less complete (shorter length) witness statements, contrasting the above studies. Macleod and Shepherd (1986) also found that females reported significantly less information about the accused than did males when they were injured. It was suggested that females may be less able to cope with stressful situations due to being more fearful and vulnerable. Males, on the other hand, have been suggested to have become habituated to such events due to being more likely to witness other incidences in the past and hence they are less susceptible to any

effect the violent incident might have on them. However, whilst this shows that stress has a negative impact on memory for females, it is unsure if there is an overall effect of memory without considering sex. Furthermore, this paper is outdated in its views as it reflects gender stereotypes. Furthermore, both studies scored a quality assessment score of 13, which is lower than studies such as Woolnough and MacLeod (2001), which scored 16. This calls into question the weight that these results can be given compared to the other papers in the review that found the opposite patterns of results.

Many of the findings from field studies (N = 3) found that increased arousal improves accuracy of memory recall for crime events. However, two studies found that, regarding the length of statements, this may in fact be decreased when there is increased arousal. Furthermore, there is a potential sex difference on the effects of crime seriousness on the length of memory reporting, but this finding should be viewed with caution due to it representing outdated stereotyped views.

Presence of a Weapon

Four studies investigated the relationship between the presence of a weapon and memory reporting of a crime. The 'weapon focus' effect describes that when there is a presence of a weapon, a witness' attention is focused to that object and subsequently has effects on the accuracy of other details of the crime (Loftus et al., 1987).

One of the four studies found no evidence of a weapon focus effect. Wagstaff et al (2003) looked at the effect of a presence of a weapon on statements of a crime by comparing witness descriptions of the perpetrator to the true description of the suspect detailed by the police. However, they concluded that there was no significant effect of presence of a weapon on the accuracy of recall of crime details.

Three of the four studies found evidence that memory was more complete when there was a weapon present. Cooper et al. (2002) also investigated the association between weapon presence on memory reporting by victims in sexual assault cases. The study found more details were recalled when a weapon was present, which contrasts with the results of laboratory studies. They suggested this is due to the complexity of actual crime scenes, as well as witness arousal levels and attention, all of which can impact memory encoding and cannot be readily replicated in the context of a laboratory.

Tollestrup et al. (1994) also examined the impact of weapons on witness statements. They found that witnesses to robberies with, as opposed to without, a weapon reported significantly more details. There was also no significant effect of weapon presence on the accuracy of statements, which was determined by comparing the witnesses' statement to the physical characteristics of suspects who had been charged with the crime. Thus, the results suggest weapon presence does not affect accuracy of details reported but does lead to an increased length in statements. However, we are unaware if the individuals charged with the crime were guilty and hence this may affect our ability to make conclusions on 'accuracy'.

Fahsing et al. (2004) looked at the effect of type of weapon used. They found that the accuracy of the descriptions was higher in crimes committed with firearms compared to knives. However, the influence of weapons compared to no weapons was not examined in this study. Despite this, the crimes that involved knives reported more basic features than firearms. Therefore, the type of weapon may influence memory for crimes as whilst

witnesses who experienced firearms provided less information, this information was more accurate compared to witnesses who experienced knives.

Despite the results from laboratory studies suggesting that weapon presence may impair eyewitness memory, the findings from this review suggest that there is no obvious effect on the accuracy of memory despite weapon presence. Interestingly, most studies ($N = 3$) concluded that weapon presence led to an increase in the length of statements.

Retention Interval

Four studies investigated the relationship between retention interval and memory reporting for a crime event. The forgetting curve would suggest that there will be a decline in memory retention over time (Ebbinghaus, 2013). Three of the four studies suggest that memory is not negatively affected by retention intervals. Christianson and Hübinette (1993) studied the effect of retention interval of memory for bank robberies, with participants filling out questionnaires about the robbery between four and 15 months after it had occurred. They concluded that there were relatively high accuracy rates even after a longer retention interval relating to details about the offender (including their actions, presence of any weapons and their clothing). However, they also found that there was low accuracy because of the retention interval for some aspects such as the offender's footwear, hair, and eye colour. They suggested that this may be due to disadvantages in viewing point and hence these details were not processed.

Odinot et al. (2009) also studied the effect of retention interval for a robbery as they conducted a research interview with witnesses three months after the event had occurred.

They concluded that about 84% of all remembered information was correct, showing that events can be retained in memory even after long time periods.

Yuille and Cutshall (1986) studied the retention interval for a shooting incident and concluded that most of the witnesses were highly accurate in their accounts which continued to be true five months after the event. It was suggested this was due to the salience and uniqueness of the event producing vivid memories. They found that individuals directly involved in the event had significantly higher accuracy ratings which suggests that details may be retained more vividly by those who participate in an event.

Contrastingly, one of the four studies found memory is negatively impacted when there is a retention interval between the crime and questioning. van Koppen and Lochun (1997) found that statements were more complete (had more details recalled) when there was a shorter time between the crime of a robbery and questioning. However, the longest retention interval in their study was three days after the crime.

Contrary to the forgetting curve hypothesis and the results from laboratory studies, the results of these studies suggest that memory for crime events tends to remain accurate despite long periods of time. However, one study that looked at the length of statements found that this decreases over time. This may be explained by the quantity accuracy trade-off framework (Koriat & Goldsmith, 1996), which would state that in unfavourable circumstances, witnesses would only volunteer answers of which they are confident on. Therefore, in the situation where a significant amount of time has passed since the crime, the witness would volunteer less information to maintain accuracy of their statement.

Effect of Witness Role

Four studies looked at relationship between witness role and memory reporting. Witness role is determined by how much exposure to the crime they have experienced. It would be suggested that those that have better viewing conditions, would have more accurate descriptions (Fahsing et al., 2004). All four studies found that having a more active role in the crime results in better memory. Fahsing et al. (2004) investigated this effect when they differentiated between bank tellers and bystanders of robberies. They found that bank tellers were able to provide significantly more accurate offender descriptions than customers/ bystanders. Christianson and Hübinette (1993) also differentiated between bank tellers and bystanders and similarly found that the bank tellers recalled significantly more accurate crime details than bystanders in robberies. Similarly, Tollestrup et al. (1994) looked at the potential impact of witness role on memory reporting. They found that victims provided significantly more total details than witnesses, as well as more physical appearance details. Wagstaff et al. (2003) also concluded that victims had higher accuracy ratings than bystander witnesses. Overall, the theory is supported in the field studies in that those witnesses who have a more active role in the crime can provide longer and more accurate statements.

Effect of Crime Duration

Two studies looked at relationship between the duration on the crime and memory reporting. Laboratory studies have shown that the longer an event is observed, the more information is encoded and recalled (Loftus, 1979). van Koppen and Lochun (1997) supported these findings as they found that the longer the duration of the crime, the increase in completeness (more details recalled) of victim statements. Fahsing et al. (2004) also found that longer observation times were associated with more accurate descriptions.

Therefore, the results of the field studies are consistent with those from the laboratory as both found that memory recall is more complete when the crime event is observed for longer.

Sex of Victim/ Witness

Two studies looked at the relationship between the victim's sex and memory reporting. Sex differences in statements has been found in laboratory studies, with females having better recall for details (Yarmey, 1993; Longstaff & Belz, 2020). This is supported by Macleod and Shepherd (1986) who found that whilst there was no overall sex difference in statement length, they found differences in the type of information reported by each sex. However, female witnesses reported significantly fewer details about the perpetrator, but significantly more details about themselves than did male witnesses. Women also reported fewer details than men when reporting information about relatively more serious crimes. Another study that looked at the effect of sex was van Koppen and Lochun (1997). They found that female witnesses were able to recall significantly more accurate details.

Overall, the results of the current review found that the number of details reported according to sex may vary depending on the type of information being reported. However, females were shown to recall more accurate details compared to males.

Number of Perpetrators

One study examined the relationship between the number of perpetrators and description quality. Previous research in the laboratory has indicated that an increase in number of offenders leads to inferior person descriptions (Clifford & Hollin, 1981). Fahsing et al.

(2004) found that observing two perpetrators compared with one was associated with less description quality. Therefore, the laboratory findings are supported by this field study.

Distance and Position

One study examined the relationship between distance and memory reporting in statements. Previous experimental studies have shown that an increased in distance leads to decreased eyewitness accuracy (Jong et al., 2005; Lampinen et al., 2014). van Koppen and Lochun (1997) found that the greater the distance between the offender and the witness, the less complete the statement. Similarly, a worse viewing position of the offender led to less completeness (less details recalled) of statements. Therefore, this supports the laboratory findings.

Use of Disguise

One study examined the relationship between the perpetrator wearing a disguise and memory reporting in statements. Laboratory research has focused on identification accuracy and has shown that a disguise negatively affects identification accuracy (e.g., Mansour et al., 2012). van Koppen and Lochun (1997) found that the use of a disguise had a detrimental effect on completeness for permanent features of the offender and for the total number of details, but this did not, understandably, influence remembering clothing of the victim. The laboratory studies are therefore supported with this field study.

Obstructions

One study examined the relationship between obstructions between the witness and the perpetrator and memory reporting. van Koppen and Lochun (1997) looked at this effect for bank robberies when there is likely to be glass separating bank tellers and the perpetrator.

They found that obstructions did not have a significant influence on the overall number of details recalled. However, witnesses with no obstructions were able to recall more 'permanent' details compared to those who had obstructions. They suggested this could be due to reflections in the glass obscuring some features or because those behind glass were likely to be direct victims of the crime which may have produced this effect.

Age

One study examined the relationship between age of the witness and memory reporting. Studies in the laboratory find that the number of details recalled for an event increases with age (Goodman et al., 2001). van Koppen and Lochun (1997) found that descriptions tended to be more complete when the witness was younger in age, which contrasts the laboratory study findings.

Misleading Questions

One study examined whether giving misleading questions influenced witnesses' memory reporting for crime events. Experts believe that an eyewitness' testimony about an event can be affected by how the questions put to that witness are worded (Kassin et al., 2001) and many laboratory studies have confirmed that misleading questions can alter testimony (e.g., Loftus & Palmer, 1974; Weinberg et al., 1986, Wang et al., 2018). However, Yuille and Cutshall (1986) found that witnesses were able to resist misleading questions about the crime event perhaps suggesting that the misinformation effect may not be as prominent in the field, but more research is needed.

Confidence

One study examined the relationship between witness confidence and memory reporting in statements. Findings in the laboratory have suggested that higher confidence is typically related to higher accuracy (Wixted & Wells, 2017). Odinot et al. (2009) concluded that correctly recalled details had higher confidence ratings than incorrectly recalled details. These results are in line with laboratory findings regarding the confidence-accuracy relationship.

Accuracy of Memory

One study investigated the general accuracy of memory of witnesses for the crimes. Granhag et al. (2013) found that there were low levels of accuracy of the reported details for a murder with only 58.5% (34.8% correct, 23.7% partly correct) of the reported details being regarded as accurate when they compared the recalled information to CCTV evidence. This leaves 41.5% of the information given by a witness considered as misleading (two out of five details reported were inaccurate). They concluded that this was due to the brief duration of the crime and exposure to the perpetrator.

Discussion

The aim of this review was to investigate how researchers have studied memory for crime events in a real-world context and review the factors that are associated with memory accuracy/ completeness. Previous reviews have been conducted on the results of laboratory studies or those that utilise lineup procedures (e.g., Christianson, 1992b; Deffenbacher et al., 2004) and hence they may not be generalisable to situations in which real witnesses monitor their memory when giving statements. Therefore, this review has examined studies of real-life witnessing situations which could help to provide advice on how to gain the best memory information when interviewing witnesses and victims.

Key Findings and Strengths

The review discussed how memory reporting has been studied in real world forensic contexts. Many of the factors investigated have contradicted laboratory findings and what experts believe to affect memory reporting (Kassin et al., 2001) and this information would be of practical use to investigators. For example, the effect of retention interval is important for knowing how long a time interval can be left to interview a witness after a crime before the memory reporting becomes negatively impacted. Laboratory studies have favoured the forgetting curve theory and suggested that longer retention interviews result in lower accuracy (Ebbesen & Rienick, 1998; Odinot & Wolters, 2006); however, the current review suggested that memory reporting for real crime events can remain vivid over time. Similarly, the current review suggests that there is no effect of presence of a weapon, whereas the weapon focus theory would suggest there would be worst recall in such situations (Loftus et al., 1987). The weapon focus theory is supported by laboratory studies (see Fawcett et al., 2013 for a meta-analysis) and hence this has contributed to the

view that there would be less accurate information from witnesses when there is a weapon present (Kassin et al., 2001). Based on the current review, it would therefore be important not to discredited witnesses who have experienced longer retention intervals, or a weapon.

However, there are also some factors investigated in the field studies that have supported the findings found in laboratory studies. One example of this is the use of the CI. In the field study by Fisher et al. (1989), they concluded that the CI allowed for an increase in investigatively relevant information. Similar results are also found in laboratory studies that have investigated the effectiveness of the CI, confirming an increased recall accuracy for CI compared to a control interview (Köhnken et al., 1999). This is important to inform on how to question witnesses and victims to obtain the best memory information. Other factors investigated in the field that confirm the results found in laboratory studies include the relation of crime seriousness, which supports the theory that states enhancement of arousal (measured through crime seriousness) results in better memory (Christianson, 1992b). There are also other factors that support the laboratory studies. For example, a shorter crime duration, the presence of more perpetrators and use of a disguise (amongst other factors) were all associated with negative effects on memory reporting.

Across studies, some factors were also investigated with regards to their association with both the length and accuracy of statements. This points to the importance of separating accuracy and length as ways of assessing statements as they can lead to different conclusions regarding the effect of a specific factor on memory reporting. For example, Tollestrup et al. (1994) examined the impact of weapons on both the length and accuracy of witness statements. They found that witnesses to robberies with a weapon reported significantly more details, but there was no significant effect of weapon presence on the

accuracy of statements. Moreover, Tollestrup et al. (1994) found that victims provided significantly more total details than witnesses, whereas Wagstaff et al. (2003) concluded that victims had higher accuracy ratings than bystander witnesses. These studies show that both weapon presence and being a victim of a crime as opposed to a witness, can lead to an increase in the amount of information recalled, but has no effect on the accuracy of memory reporting. However, when considering sex, Macleod and Shepherd (1986) found that female witnesses reported significantly fewer details about the perpetrator and reported fewer details than men when reporting information about more serious crimes. However, van Koppen and Lochun (1997) found that female witnesses were able to recall significantly more accurate details. In this instance, whilst females recalled less information compared to males, this information was more accurate.

Regarding the types of factors investigated by field research, the most common factor investigated was the relationship between interview type and memory reporting. This is important because it is useful to determine which interview types and techniques are best used in obtaining accurate information from victims and witnesses. Studies were able to show positive effects on memory reporting for different interview procedures, including CI, eye-closure interview and even hypnosis. The results suggest that such methods lead to improved memory reporting from witnesses compared to traditional methods of interviewing. Perhaps most importantly, Odinot et al. (2009) found that details provided in initial free recall are more accurate than details recalled in the questioning phase. This is consistent with laboratory work and can be used to advise interviewing procedures currently used, such as the ABE, to suggest that free recall accounts be prioritised with a reduction in the amount of questioning in the questioning phases, to increase the overall accuracy of the memory information received.

It is clear from the results that most studies focused on violent and major crimes including robbery, sexual assault, and murders. This is perhaps because these crimes will have the most impact on the witness and victim and hence are considered the most valuable to study. Furthermore, the present review has suggested that most individuals who are involved in crime situations are not of university age, which most laboratory-based studies are, possibly affecting their generalisability (Pozzulo, 2006). Such laboratory research should therefore be treated with caution when trying to use their results to come to conclusions about the effect of different witnessing factors on memory reporting.

Overall, whilst some of the findings from laboratory studies have been supported with the current reviews results, there are some contradictory findings. This information can be used to inform how to question witnesses and victims to produce the most reliable memory reporting. However, more field studies need to be conducted to replicate these findings and come to more robust conclusion. Whilst there are factors that could potentially weaken memory as in line with what experts believe (Kassin et al., 2001), it also needs to be communicated that some factors do not appear to produce a memory compromising effect in the field despite the laboratory research findings, and hence memory reporting in these situations may not be less accurate. It is also important to separate accuracy and length when assessing statements as they can lead to different conclusions regarding the effect of a specific factor on memory reporting. It is suggested that assessing accuracy rather than length can provide more forensically relevant conclusions on the effect of factors on memory reporting.

Limitations

A computer-based search of a range of databases provided 12 papers that met the inclusion criteria and quality requirements. An additional four papers were identified through manual searches of reference lists. This may mean that some papers have been missed; however, the papers that were identified were relevant and met the criteria despite not being identified in the initial search and hence were included. Human error is present throughout the process of conducting systematic literature reviews, from determining the search terms to screening and quality assessment procedures, and hence 100% accuracy of the search is unlikely (Wang et al., 2020).

There are some inevitable limitations to this review. Firstly, there is no objective verification of information given in statements (Chae, 2010). Out of all the ways of assessing statements that were discussed the method of comparing the statements to other evidence from the crime can be considered the most conclusive method for accuracy. However, as the crime event was not controlled as are in laboratory experiments, there is not definitive way of determining accuracy. Therefore, whilst conclusions can be drawn, the causal effects of the factors studied cannot be determined and hence this must be considered when generalising results.

Furthermore, there are several different variables that are involved in actual crime events which can ultimately affect eyewitness performance (Chae, 2010). There are great contrasts between the different crimes studied even if they are of the same type. This means that there would be different variations of extraneous variables. For example, even though multiple studies reviewed the relationship between retention interval on memory reporting for robberies (e.g., Odinot et al., 2009; Christianson & Hubinette, 1993), the

details of these robberies were drastically different. Furthermore, even within the same crime, all witnesses and victims will have different variations in lighting, distance, etc. which will create differences in memory reporting for the same crime. In laboratory studies this can be controlled for. This complexity in the field makes it difficult to be able to untangle the effects of the factors being studied. For example, Woolnough and MacLeod (2001) concluded that victims were able to report more details than bystanders which could be due to their central role and increased stress levels; however, this may be confounded by the fact that victims had better viewing conditions. As a result, many of the reviewed studies have used an associational design and no causal conclusions can be made. However, the nature of a real-life crime will always involve extraneous variables and hence it is still important that we study the effects on memory without controlling for these.

Lastly, the sample sizes in some of the studies in the review are small. This may increase the margin of error and therefore effect ability to come to conclusions on the factors investigated (Faber & Fonseca, 2014). Low power also increases the risk of spurious effects and type two errors. However, this area of study is niche and hence it would be difficult to obtain larger numbers.

Future Recommendations

There are avenues that should be studied in the field to further our knowledge of eyewitness memory in real crime events. Firstly, most crimes presented in this review have been violent, including robbery and sexual offences. However, the prevalence of these crimes has reduced, whereas crimes such as drug offences and public order offences

have increased in the year ending December 2020 (Office for National Statistics, 2020). Memory of these non-violent crimes have not been captured by the discussed studies, possibly due to the lack of impact on the witness and hence may not be considered as valuable to study. However, this needs to be considered as the current literature may not provide information that can be generalised to witnesses and victims of other nonviolent crimes.

Furthermore, there are also several important factors that may affect eyewitness memory in statements that have little to no research in the field. These include the effect of alcohol and drugs, and own-race bias, among other factors. The negative effect of alcohol on eyewitness memory is present throughout laboratory research findings (Yuille & Tollestrup, 1990; Hagsand et al., 2013). Furthermore, the own-race effect states that individuals are better able to recognise faces of their own race compared to faces of another (Brigham et al., 2007). However, Kuehn (1974) suggested that there was no effect of alcohol on completeness of statements and there was no evidence of the own-race effect in real-life crimes. This study was not used in the current review due to not meeting the quality assessment requirements but is worth mentioning as it provides opposing results from findings of laboratory studies. These are important factors to investigate further because these factors are believed by experts to effect eyewitness testimony (Kassin et al., 2001), victims are commonly under the influence of alcohol especially during sexual assaults (up to 74% of these crimes involve intoxication; Bureau of Justice Statistics, 2005) and eyewitnesses often must describe perpetrators of a different race / ethnicity as themselves (Sporer, 2001). Future research should focus on determining the effects of these factors on victim memory reporting in the field.

Additionally, there has been a wide range of methods which were used to assess the statements, with there being no standardised method. It is important that future research is conducted on the most efficient method to replicate results and find the most effective method of assessing statements. Sporer (1996) distinguishes between quantitative and qualitative aspects of statements. From this review, the most widely used method of assessing statements was verifying the statement information using other information, closely followed by the amount of information given in a statement. These quantitative measures are more widely used due to the fact it can be objectively and reliably measured (Sporer, 1996). Nonetheless, a qualitative method that shows promise is comprehensiveness (i.e., detail) of memory information. Whilst one study (Vredevelt et al., 2015) utilised this method to assess their statements, the rest did not. It may be that studying comprehensiveness is more relevant as this will both help to understand how more specific memory information can be obtained and as a result will provide more forensically relevant information to help an investigation. This needs to be considered for future research.

Lastly, the studies have not discussed a theoretical basis for how memory is monitored. As was determined by laboratory studies, the quantity accuracy trade-off framework (Koriat & Goldsmith, 1996) has been discussed as a possible mechanism, but it is unclear and not communicated in the studies included in this review whether individuals use a similar mechanism in real-life situations. The quantity accuracy trade-off framework would suggest that when there are memory weakening factors, an eyewitness would volunteer less information to maintain accuracy. In line with this, the current review has suggested that the presence of some memory weaking factors are not necessarily associated with

poorer memory accuracy for crimes. However, no studies have formally tested this hypothesis. Therefore, this is another aim for future research.

Gaining an understanding into the mechanisms that underlie memory reporting for real world crimes can allow for strategies to be recommended to police and court room interview procedures to gain the most forensically relevant and accurate information. For example, it may be appropriate to recommend free recall techniques when taking statements rather than asking questions. These further questions may mean that witnesses give information that they are less confident on which may not be accurate and ultimately may divert the course of the investigation. This also helps to know how much weight to place on eyewitness testimony, despite witnessing factors.

Conclusion

The aim of this review was to investigate how researchers have studied memory for crime events in real-world witnessing contexts and review the witnessing factors that are associated with memory accuracy and completeness. Laboratory studies have suggested that the presence of memory weakening factors negatively affect the memory of witnesses for a crime. However, the current review showed that many of the factors investigated have contradicted these laboratory findings in that they are not necessarily associated with poorer memory accuracy. However, future research needs to focus on the mechanisms that underlie these findings, particularly focusing on the quantity accuracy trade-off framework (Koriat & Goldsmith, 1996).

Chapter Three

Using the Quantity Accuracy Tradeoff Framework to Understand Memory Reporting in Real World Witnesses

Abstract

Research in the laboratory shows that certain factors of a crime situation (e.g., witness intoxication, long retention interval between the crime and questioning) may compromise memory, but there is limited research with real witnesses. The aim of this study is to analyse how the number and accuracy of details witnesses provide in their description of perpetrators vary in relation to the presence of memory compromising factors. It is a secondary data analysis using police crime incident reports which included information such as crime details (e.g., weapon used, nature of any injury, lighting, duration of crime, distance from crime, etc.) and the suspect's physical appearance (including age, gender, race, height, weight, build, eye colour, hair colour, hair length, hair type, facial hair, complexion). An identical physical appearance checklist was also completed by the police when a suspect was arrested allowing for correspondence between the witness' description of the perpetrator and the suspect's appearance to be determined (coded as either 'exact match' or 'no match'). The results showed that, together, the presence of memory compromising factors did affect the number of exact and no match features, meaning that statements may be briefer under circumstances that are assumed to compromise memory. The presence of memory compromising factors also affected the accuracy for exact match features but not no match features suggesting that correct features reported may be less specific under these conditions. A greater number of memory compromising factors did not necessarily lead to less detailed or less accurate memory, and some were in the opposite direction to predicted (e.g., some were associated with better memory). The findings largely support the quantity-accuracy trade-off framework, which states that individuals strategically regulate their memory reporting and only volunteer information in

which they are confident. Implications of the findings for practice are discussed, including

suggestions for police interviewing victims when potentially memory compromising

factors are present.

Introduction

Issues with Interviewing Witnesses

Witnesses and victims of violent offences have often expressed concerns about reporting their experience of a crime to the police (Champion et al., 2021). Such worries focus on confidentiality and a fear of negative reactions such as blame and stigmatisation (Champion et al., 2021) and witnesses report that giving testimony to the police can be an extremely stressful process. This indicates a view that those who took the witness statements did not believe the witness due to a perceived lack of memory relating to aspects of the crime (McMillan & Thomas, 2009). Interviewers have been shown to exert dominance leading to witness responses being fueled by anxiety which is associated with omissions critical information (Holmberg, 2004). Witnesses feeling invalidated in this way leads ultimately to the prevention of disclosing complete information of the crime thereby preventing justice being served.

Furthermore, inaccurate testimonies from witnesses can also have an effect on the accused. Loftus (1975) found that eyewitness corroboration of circumstantial evidence nearly quadrupled conviction rates by mock jurors even when it was known that the eyewitness had poor visibility of the crime. Similarly, weak testimony evidence was found to be a major factor involved in the wrongful convictions of 250 DNA exoneration cases (Garrett, 2011). This suggests that the procedures used by investigators to collect and evaluate memory reports may be inadequate. Jurors are not told of the confidence level of the witness (Garrett, 2011) and hence this leads to jurors overestimating the importance of eyewitness testimony, potentially leading to tragic real-world implications, such as wrongful convictions. It is imperative therefore to investigate factors associated with memory reporting accuracy to ensure fair justice outcomes.

Factors Affecting Memory

Memory accuracy can be affected during encoding and retrieval conditions. Both laboratory and field studies have been utilised to examine the effects of certain factors on memory reporting. Laboratory studies are defined as those that are carried out in controlled conditions, whereas field experiments are those that are carried out in a natural setting and include real witnesses (i.e., is not a staged event). As discussed in Chapter Two, the findings from the laboratory studies do not always translate to the field hence meaning that further research is required. There is a belief amongst eyewitness experts that certain factors can affect memory encoding and retrieval, affecting the accuracy of testimony given by a witness (Kassin et al., 2001). The study by Kassin el al. (2001) asked 64 psychologists about their opinions on 30 eyewitness phenomena. There was a strong consensus that the effect of the following phenomena was reliable to present in court: age of the witness, ethnicity of the perpetrator and the witness (i.e., own-race bias), retention interval, alcohol and drug intoxication, duration of exposure, relationship of the perpetrator to the eyewitness, lighting, distance, and stress. Furthermore, the Turnbull Guidelines (R v Turnbull and others, 1976), in which a judge instructs the jury to consider the circumstances of each witness's identification, instructs jury members to consider similar factors to those mentioned, including distance, lighting and retention interval, when interpreting eyewitness memory evidence. Most of these factors have also been in investigated in laboratory studies, which find these factors have negative effects on memory. The factors will now be discussed in turn.

Age of Witness

Experts generally believe that young children are less accurate witnesses than adults and are more vulnerable than adults to interviewer suggestion, peer pressures, and other social influences that could decrease the accuracy of testimony (Kassin et al., 2001). Brain regions involved with eyewitness performance include the hippocampus, prefrontal cortex, temporal lobe, and amygdala which continue to develop throughout childhood. For example, the hippocampus is an area important for the formation of long-term memory but is not completely developed until between two and eight years of age (Noa, 2009). Similarly, the prefrontal cortex is involved in creating vivid memories with contextual detail and this area is one of the last regions to develop (Noa, 2009). The temporal lobe is a location involved in memory storage and is associated with memory skills; however, the grey matter continues to develop until its peak at 16 years of age (Gogtay et al., 2004). Lastly, the amygdala is involved in the acquisition and retrieval of highly salient events, and it does not stop developing until late adolescence (Schumann, 2004). Underdevelopment of these areas can affect both the amount and specificity of information recalled by children. Studies in the laboratory indeed find that the number of details recalled for an event increases with age, as does the ability to answer specific questions, identify suspects, and resist suggestions (e.g., Goodman et al., 2001). This finding is also supported in a study where 10-year-olds perform more poorly in identifying a suspect in a staged crime than did older individuals aged 16 and above (Brigham et al., 1986).

However, there is also a point where, as age increases, memory starts to decline. Studies have shown that this can occur from as early as 45 years old (Ferrie, 2002). This has led to experts to believe that elderly eyewitnesses are less accurate than are younger adults (Kassin et al., 2001). Although the brain shrinks with age due to expansion of ventricles, it is difficult to associate the shrinking brain with loss of memory (Baddeley et al., 2020). It is well-known, however, that with ageing, memory can hold less information (Abrams & Farrell, 2012). This

means that an individual typically becomes more limited in the amount of information they can retain and subsequently recall. There are also age-related declines in attentional resources meaning that older adults are less able attend to multiple pieces of information due to a reduction in mental energy (Glisky, 2007). This can also reduce the amount of information that can be remembered from an event. Witnessing and subsequently recalling a crime would involve both the need to hold information in mind and attend to many aspects of the crime which an older adult may find more difficult. This has been confirmed in mock-eyewitness studies in the laboratory which found that older adults were more likely to be misled by false information and provided less complete statements compared to younger adults (List, 1986; West & Stone, 2014). However, it was also found that their statements were no less accurate despite the lack of completeness (List, 1986). This is also shown in the field where older adults (aged 45-65) were least superior in recalling physical characteristics of a confederate they had spoken to two minutes earlier (Yarmey, 1993). However, there is a lack of research in real-life crime situations comparing the effect of different ages on memory monitoring and recall.

Own-race Bias

The own-race bias is the phenomenon whereby people are better able to recognise faces of their own race compared to other races (Shriver et al., 2007). In the context of a crime, this would mean that if the perpetrator is of a different race to the eyewitness, the eyewitness may have more difficulty remembering features and identifying a perpetrator compared to if the perpetrator was of the same race. Multiple theories have been developed to explain why the own-race bias exists, including social cognition and perceptual expertise explanations. The social cognition theory states that people think categorically about out-group members compared to in-group members and do not notice subtle variations in a different race (Young et al., 2011). The perceptual expertise hypothesis, on the other hand, states that extensive

exposure facilitates perceptual mechanisms that allow people to be more proficient at recognising faces regardless of if the race matches one's own (Herzmann et al., 2011). The perceptual expertise hypothesis is supported by a relatively recent study, for example, which found that there was no evidence of the own-race bias for Chinese faces in a Malaysian non-Chinese population, ostensibly because Malaysian individuals are largely exposed to Chinese faces, as Chinese faces are one of the predominant racial groups in Malaysia (Estudillo et al., 2019). However, no theory has been able to fully account for the full body of evidence for the own-race bias and hence multiple theories are still considered viable accounts (Young et al., 2011).

Most of the research into the own-race bias in eyewitness testimony has focused on identification procedures. An identification procedure involves showing a witness lineup member, including the suspect(s), either simultaneously (at one time) or sequentially (one after the other) so they can attempt to identify who they saw commit the crime. A meta-analysis concluded that the own-race bias reduced the number of correct identifications and increased the number of false identifications showing its potential prominence in criminal justice outcomes (Meissner & Brigham, 2001). However, identification uses different memory processes compared to giving testimony. Ellis et al. (1975) did not find a clear own-race bias in the verbal descriptions produced when participants were asked to describe faces of a different race and hence there are still unanswered questions on whether this effect still stands when an eyewitness is giving testimony and providing details about the perpetrator. Furthermore, there is also a lack of research into this effect outside of laboratory settings on real-life crimes and so it is not clear the extent to which the own race bias is observable in real crimes.

Retention Interval

The forgetting curve (Ebbinghaus, 1964) refers to the tendency for people to more accurately retrieve newly created memories compared to older ones. Experts also believe that forgetting is a factor that affects eyewitness memory testimonies and therefore the retention interval (i.e., the time between the crime and questioning of the eyewitness) is an important determinant of memory accuracy (Kassin et al., 2001). The retention interval is also mentioned in the Turnbull Guidelines (R v Turnbull and others, 1976) in which juries are instructed to consider the time that has elapsed between the crime and the recall. It has been suggested that amount of accurate correct details recalled can decrease as the retention interval increases and this leads to an increased proportion of errors (Shapira & Pansky, 2019). This is exemplified in mock-crime laboratory research, as participants recall more details and are more accurate during immediate compared to delayed recall (e.g., Hagsand et al., 2016).

However, these findings contrast with studies which were conducted in the field on real crime events which have found that memory for real crime events remained highly accurate even after retention intervals of five months (Yuille & Cutshall, 1986) and 15 months (Christianson & Hubinette, 1993; see Chapter Two). Whilst experts believe statements given after an extended period of time would be less accurate and hence diminish their credibility (Kassin et al., 2001), the current studies in the field do not support this. However, the speed of forgetting can be reduced when there is a better memory representation and when there is repetition based on active recall (Ebbinghaus, 1964). This suggests that if a witness was required to remember the event often, they would experience less forgetting. Indeed, there is research in the laboratory that shows that repeated retrieval attempts help to consolidate memory and reduce forgetting (Karpicke & Roediger, 2007). More research regarding retention interval is needed.

Alcohol and drug intoxication

Alcohol and drug intoxication is another factor cited by experts that negatively affect people's ability to remember crime events (Kassin et al., 2001). This belief is likely a result of both laboratory studies which have shown that acute alcohol administration impairs episodic memory (Mintzer, 2007) and popular culture where movies, books and television shows show characters drinking alcohol to the point of memory loss. However, recent studies have shown that witnesses who were alcohol-intoxicated were no more likely to recall incorrect information compared to those that were sober (Jores et al., 2019; Flowe et al., 2020). Indeed, levels of incorrect reporting are similar in those who were alcohol intoxicated at encoding compared to those who were not (Flowe et al., 2019). Alcohol has been linked to the disruption of hippocampal function, especially impacting on long-term potentiation (LTP) which is associated with learning and memory (Bliss & Collingridge, 1993). This could narrow the memory to important central details (Goodwin et al., 1970) but this does not necessarily affect accuracy of the details recalled (Flowe et al., 2019). Therefore, whilst intoxicated witnesses may not be able to remember as much (i.e., their testimony may be less complete), this does not mean that their memories are less accurate. However, these studies were conducted in laboratory conditions which, due to ethical limitations, meant that individuals were only intoxicated to a blood alcohol content (BAC) of 0.08% (the UK legal limit for driving, Drinkaware, n.d.). It is likely that blood alcohol concentration is often higher in a real-life situation and hence results may not be generalizable to such situations. Studies that have tested people self-intoxicating to higher BAC levels have found that intoxication impacts completeness of reports for a live interaction but did not test for accuracy (e.g., van Oorsouw et al., 2019)

The effects of taking illicit drugs on memory are less well researched than alcohol intoxication. Cannabis has been found to have a similar effect on memory as when people are

under the influence of alcohol, as those under the influence of cannabis have been shown to remember fewer correct details of a mock-crime but there not more likely to report incorrect details, than those who were sober (Vredeveldt et al., 2018). Therefore, like alcohol, cannabis appears to affect encoding of information into memory but not accuracy of the details that are recalled. However, there has been a lack of studies on cannabis conducted outside of laboratory conditions and on the effect of other illicit drugs on eyewitness memory.

Duration of Exposure

Another factor which can affect the accuracy of memory recall is the duration that the individual is exposed to the crime. Experts believe that the less time an eyewitness has to observe an event, the less well they will remember it (Kassin et al., 2001). This factor is also mentioned in the Turnbull Guidelines (R v Turnbull and others, 1976) in which juries are instructed to consider the amount of time the witness observed the suspect. Consistent with this idea, laboratory studies have found that the longer an event is observed, the more information is encoded and recalled (Loftus, 1979; Memon et al., 2003). This finding has also been supported by studies of real-life crimes, which show that longer exposure increases both the completeness and accuracy of memory reporting (van Koppen & Lochun, 1997; Fahsing et al., 2004; see Chapter Two) and identification accuracy (Valentine et al., 2003).

Relationship of Perpetrator to Eyewitness

One could assume that if an eyewitness knew the perpetrator before the crime, it would be easier to remember them at a later date. This factor is considered in the Turnbull Guidelines (R v Turnbull and others, 1976) in which juries are instructed to consider if the witness has ever seen the suspect before and how often. There is a lack of research into the effect of the

relationship between a perpetrator and the eyewitness on witness memory reporting. However, there is much research to show that facial recognition ability and face matching ability is much poorer for previously unfamiliar faces compared to familiar faces (Hancock et al., 2000; Megreya & Burton, 2006). However, these studies are related to identification, and research has not considered the effect of prior familiarity on witness statements.

Lighting

Low levels of lighting at encoding are thought to reduce the accuracy of eyewitness identification (Kassin et al., 2001) and as a result experts are sometimes asked to consider the lighting conditions present at the crime when evaluating the testimony of an eyewitness (Papailiou et al., 2014). This is because in relatively low light conditions, a face may not be perceived well enough to be recognised later (Wells & Olson, 2003). This factor is also mentioned in the Turnbull Guidelines (R v Turnbull and others, 1976) in which juries are instructed to consider in what light the witness viewed the suspect. Valentine et al. (2003) conducted a field study using Metropolitan Police data which investigated the effect of good (daylight or good indoor lighting) and poor (poor indoor lighting, twilight, or night) lighting conditions on whether a suspect was identified in a lineup. The authors found that there was no significant independent effect of lighting on being able to correctly identify a suspect from a lineup. This indicates that lighting may not have an effect on the ability of a witness recognise a perpetrator. However, such research has focused on identification of an offender. There is a lack of research into the effect of lighting on statements from eyewitnesses.

Distance

Experts were not asked about the effect of distance on memory accuracy in the Kassin et al. study (2001). It is, however, considered in the Turnbull Guidelines (R v Turnbull and others, 1976) in which juries are instructed to consider at what distance the witness viewed the

suspect. There has been laboratory research on the effect of distance on eyewitness memory; for example, Lampinen et al. (2014) measured participants ability to identify targets from a simultaneous lineup at distances ranging from five metres to 37 metres and found that discriminability decreased with distance. However, this finding has not been replicated in field research involving real crimes. Valentine et al. (2003) found that distance did not make any significant contributions to whether eyewitnesses identified suspects or a foil in real crime cases. Studies on the effect of distance on witness memory reporting have also been conducted in the field. For example, van Koppen and Lochun (1997) found greater distance between the offender and the witness had a negative effect on how complete the witness's statements were. Further research therefore is needed to better understand the effect of distance on memory accuracy for statements in real-life crime situations.

Stress (Severity of the Crime)

Stress is considered to affect the encoding of memory (Cowan, 2006). Whereas stress in small amounts can aid memory when stress hormones are released by the amygdala to promote the consolidation of emotional memory, stress in high amounts may hinder memory performance (McGaugh, 2004). A common assumption behind this view is based on the Yerkes-Dodson law (Yerkes & Dodson, 1908) that states there is a decrease in processing capacity in high states of emotional arousal which leads to less efficient memory processing. This suggests that crime events would be less well remembered due to the high amount of stress that is induced. Experts believe that very high levels of stress impair the accuracy of eyewitness testimony (Kassin et al., 2001). Contrasting this, a meta-analysis on both laboratory and field studies conducted on emotional stress and eyewitness memory concluded that highly negative events are relatively well retained; however, peripheral details are less accurately retained (Christianson, 1992b). Chapter Two discussed the findings of field studies that have looked at the effect of stress on memory reporting and found mixed results, with

some concluding that increased arousal leads improves accurate memory reporting whereas others found that it can have a negative effect. Due to these contrasting results, it is important that this effect is investigated further in the field.

Weapon Focus

Stress has been measured in various ways across studies. Weapon presence is used as a proxy variable for how much stress may have been experienced, assuming that a weapon would typically be associated with higher stress levels (Christianson, 1992b). The presence of a weapon at encoding is a factor believed by experts to affect an eyewitness' ability to accurately identify and describe a perpetrator's face (Kassin et al., 2001). The presence of a weapon at encoding is thought to result in the weapon focus effect; the idea that the witness will focus on the central detail of the weapon and will lose focus of the peripheral details which will result in worse recall of the perpetrator or other details of the scene (Loftus et al., 1987; Carlson et al., 2016). The arousal/ threat hypothesis discussed by Fawcett et al. (2013) is based on the Yerkes-Dodson law (Yerkes & Dodson, 1908) which suggests that in high stress situations where a weapon is present, there is a reduction in mental resources. Therefore, the range of cues that a subject can attend to is reduced leading to focusing on the object of arousal (the weapon) and miss peripheral details (such as the identity of the offender) (Easterbrook, 1959). Several early studies have supported this notion, showing that individuals give greater attention to weapons, and this subsequently has negative effects on recall (Ross et al., 1994). However, this effect is typically tested in laboratory settings and hence has been argued to be too small to be detected in complex, real-life crimes (Griesel & Yuille, 2012). It has, in fact, been seen to enhance identification accuracy for offenders in mock crime situations (Harvey et al., 2020). The absence of this effect in real life scenarios is also supported by field studies on crimes which have suggested that there is no significant

effect on memory despite the presence of a weapon (Tollestrup et al., 1994; Cooper et al., 2002; Wagstaff et al., 2003; see Chapter Two).

Injury to the Witness

Another proxy variable for how much stress may have been experienced includes whether the witness was injured (Woolnough & Macleod, 2001). The amount of arousal experienced by a witness is hypothesised to increase with increasing levels of violence and this arousal is thought to effect memory encoding and retrieval (Christianson, 1992a) in line with the Yerkes-Dodson law (Yerkes & Dodson, 1908). Several studies have investigated the effect of injury to the witness in the field and, despite the theory, it was concluded that enhancement of arousal (measured through incident seriousness) was associated with more accurate memory (Woolnough & MacLeod, 2001; Wagstaff et al., 2003; see Chapter 2). Research into this factor is often mixed between the laboratory and field and there is limited research to come to conclusions on how it influences witness memory reporting.

Monitoring Memory

There is evidence that individuals strategically regulate their memory reporting which affects the amount of information given in testimonies. According to the quantity accuracy trade-off framework, witnesses volunteer to the police only memories about which they are confident (Koriat & Goldsmith, 1996). The quantity accuracy trade-off states that when individuals are questioned by the police, they use a monitoring mechanism that assesses the correctness of a potential answer, and a control mechanism that determines whether to volunteer it. The participant sets a response criterion, and they report the candidate response if it exceeds this criterion. The quantity accuracy trade-off also considers the gain for providing correct information relative to the cost of providing wrong information which would further alter the criterion (Koriat & Goldsmith, 1996). For example, as there are

potentially detrimental consequences of getting information wrong, such as wrongful conviction, the eyewitness would only report information for which they are very sure (and therefore likely accurate).

The quantity accuracy trade-off was examined in relation to eyewitness memory by Palmer et al., (2013) who proposed that eyewitnesses make 'theory-based' confidence judgements when they are aware that there are factors that can weaken their memory. They state that an individual will take information about possible memory compromising factors into account when evaluating their confidence on the accuracy of memory and recall information based on this newly calibrated confidence. Therefore, if encoding conditions are unfavourable (i.e., factors that can weaken memory are present), individuals would be less confident and therefore volunteer less information. Critically, though, this does not necessarily mean that the reported information would be less accurate. Therefore, whilst there are factors that influence memory accuracy (as described above), people sometimes know that their memory is inaccurate (i.e., they monitor memory) and withhold answers of which they are unsure.

 Consistent with the idea that people can monitor their memory, research finds that people can communicate their likely memory accuracy using appropriate confidence judgements (i.e., are highly confident when they are highly accurate, and less confident when they are less accurate e.g., Wixted & Wells, 2017). For example, Wixted et al., (2016) found that even after very long retention intervals, the relationship between confidence and accuracy on an identification task was the same as for short intervals. That is, regardless of the retention interval, high confidence identifications of a suspect were associated with high accuracy. Similarly, Colloff et al. (2017) found that even though older adults had poorer

memory discrimination accuracy compared to younger adults, they seemed to adjust their confidence reports accordingly and so were still as accurate as younger adults at when they were highly confident. This finding has also been found using child witnesses (Winsor et al., 2021), intoxicated witnesses (Flowe et al., 2017) and when the suspect was of a different race (own-race bias; Nguyen et al., 2017). Therefore, regardless of whether there was a memory compromising factor present at encoding, high confidence is typically associated with high accuracy in suspect identification (although there are some exceptions to this general rule, see Sauer et al., 2010).

Importantly, the interviewing procedures used in practice have the potential to change a witness' quantity accuracy tradeoff (Wixted et al., 2018). Some practices—such as interrupting the witness, asking too many short answer questions and inappropriate sequencing of questions—can affect the confidence of the witness and contribute to inaccurate recollections, therefore reducing the confidence-accuracy relationship (Fisher, 1995). As a result, the mistakes that made by eyewitnesses can be considered as a product of inappropriate legal system procedures rather than memory being inherently faulty (Wixted et al., 2018). In the UK, the Achieving Best Evidence (ABE) guidance states that police interviewers should ask witnesses to provide a free recall account, whereby they report everything in their own words without interruption (Ministry of Justice, 2011). In this phase, the quantity accuracy tradeoff mechanism is likely to be involved and hence determines how much, and the specificity of information given by the individual. In other words, the witness can decide what to report. Thereafter, however, the questioning phase takes place, in which witnesses are "systematically probed using open-ended and specific-closed questions until all the relevant material relating to it has been obtained" (Ministry of Justice, 2011, p.76). These following questions may lead witnesses to lower their report

criterion, and report details they are unsure about that they otherwise would not have reported, thereby reducing the accuracy of their memory evidence overall. Indeed, accuracy of testimony is higher during the free recall phase than the questioning phase (Flowe et al., 2019).

Together this research suggests that in real life crimes, when witnesses know the costs of providing inaccurate information (so set a high threshold for reporting) and are given the option of what they want to report, the presence of memory compromising factors at encoding might not affect accuracy, they may only reduce the amount of information reported. However, it may be that the procedures used to collect memory information make witnesses lower their report criterion, hence leading to inaccurate memory reports.

The Current Research

Research on the conditions that affect witness memory has been largely undertaken utilizing suspect identification procedures in the laboratory instead of in the field with actual witnesses. This is because memory accuracy can be measured in controlled laboratory study, whereas ground truth is usually unknown in actual criminal cases, unless the crime is captured on CCTV, for example. The benefits and limitations of laboratory vs real-world research is more widely discussed in Chapter Two. Nevertheless, there is some literature that has explored whether the conditions that have been found to affect remembering in the laboratory are associated with the quality of memory reports given by actual witnesses. In Chapter Two of this thesis, a systematic review of this literature is presented. One of the main findings from the review is that most field research analyses the quantity and type of details reported. The present investigation presented in this chapter extends this literature

by analysing descriptions reported to the police by witnesses in actual criminal cases and making predictions using the quantity-accuracy framework. The analyses explore whether the level of detail (i.e., number of features in statements) or accuracy (i.e., correspondence between recalled features and actual perpetrator features) varies in relation to the conditions in which witnesses encode the crime. If real world witnesses consider that some conditions lead to relatively poor memory (e.g., weapon present, low lighting, own-race perpetrator, long retention interval), then the quantity accuracy tradeoff framework predicts that they should give descriptions of perpetrators that are briefer, but not less accurate.

Aims

The aim of this study was to analyse how the number and accuracy of details witnesses provide in their description of the perpetrator vary in relation to the presence of memory compromising factors. Based on previous research, the memory compromising factors to be studied include: age of witness (child and older adults versus other age categories); own race bias; long retention interval between the crime and the police interview; acute intoxication with alcohol and other drugs; short duration of exposure; the perpetrator being a stranger; poorer lighting conditions; longer distance between the witness and the crime, presence of a weapon, and injury to the witness). Specifically, the first aim was to examine whether the presence of each of these factors was associated with the witness providing a less detailed (i.e., the number of features given) and accurate (i.e., the proportion of correct features) description of the perpetrator. A second aim was to examine if the total number of memory compromising factors in a case is associated with witnesses providing a less detailed and accurate perpetrator description. These hypotheses will be described in more detail in the method section after the variables are operationalized.

Method

Sample

This study is a secondary data analysis of data collected by Flowe et al. (2011). This included randomly sampled felony cases (rape n=196, robbery n=195, assault n=76 and unlabelled crime type n=41) from San Diego District Attorney's closed case archives between the years of 1991 and 2000 comprising of adult suspects who were arrested for these crimes.

Not all cases had complete information regarding witness demographics. However, from the information available, the mean age was 29.66 years (range between four and 86 years). Furthermore, there was information regarding the sex (females=288 and males=211) and the ethnicity (White=234, Hispanic=146, Black=76, Asian=19, Unknown=10 and Other=3) of witnesses.

Materials

The Flowe et al. (2011) study focused on the coding of closed case files and the prevalence of the measures across the sample. The case files obtained included police records, preliminary trial transcripts, probation officer's assessment of the defendant's personal history, and the sentencing agreement. Police crime incident reports included eyewitness descriptors of crime information (including weapon used, nature of any injury, lighting, duration of crime, distance from crime, etc.) and of the perpetrator's physical appearance (including age, gender, race, height, weight, build, eye colour, hair colour, hair length, hair type, facial hair, complexion). An identical physical appearance checklist was completed by the police when a suspect was arrested, allowing for correspondence between the witness' description of the perpetrator and the person later convicted of the offence's appearance to

be determined. In the Flowe et al. (2011) study, the cases in the sample were coded for the measures described below by 46 thoroughly trained research assistants.

Measures

Types of information coded for in the Flowe et al. (2011) study included suspect and victim characteristics, crime incident factors and eyewitness variables. Flowe et al. (2011) also coded for whether the suspect's features recalled by the eyewitness were an 'exact match', 'match' or 'no match' to the apprehended suspect's characteristics.

In the current study, the suspect and victim characteristics, crime incident factors and eyewitness variables available in the Flowe et al. (2011) data were coded as to whether they were considered a memory compromising factor or not. The variables selected for investigation in the current study were predominantly based on the factors believed to affect eyewitness memory as stated in Kassin et al. (2001). Namely, age of witness, own-race bias, retention interval, alcohol and drug intoxication, lighting, duration of exposure, and stress. Distance was also included in the current study due to the association between increased distance and poorer eyewitness memory observed in controlled experimental studies and real crime events (e.g., van Koppen & Lochun, 1997; Valentine et al., 2003). The relationship of the perpetrator to the eyewitness was also included due to laboratory research showing that facial recognition ability is much poorer for previously unfamiliar faces compared to familiar faces (e.g., Megreya & Burton, 2006). Table 1 summarises how the variables were operationalized for the purposes of the current study. In short, the presence of a possible memory compromising factor was coded as 1 and the absence of a memory compromising factor was coded as 0.

Table 2 Variables coded in the current study.

Variable	Definition
Exploratory Variables	
Age of Witness	If the eyewitness was younger than age 16 years or older than age 45, this was coded as 1; any other age was coded as 0 (Note: previous research has used these age categories for children (Brigham et al., 1986) and older adults (Yarmey, 1993)).
Own-race Bias	If the eyewitness was of the same race as the suspect, this was coded as 0; if the eyewitness was a different race to the perpetrator, this was coded as 1.
Retention Interval	If the suspect was questioned less than 24 hours after the crime, this was coded as 0; if the suspect was questioned over 24 hours after the crime, this was coded as 1 (Note: previous studies, especially those in the field have used longer retention intervals. However, Shapira and Pansky (2019) found that memory retention has a relatively large initial decline over the first 24 hours and a smaller non-significant decline thereafter).
Alcohol and Drug Intoxication	If the eyewitness was not under the influence, this was coded as 0; if they were under the influence, this was coded as 1.
Eyewitness Knew Suspect (Relationship of Suspect to Eyewitness)	If the suspect was known to the eyewitness, this was coded as 0; if the suspect was a stranger, this was coded as 1.

Exposure to Suspect Before the Crime (Relationship of Suspect to Eyewitness)	If the eyewitness first saw the suspect during the crime, this was coded as 1; if they knew or interacted with the suspect before the crime, this was coded as 0.
Lighting	If the crime occurred in daylight or indoor light, this was coded as 0; if it occurred in any other lighting, this was coded as 1.
Distance	If eyewitness was less than the median distance (15 inches), this was coded as 0; if the eyewitness greater than the median distance, this was coded as 1.
Crime Length (Duration of Exposure)	If the duration of the crime was more than the median (5 minutes), this was coded as 0; if it was less than the median, this was coded as 1.
Exposure to Suspect's Face (Duration of Exposure)	If the duration of exposure to the suspect's face was more than the median (6 minutes); this was coded as 0, if it was less than the median, this was coded as 1.
Witness Experienced Violence	If the eyewitness did not experience violence, this was coded as 0; if they did experience violence, this was coded as 1.
Weapon Present	If the perpetrator did not use a weapon, this was coded as 0; if they did use a weapon, this was coded as 1.
Witness Injured	If the eyewitness was not injured, this was coded as 0; if they were injured. this was coded as 1.

Number of Memory Compromising Factors	The number of compromising factors (those above coded as 1) for each case were totalled. The maximum number of possible memory compromising factors was 13.
Dependent Variables	
Number Exact Match	The number of features recalled by the eyewitness that were an exact match to suspect characteristics were totalled for each case. (Note: only features that are relatively stable were used in the analysis, including age, gender, race, height, weight, build, eye colour, hair colour, hair length, hair type, facial hair, complexion. Other features were also present in the data set, for example, demeanor, voice, and clothing but these were not included due to the likelihood of these being changed from when the witness gives testimony and the arrest of the suspect). The maximum number of features was 13.
Proportion Exact Match	The number of features recalled by the eyewitness that were an exact match to suspect characteristics (as noted by the police) was divided by the overall number of features recalled by the eyewitness for each case.
Number Non-Match	The number of features recalled by the eyewitness that were not a match to suspect characteristics were totalled for each case.
Proportion Non-Match	The number of features recalled by the eyewitness that were not a match to suspect characteristics (as noted by the police) was divided by overall number of features recalled by the eyewitness for each case.

It is important to add that the current study assumes that if the suspect characteristics matches the eyewitness description that this equates to accurate eyewitness memory.

However, it is also possible that the suspect may not be guilty of the crime (i.e., the ground truth is never certain).

Hypotheses

The aims of the current study are to examine whether the presence memory compromising factors was associated with the witness providing a less detailed (i.e., the number of features given) and accurate (i.e., the proportion of correct features) description of the perpetrator. A second aim was to examine if the total number of memory compromising factors is associated with witnesses providing a less detailed and accurate perpetrator description. In terms of the current data set, this translates to the following hypotheses:

1) The presence of each memory compromising factor is predicted to be associated with a decrease in the *number* of 'exact match' features. Previous research has shown that in situations where there are factors that can weaken their memory, participants would only volunteer answers when they were likely to be correct (Palmer et al., 2013), meaning they are less likely to report features of which they are unsure.

2) Experts and some research suggest that memory compromising factors can affect memory for crimes (Kassin et al., 2001) and hence the presence of memory compromising factors may lead to an increase in errors, that is, an increase in the *number* of 'no match' features. However, the quantity-accuracy tradeoff framework suggests that, when given the choice, individuals only volunteer answers when they were likely to be correct (Palmer et al., 2013) and so memory compromising factors may not be associated with an increase in the number of 'no match' features in real-world witnesses.

3) Experts and some research suggest that memory compromising factors can affect memory for crimes (Kassin et al., 2001) and hence the presence of memory

compromising factors may lead to an increase in errors, that is, a decrease in the *proportion* of 'exact match' features and an increase in the *proportion* of 'no match' features. However, as discussed in the introduction, individuals, when given the opportunity, can often strategically regulate their memory reporting (Koriat & Goldsmith, 1996) meaning that the overall accuracy of statements may not be affected by memory compromising factors in real worl witnesses. That it, for real-world witnesses, there may be no association between the presence of memory compromising factors and the *proportion* of 'exact match' and 'no match' features. In the current data set, accuracy can be operationalized by the proportion of exact and no match features, as if there is a larger amount of exact match features compared to the overall number of features recalled, the statement can be seen as more 'accurate'. Similarly, if there is a larger amount of no match features compared to the overall number of features recalled, the statement can be seen as more 'inaccurate'.

4) A negative correlation between the total number of memory compromising factors and the number of 'exact match' features is predicted. Experts and some research suggest that memory compromising factors can affect memory for crimes (Kassin et al., 2001) and hence an increase in the total number of memory compromising factors may lead to a decrease in the number 'exact match' features (i.e., completeness).

5) With similar reasoning to above, a positive correlation between the total number of memory compromising factors and the number of 'no match' features are predicted, as an increase in the total number of memory compromising factors may lead to an increase in 'no match' features.

6) Experts and some research suggest that memory compromising factors can affect memory for crimes (Kassin et al., 2001) and hence an increase in the total number of memory compromising factors may lead to an increase in errors, that is, a negative correlation with the *proportion* of 'exact match' features and a positive correlation with the *proportion* of 'no match' features. However, as discussed, the quantity accuracy trade-off suggests that individuals, when given the opportunity, can often strategically regulate their memory reporting (Koriat & Goldsmith, 1996) meaning that the overall accuracy of statements of witnesses in the real-world may not differ in relation to the total number of memory compromising factors present. That is, there would be no correlation between the total number of memory compromising factors and the *proportion* of both 'exact' and 'no match' features.

Ethical Considerations

Ethical conditions have been considered in the design of the present study. The study uses data already collated by Flowe et al. (2011). In their study they used public data, but this has been de-identified before release to the current researcher and hence no information regarding the person or case is able to be identified.

Results

To test the hypotheses, the main analyses were multiple regressions to determine whether the presence of memory compromising factors was associated with the witness providing a less detailed (i.e., the number of features given) and accurate (i.e., the proportion of correct features) description of the perpetrator. Table 2 shows the descriptive statistics for each of the exploratory variables with respect to the dependent variables. The average number of exact match features given by witnesses was 6.12 (SD = 3.76; range = 13) and the average number of no match features given was 0.95 (SD = 1.28; range = 6).

Table 3 Descriptive statistics for the variables

Memory compromising factor and if it was present or absent		Number exact match		Proportion exact match		Number no match		Proportion no match	
	N	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Witness Age									
Absent	374	6.30	3.66	0.69	0.26	0.99	1.30	0.10	0.13
Present	106	5.93	3.83	0.72	0.26	0.88	1.23	0.10	0.13
Own-Race Bias									
Absent	242	7.38	3.54	0.74	0.21	1.09	1.33	0.10	0.12
Present	201	5.82	3.29	0.70	0.23	0.93	1.26	0.10	0.14
Alcohol and Drug Intoxication									
Absent	392	5.96	3.75	0.68	0.27	0.96	1.33	0.10	0.13
Present	78	7.33	3.84	0.74	0.24	0.88	1.12	0.09	0.12

Retention Interval									
Absent	284	6.64	3.7	0.7	0.15	1.08	1.37	0.11	0.14
Present	80	7.96	3.15	0.72	0.22	1.26	1.42	0.11	0.12
Eyewitness Knew Suspect									
Absent	286	7.48	3.62	0.72	0.23	1.03	1.27	0.10	0.11
Present	222	4.37	3.19	0.65	0.30	0.83	1.28	0.10	0.15
Exposure to Suspect Before the Crime									
Absent	312	7.03	3.72	0.71	0.26	1.05	1.39	0.10	0.12
Present	50	4.76	3.04	0.65	0.25	1.20	1.29	0.14	0.16
Lighting									
Absent	272	6.39	3.86	0.70	0.27	0.95	1.32	0.10	0.13
Present	187	6.12	3.62	0.69	0.26	0.97	1.28	0.10	0.13
Distance									
Absent	419	6.53	3.78	0.7	0.26	0.99	1.3	0.10	0.13
Present	89	4.21	3.09	0.67	0.29	0.75	1.14	0.09	0.13
Crime Length									
Absent	242	6.43	3.72	0.71	0.26	0.93	1.26	0.09	0.13
Present	266	5.84	3.79	0.68	0.26	0.96	1.30	0.10	0.13
Exposure to Suspect Face									
Absent	148	7.66	3.53	0.71	0.24	1.15	1.41	0.11	0.13
Present	132	5.81	3.43	0.70	0.23	0.93	1.34	0.10	0.13
Witness Experienced Violence									
Absent	102	4.28	3.4	0.68	0.30	0.53	0.88	0.08	0.14
Present	362	6.73	3.71	0.70	0.26	1.07	1.37	0.10	0.13
Weapon Present									
Absent	153	4.72	3.34	0.69	0.28	0.66	0.96	0.85	0.13
Present	355	6.72	3.70	0.69	0.26	1.07	1.38	0.10	0.13
Witness Injured									
Absent	4	9.75	4.72	0.94	0.12	0.50	1.00	0.04	0.08

Present	187	7.65	3.42	0.71	0.23	1.17	1.42	0.11	0.12

Preliminary Analyses

Preliminary analyses were conducted to identify which variables had excessive missing data that would affect the regression analyses. If a variable has a lot of missing data, it affects all the variables because a whole case will not be entered into the regression if it has a single missing value, leading to a substantial loss of data. Variables that had less than 25% missing data were entered as predictors in the multiple regression model. The nine predictor variables were witness age, own-race bias, alcohol and drug intoxication, if the eyewitness knew the suspect, lighting, distance, crime length, if the witness experienced violence, and weapon presence. Retention interval, exposure to the suspect before the crime, exposure to the suspect face and witness injured were not entered into the regression as predictors variables due to missing data. The dependent variables were number of exact match features, proportion of exact match features, number of no match features, and proportion of no match features. This resulted in a final sample size of 387 cases which is appropriate for multiple regression with nine predictor variables, according to Tabachnick and Fidell (2019; see Appendix D).

Regression Analyses

Multiple regressions with enter method were used to predict whether the presence of each of memory compromising factor was associated with the witness providing a less detailed (i.e., the number of features given) and less accurate (i.e., the proportion of correct features) description of the perpetrator. A separate multiple regression was conducted for each of the four dependent variables.

When performing a hypothesis test with multiple comparisons, a Bonferroni test is commonly used to prevent a result that demonstrates statistical significance when there is none (Type 1 error; Bonferroni, 1936). However, most of the current study's hypotheses predict no statistical significance and hence using the Bonferroni test would be making it easier to achieve evidence for the null. Therefore, Bonferroni test will not be utilized in the current study.

Number of Exact Match Features

A multiple regression was conducted to predict number of exact match features from the nine predictor variables (i.e., witness age, own-race bias, alcohol and drug intoxication, eyewitness knew suspect, lighting, distance, crime length, witness experienced violence, weapon present). The results are shown in Table 3. The overall model was statistically significant, $F(9,378) = 15.12$, $p < .001$, $R^2 = 0.27$. The R^2 coefficient shows that 27% of the variability in number of exact match features was accounted for by the predictors in the model. Three variables were statistically significant to the prediction of the number of exact match features: own-race bias, eyewitness knew the suspect, and distance. This means that when the perpetrator is of a different race to the witness (own-race bias is present), there is an average decrease of 0.92 exact match features reported (B = -0.92, p = .01). Similarly, when the suspect was not known to the eyewitness, there was an average decrease in 2.67 exact match features (B = -2.67, $p < .001$). Lastly, when the eyewitness was more than the median distance (15 inches), there was an average decrease in 1.25 exact match features [B = -1.25, p = .01]. Whilst the other predictor variables were not significant to the model, consideration of the B values indicated that, descriptively speaking, most showed the direction expected for when a memory compromising factor is present. That is, young or old witness age, own-race bias, the eyewitness knowing the suspect before the crime, poor lighting, longer distance, and

a weapon being present, were associated with a decrease in the number of exact match features. However, descriptively speaking, alcohol and drug intoxication, decrease in crime length, and witness experiencing violence were associated with a small (non-significant) increase in the number of exact match features. Overall, in accordance with hypothesis one and previous research, this suggests that the presence of memory compromising factors at encoding is associated with the witness providing less 'exact match' features. However, only three variables were statistically significant.

Table 4. Output of the Multiple Regression for Number of Exact Match Features

	B	*p*
Witness Age	-0.37	.32
Own-Race Bias	-0.92	.01*
Alcohol and Drugs	0.14	.73
Eyewitness Knew Suspect	-2.67	<.001*
Lighting	-0.03	.94
Distance	-1.25	.01*
Crime Length	0.03	.93
Witness Experienced Violence	1.20	.35
Weapon Present	-0.79	.53

Number of No Match Features

A multiple regression was conducted to predict the number of no match features from the nine predictor variables (i.e., witness age, own-race bias, alcohol and drug intoxication, eyewitness knew suspect, lighting, distance, crime length, witness experienced violence, weapon present). The results are shown in Table 4. The overall model was statistically significant, $F(9,378) = 1.96$, $p = .04$, $R^2 = 0.05$. The R^2 coefficient shows that only 5% of the variability in number of no match features was accounted for by the predictors in the

model. None of the individual predictor variables were statistically significant in the model. However, descriptively speaking, four of the variables (increase in distance, decrease in crime length, witness experiencing violence, weapon present) were in the direction expected if memory compromising factors result in more memory errors, meaning that the presence of the memory compromising variable led to a small non-significant increase in the number of no match features. However, descriptively speaking, five of the variables showed the opposite direction that would be expected. That is, young or old witness age, own-race bias, alcohol and drug intoxication, eyewitness not knowing the suspect, and poor lighting (i.e., the presence of the memory compromising variable) led to a non-significant decrease in number of no match features. However, these are very small trends (the B vales are close to 0). Two alternate hypotheses (hypothesis two) were predicted for this analysis. On the one hand, the presence of memory compromising factors may lead to an increase in errors, that is, in the number of 'no match' features. On the other hand, the strategic regulation of memory predicts that individuals only volunteer answers when they were likely to be correct (Palmer, Brewer, Weber & Nagesh, 2013) and so memory compromising factors may not be associated with an increase in the number of 'no match' features. The overall model was statistically significant, but as there were no individual factors that were statistically significant and the R^2 was very small, neither hypothesis can be convincingly supported.

Table 5. *Output of the Multiple Regression for Number of No Match Features*

	B	*p*
Witness Age	-0.13	.42
Own-Race Bias	-0.18	.21
Alcohol and Drugs	-0.32	.09
Eyewitness Knew Suspect	-0.04	.80
Lighting	-0.03	.84

Distance	0.05	.80
Crime Length	0.12	.39
Witness Experienced Violence	0.54	.34
Weapon Present	0.09	.87

Proportion of Exact Match Features

A multiple regression was conducted to predict the proportion of exact match features from the predictor variables (i.e., witness age, own-race bias, alcohol and drug intoxication, eyewitness knew suspect, lighting, distance, crime length, witness experienced violence, weapon present). The results are shown in Table 5. The overall model was statistically significant, $F(9,378) = 2.08$, $p = .03$, $R^2 = 0.05$. The R^2 coefficient shows that only 5% of the variability in proportion of exact match features was accounted for by the predictors in the model. Only one variable was statistically significant to the prediction of proportion of the exact match features: witness age. This means that when the witness was younger (under 16 years old) or older (over 45 years old), there is an average increase of 0.06 proportion exact match features (B = 0.06, $p = .03$). Therefore, this effect is in the opposite direction than would be expected if memory compromising factors (i.e., being a child or older) result in more memory errors. However, the B value for age (and the other factors) were very small and close to zero. In accordance with hypothesis three, this overall significance of the model suggests that the presence of memory compromising factors can affect the overall accuracy of exact match features which does not support the quantity accuracy tradeoff theory. However, as the effect size of the model (R^2) was small, and there were no factors that significantly reduced the number of exact match features, it is not strong evidence for the hypothesis.

To explore the age finding in more detail, mean exact match memory accuracy was calculated separately for children, adults, and older adults. Descriptively speaking, older adults had higher proportion exact match features (0.74) than adults (0.68) and children (0.69), suggesting the significant effect of age in the regression analysis was due to superior performance of older adults. Children and adults appeared to perform similarly.

Table 6. Output of the Multiple Regression for Proportion of Exact Match Features

	B	*p*
Witness Age	0.06	.03*
Own-Race Bias	-0.03	.17
Alcohol and Drugs	0.04	.15
Eyewitness Knew Suspect	-0.04	.13
Lighting	-0.02	.30
Distance	0.01	.81
Crime Length	-0.00	.88
Witness Experienced Violence	0.06	.54
Weapon Present	-0.11	.23

Proportion of No Match Features

A multiple regression was conducted to predict proportion of no match features from the predictor variables (i.e., witness age, own-race bias, alcohol and drug intoxication, eyewitness knew suspect, lighting, distance, crime length, witness experienced violence, weapon present). The results are shown in Table 6. The overall model was not statistically significant, $F(9,378) = 1.23$, $p = .27$, $R^2 = 0.03$. As shown in Table 5, B values are small which is consistent with the idea that there is no meaningful effect of memory compromising variables on (in)accuracy. In contrast with hypothesis three, this suggests that the overall accuracy of no match features is not affected by the presence of memory compromising

factors, possibly because individuals can strategically regulate their memory reporting

(Koriat & Goldsmith, 1996). This supports the quantity accuracy tradeoff theory.

Table 7. Output of the Multiple Regression for Proportion of No Match Features

	B	*p*
Witness Age	0.00	.80
Own-Race Bias	-0.01	.61
Alcohol and Drugs	-0.03	.13
Eyewitness Knew Suspect	0.01	.33
Lighting	0.00	.99
Distance	0.01	.61
Crime Length	0.01	.30
Witness Experienced Violence	0.05	.37
Weapon Present	0.00	1.00

Correlation Analyses

Spearman's rho correlation coefficient was used to examine if there was a relationship

between the total number of memory compromising factors and both the number and

proportion of features provided by the witness about the perpetrator. Unlike the multiple

regression analysis, no predictor variables were excluded from the correlation analyses,

because missing variables do not affect correlational analyses. Therefore, there was a total of

12 possible memory compromising factors in these analyses. Correlation analyses were

conducted for both number exact match features and no match features. Cohen gives the

following interpretations for r values in correlations: small, $0.10 - < 0.30$; medium, $0.30 -$

<0.50, and large, ≥ 0.50 (Cohen, 1988).

Number of Exact Match Features

The results indicate a very small non-significant positive relationship between total number of memory compromising factors and number of exact match features, $r(508) = 0.04$, $p = .36$. Figure 1 shows a scatter graph of the results. Contrary with hypothesis four, this suggests that is no association between the total number of memory compromising factors and number of exact match features.

Figure 2. Scatter graph of the relationship between total number of memory compromising factors and number of exact match features
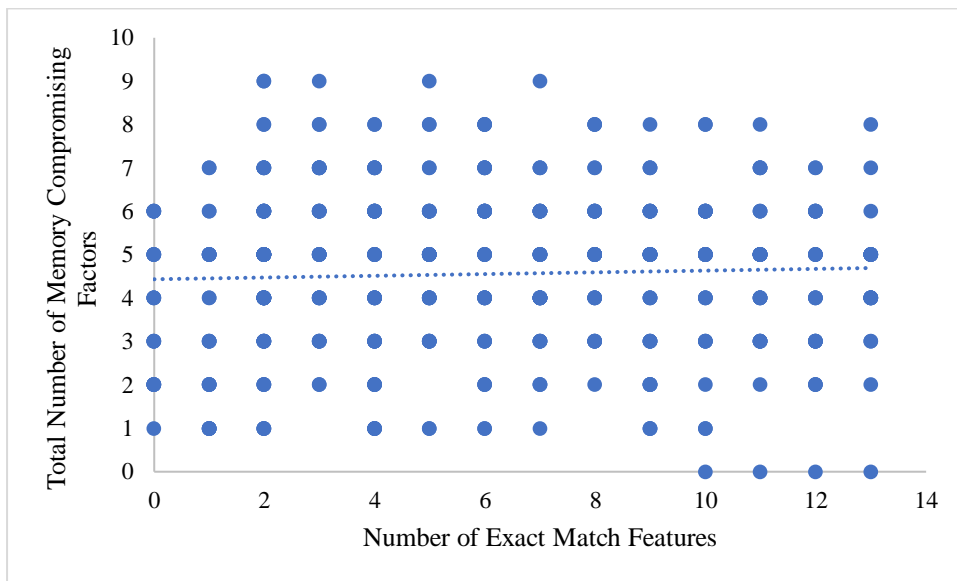
Number of No Match Features

There was a significant small positive correlation between total number of memory compromising factors and number of no match features, $r(508) = 0.10$, $p = .03$. Figure 2 shows a scatter graph of the results. In accordance with hypothesis five and previous research, this suggests that the total number of memory compromising factors may lead to a small increase in 'no match' features.

Figure 3. Scatter graph of the relationship between total number of memory compromising factors and number of no match features

Total Number of Memory Compromising Factors

10
9
8
7
6
5
4
3
2
1
0

0   1   2   3   4   5   6   7   8   9

Number of No Match Features

Proportion of Exact Match Features

The results indicate a very small non-significant negative relationship between the total number of memory compromising factors and the proportion of exact match features, $r(508) = -0.01$, $p = .85$. Figure 3 shows a scatter graph of the results. In contrast with hypothesis six, this suggests that there is no association between the total number of memory compromising factors and proportion of exact match features. This supports the quantity accuracy trade-off which suggests that individuals can strategically regulate their memory reporting to maintain accuracy (Koriat & Goldsmith, 1996).
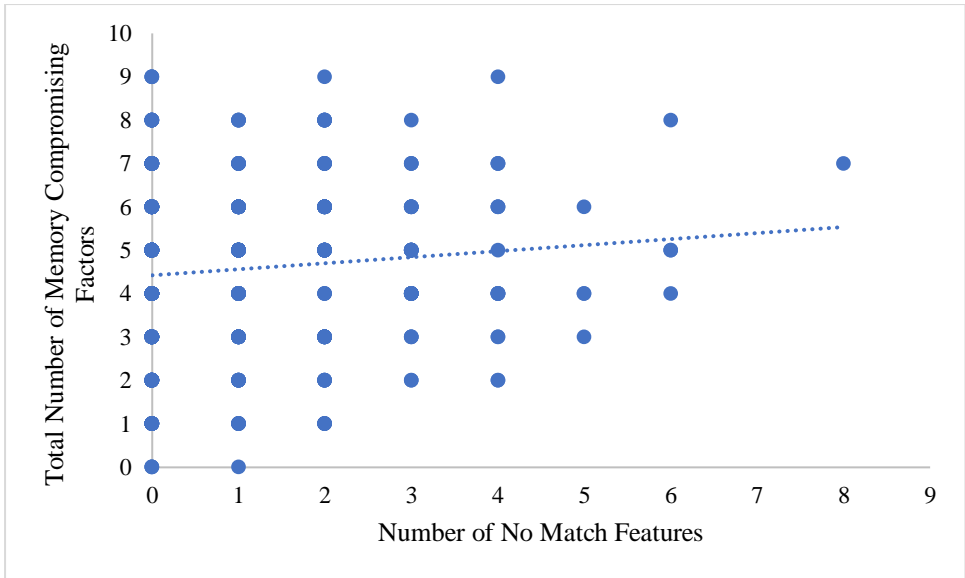
Figure 4. Scatter graph of the relationship between total number of memory compromising factors and proportion of exact match features

Proportion of No Match Features

There was a very small non-significant positive correlation between the total number of memory compromising factors and the proportion of no match features, $r(508) = 0.08$, $p = .07$. Figure 2 shows a scatter graph of the results. In contrast with hypothesis six, this suggests that there may be no association between the total number of memory compromising factors and proportion of no match features. This supports the quantity accuracy trade-off which suggests that individuals can strategically regulate their memory reporting to maintain accuracy (Koriat & Goldsmith, 1996).
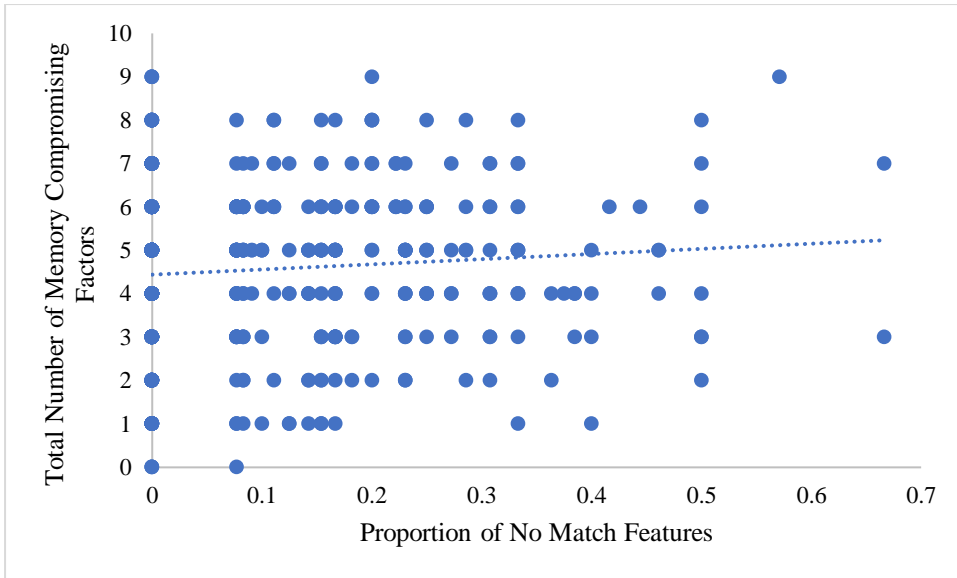
Figure 5. Scatter graph of the relationship between total number of memory compromising factors and proportion of exact no features

Discussion

The aim of this study was to examine how the number and accuracy of details witnesses provide in their description of the perpetrator vary in relation to the presence of factors that are commonly believed to compromise memory (hereafter, memory compromising factors). Specifically, the first aim was to examine whether the presence of memory compromising factors was associated with the witness providing a less detailed (i.e., the number of exact match features given) and accurate (i.e., the proportion of correct versus not correct features) description of the perpetrator. The results showed that the presence of memory compromising factors was associated with the number of exact and no match features. The finding of an association between memory compromising factors and the number of exact match features means that statements may be briefer under these circumstances. However, only a few memory compromising factors, on their own, were significantly associated with the number of details reported; many were not significantly associated with the number of details reported. Furthermore, the presence of memory compromising factors was also associated with the accuracy for exact match features but not no match features meaning that specificity of correct features reported may be affected under these conditions. A second aim of the study was to examine if the total number of memory compromising factors is associated with witnesses providing a less detailed and accurate perpetrator description. The results showed that more memory compromising factors did not necessarily lead to less detailed or accurate memory reports.

In our analyses we first considered the number of details reported by witnesses. Both regression analyses for number of exact match and number of no match features were significant, suggesting that the detail of perpetrator descriptions can be associated with the presence of the nine memory compromising factors. However, the association with memory

compromising factors was clearer on the number of exact match features, compared to the number of no match features. Considering the number of exact match features, there were three individual memory compromising factors that were statistically significant in the model, and all followed the predicted direction; the presence of memory compromising factors was associated with a decrease in exact match features, with 27% of the variance accounted for. Considering the number of no match features, however, there were no individual memory compromising factors that were statistically significant, and the direction of the (non-significant) effects were all different. That is, the presence of some memory compromising factors was associated with an increase, and some was associated with a decrease in no match features. The 'no match' model also only just reached statistical significance and only 5% of the variability in number of no match features was accounted for by the predictors in the model. From these results, it is not clear if the significant difference in the number of no match features in the presence of memory compromising factors was due to an overall decrease in features (since the direction of the effects were different depending on the memory compromising factor variable). This could still be explained by the quantity accuracy trade-off which states that witnesses would only volunteer answers when they were likely to be correct (Palmer et al., 2013), meaning they are less likely to report features of which they are unsure, leading to an overall decrease in features under some circumstances. Future research needs to be conducted to be able to come to more robust conclusions regarding if and how memory compromising factors affect the number of no match features. At the very least, we can conclude that here, there is no evidence that memory compromising factors result in an increase in the number of no match features reported, as might be considered to be true by experts in the field (e.g., Kassin et al., 2001).

Although the regression models were found to be statistically significant, not all the individual memory compromising factors appeared to be important in predicting the number of details reported. In fact, there were only three significant predictors for the number of exact match features and none for the number of no match features. The individual variables that were significant in predicting a decrease in the number of exact match features were: when the witness and perpetrator were of different races (own-race bias), when the eyewitness did not know the suspect, and increased distance between the witness and the crime. This suggests that the presence of these factors in a crime scenario may be associated with a reduction in the number of correct details of the perpetrator given by a witness. Each of those factors will now be considered in turn.

The presence of own-race bias was associated with a decrease in the number of exact match features. This supports the findings from a meta-analysis conducted on laboratory studies on own-race bias (Meissner & Brigham, 2001), and illustrates that the own-race bias can also apply to real-life settings. This is supported by another field study in which individuals were approached by a confederate who was either the same or a different race to their own and later asked to identify them in a sequential line-up (Wright et al., 2001). The study found that individuals were better able to recognize people of their own race, supporting the current study's finding that eyewitnesses are better able to describe perpetrators that are the same race to their own. This also suggests that the own-race effect is applicable to both identification and verbal procedures in the criminal justice system, contrasting Ellis et al. (1975) who did not find a clear own-race bias in verbal descriptions. From these findings, own-race bias should therefore be considered in the Turnbull Guidelines used in the UK (R v Turnbull and others, 1976) to tell juries that this factor should be considered on whether this may have impacted the testimony completeness.

Another variable that was significant in predicting the number of exact match features of perpetrator descriptions was if the eyewitness knew the suspect. The findings indicated that if the eyewitness did not know the perpetrator before the crime, this resulted in less exact match features reported. This is consistent with the already robust finding in the literature that facial recognition ability is poorer for unfamiliar faces compared to familiar faces (Hancock et al., 2000; Megreya & Burton, 2006). Furthermore, as most previous research has been conducted with identification procedures, this shows that this effect is also applicable to verbal descriptions. This variable has also not previously been considered in the context of eyewitness memory for crimes. However, this factor is mentioned in the Turnbull Guidelines (R v Turnbull and others, 1976), in which a judge should instruct the jury to consider whether the witness knew the suspect. The current results also suggest that the eyewitness knowing the suspect may have a significant implication on eyewitnesses' detail in their statements of the perpetrator.

Distance was the final variable that significantly predicted the number of exact match features in perpetrator descriptions, with eyewitnesses who were further from the crime giving less exact match features than those who were closer. This supports both laboratory (Lampinen et al., 2014) and field (van Koppen & Lochun, 1997) study findings. When a face is moved away from an observer, the details of the face become progressively coarser (Loftus & Harley, 2005) and therefore are more difficult to identify and recall. This appears to be a robust finding and is also reflected in the Turnbull Guidelines (R v Turnbull and others, 1976), where judges are instructed to direct the jury to examine the circumstances in which the identification by the witness came to be made, including at what distance. However, it is also important to add that the Turnbull Guidelines (R v Turnbull and others, 1976) are

utilised for visual recognition evidence (i.e., when identification is in dispute) and hence might not be read when witness descriptions are being interpreted in court.

It is interesting that the findings of this study may not be consistent with what psychology experts, or the legal field believe to be true about the effect of suboptimal encoding conditions on eyewitness memory. For example, Kassin et al. (2001) mention that lighting and crime length may affect memory reporting and these factors are also mentioned in the Turnbull Guidance (R v Turnbull and others, 1976), in which juries are instructed to consider in what light and how long the witness observed the suspect. However, according to the findings in the current study, these factors do not have large effects on recall. Only one of the factors that was associated with memory reporting, own-race bias, was asked by Kassin et al. (2001) in their study. The other variables explored in their study that experts believed affect memory for crimes did not significantly affect memory completeness in the current study, and, descriptively speaking, some were not even in the direction expected by experts. These include the witness being a young or old age, the witness being under the influence of alcohol or drugs, poor lighting, shorter crime length, the witness experiencing violence and the crime involving a weapon (weapon focus). Therefore, longstanding held beliefs by experts are not supported by the current data. For example, whilst alcohol intoxication has been shown to impair episodic memory (Mintzer, 2007), the current results support more recent research that suggests that those under the influence of alcohol are no more likely to recall incorrect information compared to those that were sober (Flowe et al., 2019) and suggests that drug intoxication works in a similar way. Furthermore, both the lack of association with poor lighting and the presence of a weapon supports field studies into both factors whereby their presence did not influence memory (Valentine et al., 2003; Wagstaff et al., 2003); these results are not found in laboratory studies. In laboratory studies, both poor lighting and

weapon presence were found to negatively affect memory reporting (Wells & Olson, 2003; and Ross et al., 1994 respectively). However, it is also important to add that the factors that were not significantly associated with changes in memory reporting may have affected memory in some individual crime cases. However, as the current study is based on group averages and not applied to individual witnesses, this was not captured.

In our analyses we also considered the accuracy of details reported by the witnesses. In the regressions that considered the accuracy of statements, the regression for the proportion of exact match features was statistically significant whilst the regression for the proportion of no match features was not significant. The quantity accuracy trade-off framework (Koriat & Goldsmith, 1996) would predict no association in either of these regressions if individuals strategically regulate their memory reporting as the overall accuracy of statements would not be affected. However, the current study's results suggest that the presence of memory compromising factors was associated with a difference in the accuracy of exact but not no match features. Though we cannot be sure of direction of effects, one possibility that explains this pattern of results, is that the correct features could be less specific under 'memory compromising' conditions. Weber and Brewer (2008) investigated the levels of detail and specificity of information given in responses. Consistent with the quantity accuracy trade-off framework, they found that participants would only volunteer precise answers when they were likely to be correct and therefore level of detail was positively related to confidence in the response. The current findings are therefore consistent with Weber and Brewer (2008) and suggest that in circumstances in which memory compromising factors are present, this does not mean that witnesses will provide *incorrect* features but that it may affect their ability to be *exactly correct.* Wogalter (1996) also found that witnesses generally do not give many specific descriptors when describing faces and hence this can explain why there are likely to

be times where witnesses provide features that 'match' but do not 'exactly match'. The current study only focused on 'exact match' features due to wanting to investigate factors that relate to the most forensically relevant information for the police to be able to apprehend a perpetrator. However, it may be that, when memory compromising factors are present, witnesses are more likely to give 'match' features, which still provide the police with applicable evidence that they can use. To further investigate this possibility, future research could code features as 'match' when the features recalled by the eyewitness were similar to the perpetrator characteristics but not an exact match.

However, it is important to again note there was only one potential memory weakening factor—the age of the witness—that was significantly associated with the accuracy of exact match features. Interestingly, the direction of the effect was *opposite* to what would be expected if memory compromising factors impair recall accuracy; in the current analysis when the witness older (over 45 years old) there was an increase in accuracy of exact match features, compared to younger (under 16 years old) or of adult age who performed similarly. This does not support both experts' beliefs (Kassin et al., 2001) and laboratory and field studies that suggest that child (Brigham et al., 1986; Goodman et al., 2001) and older adults (List, 1986; Yarmey, 1993) are less accurate at recall than young or middle-aged adult witnesses. The current study's finding is in line with the findings that both older adults (Colloff et al., 2017) and children (Winsor et al., 2021) can know circumstances in which their memory is inaccurate (i.e., they monitor memory) and can withhold answers of which they are unsure compared to young adults. This again is consistent with the quantity accuracy trade-off theory (Koriat & Goldsmith, 1996). An explanation to why younger witnesses had similar accuracy to young adults is consistent with studies that have shown that children can be as good as adults in recall when developmentally appropriate procedures are used (Ingham

et al., 2019). The results showed that older adults gave more proportion exact match features compared to both adults and children. One theory is that we learn through error feedback training when our memories are likely to be accurate or not (Mickes et al., 2011). Older adults would have received much more error feedback training and hence be more aware when their memory may be inaccurate, thereby appropriately monitoring their memory and recalling only features on which they are most confident as predicted by the quantity accuracy trade-off framework (Koriat & Goldsmith, 1996).

We also considered how the number of memory compromising factors influenced witness's descriptions of a perpetrator. Three out of four correlation analyses were not significant, suggesting that an increase in the total number of memory compromising factors was not associated with a less detailed or accurate statements. However, a significant, small positive relationship was found between total number of memory compromising factors and number of no match features suggesting that the presence of a greater number of memory compromising factors can slightly increase the number of no match features reported. However, as the regression analyses taken together did not strongly support this view, it may be that only when there are multiple memory compromising factors present during a crime, this is when memory starts to become (slightly) less accurate. No previous study has examined the effect of total number of memory compromising factors and hence this needs to be replicated before robust conclusions can be made. Furthermore, laboratory studies are controlled and so are usually focus on one variable (Lane, 2006; Chae, 2010). They do not typically consider the combination of factors that co-occur and may interact in real life, which is also an important factor to take into consideration, based on the results here.

The findings of this study have allowed for more understanding on factors that may impact eyewitness memory retrieval. This therefore points to the need for education to criminal justice practitioners to prevent beliefs that certain testimonies may be less 'accurate' under certain circumstances. This is especially relevant for the Turnbull Guidance (R v Turnbull and others, 1976), which, as mentioned, states judges are instructed to direct the jury to examine the circumstances in which the identification by the witness came to be made, including at what distance and lighting. However, these factors were not found to affect memory reporting in the current findings. This is also important for police who may unconsciously (or consciously) affect the interviewing process. For example, Crossland et al. (2018) found that police in the UK would only take a 'holding statement', which is a less formal, non-evidential account involving only basic event details from witnesses who were intoxicated with the view to complete a sober interview later. It is only the sober account that is formally recorded and presented in court (Crossland et al., 2018). However, from the current findings it shows that alcohol intoxication does not have a large effect on memory reporting and hence the initial statement may be just as accurate. This is in line with research which states that alcohol intoxication does not affect accuracy of details recalled (Flowe et al., 2019). Furthermore, this is important to communicate to criminal justice professionals and especially to juries. Juror perception research has shown that many of the factors believed by experts are also believed by jurors to affect memory. For example, there are beliefs in jurors that alcohol and drugs have large negative effects on memory (Monds et al., 2022) and that child witnesses are less credible than adults (Pozzulo & Dempsey, 2009), but both factors were not found to effect memory in the current findings. Therefore, the current study's findings need to be communicated to prevent beliefs affecting the view of the witness's statement which may ultimately affect the outcome of the case.

Furthermore, the study results are largely in support of the quantity accuracy trade-off framework which states that individuals strategically regulate their memory reporting and hence they would volunteer to the police only memories about which they are confident (Koriat & Goldsmith, 1996). Greater theoretical understanding about memory reporting is important as it can help to guide procedures when dealing with witnesses of crimes. For example, the current procedure of interviewing in the UK- the ABE procedure- utilises a free-recall phase and a questioning phase. If the quantity accuracy trade-off framework is correct, questioning witnesses makes witnesses lower their criterion for reporting, and therefore a long and detailed questioning phase could increase the number of inaccurate details reported and reduce the overall accuracy of witness's memory evidence (Ingham et al., 2019). It could therefore be recommended to police and courtroom procedures to utilise a free recall phase and reduce questioning phases unless necessary to maximise the most accurate information from the witness. Such a procedure could serve to reduce the adverse effects that are associated with poor testimony on both witnesses and wrongfully accused.

There are some limitations to the study. First, it is possible that the suspect may have changed their appearance from the time of the crime to when they were apprehended which would mean that there would be discrepancies between the witness statement and the police checklist (i.e., coded as 'no match' features), but this was not the result of inaccurate witness memory. Pike et al. (2002) found that out of 702 cancelled identification parades, 1% were cancelled because the suspect had changed their appearance. These represent the appearance changes that are known about and hence likely represent an underestimation of the reality. However, it is important to keep in mind that the current results do not take any appearance changes into account and so there may be some circumstances that a description is coded as an inaccurate memory when this was not the case.

Furthermore, due to the secondary nature of the data, it is unknown how the witness interviews were conducted. As previously outlined, the way interviews are conducted have the potential to alter memory information given. Field studies have suggested that investigators make common mistakes when interviewing a witness. These include, for example, interrupting the interviewee, which has the effect of stopping the interviewee before they have the chance to add potentially valuable information (Milne & Bull, 1999) and asking closed questions which also inhibit the amount of information available (Milne & Bull, 1999). Investigators have also been shown to ask leading questions, whether intentional or not, which may interfere with a witness's ability to correctly recall events (e.g., Fisher, 1995). Without knowledge on how the interviews were conducted, it is unknown whether any inaccuracies in memory are due to the memory compromising factors present in the case or due to inadequate criminal justice procedures (Wixted et al., 2018).

There are also constraints of generalisability of the current results. For example, the study utilises one sample that was taken from a one police jurisdiction in the USA. This may lack generalisability to UK practice due to the differences between US and UK policing and culture. For example, in the USA, the availability of guns is much higher, where it is estimated that there are 120.5 legal and illegal firearms per 100 people, compared to the UK, where it is estimated that there are 4.6 legal and illegal firearms per 100 people (United States Congressional Research Service, 2017). This would affect the types of crimes that occur in the USA and hence are present in the current data set. Furthermore, the amount of training and education varies across forces in the USA, whereas considerable training and educational requirements are in practice in the UK (Evans, 2007). Therefore, the current checklist used to gather the current dataset may have been completed differently if the study

was based in the UK. Together, this means it may be difficult to generalise the current results to those in the UK.

In conclusion, the current research has largely supported the quantity accuracy trade-off framework, which states that individuals strategically regulate their memory reporting and only volunteer information in which they are confident. The longstanding beliefs of experts and the legal system that certain factors affect the accuracy of recall memory for crime has not been supported in the current study. This suggests that education for criminal justice practitioners is required. Furthermore, the current police interviewing procedure needs to reflect the quantity accuracy trade-off mechanism through utilising an extensive free recall phase and reducing questioning to maximise the accuracy of information from witnesses.

Chapter Four

A Psychometric Critique of Criteria-Based Content Analysis (CBCA)

Abstract

Whilst inaccurate memory reporting might be due to factors outside of the witness's control (e.g., crime factors such as lighting or distance), it might also be that the witness is deliberately reporting inaccurately. Statement Validity Analysis (SVA) is a procedure that was designed to determine credibility of statements. A component of this process is Criteria-Based Content Analysis (CBCA) which analyses the content of statements to determine if they are an attempt at memory recall or fabricated. CBCA is used as evidence in court in some countries (e.g., The Netherlands, Germany, and Sweden) and hence it has impactful real-world consequences. This critique evaluates CBCA criteria by reviewing evidence into its reliability and validity and discussing future research directions. Overall, CBCA criteria has demonstrated good reliability as well as face and construct validity. However, other techniques may be more appropriate for use on adult populations. Furthermore, there are concerns about the use of this tool on non-Western cultures and statements of certain crimes. The implications for this in practice are discussed.

Introduction

Much experimental psychology research has been concerned with the capacity of eyewitnesses to remember a crime. Early laboratory studies into this concluded that "…completely accurate recollection is not the rule but the exception" (Stern, 1902, p. 327). This has later been disputed with multiple field studies concluding that eyewitness testimony can, in fact, be accurate despite memory weakening factors being present (see Chapter Two for a systematic review of this literature and Chapter Three for an empirical study into this topic). Most of this literature, however, assumes that witnesses are intending to provide honest accounts.

In allegations of sexual abuse there is often no physical evidence and no witnesses to the crime meaning that a victim account can be the sole evidence for the case (Vrij, 2015). Furthermore, victims can give contradictory statements and children, in particular, are seen as more suggestible and less accurate in their testimonies (Kassin et al., 2001). With the importance placed on testimonies in these cases in investigations and court proceedings, an area of interest has been developing measures to determine whether statements can be proved to be credible (i.e., a true reflection of the event proposed)? Credibility is comprised of two components: a cognitive side and a motivational side (Undeutsch, 1989). The cognitive side refers to the eyewitness' ability to report the details of an observed event correctly, whereas the motivational side refers to the willingness of the eyewitness to tell the truth (Undeutsch, 1989). Before the development of SVA, no measure had captured both components.

Statement Validity Assessment (SVA) (Undeutsch, 1984; 1989) was developed in Germany to determine the creditability of child witness testimonies in sexual abuse trials. However, the assessment has also been used to determine the credibility of adults' statements, offenders'

statements, and crimes other than sexual offences (Yuille & Cutshall, 1989; Porter & Yuille, 1994). SVA is based on the notion that truthful accounts significantly differ from fabricated accounts, known as the Undeutsch hypothesis (Steller, 1989). SVA provides an assessment of the credibility of the statement, not of the individual. The assessment consists of three main stages: a) a structured interview; b) a criteria-based content analysis (CBCA), which assesses the content quality of the statement; and c) the Validity Checklist, which evaluates the outcome of the CBCA in relation to other evidence and interview factors (Raskin & Esplin, 1991). Each of these stages will now be discussed.

Semi-Structured Interview

This stage of SVA involves the witness providing their account of the event. This should initially involve the interviewer eliciting a free narrative from the witness in which the witness should not be interrupted in their testimony. Open and closed-ended questions then follow the free narrative phase to elicit additional information.

Criteria-Based Content Analysis (CBCA)

The purpose of the CBCA phase is to determine if the contents of the statement are indicative of an attempt to recall an actual memory or the result of fabrication. The interviews are transcribed, and trained evaluators judge for the presence or absence of 19 criteria (see Table 1; see further Raskin & Esplin, 1991). The CBCA assessment assumes that a statement derived from memory of an actual experience will differ in content and quality from a statement that is invented (the Undeutsch hypothesis; Steller, 1989). Therefore, a truthful statement will contain more of the CBCA criteria than false statements.

Table 8. CBCA Criteria (Vrij, 2015)

General characteristics

1. Logical structure	The statement is coherent and does not contain inconsistencies or contradictions
2. Unstructured production	The presentation of information in the statement is in a chronological order
3. Quantity of details	The statement includes descriptions of place, time, persons, objects, and events

Specific contents

4. Contextual embedding	The events described in the statement are placed in time and location. The actions described in the statement relate to other daily activities
5. Descriptions of interactions	The statement includes information that interlinks the alleged perpetrator and witness
6. Reproduction of conversation	The statement includes conversations that are reported in their original form
7. Unexpected complications during the incident	The statement includes elements that are unexpected
8. Unusual details	The statement includes details of people, objects or events that are unusual but meaningful in the context

9. Superfluous details	The statement contains details that are not essential for the accusation
10. Accurately reported details misunderstood	The statement mentions details that are beyond the interviewee's comprehension
11. Related external associations	The statement contains reported events that are not part of the alleged offence
12. Accounts of subjective mental state	The statement includes accounts of feelings and thoughts experienced at the time of the incident.
13. Attribution of perpetrator's mental state	The statement includes descriptions of the alleged perpetrator's feelings, thoughts, or motives during the incident
Motivation-related contents	
14. Spontaneous corrections	The statement contains corrections that are made to material previously provided without having been prompted by the interviewer
15. Admitting lack of memory	The statement contains an admission of a lack of memory
16. Raising doubts about one's own testimony	The statement indicates doubts about his or her description
17. Self-deprecation	The statement includes personally unfavourable or self-incriminating details

18. Pardoning the perpetrator	The statement fails to blame the perpetrator or excuses his or her behaviour

Offence-specific elements	
19. Details characteristic of the offence	The statement contains descriptions of elements of the crime that are known by professionals to be typical for the type of crime but that are not publicly known

Validity Checklist

The validity checklist draws conclusions from the CBCA by considering other factors that may have influenced the statement. There are different versions of the Validity Checklist available. However, a commonly used version was published by Raskin et al. (1992). This Validity Checklist comprises of 13 issues that may affect CBCA scores. These items are considered and rated as affirmative or negative, with more affirmative responses suggesting that the CBCA scores may be influenced by other factors and hence other interpretations (than fabrication) need to be considered.

Table 9. Validity Checklist (Raskin et al., 1992)

Psychological characteristics	
1. Cognitive-emotional limitations	There are indications of limited cognitive abilities, unwillingness to discuss events or discomfort during the interview that interfered with obtaining adequate information

2. Language and knowledge	The use of language and display of knowledge was beyond the normal capacity for an individual of that age and experience
3. Affect during the interview	The individual displayed inappropriate affect or there was an absence of affect that would be expected
4. Suggestibility	The individual demonstrated susceptibility to suggestion or asked questions during the interview to attempt to obtain clues on what to say
Interview characteristics	
5. Interview procedures	The interview was inadequate: the interviewer introduced distractions, failed to established rapport, inadequately attempted to elicit a free narrative, failed to use appropriate follow up questions, or failed to attempt to resolve any inconsistencies
6. Influence on statement contents	There was leading/ suggestive questioning, pressure or coercion employed in the interview
Motivational factors	

7. Motives for reporting	The individual's relationship or other variables suggest possible motives to make a false allegation
8. Context of Disclosure	There are questionable elements in the content of the original disclosure
9. Influence by others	There are indications that others suggested, coached, pressured, or coerced the individual to make a false report
Investigative Questions	
10. Lack of realism	The events described are unrealistic and are contrary to the laws of nature
11. Inconsistent statements	There are elements in the statement that are inconsistent by another statement made by the individual
12. Contradictory evidence	There are elements in the statements that are contradicted by physical evidence
13. Characteristics of the offence	The description of the alleged offence is lacking in the normal details and general characteristics of this type of offence

Much of the research into the reliability and validity of SVA has focused on the CBCA stage and has been conducted in both laboratory and field settings. There has been some research that has focused on the Validity Checklist (for example, Lamers-Winkelman, 2021),

however, the focus of this review will be on the CBCA criteria considering reliability, validity, and other issues relating to the measure.

A critique of CBCA criteria is necessary as the technique is not validated and no previous review has considered all forms of reliability and validity. Furthermore, CBCA is used as evidence in courtrooms in several countries (e.g., The Netherlands, Germany, and Sweden) and can therefore influence case outcomes. With the potential for miscarriages of justice, it is important that the measure is well-constructed and accurate.

Reliability

"Reliability concerns the extent to which measurements are consistent and repeatable" (Zeller & Carmines, 1980, p.17). There are four ways to assess reliability (Institute of Medicine et al., 2015): inter-rater reliability (examines the consistency across different independent judges), test-retest (examines the consistency of the same test across different points in time), internal (examines the consistency of different items within the same test), and parallel or alternate forms (examines the consistency of different forms of the same test).

Different statistics can be used to measure reliability, including proportion agreement rates, correlations, Cohen's kappas, Maxwell's random error coefficient of agreement (RE) or Finn co-efficients. However, regardless of the method used a score of .50 or higher is typically considered adequate reliability (Fleiss, 2014). Scores between .60 and .75 are considered good and scores over .75 are considered as excellent (Fleiss, 2014).

Inter-rater Reliability

Inter-rater reliability refers to the degree of agreement among independent observers when rating the same phenomenon (Saal et al., 1980). When considering the CBCA criteria, the

inter-rater reliability can be obtained by independent raters analysing the same transcript using the 19 CBCA criteria. The scores given for each criterion can then be compared across raters. Vrij (2005) conducted a qualitative review of the first 37 studies into SVA published from 1988 to 2005, including both laboratory and field studies, and child and adult participants. 16 of the 37 studies reviewed commented on the inter-rater reliability scores of the individual CBCA criteria as measured using various indices. For 16 of the 19 criteria, good inter-rater scores (.60 or higher) were obtained (except for criteria two and four; see Table 1) and many agreement rates were above .75. Three studies also commented on the inter-rater reliability of total CBCA score (Horowitz et al., 1997; Höfer et al., 1996; Vrij et al., 2004), all of which fell in the excellent range. These findings suggest that total CBCA scores are more reliable for assessing statements than scores for the individual criteria.

A more recent meta-analysis (Hauch et al., 2017) also investigated the inter-rater reliability of CBCA criteria. This included 74 studies published between 1982 and 2014. The meta-analysis revealed acceptable (above .50) inter-rater reliabilities for 17 of 19 CBCA criteria, as measured with various indices. Criteria two and nine had low reliabilities regardless of the coefficient used. However, results were largely heterogeneous across studies which the authors concluded was due to the variety in how the measure is utilized, as CBCA criteria can be coded in different ways, including dichotomous ratings (present versus not present) and Likert scales (e.g., 1-5). Furthermore, how much training each rater received also differs across studies in both the laboratory and the field. There is no standardised manual for SVA/CBCA use and hence differences in the way the CBCA are judged could account for the diverse findings across studies.

Considering the results of both reviews, most criterion in the CBCA can reach good levels of inter-rater reliability, with most reaching excellent scores. The measure is most reliable when using total CBCA scores compared to individual criteria. However, it is important that use of CBCA criteria is standardised, to be able to better compare across studies and hence come to overall conclusions that can be useful for field use.

Test-retest Reliability

There is only one known study on the test-retest reliability of CBCA criteria. Test-retest reliability is not usually investigated because truth-telling must consider random event characteristics, whereas lies are constructed by the individual and hence truth-telling would be less consistent across time (Shemmel et al., 2019). However, Schemmel et al. (2019) found that correlations between true, experience-based statements' CBCA scores was .44 compared to .61 for fabricated statements' scores, although the difference was not statistically significant. The fact that the correlations between true statements did not reach 'acceptable' level, it can be explained by the fact that truth-telling is less consistent across time (Shemmel et al., 2019) rather than failures in the CBCA criteria. However, the direction shows that fabricated accounts are more consistent across time compared to truthful ones which is what would be expected. The CBCA criteria are commonly used on statements at a single point in time and hence test-retest reliability is not usually appliable; however, the findings from this study show that it is able to reliably tell if someone is fabricating their statement to an acceptable standard which is a goal of the measure.

Internal Reliability

There is a lack of research into internal reliability of CBCA. Internal reliability describes the consistency of results across items within a measure (McLeod, 2007). It is therefore it is currently unknown whether the CBCA criteria are all measuring the same construct.

Alternate Forms Reliability

There has been no research into alternate forms reliability of CBCA as there is only one form of the assessment. However, there are different ways of coding the CBCA criteria, including dichotomous ratings (present versus not present) and Likert scales (e.g., 1-5). As stated, Hauch et al. (2017) found that results were largely heterogeneous across studies in their meta-analysis due to the variety in how the measure is utilized suggesting that using different rating scales could affect overall scores of the same statement. This is problematic, as when CBCA is used as evidence, it may influence case outcomes. Future research could consider examining alternate forms reliability by examining the different ways of coding the criterion to see how the outcome of the test changes when different methods are used.

Validity

Validity is also important as a reliable measurement is not always valid. A valid test ensures that results are an accurate reflection of what is being measured (Cizek, 2012). Therefore, whilst results of CBCA may be consistent (i.e., are reliable), that does necessarily mean they are correct (i.e., are valid). Validity can be demonstrated by a clear relationship between the test and what it is meant to measure. There are four types of validity: content validity, face validity, construct validity, and criterion-related validity.

Content validity

Content validity measures whether a test represents the entire range of possible items the test should cover (Institute of Medicine et al., 2015). This can be determined by looking at empirical evidence that underpins the Undeutsch hypothesis. When the hypothesis was developed, there was limited theory to underpin it (Tully, 1999) which is concerning as the CBCA criteria has been developed based on the hypothesis. However, there has since been studies that support the Undeutsch hypothesis. For example, research has found that truthful statements were longer in length compared to fabricated statements (Porter & Yuille, 1994) which lends support to the quantity of details criterion. Furthermore, there have been formal applications of the Undeutsch hypothesis to actual crime events. For example, Yuille and Cutshall (1989) found that a truthful account of witnessing a murder included logical consistency, a relationship of the account to the general pattern of the murders, a spontaneous nature of descriptions, and a wealth of details, supporting the Undeutsch hypothesis and hence the validity of criteria such as logical consistency, contextual embedding, spontaneous corrections, and quantity of details. However, a limitation is that these studies do not assess how valid the hypothesis is by using measures of validity (e.g., face or construct validity) nor do they examine whether there are other facets to differentiate a truthful and fabricated statement that are not currently considered in the Undeutsch hypothesis, or in the CBCA criteria.

Face Validity

Face validity considers whether a test *appears* to measure its intended variables (Johnson, 2012). This can be investigated by considering whether CBCA is able to correctly classify truths and lies better than chance (50%). For example, in a laboratory study, Vrij (2008) found that 71% of truths and 71% of lies were correctly classified by CBCA assessments. A similar finding also shown in the field; Akehurst et al. (2011) asked two experts to apply the

CBCA criteria: one classified 81% and the other classified 60% of the fabricated cases correctly. These findings suggest that CBCA criteria can correctly differentiate between fabricated and truthful statements better than chance level. However, questions remain about whether this is acceptable enough for a court of law.

Construct Validity

Construct validity is when test scores have an association with the prediction of an outcome (Ginty, 2012). Vrij's (2005) review investigated whether the CBCA could differentiate between truth tellers and liars in both children and adults. The Undeutsch hypothesis states that a statement derived from memory of an actual experience will differ in content and quality from a statement that is invented and therefore a truthful statement will contain more of each CBCA criteria. Therefore, to test CBCA validity, studies have compared the CBCA scores of both truthful and fabricated statements. In the field studies reviewed by Vrij (2005) ($N =5$), confirmed cases were compared with doubtful cases (e.g., cases that were dismissed by a judge). In all studies, the differences between statements in confirmed and non-confirmed studies were in the direction predicted by the Undeutsch hypothesis. For example, in Esplin et al.'s (1988) study, confirmed cases received a mean CBCA score of 24.8, compared to a mean score of 3.6 in the doubtful cases. However, the criteria for establishing that a statement was doubtful was vague (e.g., judicial dismissal, no confessions made by the accused, persistent denial by the accused) and hence the results may not be generalized to real life. This limitation can be overcome through using laboratory studies in which truthful and fabricated statements can be controlled. Support for the Undeutsch hypothesis was also shown in the laboratory studies reviewed by Vrij (2005) ($N =18$). Almost all differences between truthful and fabricated statements were in the expected direction, with the CBCA criteria occurring more frequently in truthful compared to fabricated statements. However,

there were fewer differences between truth tellers and liars than shown in the field studies. Therefore, the Undeutsch hypothesis was supported as there was significant differences in the CBCA scores of truthful compared to fabricated statements.

A more recent meta-analysis (Amado et al., 2015) further investigated whether the CBCA criteria could discriminate between memory of real and of false accounts in children. Results reported a positive effect size that is considered large (0.79) (Lakens, 2013) meaning that there is a strong relationship between higher CBCA values and truthful statements. Furthermore, a significant positive effect size was also observed in all the criterion when they were considered separately. Therefore, the results corroborated the previous review's results and hence established the validity of the Undeutsch hypothesis and the CBCA criteria to be able to determine the credibility of a statement.

Another meta-analysis considered only adult literature (Amado et al., 2016) and found the CBCA to be a valid technique. However, they found that criterion 17 and 18 (self-deprecation and pardoning the perpetrator, respectively) failed to discriminate between real and fabricated statements. This may be related to adults having more developed morality compared to children (Piaget, 2015) and hence they know that the perpetrator was wrong, which younger children may not understand yet. They also concluded that their results were not generalisable to future samples and that the CBCA was better able to discriminate between statements of certain crimes, including sexual offences and intimate partner violence cases, compared to other self-experienced or witnessed events. In other events, there was a higher likelihood of false positives. This may mean that caution must be taken when using the CBCA on adult statements and for some crime types.

Concurrent Validity

Concurrent validity is a type of criterion-related validity. It is the ability for a test to estimate a current phenomenon (Lin & Yao, 2014). This can be assessed by comparing the measure to other ways of measuring the same phenomenon or by observing other factors associated with the phenomenon being measured.

Reality Monitoring (RM). RM is the process by which a person attributes a memory to an actual experience or their imagination (Johnson et al., 1988). Similar to CBCA, RM is based on the assumption that memories based on real experiences differ in quality from memories based in imagination (Johnson & Raye, 1981). There is no standardised set of RM criteria. However, the criteria in Table 3 are followed in most assessments.

Table 10. RM Criteria (Vrij, 2008)

Sensory (perceptual) information	The statement contains details about what the interviewee saw, heard, smelled, touched, or tasted
Spatial detail	The statement contains information about locations to the spatial arrangement of people and/or objects
Temporal detail	The statement contains information about when the event happened or describes a sequence of events
Affect	The statement contains information about how the interviewee felt during the event

Reconstructability of the story	The event can be reconstructed based on the information that is given in the statement
Realism	The story given in the statement is plausible, realistic and makes sense
Cognitive operations	The statement does not contain descriptions of inferences as this suggests deceit

The average total accuracy rate for RM (68.13%) was slightly higher than for CBCA (63.63%) across eight laboratory studies that investigated both RM and CBCA's ability to discriminate between truth tellers from liars (Vrij, 2015). However, Oberlader et al. (2016) compared CBCA and RM in a meta-analysis of 56 laboratory studies and found that there was no significant difference in the effectiveness of CBCA and RM. Nevertheless, RM cannot be used with young children. This is because children do not differentiate between fact and fantasy as well as adults do and, due to a richer imagination, they are better able to imagine and describe scenes with the RM criteria present (Lindsay, 2002). It is also difficult to use RM when assessing statements for events that happened a long time ago. Over time the cognitive operations criteria may appear in statements as this facilitates the remembering of events (Roediger, 1996). Therefore, whilst they are similar in accuracy, CBCA better takes into consideration these factors.

Scientific Content Analysis (SCAN). Another technique is SCAN. Similar to CBCA and RM, written statements are analysed based on a list of criteria and truthful statements are thought to contain more of these criteria compared to deceptive statements (Sapir, 1987, 2000).

Despite this technique being used worldwide, in federal law enforcement (including the

Federal Bureau of Investigation (FBI)), military agencies and secret services (including the

Central Intelligence Agency (CIA)), there is no theoretical justification behind the criteria in

SCAN. The most used SCAN criteria are presented in Table 4.

Table 11. SCAN criteria (Vrij, 2008)

1. Denial of allegations	The statement contains denial of the allegations. Truthful interviewees are more likely to do this.
2. Social introduction	The statement includes unambiguous introductions of persons (e.g., stating 'my friend by the name of X' rather than just 'we'). Ambiguous introductions indicate deceit.
3. Spontaneous corrections	The statement does not contain corrections, such as crossing out what has been written, as this indicates deceit
4. Lack of conviction and memory	The statement is not vague (e.g., states 'I can't remember' a lot), as lack of memory indicates deceit
5. Structure of statement	The statement's structure is balanced. It is thought that in a truthful statement the first 20% us used to describe activities leading up to the event, the next 50% to describe the actual event

and the fine 30% to discuss what happened after the event.

6. Emotions	The statement contains descriptions of emotions at particular points. Truth tellers are thought to state emotions *throughout* the story of the event and *after* the climax of the event, whereas liars are thought to only give emotions just *before* the climax of the story
7. Objective and subjective time	Objective and subjective time will correspond with each other in the statement. (Objective time is the actual duration of events described in the statement, whereas subjective time is the number of words spent to describe these events)
8. Out-of-sequence and extraneous information	The statement recounts events in chronological order and does not contain information that does not seem relevant (extraneous information). Extraneous information indicates deceit as they are included to hide more important information
9. Missing information	The statement does not include words that indicate some information has been left out,

	such as 'sometime after' or 'later on' as this indicates deceit
10. First person singular tense	The statement is written in the first person singular, past tense. This describes an event that has taken place
11. Pronouns	The statement uses pronouns (e.g., 'I', 'my', 'he', 'his', 'they', 'their', etc) as this signals commitment, responsibility, and possession. Omitting pronouns suggests reluctance of the interviewee to commit themselves to the described action and thus indicates deceit
12. Change in language	The statement includes a change of terminology or vocabulary as this indicates that something has altered in the mind of the interviewee

Some of the SCAN criteria contradict the CBCA criteria. For example, SCAN criterion three and CBCA criterion 14 are both concerned with spontaneous corrections. However, CBCA states that spontaneous corrections indicate truthfulness whereas SCAN interprets it as a sign of deceit. This may be because SCAN is more often used to interview suspects of crime to determine their deceit (much like a polygraph, which is not used in the UK but is in other countries), whereas RM and CBCA are more often used on witnesses to determine their truthfulness. Regardless of this, studies have shown that SCAN is able to correctly classify between 71% and 95% of truth tellers and liars (Driscoll, 1994; Smith, 2001; Nahari et al., 2012). However, the method has been criticised as detectives untrained in SCAN were also

able to classify just as well (Smith, 2001) and some of the criteria did not differ in quantity in truthful and deceptive statements (Porter & Yuille, 1996). Furthermore, there have been no direct comparisons between SCAN and CBCA; although one study into the validity of SCAN decided to use CBCA criteria on a subset of their data and found that it was able to discriminate between truthful and fabricated statements, whilst SCAN was not (Bogaard et al., 2016). There is also a lack of research into the validity of SCAN (Komel et al., 2020). Future research is needed to explore SCAN's superiority over CBCA, especially because SCAN is already widely used in criminal justice settings across the world.

Multivariable Adults' Statements Assessment Model (MASAM). A final technique used is the MASAM. This is a more recent tool designed to judge the credibility of adult witnesses' statements (Wojciechowski, 2015). The hypothesis of MASAM is the same as CBCA (i.e., that a statement that is derived from memory of an actual experience will differ in content and quality from a statement that is invented); however, there are additional assumptions in MASAM. Firstly, it considers the fact that in each statement, true and false information can be found. Furthermore, if a witness's intention is to give an untruthful statement, certain differences appear in the form and content of the statement. Lastly, according to MASAM, it is important to analyse the circumstances of the event, establish the witness' characteristics, and to assess how the witness was interviewed, as there will be coherence between these factors in truthful statements (Wojciechowski, 2015). There are 21 MASAM criteria used to analyse statements (Table 5). An algorithm is used to provide raters with accurate guidelines in how the criteria should be interpreted.

Table 12. MASAM criteria (Wojciechowski, 2015)

1. Internal coherency	The statement has an internal structure, lack of contradictory remarks and self-contradictory elements
2. Coherence with other statements	The statement has a lack of contradictions to other statements given by other individuals
3. Coherence with other evidence	The statement has a lack of contradictions to information established based on other evidence
4. Volume of statement	The statement contains a vast amount of information, details, and descriptions
5. Description language	The statement features language used by the suspect that is accurate, unique, and related to their linguistic capability
6. Structure of statement	The statement is hierarchical, includes cause-effect statements, and is chronological
7. Statement linguistic function	The statement is descriptive, expressive, and persuasive
8. Character and types of details	The statement does not contain irrelevant details that are unexpected
9. Interactions descriptions	The statement includes cause-effect chains of interactions

10. Consequences	The statement contains information on consequences of described events or suspects' awareness of consequences
11. Contextual setting and external associations	The statement contains information on circumstances in which the event took place
12. Sensory data	The statement contains sensory data (visual, auditorial, smell, taste, sensational)
13. Source of statement	The statement is based on data originating from different senses
14. Description of internal states	The statement contains emotions and/or thoughts and describes their character and level of intensity
15. Descriptions of relations	The statement describes one's relation to the event, to people involved in the event and to the cause of the event
16. Readiness to depose	The individual's attitude is characterised with a readiness to describe and recollect the main plots
17. Readiness to search, identify, and reproduce memory traces	The individual has a willingness to search one's memory to find information necessary to answer questions

18. Level of confidence	The individual has internal doubts about their memories
19. Complementing	The individual has a readiness to complement the statement through answering detailed questions
20. Memory loses	The statement includes the individual's awareness of their memory deficits
21. Search for acceptance	The statement includes a phrasing which reveals the individual's need of being accepted or understood by the interviewer

With the use of the algorithm, the level of accuracy reached between 90-99.95% for MASAM distinguishing between truthful and fabricated statements (Wojciechowski, 2015). Comparing it with CBCA, whilst both assessments were able to significantly distinguish between truthful and false statements compared to lay evaluators, only three CBCA criteria (namely, self-deprecation, pardoning the perpetrator and details characteristic of the offence) were able to differentiate between the statements compared to the vast majority of MASAM criteria (only character and types of details, and source of statement failed to discriminate between truthful and fabricated accounts; Wojciechowski et al., 2018). However, this procedure is recommended to be used on adult statements, unlike CBCA which was developed to also be used on children.

Regarding the overall concurrent validity of CBCA as compared to other measures, whilst there appears to be no significant different in the effectiveness of CBCA and RM, CBCA has a been shown to be better able to discriminate between truthful and fabricated statements compared to SCAN, a widely used technique in the criminal justice sector. Furthermore, RM and SCAN have their own limitations that may render them unusable in some circumstances, for example, on child populations. However, MASAM has shown to be more effective for use on adult statements compared to CBCA.

As mentioned, another way of assessing concurrent validity is to observe other factors associated with the phenomenon the tool attempts to measure.

Age. It is well known that cognitive and language abilities develop with age (Hoff, 2020) and hence this would affect the presence of certain CBCA criteria in an individual's statement. Much research has shown that, as expected, CBCA scores are positively correlated with age (for example, Anson et al., 1993; Buck et al., 2002; Vrij et al., 2002).

Coaching of interviewee. Another factor is coaching of the interviewee. It can be suggested that if an individual was taught how to give a convincing interview, that this would yield higher CBCA scores. Research that has tested this assumption by coaching participants on the CBCA criteria has found support for this (Joffe & Yuille, 1992; Vrij et al., 2002). These findings lend more support to concurrent validity of CBCA. It also raises a concerning point, as individuals could better fake their lies as truths by becoming aware of the criteria.

Normative data

There were no normative data studies published for CBCA. Most countries in which SVA is employed have a western culture (Cacuci et al., 2021), and the studies discussed in this review also reflect this. This means that there is a lack of knowledge on how this method might vary depending on cultural and ethnic factors. Immigrants, asylum seekers and foreign visitors are frequently faced with situations in which the credibility of their statements is questioned (Bull, 2018), yet the assessment is not empirically verified for use on these populations. There are cultural differences in communication patterns that would affect CBCA results. For example, some cultures use more words compared to other cultures (Leal et al., 2018) which would affect the quantity of details criterion. Furthermore, collectivist societies are more likely to refer to societal roles rather than to the individual and speaking about one's own mental states is not as endorsed (Nisbett & Masuda, 2007). This would affect criterion such as accounts of subjective mental state and attribution of perpetrator's mental state. Many psychometric instruments require adaptation and validation for different cultures (Bravo, 2003) and it is likely that this is also the case for CBCA. Whilst there may be support for use on Western cultures, it may mean that caution must be taken when using this assessment on those from different cultures.

Furthermore, there has been some evidence to show that CBCA is not appropriate to use on all types of statements (Jagodzinski, 2018). Jagodzinski (2018) found that the criteria are not appropriate to use for public appeal statements for missing individuals as the context of statement production is different and crime victimisation narratives were absent, rendering some criteria irrelevant. This is also supported by Amado et al.'s (2016) meta-analysis as the authors concluded that CBCA was better able to discriminate between statements of certain crimes, including sexual offences and intimate partner violence cases, compared to others.

Therefore, CBCA may only be relevant to certain crime statements and future research is needed to determine for which crimes it is valid.

Limitations

There are some other limitations to the CBCA method that need to be considered. As mentioned, there is no manual and hence no standardised way of conducting an SVA assessment using CBCA criteria. Across studies, the type of rating scale used to measure the criteria differs and there are differences in the intensity of training that raters receive (Hauch et al., 2017). Furthermore, there is no clear threshold on if a criterion has been met, nor a threshold for what is considered a 'true' statement and what is considered a 'false' statement. Undeutsch stated that the examiner must consider the 'intensity 'of each criterion and how pronounced it is (Tully, 1999) which lacks objectivity. With vast variability in how the research has been conducted and clear gaps still missing in the field relating to CBCA research, this poses the question of whether study findings can be generalised to support its use in the criminal justice system.

Another limitation is that CBCA criteria consider the statement as a whole; however, only part of the statement may be false. This may invalidate the entire testimony when the actual accusation may be true (Köhnken et al., 2015). If the entire case is based on an individual's testimony, as is the case in many sexual assault claims, a non-credible CBCA score may influence the judge in favour of the defence, hence preventing justice. Therefore, the MASAM method might be more appropriate to use in practice as it considers that statements contain both truthful and false information.

Lastly, it is difficult to test the validity of CBCA in the field. Whilst the research has shown that CBCA is able to discriminate between statements that were dismissed (assumed fabricated statements) and those that were prosecuted (assumed truthful statements). It may be that individuals who were dismissed were guilty of the crime but there was not enough evidence to support conviction, or the individual that was prosecuted was not guilty. Without knowledge of the ground truth, the validity from field studies must be viewed with caution. However, there are similar results in the laboratory when this the ground truth is known (i.e., truthful and fabricated statements are experimentally manipulated) and hence these findings have been used to give support to CBCA use in the field.

Conclusions

CBCA is the core of SVA, a technique created to evaluate children's accounts in alleged sexual abuse cases. It is the most popular instrument currently used to assess the veracity of children's verbal statements (Cacuci et al., 2021). Despite the popularity and importance of this instrument in legal settings, there is no manual and hence no standardised way of completing the assessment. Whilst meta-analyses have been able to come to conclusions on both reliability and validity of the measure, results were largely heterogeneous across studies, raising the question as to whether the success of the tool also varies across real cases. Furthermore, it is unclear how much the different ways of rating the criteria in the studies have affected the results. Standardising CBCA and SVA use would potentially reduce differences between studies and maximise reproducibility of any results, allowing for more solid conclusions to be made on the reliability and validity of the measure.

Using the current research, overall, CBCA has demonstrated good reliability which has been confirmed across separate meta-analyses (Vrij, 2005; Hauch et al., 2017). However, whilst CBCA showed good face and construct validity, it appears that CBCA may be more valid for

use with child populations compared with adults, as validity measures were higher on use with children and there are other methods (for example, MASAM) that may be more valid for use with adults.

It would be beneficial for future research to focus on internal reliability, to determine whether all criteria are consistent across the assessment, and also on alternate forms reliability, to see if the different ways of rating CBCA criteria are consistent. Further to this, more research into the criterion validity of CBCA criteria would be valuable to understand whether the Undeutsch hypothesis captures all the facets of determining the difference between a truthful and a fabricated statement. This is supported by the fact that SCAN (another credibility tool) has contradicting criteria compared to CBCA and hence more research is needed into the underlying hypotheses for these tools. Lastly, more robust research, that calculates measures of validity (e.g., Cronbach's alpha), would be beneficial. This would be useful as it is currently difficult to reach conclusions on the CBCA assessment and compare it to other methods, without the use of standardised validity measures that are widely used for assessing psychometric tools and psychological measurement.

Currently, caution needs to be taken when using this assessment on witnesses from non-western cultures and on some types of statements (that are not related sexual offences or intimate partner violence) as there is evidence to suggest that the criteria may not be appropriate for these. Future research is needed on different cultures to understand whether the assessment needs to be adapted for use on those outside of Western cultures. It is also needed on different types of statements to determine which crime types CBCA is appropriate for.

It is understandable that a tool such as CBCA was developed due to the nature of sexual abuse cases and hence needing to examine the veracity of statements which may be the only form of evidence. However, the process of assessing the veracity of a statement may be invalidating to witnesses if they are told that their statement is to be analysed for credibility. There is much research to show that the process of being interviewed is already invalidating to witnesses of a crime (Champion et al., 2021) and the use of this assessment may cause further distress and prevent witnesses from disclosing crimes. Whilst there appears to be good reliability and validity for use the CBCA's use with children, it could be recommended that CBCA should only be used if *completely necessary*, especially due to the current gaps in the literature discussed.

Chapter Five

Discussion

This chapter provides an overview of the findings of each chapter and the conclusions made across the thesis. The work in this thesis could be used to inform theory, policy, and practice.

Overall Aims

The overall aim of this thesis was to better understand the accuracy of witness statements in the real world. This was achieved through a systematic literature review that explored the current knowledge gained from studies of witness memory reports of real-world crimes; a secondary data analysis that attempted to test a potential mechanism accounting for memory recall in witnesses; and a critique of a measure that attempts to identify dishonest statements in real cases. Together, this work aimed to increase knowledge of factors that can affect accuracy for memory reporting in real life crime events, which could inform practice, including developing and recommending strategies for how police should interview witnesses.

Summary of Findings

Chapter One introduced the basis for the current thesis, explaining what eyewitness memory is and why it is important in criminal investigations, including the fact that many wrongful convictions cases contained poor eyewitness testimony (Innocence Project, 2020). Furthermore, it discussed the findings of laboratory studies and how these have been used to inform practice in the criminal justice system, such as the Achieving Best Evidence (ABE) procedure used in the UK to interview witnesses and Statement Validity Analysis (SVA) used to determine the credibility if witnesses. Lastly, it discussed the need

for field research as, whilst laboratory research has strengths, the findings may not be applicable to real world situations. Lastly, the quantity-accuracy trade-off framework (Koriat & Goldsmith, 1996) was introduced which may be a mechanism to explain that can explain differences between laboratory and field research and is discussed throughout the thesis.

Chapter Two presented a systematic literature review to look at the current findings from field research on real crime situations. The review considered 16 studies and discussed the types of participants and crimes they investigated, the types of statements collected and how these were analysed, and the factors the studies investigated and their associations with memory reporting. The review found that most studies examined violent crimes and statements from participants (i.e., witnesses) that were older than typically university age (in comparison to laboratory studies which predominantly test undergraduate students). Furthermore, most studies used police statements which were analysed by looking at both the length of the statements and their accuracy, determined by comparing witness statements to "facts" about the crime. It was found that memory weakening factors can affect length differently to accuracy. Lastly, the memory weakening factors that were investigated in each study were discussed. Overall, it was found that the findings from laboratory and field studies did not necessarily correspond. This points to the importance of utilizing field research to also inform practice. However, a limitation of the current field studies, is that they do not consider potential mechanisms that underlie memory reporting in the real world. This was noted as a consideration for future research.

Chapter Three introduced a study that aimed to fill this gap in the field. The study was a secondary data analysis of data collected by Flowe et al. (2011). The data were randomly sampled felony cases from San Diego District Attorney's closed case archives. Police

crime incident reports were analysed, which included eyewitness descriptors of crime information (including weapon used, nature of any injury, lighting, duration of crime, distance from crime, etc.) and of the perpetrator's physical appearance (including age, gender, race, height, weight, build, eye colour, hair colour, hair length, hair type, facial hair, complexion). An identical physical appearance checklist was completed by the police when a suspect was arrested, allowing for correspondence between the witness' description of the perpetrator and the person later convicted of the offence's appearance to be determined. The aim of this study was to examine how the number and accuracy of details witnesses provide in their description of the perpetrator vary in relation to the presence of factors that are commonly believed to compromise memory. Specifically, the first aim was to examine whether the presence of each of the memory compromising factors was associated with the witness providing a less detailed (i.e., the number of exact match features given) and accurate (i.e., the proportion of correct versus not correct features) description of the perpetrator. The results showed that the presence of memory compromising factors did affect the number of 'exact match' and 'no match' features. The association of memory compromising factors with the exact match features means that statements may be briefer under circumstances when memory compromising factors are present. Furthermore, the presence of memory compromising factors was also associated with the accuracy for exact match features but not no match features meaning that specificity of correct features reported may be altered under conditions where memory compromising factors are present. A second aim of the study was to examine if the total number of memory compromising factors is associated with witnesses providing a less detailed and accurate perpetrator description. The results showed that more memory compromising factors did not necessarily lead to less detailed or accurate memory. The study's results are largely in support of the quantity accuracy trade-off framework which

states that individuals strategically regulate their memory reporting and hence they would volunteer to the police only memories about which they are confident (Koriat & Goldsmith, 1996). That is, although there are memory compromising factors present, witnesses may report less information or less precise information, but maintain accuracy. The study findings also support the findings from the systematic literature review that the conclusions from laboratory and field studies are not always the same. Namely, the study did not support laboratory findings that suggest that certain factors affect the accuracy of memory reporting for crimes, because the study found that the presence of memory compromising factors is not necessarily associated with less accurate memory reporting.

Chapter Four critiqued the Statement Validity Analysis (SVA) procedure, focusing on the Criteria-Based Content Analysis (CBCA) component. This procedure is used to determine the credibility of witnesses and can be used in court in some North American and West European countries. Overall, CBCA demonstrated good reliability which has been confirmed across two meta-analyses (Vrij, 2005; Hauch et al., 2017), as well as good face and construct validity. However, there are some limitations to the measure which should be addressed due to its wide use in criminal proceedings. Firstly, there is no manual and hence no standardised way of completing the assessment which ostensibly led to heterogenous results across the studies reviewed (and therefore might also results in variability when it is applied in practice). Furthermore, CBCA may be more valid for use with child populations compared with adults, as there were other methods (for example, MASAM; Wojciechowski, 2015) that may be more valid for use with adults. Caution also needs to be taken when using this assessment on cultures outside of the West and on some types of crime as there is evidence to suggest that the criteria may not be appropriate for these. There are also concerns about the general use of this tool as it may be invalidating to witnesses if they are told that their statement is to be analysed for credibility.

Research and Theoretical Implications

Much of the early research into eyewitness memory has been completed in the laboratory. This allows researchers to manipulate factors of interest and observe if those manipulations effect memory reporting (Hulley et al., 2013). However, field research allows for studying experiences in their natural form (Fisher & Wood, 2007). This is important in the current arena of research as real-life crimes involve extraneous variables often not present in laboratory studies, and hence it is still important that we study the effects on memory without controlling for these extraneous variables. Indeed, Bartels et al. (2018) state that hypotheses about human behaviour discovered in laboratory studies often languish in the field. This is supported in the systematic literature review and the research chapter in this thesis as these both concluded that the findings from research conducted in the field can be different to findings from the laboratory. Bartels et al. (2018) suggest that phenomena and theoretical hypotheses should first start in the laboratory and then be tested in field settings. As there has been much research into eyewitness memory in the laboratory, this suggests that future research needs to focus on field studies into eyewitness testimony for real crime situations, which currently lack in number compared to laboratory studies.

Furthermore, the results of the research chapter have supported the quantity accuracy trade-off framework (Koriat & Goldsmith, 1996) in explaining memory reporting in real world crime situations. Whilst some previous studies have discussed it as a potential mechanism that underlies their findings (for example, alcohol-induced witnesses provide less but still accurate information to the police; Flowe et al., 2015), the current research derived hypotheses from the theory, tested those hypotheses, and found support for the theory. The study (Chapter Three) examined multiple memory compromising factors,

suggesting that this theoretical basis may be generalizable to many situations. Whilst this is a positive step, future research should also focus on discussing and testing the theoretical bases for their own findings. Furthermore, more research is needed on the quantity-accuracy framework in real world crime situations to replicate the findings of the current thesis. It is important to test the theory because there have already been practical applications of the quantity accuracy trade-off framework (Koriat and Goldsmith, 1996), such as the Self-Administered Interview (SAI) (Gabbert et al., 2009) which was been developed based on this framework. The SAI provides the witness with retrieval support and encourages metacognitive monitoring and control without the need for an interviewer. Meta-analyses of the SAI found that subsequent memory recall, after using an initial SAI, were more detailed and accurate showing that the SAI can be an effective tool for preserving witness memory (Horry et al., 2020). This again suggests that the quantity-accuracy theory is a useful theoretical framework for understanding real world witness memory.

Practical Implications

The findings of this thesis can also be used in practice. Firstly, the findings suggest that witnesses need to be treated with care during interviewing. For example, the current procedure of interviewing in the UK—the ABE procedure (Ministry of Justice, 2011) — utilises a free-recall phase and a questioning phase. Chapter Three explains that if the quantity-accuracy trade-off framework is correct, excessively questioning witnesses makes it harder for witnesses to strategically control what they are reporting, and therefore a long or detailed questioning phase could increase the number of inaccurate details reported and reduce the overall accuracy of witness's memory evidence (Ingham et al., 2019). Furthermore, it has been communicated that witnesses will recall more accurately if they refrain from guessing which would be more likely to happen if police are asking detailed

questions of which the witness is not sure (Fisher & Geiselman, 2010). The questioning phase of an interview is also associated with miscommunication (Filipović, 2022) and suggestibility (Gudjonsson, 2013). Questioning phases are also an issue when interviewing those with mental health or learning disabilities, as police often adapt their communication (e.g., to suit the needs of the vulnerable suspect) but this leads to best practice not being adhered to, such as not using open questions (Farrugia & Gabbert, 2019). It could therefore be recommended for police and courtroom procedures to prioritise the free recall phase, and reduce the questioning phase unless necessary, to maximise the accuracy of the information from the witness.

The findings of the thesis also have implications for the Turnbull Guidelines (R v Turnbull and others, 1976), where judges are instructed to direct the jury to examine the circumstances in which the identification by the witness came to be made. This includes how long the witness viewed the accused for, if the witness knew the accused, and the retention interval between the crime and police questioning. However, none of these factors were found to affect accuracy of recall memory in the thesis research. This may therefore mean that these guidelines need to be revisited and updated based on new research. However, as previously discussed, the Turnbull Guidelines (R v Turnbull and others, 1976) are utilised when there has been identification evidence and hence they may not be applicable to recall evidence as it may be that different mechanisms underpin this. Nonetheless, it may be suggested that jurors would also apply this guidance to all eyewitness evidence presented in court. Therefore, the thesis findings suggest that these guidelines should be updated with current findings.

Furthermore, it is also important that witnesses should be treated with respect as this may affect the quality of their memory reporting. As explained in Chapter Three, witnesses and victims of violent offences have often expressed reporting their experience of a crime to the police as a stressful process, including having worries about confidentiality and a fear of negative reactions such as blame and stigmatisation (Champion, Lock, Puntan & Hendra, 2021). Furthermore, there is evidence that current interviewing techniques do not take into consideration the emotional needs of the interviewee (e.g., see Parr & Stevenson, 2013; Gabbert et al., 2020). It is possible that negative feelings are further enhanced using tools such as the SVA, where assuming that witnesses are lying may lead to feelings of invalidation and prevent disclosure of crimes. As the aim of police procedures is to increase the accuracy of testimony, by undermining witnesses, this will prevent accurate information from being achieved in the future. The potential invalidation of witnesses needs to be considered throughout the criminal justice system.

Conclusion

To conclude, the thesis aimed to better understand the accuracy of witness statements for real world crimes. This was achieved through a systematic literature review which discussed the current findings of field research and suggested that the findings of laboratory and field research do not always correspond with each other. It also suggested that future field research focus on the theoretical bases underlying the findings (Chapter Two). A theoretical basis was empirically tested in Chapter Three using a secondary data analysis on real world crimes and the quantity accuracy trade-off framework (Koriat & Goldsmith, 1996) was found to be a potential underlying theory of how witnesses monitor their memory during questioning. Lastly, CBCA, a component of SVA, was critiqued and future recommendations for the use of this measure were suggested (Chapter Four).

Together, this thesis suggests that research should focus on conducting theoretically driven field studies on factors associate with witness memory reporting. This thesis was the first empirical test of the quantity accuracy trade-off framework in field studies (Koriat & Goldsmith, 1996) as a potential mechanism that may underlie witness reporting and hence the quantity accuracy tradeoff framework needs to be examined further in research. The thesis also suggests that police interviews should focus on free recall accounts as the questioning phase used in UK practice (ABE procedure, Ministry of Justice, 2011) may inadvertently reduce the overall accuracy of memory reporting. Lastly, witnesses need to be treated with respect as the negative experiences encountered during questioning may prevent future crimes from being reported.

References

Abrams, L., & Farrell, M. T. (2011). Language processing in normal aging. In J. Guendouzi, F. Loncke, & M. J. Williams (Eds.), *The handbook of psycholinguistic and cognitive processes: Perspectives in communication disorders* (pp. 49–73). Psychology Press. https://doi.org/10.4324/9780203848005.ch3

Akehurst, L., Manton, S., & Quandte, S. (2011). Careful calculation or a leap of faith? A field study of the translation of CBCA ratings to final credibility judgements. *Applied Cognitive Psychology*, *25*(2), 236–243. https://doi.org/10.1002/acp.1669

Amado, B. G., Arce, R., & Fariña, F. (2015). Undeutsch hypothesis and Criteria Based Content Analysis: A meta-analytic review. *The European Journal of Psychology Applied to Legal Context*, *7*(1), 3–12. https://doi.org/10.1016/j.ejpal.2014.11.002

Amado, B. G., Arce, R., Fariña, F., & Vilariño, M. (2016). Criteria-Based Content Analysis (CBCA) reality criteria in adults: A meta-analytic review. *International Journal of Clinical and Health Psychology*, *16*(2), 201–210. https://doi.org/10.1016/j.ijchp.2016.01.002

Anson, D. A., Golding, S. L., & Gully, K. J. (1993). Child sexual abuse allegations: Reliability of criteria-based content analysis. *Law and Human Behavior*, *17*(3), 331–341. https://doi.org/10.1007/bf01044512

Baddeley, A., Eysenck, M. W., & Anderson, M. C. (2020). *Memory*. Routledge.

Bliss, T. V. & Collingridge, G. L. (1993). A synaptic model of memory: Long-term potentiation in the hippocampus. *Nature, 361(6407),* 31-39. https://doi.org/10.1038%2F361031a0

Bartels, D. M., Hastie, R., & Urminsky, O. (2018). Connecting Laboratory and Field Research in Judgment and Decision Making: Causality and the Breadth of External

Validity. *Journal of Applied Research in Memory and Cognition*, *7*(1), 11–15.

https://doi.org/10.1016/j.jarmac.2018.01.001

Bogaard, G., Meijer, E. H., Vrij, A., & Merckelbach, H. (2016). Scientific Content Analysis
(SCAN) Cannot Distinguish Between Truthful and Fabricated Accounts of a Negative
Event. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00243

Bonferroni, C. (1936). Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni
del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze*, *8*, 3-62.

Bravo, M. (2003). Instrument development: Cultural adaptations for ethnic minority research.
In Bernal, G. (Ed.) *Handbook of Racial and Ethnic Minority Psychology* (pp. 220-
236). SAGE.

Brigham, J. C., & Bothwell, R. K. (1983). The ability of prospective jurors to estimate the
accuracy of eyewitness identifications. *Law and Human Behavior*, *7*(1), 19–30.
https://doi.org/10.1007/bf01045284

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., & Weber, N. (2008). Eyewitness confidence and latency: Indices of memory
processes not just markers of accuracy. *Applied Cognitive Psychology*, *22*(6), 827–
840. https://doi.org/10.1002/acp.1486

Brigham, J. C., Verst, M. V., & Bothwell, R. K. (1986). Accuracy of Children's Eyewitness
Identifications in a Field Setting. *Basic and Applied Social Psychology*, *7*(4), 295–
306. https://doi.org/10.1207/s15324834basp0704_4

Brigham, J., Bennett, L., Meissner, C., & Mitchell, T. (2007). The influence of race on
eyewitness memory. In: Lindsay R, Ross D, Read J, Toglia M. (Eds.), *Handbook of
Eyewitness Psychology Memory for People.* Routledge.

Brónnimann, R., Herlihy, J., Müller, J., & Ehlert, U. (2013). Do testimonies of traumatic

    events differ depending on the interviewer? *The European journal of psychology*

    applied to legal context, *5*(1), 97-121.

Buck, J. A., Warren, A. R., Betman, S. I., & Brigham, J. C. (2002). Age differences in

    Criteria-Based Content Analysis scores in typical child sexual abuse interviews.

    Journal of Applied Developmental Psychology, *23*(3), 267–283.

    https://doi.org/10.1016/s0193-3973(02)00107-7

Bull, R. (2018). Credible accounts: What they are and how to obtain them. *Linguistics and*

    the Human Sciences, *12*(2-3), 166–176. https://doi.org/10.1558/lhs.36990

Bureau of Justice Statistics. (2005). *Sourcebook of Criminal Justice, 2003* (NCJ Publication

    No. 208756). Washington, DC: Bureau of Justice Statistics.

Cacuci, S. A., Bull, R., Huang, C. Y., & Visu-Petra, L. (2021). Criteria-based content

    analysis in child sexual abuse cases: a cross-cultural perspective. *Child abuse*

    review, *30*(6), 520-535. https://doi.org/10.1002/car.2733

Carlson, C. A., Young, D. F., Weatherford, D. R., Carlson, M. A., Bednarz, J. E., & Jones, A.

    R. (2016). The Influence of Perpetrator Exposure Time and Weapon Presence/Timing

    on Eyewitness Confidence and Accuracy. *Applied Cognitive Psychology*, *30*(6), 898–

    910. https://doi.org/10.1002/acp.3275

Chae, Y. (2010). Application of Laboratory Research on Eyewitness Testimony. *Journal of*

    Forensic Psychology Practice, *10*(3), 252–261.

    https://doi.org/10.1080/15228930903550608

Champion, H., Lock, K., Puntan, L., & Hendra, H. (2021). *Evaluation of Rape Survivors'*

    Experience of the Police & Other Criminal Justice Agencies. Criminal Justice Joint

    Inspection. https://www.justiceinspectorates.gov.uk/hmicfrs/wp-

content/uploads/evaluation-of-rape-survivors-experience-of-the-police-and-other-criminal-justice-agencies.pdf

Christianson S-AÊ. (1992a). *Handbook of Emotion and Memory: Current Research and Theory.* Psychology Press.

Christianson, S.-Å. (1992b). Emotional stress and eyewitness memory: A critical review. *Psychological Bulletin*, *112*(2), 284–309. https://doi.org/10.1037/0033-2909.112.2.284

Christianson, S.-Å., & Hübinette, B. (1993). Hands up! A study of witnesses 'emotional reactions and memories associated with bank robberies. *Applied Cognitive Psychology*, *7*(5), 365–379. https://doi.org/10.1002/acp.2350070502

Cizek, G. J. (2012). Defining and distinguishing validity: Interpretations of score meaning and justifications of test use. *Psychological Methods*, *17*(1), 31–43. https://doi.org/10.1037/a0026975

Clifford, B. R., & Hollin, C. R. (1981). Effects of the type of incident and the number of perpetrators on eyewitness memory. *Journal of Applied Psychology*, *66*(3), 364–370. https://doi.org/10.1037/0021-9010.66.3.364

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences, 2nd Edition.* Routledge.

Colloff, M. F., Wade, K. A., Wixted, J. T., & Maylor, E. A. (2017). A signal-detection analysis of eyewitness identification across the adult lifespan. *Psychology and Aging*, *32*(3), 243–258. https://doi.org/10.1037/pag0000168

Cooper, B. S., Kennedy, M. A., Hervé, H. F., & Yuille, J. C. (2002). Weapon focus in sexual assault memories of prostitutes. *International Journal of Law and Psychiatry*, *25*(2), 181–191. https://doi.org/10.1016/s0160-2527(01)00107-8

Coupe, T., & Griffiths, M. (1996). *Solving residential burglary* (Crime Detection and

Prevention Series No. 77). Home Office.

Cowan, N. (2006). *The development of memory in childhood.* Psychology Press.

Craig, R. A., Scheibe, R., Raskin, D. C., Kircher, J. C., & Dodd, D. H. (1999). Interviewer

questions and content analysis of children's statements of sexual abuse. *Applied

Developmental Science*, *3*(2), 77-85. https://doi.org/10.1207/s1532480xads0302_2

Crossland, D., Kneller, W., & Wilcock, R. (2018). Intoxicated eyewitnesses: prevalence and

procedures according to England's police officers. *Psychology, Crime & Law*, *24*(10),

979–997. https://doi.org/10.1080/1068316x.2018.1474216

Crown Prosecution Service (CPS). (2021). *Prosecution Statistics Published for 2021-21*.

Crime Prosecution Service. https://www.cps.gov.uk/cps/news/prosecution-statistics-

published-2020-21

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 and

Human Behavior*, *28*(6), 687–706. https://doi.org/10.1007/s10979-004-0565-x

Drinkaware. (n.d.). *Drink driving and the legal alcohol limit.* Drinkaware.

https://www.drinkaware.co.uk/facts/alcohol-and-the-law/drink-driving-and-the-legal-

alcohol-limit

Driscoll, L. N. (1994). A validity assessment of written statements from suspects in criminal

investigations using the SCAN technique. *Police Studies*, *17*, 77-88.

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of

behavior. *Psychological Review*, *66*(3), 183–201. https://doi.org/10.1037/h0047707

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

Ebbinghaus, H. (2013). Memory: A Contribution to Experimental Psychology. *Annals of Neurosciences*, *20*(4). https://doi.org/10.5214/ans.0972.7531.200408

Ebbinghaus, H., (1964). *Memory: A contribution to experimental psychology.* Dover Publications.

Ellis, H. D., Deregowski, J. B., & Shepherd, J. W. (1975). Description of White and Black Faces by White and Black Subjects. *International Journal of Psychology*, *10*(2), 119–123. https://doi.org/10.1080/00207597508247325

Esplin, P. W., Boychuk, T., & Raskin, D. C. (1988). A field validity study of criteria-based content analysis of children's statements in sexual abuse cases. *NATO advanced study institute on credibility assessment, Maratea, Italy.*

Estudillo, A. J., Lee, J. K. W., Mennie, N., & Burns, E. (2019). No evidence of other-race effect for Chinese faces in Malaysian non-Chinese population. *Applied Cognitive Psychology*, *34*(1), 270–276. https://doi.org/10.1002/acp.3609

Evans, P. (2007). Policing in the UK and USA: A Brief Comparison. *Economic Affairs*, *27*(4), 44–45. https://doi.org/10.1111/j.1468-0270.2007.00779.x

Faber, J. & Fonseca, L. M. (2014). How sample size influences research outcomes. *Dental Press Journal of Orthodontics, 19*(4), 27-29. https://dx.doi.org/10.1590%2F2176-9451.19.4.027-029.ebo

Fahsing, I. A., Ask, K., & Granhag, P. A. (2004). The Man Behind the Mask: Accuracy and Predictors of Eyewitness Offender Descriptions. *Journal of Applied Psychology*, *89*(4), 722–729. https://doi.org/10.1037/0021-9010.89.4.722

Farrington, D. P., & Lambert, S. (1997). Predicting offender profiles from victim and witness descriptions. In J. L. Jackson & D. A. Bekerian (Eds.), *Offender profiling: Theory, research, and practice* (pp. 133– 158). Wiley.

Farrugia, L., & Gabbert, F. (2019). Vulnerable suspects in police interviews: Exploring

current practice in England and Wales. *Journal of Investigative Psychology and

Offender Profiling*, *17*(1). https://doi.org/10.1002/jip.1537

Fawcett, J. M., Russell, E. J., Peace, K. A., & Christie, J. (2013). Of guns and geese: a meta-

analytic review of the "weapon focus" literature. *Psychology, Crime & Law*, *19*(1),

35–66. https://doi.org/10.1080/1068316x.2011.599325

Fawcett, J. M., Russell, E. J., Peace, K. A., & Christie, J. (2013). Of guns and geese: a meta-

analytic review of the "weapon focus" literature. *Psychology, Crime & Law*, *19*(1),

35–66. https://doi.org/10.1080/1068316x.2011.599325

Ferrie, J. E. (2002). Change in health inequalities among British civil servants: the Whitehall

II study. *Journal of Epidemiology & Community Health*, *56*(12), 922–926.

https://doi.org/10.1136/jech.56.12.922

Filipović, L. (2022). The Good, the Bad and the Ugly: Miscommunication in UK Police

Interviews and US Police Interrogations. *Journal of Police and Criminal Psychology*.

https://doi.org/10.1007/s11896-022-09495-w

Fisher, C. G., & Wood, K. B. (2007). Introduction to and Techniques of Evidence-Based

Medicine. *Spine*, *32*(Supplement), S66–S72.

https://doi.org/10.1097/brs.0b013e318145308d

Fisher, R. P. (1995). Interviewing victims and witnesses of crime. *Psychology, Public Policy,

and Law*, *1*(4), 732–764. https://doi.org/10.1037/1076-8971.1.4.732

Fisher, R. P., & Geiselman, R. E. (2010). The Cognitive Interview method of conducting

police interviews: Eliciting extensive information and promoting Therapeutic

Jurisprudence. *International Journal of Law and Psychiatry*, *33*(5-6), 321–328.

https://doi.org/10.1016/j.ijlp.2010.09.004

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

Fleiss, J. L. (2014). *Statistical methods for rates and proportions.* Wiley.

Flin, R., Boon, J., Knox, A., & Bull, R. (1992). The effect of a five-month delay on children's and adults 'eyewitness memory. *British Journal of Psychology*, *83*(3), 323–336. https://doi.org/10.1111/j.2044-8295.1992.tb02444.x

Flowe, H. D., Takarangi, M. K. T., Humphries, J. E., & Wright, D. S. (2015). Alcohol and remembering a hypothetical sexual assault: Can people who were under the influence of alcohol during the event provide accurate testimony? *Memory*, *24*(8), 1042–1061. https://doi.org/10.1080/09658211.2015.1064536

Flowe, H. D., Carline, A., & Karoğlu, N. (2018). Testing the reflection assumption. *The International Journal of Evidence & Proof*, *22*(3), 239–261. https://doi.org/10.1177/1365712718782996

Flowe, H. D., Colloff, M. F., Karoğlu, N., Zelek, K., Ryder, H., Humphries, J. E., & Takarangi, M. K. T. (2017). The Effects of Alcohol Intoxication on Accuracy and the Confidence-Accuracy Relationship in Photographic Simultaneous Line-ups. *Applied Cognitive Psychology*, *31*(4), 379–391. https://doi.org/10.1002/acp.3332

Flowe, H. D., Colloff, M. F., Kloft, L., Jores, T., & Stevens, L. M. (2020). *Impact of alcohol and other drugs on eyewitness memory* (pp. 149-162). Routledge.

Flowe, H. D., Humphries, J. E., Takarangi, M. K., Zelek, K., Karoglu, N., Gabbery, F., & Hope, L. (2019). An experimental examination of the effects of alcohol consumption and exposure to misleading postevent information on remembering a hypothetical rape scenario. *Applied Cognitive Psychology, 33*(3), 393-413. https://doi.org/10.1002/acp.3531

Flowe, H. D., Mehta, A., & Ebbesen, E. B. (2011). The role of eyewitness identification evidence in felony case dispositions. *Psychology, Public Policy, and Law, 17*(1), 140-159. https://doi.org/10.1037/a0021311

Flowe, H., Finklea, K., & Ebbesen E. (2009). Limitations of expert psychology testimony on eyewitness identification. In B. Cutler (Ed.) *Expert testimony on the psychology of eyewitness identification.* Oxford University Press.

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), 298–307. https://doi.org/10.1007/s10979-008-9146-8

Gabbert, F., Tamonyte, D., Apps, J., Caso, A., Woolnough, P., Hope, L., ... & Waterworth, G. (2020). Examining the efficacy of a self-administered report form in missing person investigations. *Legal and Criminological Psychology*, *25*(1), 1-16.

Garrett, B. (2011). *Convicting the innocent*. Harvard University Press.

Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1985). Eyewitness memory enhancement in the police interview: Cognitive retrieval mnemonics versus hypnosis. *Journal of Applied Psychology*, *70*(2), 401–412. https://doi.org/10.1037/0021-9010.70.2.401

Ginty, A. T. (2012). Construct Validity. In Gellman, M. & Turner, J. R. (Eds.) *Encyclopedia of Behavioral Medicine* (pp. 487). Springer.

Glenberg, A. M., Schroeder, J. L., & Robertson, D. A. (1998). Averting the gaze disengages the environment and facilitates remembering. *Memory & Cognition*, *26*(4), 651–658. https://doi.org/10.3758/bf03211385

Glisky, E. (2007). Changes in Cognitive Function in Human Aging. In Riddle, D. R. (Ed.). *Brain Aging: Models, Methods, and Mechanisms.* CRC Press.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., Nugent, T. F., Herman, D. H., Clasen, L. S., Toga, A. W., Rapoport, J. L., & Thompson, P. M. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences*, *101*(21), 8174–8179. https://doi.org/10.1073/pnas.0402680101

Goldsmith, M., Pansky, A., & Koriat, A. (2014). Metacognitive control of memory reporting. In Perfect, T.J. & Lindsay D. S. (Eds.) *The SAGE handbook of applied memory*, (pp. 481-500). SAGE Publications Ltd.

Goldstein, A. G., Chance, J. E., & Schneller, G. R. (1989). Frequency of eyewitness identification in criminal cases: A survey of prosecutors. *Bulletin of the Psychonomic Society*, *27*(1), 71–74. https://doi.org/10.3758/bf03329902

Goodman, G. S., Bottoms, B. L., Rudy, L., Davis, S. L., & Schwartz-Kenney, B. M. (2001). Effects of past abuse experiences on children's eyewitness memory. *Law and Human Behavior*, *25*(3), 269–298. https://doi.org/10.1023/a:1010797911805

Goodman, G. S., & Melinder, A. (2007). Child witness research and forensic interviews of young children: A review. *Legal and Criminological Psychology*, *12*(1), 1–19. https://doi.org/10.1348/135532506x156620

Goodwin, D. W., Othmer, E., Halikas, J. A., & Freemon, F. (1970). Loss of Short Term Memory as a Predictor of the Alcoholic "Blackout." *Nature*, *227*(5254), 201–202. https://doi.org/10.1038/227201a0

Granhag, P. A., Ask, K., Rebelius, A., Öhman, L., & Mac Giolla, E. (2013). "I saw the man who killed Anna Lindh!" An archival study of witnesses 'offender descriptions. *Psychology, Crime & Law*, *19*(10), 921–931. https://doi.org/10.1080/1068316x.2012.719620

Griesel, D., & Yuille, J. C. (2012). Sex trade workers 'narratives of sexual violence: A field

    investigation. *Memory*, *20*(3), 236–253.

    https://doi.org/10.1080/09658211.2012.654797

Gudjonsson, G. H. (2013) Interrogative Suggestibility and Compliance. In Ridley, A. M.,

    Gilbert, F. & LaRoney, D. J. (Eds.) *Suggestibility in Legal Contexts: Psychological*

    Research and Forensic Implications (pp. 45-61). Wiley Blackwell.

Hagsand, A. V., Roos af Hjelmsäter, E., Granhag, P. A., Fahlke, C., & Söderpalm Gordh, A.

    (2016). Witnesses stumbling down memory lane: The effects of alcohol intoxication,

    retention interval, and repeated interviewing. *Memory*, *25*(4), 531–543.

    https://doi.org/10.1080/09658211.2016.1191652

Hagsand, A., Hjelmsäter, E. R. A., Granhag, P. A., Fahlke, C., & Söderpalm-Gordh, A.

    (2013). Bottled memories: On how alcohol affects eyewitness recall. *Scandinavian*

    Journal of Psychology, *54*(3), 188–195. https://doi.org/10.1111/sjop.12035

Hancock, P. J. B., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces.

    Trends in Cognitive Sciences, *4*(9), 330–337. https://doi.org/10.1016/s1364-

    6613(00)01519-9

Harvey, A. J., Shrimpton, B., Azzopardi, Z., O'Grady, K., Hicks, E., Hirst, E., & Atkinson-

    Cox, K. (2020). The influence of alcohol and weapon presence on eyewitness

    memory and confidence. *Applied Cognitive Psychology*, *34*(2), 489–503.

    https://doi.org/10.1002/acp.3636

Hauch, V., Sporer, S. L., Masip, J., & Blandón-Gitlin, I. (2017). Can credibility criteria be

    assessed reliably? A meta-analysis of criteria-based content analysis. *Psychological*

    Assessment, 29(6), 819–834. https://doi.org/10.1037/pas0000426

Herzmann, G., Willenbockel, V., Tanaka, J. W. & Curran, T. (2011). The neural correlates of memory encoding and recognition for own-race and other-race faces. *Neuropsychologia, 49,* 11, 3103–3115. doi:10.1016/j.neuropsychologia.2011.07.019

Höfer, E., Akehurst, L., & Metzger, G. (1996). Reality monitoring: A chance for further development of CBCA. In *annual meeting of the European Association on Psychology and Law, Sienna, Italy.*

Hoff, E. (2020). *Language development.* Wadsworth.

Holleman, G. A., Hooge, I. T. C., Kemner, C., & Hessels, R. S. (2020). The "Real-World Approach" and Its Problems: A Critique of the Term Ecological Validity. *Frontiers in Psychology*, *11*(721). https://doi.org/10.3389/fpsyg.2020.00721

Holmberg, U. (2004). *Police Interviews with Victims and Suspects of Violent and Sexual Crimes; Interviewees' Experiences and Interview Outcomes.* [Doctoral Dissertation, Stockholm University].

Horowitz, S. W., Lamb, M. E., Esplin, P. W., Boychuk, T. D., Krispin, O., & Reiter-Lavery, L. (1997). Reliability of criteria-based content analysis of child witness statements. *Legal and Criminological Psychology*, *2*(1), 11-21. https://doi.org/10.1111/j.2044-8333.1997.tb00329.x

Horry, R., Hughes, C., Sharma, A., Gabbert, F., & Hope, L. (2020). A meta-analytic review of the Self-Administered Interview©: Quantity and accuracy of details reported on initial and subsequent retrieval attempts. *Applied Cognitive Psychology*, *35*(2), 428–444. https://doi.org/10.1002/acp.3753

Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D. G., & Newman, T. B. (2013). *Designing clinical research* (4th ed.). Wolters Kluwer/Lippincott Williams & Wilkins.

Ingham, M. P., Gibbs, B., Colloff, M. F., Stevens, L., Rockowitz, S. R., Hayre, R. K., Butt, M., & Flowe, H. D. (2019). *Harnessing meta-memory processes in police interviews to assess recall accuracy: A secondary analysis of interviews from Flowe et al. (2019).* [Unpublished manuscript]

Innocence Project (2020). *DNA Exonerations in the United States (1989-2020).* Innocence Project. https://innocenceproject.org/dna-exonerations-in-%20the-%20united-%20states/

Institute of Medicine; Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations; Board on the Health of Select Populations. (2015). *Psychological Testing in the Service of Disability Determination.* National Academies Press.

Jagodzinski, R. (2018). *A Critical Evaluation of the Usefulness of Criteria-Based Content Analysis for Judging the Credibility of Public Appeals for Missing Persons* [Doctoral dissertation, University of Central Lancashire]. http://clok.uclan.ac.uk/23955/1/23955%20Jagodzi%C5%84ski%20Rebecca%20Final%20e-Thesis%20%28Master%20Copy%29.pdf

Joffe, R. & Yuille, J. C. (1992). *Criteria-Based Content Analysis: An experimental investigation.* Paper presented at the NATO Advanced Study Institute on the Child Witness in Context: Cognitive, Social and Legal Perspectives, Lucca, Italy.

Johnson, E. (2012). Face Validity. In Volkmar, F. R. (Ed.) *Encyclopedia of Autism Spectrum Disorders* (pp. 1226-1227). Springer.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*(1), 67–85. https://doi.org/10.1037/0033-295X.88.1.67

Johnson, M. K., Foley, M. A., Suengas, A. G., & Raye, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events.

Journal of Experimental Psychology: General, *117*(4), 371–376.

https://doi.org/10.1037/0096-3445.117.4.371

Jong, M. D., Wagenaar, W. A., Wolters, G., & Verstijnen, I. M. (2005). Familiar face

recognition as a function of distance and illumination: a practical tool for use in the

courtroom. *Psychology, Crime & Law*, *11*(1), 87–97.

https://doi.org/10.1080/10683160410001715123

Jores, T., Colloff, M. F., Kloft, L., Smailes, H., & Flowe, H. D. (2019). A meta-analysis of

the effects of acute alcohol intoxication on witness recall. *Applied Cognitive

Psychology*, *33*(3), 334–343. https://doi.org/10.1002/acp.3533

Junnarkar, M. and Lakhani, S., (2021). Investigating the Eyewitness: Accuracy and Fallacies

of Memory. In Sahni, S.P. & Bhadra, P. (Eds.) *Criminal Psychology and the Criminal

Justice System in India and Beyond* (pp. 203-213). Springer.

Karpicke, J. and Roediger III, H. (2007). Repeated retrieval during learning is the key to

long-term retention. *Journal of Memory and Language*, 57(2), pp.151–162.

https://doi.org/10.1016/j.jml.2006.09.004

Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the "general acceptance"

of eyewitness testimony research: A new survey of the experts. *American

Psychologist*, *56*(5), 405–416. https://doi.org/10.1037/0003-066x.56.5.405

Köhnken, G., Manzanero, A. L., & Scott, M. T. (2016). Statement validity assessment: myths

and limitations. *Ene*, *12*, 27. http://dx.doi.org/10.1016/j.apj

Köhnken, G., Milne, R., Memon, A., & Bull, R. (1999). The cognitive interview: A meta-

analysis. *Psychology, Crime & Law*, *5*(1-2), 3–27.

https://doi.org/10.1080/10683169908414991

Komel, M., Šterk, K., Župan Galunić, A., & Savić, G. (2020). SCAN Revisited Through Linguistic Psychoanalysis. *Kriminalistička teorija i praksa*, *7*(1/2020.), 7-30. https://hrcak.srce.hr/file/367360

Koriat, A. & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory performance. *Psychological Review, 103,* 490-517.

Kroger, W. S., & Doucé, R. G. (1980). Forensic Uses of Hypnosis. *American Journal of Clinical Hypnosis*, *23*(2), 86–94. https://doi.org/10.1080/00029157.1980.10403245

Kuehn, L. L. (1974). Looking down a Gun Barrel: Person Perception and Violent Crime. *Perceptual and Motor Skills*, *39*(3), 1159–1164. https://doi.org/10.2466/pms.1974.39.3.1159

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*(863). https://doi.org/10.3389/fpsyg.2013.00863

Lamers-Winkelman, F. (2021). Statement validity analysis: Its application to a sample of Dutch children who may have been sexually abused. In Faller, K. C., & Vanderlaan, R. (Eds.) *Maltreatment in early childhood: tools for research-based intervention* (pp. 59-81)*.* Routledge.

Lampinen, J. M., Erickson, W. B., Moore, K. N., & Hittson, A. (2014). Effects of distance on face recognition: implications for eyewitness identification. *Psychonomic Bulletin & Review*, *21*(6), 1489–1494. https://doi.org/10.3758/s13423-014-0641-2

Lane, S. M. (2006). Dividing attention during a witnessed event increases eyewitness suggestibility. *Applied Cognitive Psychology*, *20*(2), 199–212. https://doi.org/10.1002/acp.1177

Law Insider. (n.d.). *Law Insider*. Retrieved June 8, 2021, from https://www.lawinsider.com/dictionary/special-investigation

Leal, S., Vrij, A., Vernham, Z., Dalton, G., Jupe, L., Harvey, A., & Nahari, G. (2018). Cross-cultural verbal deception. *Legal and Criminological Psychology*, *23*(2), 192–213. https://doi.org/10.1111/lcrp.12131

Lerner, Y., Papo, D., Zhdanov, A., Belozersky, L., & Hendler, T. (2009). Eyes Wide Shut: Amygdala Mediates Eyes-Closed Effect on Emotional Experience with Music. *PLoS ONE*, *4*(7), e6230. https://doi.org/10.1371/journal.pone.0006230

Lieberman, J. D., Carrell, C. A., Miethe, T. D., & Krauss, D. A. (2008). Gold versus platinum: Do jurors recognize the superiority and limitations of DNA evidence compared to other types of forensic evidence? *Psychology, Public Policy, and Law*, *14*(1), 27–62. https://doi.org/10.1037/1076-8971.14.1.27

Lin, W-L. & Yao, G. (2014). Concurrent Validity. In Michalos, A. C. (Ed.) *Encyclopedia of quality of life and well-being research* (pp.1184-1185). Dordrecht Springer.

Lindsay, R. C., Wells, G. L., & Rumpel, C. M. (1981). Can people detect eyewitness-identification accuracy within and across situations? *Journal of Applied Psychology, 66*(1), 79–89. https://doi.org/10.1037/0021-9010.66.1.79

Lindsay, D. S. (2002). Children's source monitoring. In H. L. Westcott, G. M. Davies, & R. H. C. Bull (Eds.) *Children's testimony: A handbook of psychological research and forensic practice* (pp. 83-98). Wiley.

List, J. A. (1986). Age and schematic differences in the reliability of eyewitness testimony. *Developmental Psychology*, *22*(1), 50–57. https://doi.org/10.1037/0012-1649.22.1.50

Loftus, E. F. (1975). Reconstructing memory: The incredible eyewitness. *Jurimetrics Journal*, *15*(3), 188-193.

Loftus, E. F. (1979). *Eyewitness testimony.* Harvard University Press.

Loftus, E. F. (1996). *Eyewitness testimony.* Harvard University Press.

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of Automobile Destruction: an Example of the Interaction between Language and Memory. *Journal of Verbal Learning and Verbal Behavior, 13*(5), 585-589. https://doi.org/10.1016/s0022-5371(74)80011-3

Loftus, E. F., Loftus, G. R., & Messo, J. (1987). Some facts about "weapon focus." *Law and Human Behavior*, *11*(1), 55–62. https://doi.org/10.1007/bf01044839

Loftus, G. R., & Harley, E. M. (2005). Why is it easier to identify someone close than far away? *Psychonomic Bulletin & Review*, *12*(1), 43–65. https://doi.org/10.3758/bf03196348

Longstaff, M. G., & Belz, G. K. (2020). Sex differences in eyewitness memory: Females are more accurate than males for details related to people and less accurate for details surrounding them, and feel more anxious and threatened in a neutral but potentially threatening context. *Personality and Individual Differences*, *164*, 110093. https://doi.org/10.1016/j.paid.2020.110093

Macleod, M. D., & Shepherd, J. W. (1986). Sex Differences in Eyewitness Reports of Criminal Assaults. *Medicine, Science and the Law*, *26*(4), 311–318. https://doi.org/10.1177/002580248602600413

Mansour, J. K., Beaudry, J. L., Bertrand, M. I., Kalmet, N., Melsom, E. I., & Lindsay, R. C. L. (2020). Impact of disguise on identification decisions and confidence with simultaneous and sequential lineups. *Law and Human Behavior*, *44*(6), 502–515. https://doi.org/10.1037/lhb0000427

McGaugh, J. L. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annual Review of Neuroscience*, *27*(1), 1–28. https://doi.org/10.1146/annurev.neuro.27.070203.144157

McLeod, S. A. (2007). *What is reliability?* Simply Psychology. www.simplypsychology.org/reliability.html

McMillan, L., & Thomas, M. (2009). Police interviews of rape victims: Tensions and

    contradictions. In Horvath, M. & Brown, J. (Eds.), *Rape: Challenging contemporary*

    thinking, (pp. 255-280). Willan.

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a

    matching task. *Memory & Cognition*, *34*(4), 865–876.

    https://doi.org/10.3758/bf03193433

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in

    memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *7*(1),

    3–35. https://doi.org/10.1037//1076-8971.7.1.3

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

Mickes, L., Hwe, V., Wais, P. E., & Wixted, J. T. (2011). Strong memories are hard to scale.

    Journal of Experimental Psychology: General, *140*(2), 239–257.

    https://doi.org/10.1037/a0023007

Milne, R., & Bull, R. (1999). *Investigative interviewing: psychology and practice.* John

    Wiley & Sons.

Ministry of Justice. (2011). *Achieving Best Evidence in Criminal Proceedings: Guidance on*

    interviewing victims and witnesses, and guidance on using special measures. Ministry

    of Justice.

    https://www.cps.gov.uk/sites/default/files/documents/legal_guidance/best_evidence_i

    n_criminal_proceedings.pdf

Mintzer, M. Z. (2007). The acute effects of alcohol on memory: A review of laboratory

    studies in healthy adults. *International Journal on Disability and Human*

    Development, *6*(4). https://doi.org/10.1515/ijdhd.2007.6.4.397

Monds, L. A., Cullen, H. J., Kloft, L., van Golde, C., Harrison, A. W., & Flowe, H. (2022). Memory and credibility perceptions of alcohol and other drug intoxicated witnesses and victims of crime. *Psychology, Crime & Law*, *28*(8), 820-840. https://a-w-harrison.github.io/articlePDFs/MemoryAndCredibilityPerceptionsIntoxicatedWitnesses.pdf

Nahari, G., Vrij, A., & Fisher, R. P. (2012). Exploiting liars 'verbal strategies by examining the verifiability of details. *Legal and Criminological Psychology*, *19*(2), 227–239. https://doi.org/10.1111/j.2044-8333.2012.02069.x

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

Nisbett, R. E., & Masuda, T. (2007). Culture and Point of View. *Intellectica. Revue de l'Association Pour La Recherche Cognitive*, *46*(2), 153–172. https://doi.org/10.3406/intel.2007.1282

Noa, O. (2009). The development of declarative memory systems in the brain. *Frontiers in Human Neuroscience*, *3*. https://doi.org/10.3389/conf.neuro.09.2009.10.020

Oberlader, V. A., Naefgen, C., Koppehele-Gossel, J., Quinten, L., Banse, R., & Schmidt, A. F. (2016). Validity of content-based techniques to distinguish true and fabricated statements: A meta-analysis. *Law and Human Behavior*, *40*(4), 440–457. https://doi.org/10.1037/lhb0000193

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., & van Koppen, P. J. (2009). Eyewitness memory of a supermarket

robbery: A case study of accuracy and confidence after 3 months. *Law and Human*

Behavior, *33*(6), 506–514. https://doi.org/10.1007/s10979-008-9152-x

Office for National Statistics (ONS). (2020). *Crime in England and Wales: year ending*

December 2020. Office for National Statistics.

https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/cri

meinenglandandwales/yearendingdecember2020

Office for National Statistics (ONS). (2021). *Crime in England and Wales: year ended*

March 2021. Office for National Statistics.

https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/cri

meinenglandandwales/yearendingmarch2021#types-of-violence

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

Papailiou, A. P., Yokum, D. V., & Robertson, C. T. (2014). The Novel New Jersey

Eyewitness Instruction Induces Skepticism But Not Sensitivity. *SSRN Electronic*

Journal. https://doi.org/10.2139/ssrn.2475217

Parr, H., & Stevenson, O. (2013). *Families living with absence: Searching for missing*

people. University of Glasgow.

http://geographiesofmissingpeople.org.uk/downloads/familiesreport.pdf

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *Linguistic Inquiry and Word Count:*

LIWC. LIWC.net

Penrod, S. D., Fulero, S. M., & Cutler, B. L. (1995). Expert psychological testimony on

eyewitness reliability before and after Daubert: The state of the law and the science.

Behavioral Sciences & the Law, *13*(2), 229–259.

https://doi.org/10.1002/bsl.2370130206

Perfect, T. J., Andrade, J., & Eagan, I. (2011). "Eye closure reduces the cross-modal memory

impairment caused by auditory distraction": Correction to Perfect et al. (2011)..

Journal of Experimental Psychology: Learning, Memory, and Cognition, *37*(5),

1286–1286. https://doi.org/10.1037/a0025052

Perfect, T. J., Wagstaff, G. F., Moore, D., Andrews, B., Cleveland, V., Newcombe, S.,

Brisbane, K.-A., & Brown, L. (2008). How can we help witnesses to remember more?

It's an (eyes) open and shut case. *Law and Human Behavior*, *32*(4), 314–324.

https://doi.org/10.1007/s10979-007-9109-5

Piaget, J. (2015). *The moral judgment of the child.* Andesite Press.

Pike, G., Brace, N., & Kynan, S. (2002). *The Visual Identification of Suspects: Procedures

and Practice.* Home Office, A Publication of the Policing and Reducing Crime Unit

Home Office Research, Development and Statistics Directorate.

http://oro.open.ac.uk/84749/1/brf202%20-

%20The%20Visual%20Identification%20of%20Suspects_%20Procedures%20and%2

0Practice%20.pdf

Porter, S., & Yuille, J. C. (1994). Credibility assessment of criminal suspects through

statement analysis. *Psychology, Crime & Law*, *1*(4), 319–331.

https://doi.org/10.1080/10683169508411969

Porter, S., & Yuille, J. C. (1996). The language of deceit: An investigation of the verbal clues

to deception in the interrogation context. *Law and Human Behavior*, *20*(4), 443–458.

https://doi.org/10.1007/bf01498980

Pozzulo, J. (2006). Person description and identification by child eyewitnesses. In R. C. L.

Lindsay, D.F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of eyewitness

psychology: Memory for people* (pp. 283-308). Lawrence Erlbaum Associates.

Pozzulo, J. D., & Dempsey, J. L. (2009). Witness Factors and Their Influence on Jurors'
Perceptions and Verdicts. *Criminal Justice and Behavior*, *36*(9), 923–934.
https://doi.org/10.1177/0093854809338450

Protogerou, C., & Hagger, M. S. (2020). A checklist to assess the quality of survey studies
in psychology. *Methods in Psychology*, *3*, 100031.
https://doi.org/10.1016/j.metip.2020.100031

R v Turnbull and others (1976). 63 Cr App R 132.

Raskin, D. C., & Esplin, P. W. (1991). Statement validity assessment: Interview procedures
and content analysis of children's statements of sexual abuse. *Behavioral Assessment,*
13, 265–291.

Raskin, D. C., Esplin, P. W., & Horowitz, S. W. (1991). *Investigative interviews and*
assessments of children in sexual abuse cases. [Unpublished manuscript].

Reno, F., Fisher, R. C., Robinson, L., Brennan, N., & Travis, J. (1999). *Eyewitness Evidence:*
A Guide for Law Enforcement. U.S. Department of Justice.
https://www.ojp.gov/pdffiles1/nij/178240.pdf

Roediger, H. L. (1996). Memory Illusions. *Journal of Memory and Language*, *35*(2), 76–100.
https://doi.org/10.1006/jmla.1996.0005

Ross, D. F., Read, J. D., & Toglia, M. (1994). *Adult Eyewitness Testimony: Current Trends*
and Developments. Cambridge University Press.

Saal, F. E., Downey, R. G., & Lahey, M. A. (1980). Rating the ratings: Assessing the
psychometric quality of rating data. *Psychological Bulletin*, *88*(2), 413–428.
https://doi.org/10.1037/0033-2909.88.2.413

Sapir, A. (1987/2000). *The LSI course on scientific content analysis (SCAN).* Laboratory for
Scientific Interrogation.

Sauer, J., Brewer, N., Zweck, T. & Weber, N. (2010). The effect of retention interval on the confidence–accuracy relationship for eyewitness identification. *Law and Human Behavior*, 34(4), pp.337–347. https://doi.org/10.1007/s10979-009-9192-x

Schacter, D. L. (1996). *Searching for memory.* Basic Books.

Schemmel, J., Maier, B. G., & Volbert, R. (2019). Verbal Baselining: Within-Subject Consistency of CBCA Scores across Different Truthful and Fabricated Accounts. *The European Journal of Psychology Applied to Legal Context*, *12*(1), 35–42. https://doi.org/10.5093/ejpalc2020a4

Schumann, C. M. (2004). The Amygdala Is Enlarged in Children But Not Adolescents with Autism; the Hippocampus Is Enlarged at All Ages. *Journal of Neuroscience*, *24*(28), 6392–6401. https://doi.org/10.1523/jneurosci.1297-04.2004

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

Shriver, E. R., Young, S. G., Hugenberg, K., Bernstein, M. J., & Lanter, J. R. (2007). Class, Race, and the Face: Social Context Modulates the Cross-Race Effect in Face Recognition. *Personality and Social Psychology Bulletin*, *34*(2), 260–274. https://doi.org/10.1177/0146167207310455

Smith, N. (2001). *Police Research Series Paper: Vol. 135. Reading between the lines: An evaluation of the scientific content analysis technique (SCAN).* UK Home Office, Research, Development and Statistics Directorate.

Sporer, S. L. (1996). Psychological aspects of person descriptions. In S. L. Sporer, R. S. Malpass, & G. Köhnken (Eds.), *Psychological issues in eyewitness identification* (pp. 53–86). Erlbaum.

Sporer, S. L. (2001). The cross-race effect: Beyond recognition of faces in the laboratory. *Psychology, Public Policy, and Law*, *7*(1), 170–200. https://doi.org/10.1037/1076-8971.7.1.170

Steller, M. (1989). Recent developments in statement analysis. In Yuille, J. C. (Ed.), *Credibility assessment* (pp.135-154). Kluwer Academic.

Stern, W. L. (1902). *Zur psychologie der aussage.* E-Content Generic Vendor.

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.

Tollestrup, P.A., Turtle, J.W., & Yuille, J.C. (1994). Actual victims and witnesses to robbery and fraud: An archival analysis. In D. Ross, D. Read, & C. Ceci (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 144-162). Press Syndicate of the University of Cambridge.

Tully, B. (1999). Statement Validation. In Canter, D. & Alison, L. (Eds.) *Interviewing and Deception* (pp. 85-103). Routledge.

Undeutsch, U. (1984). Courtroom evaluation of eyewitness testimony. *Applied Psychology*, *33*(1), 51–66. https://doi.org/10.1111/j.1464-0597.1984.tb01416.x

Undeutsch, U. (1989). The development of statement reality analysis. In Yuille, J. C. (Ed.), *Credibility assessment* (pp.101-120). Kluwer Academic.

United States Congressional Research Service (2017). *Public Trust and Law Enforcement: A Discussion for Policymakers*. Congressional Research Service. https://fas.org/sgp/crs/misc/R43904.pdf

Valentine, T., Pickering, A., & Darling, S. (2003). Characteristics of eyewitness identification that predict the outcome of real lineups. *Applied Cognitive Psychology*, *17*(8), 969–993. https://doi.org/10.1002/acp.939

van Koppen, P. J., & Lochun, S. K. (1997). Portraying perpetrators: The validity of offender descriptions by witnesses. *Law and Human Behavior*, *21*(6), 661–685. https://doi.org/10.1023/a:1024812831576

van Oorsouw, K., Broers, N.J. & Sauderland, M. (2019). Alcohol intoxication impairs eyewitness memory and increases suggestibility: Two field studies. *Applied Cognitive Psychology, 33*(3), pp. 439-455. https://doi.org/10.1002/acp.3561.

Vredeveldt, A., Charman, S. D., den Blanken, A., & Hooydonk, M. (2018). Effects of cannabis on eyewitness memory: A field study. *Applied Cognitive Psychology*, *32*(4), 420–428. https://doi.org/10.1002/acp.3414

Vredeveldt, A., Tredoux, C. G., Nortje, A., Kempen, K., Puljević, C., & Labuschagne, G. N. (2015). A field evaluation of the Eye-Closure Interview with witnesses of serious crimes. *Law and Human Behavior*, *39*(2), 189–197. https://doi.org/10.1037/lhb0000113

Vrij, A. (2005). Criteria-Based Content Analysis: A Qualitative Review of the First 37 Studies. *Psychology, Public Policy, and Law*, *11*(1), 3–41. https://doi.org/10.1037/1076-8971.11.1.3

Vrij, A. (2008). *Detecting Lies and Deceit: Pitfalls and Opportunities, 2nd Edition.* John Wiley & Sons.

Vrij, A. (2015). Verbal lie detection tools: Statement validity analysis, reality monitoring and scientific content analysis. In Granhag, P. A., Vrij, A. & Verschuere, B. (Eds.) *Detecting deception current challenges and cognitive approaches* (pp. 1-35). Wiley.

Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2002). Will the truth come out? The effect of deception, age, status, coaching, and social skills on CBCA scores. *Law and Human Behavior*, *26*(3), 261–283. https://doi.org/10.1023/a:1015313120905

Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2004). Detecting deceit via analyses of verbal

    and nonverbal behavior in children and adults. *Human communication research*,

    30(1), 8-41. https://doi.org/10.1111/j.1468-2958.2004.tb00723.x

Wagstaff, G. F., MaCveigh, J., Boston, R., Scott, L., Brunas-Wagstaff, J., & Cole, J. (2003).

    Can Laboratory Findings on Eyewitness Testimony Be Generalized to the Real

    World? An Archival Analysis of the Influence of Violence, Weapon Presence, and

    Age on Eyewitness Accuracy. *The Journal of Psychology*, *137*(1), 17–28.

    https://doi.org/10.1080/00223980309600596

Wagstaff, G. F., MaCveigh, J., Boston, R., Scott, L., Brunas-Wagstaff, J., & Cole, J. (2003).

    Can Laboratory Findings on Eyewitness Testimony Be Generalized to the Real

    World? An Archival Analysis of the Influence of Violence, Weapon Presence, and

    Age on Eyewitness Accuracy. *The Journal of Psychology*, *137*(1), 17–28.

    https://doi.org/10.1080/00223980309600596

Wang, Z., Nayfeh, T., Tetzlaff, J., O'Blenis, P., & Murad, M. H. (2020). Error rates of human

    reviewers during abstract screening in systematic reviews. *PLOS ONE*, *15*(1),

    e0227742. https://doi.org/10.1371/journal.pone.0227742

Wang, J., Otgaar, H., Smeets, T., Howe, M. L., Merckelbach, H., & Zhuo, C. (2018).

    Consequences of False Memories in Eyewitness Testimony: A Review and

    Implications for Chinese Legal Practice. *Psychological Research on Urban Society*,

    1(1), 12. https://doi.org/10.7454/proust.v1i1.15

Weber, N., & Brewer, N. (2008). Eyewitness recall: Regulation of grain size and the role of

    confidence. *Journal of Experimental Psychology: Applied*, *14*(1), 50–60.

    https://doi.org/10.1037/1076-898x.14.1.50

Weinberg, H. I., Wadsworth, J., & Baron, R. S. (1983). Demand and the impact of leading questions on eyewitness testimony. *Memory & Cognition, 11*(1), 101-104. https://doi.org.10.3758/bf03197667

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual review of Psychology*, *54*(1), 277-295. https://capitalpunishmentincontext.org/files/resources/innocence/annual_review_2003 .pdf

Wells, G. L., Kovera, M. B., Douglass, A. B., Brewer, N., Meissner, C. A., & Wixted, J. T. (2020). Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. *Law and Human Behavior*, *44*(1), 3–36. https://doi.org/10.1037/lhb0000359

West, R. L. & Stone, K. R. (2014) Age Differences in Eyewitness Memory for a Realistic Event. *The Journals of Gerontology, 69(3). pp.* 338–347. https://doi.org/10.1093/geronb/gbt014

Winsor, A. A., Flowe, H. D., Seale-Carlisle, T. M., Killeen, I. M., Hett, D., Jores, T., Ingham, M., Lee, B. P., Stevens, L. M., & Colloff, M. F. (2021). Child witness expressions of certainty are informative. *Journal of Experimental Psychology: General*, *150*(11), 2387–2407. https://doi.org/10.1037/xge0001049

Wise, R. A., Sartori, G., Magnussen, S., & Safer, M. A. (2014). An Examination of the Causes and Solutions to Eyewitness Error. *Frontiers in Psychiatry*, *5*(102). https://doi.org/10.3389/fpsyt.2014.00102

Wixted, J. T. (2018). Time to exonerate eyewitness memory. *Forensic Science International*, *292*, e13–e15. https://doi.org/10.1016/j.forsciint.2018.08.018

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., & Fisher, R. P. (2018). Rethinking the Reliability of Eyewitness Memory. *Perspectives on Psychological Science*, *13*(3), 324–335. https://doi.org/10.1177/1745691617734878

Wixted, J. T., Read, J. D., & Lindsay, D. S. (2016). The Effect of Retention Interval on the Eyewitness Identification Confidence–Accuracy Relationship. *Journal of Applied Research in Memory and Cognition*, *5*(2), 192–203. https://doi.org/10.1016/j.jarmac.2016.04.006

Wogalter, M. S. (1996). Describing Faces from Memory: Accuracy and Effects on Subsequent Recognition Performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *40*(11), 536–540. https://doi.org/10.1177/154193129604001103

Wojciechowski, B. W. (2015). *Analiza i ocena zeznań świadków [Analysis and assessment of witnesses 'statements]*. Gdańskie Wydawnictwo Psychologiczne.

Wojciechowski, B. W., Gräns, M., & Lidén, M. (2018). A true denial or a false confession? Assessing veracity of suspects 'statements using MASAM and SVA. *PLOS ONE*, *13*(6), e0198211. https://doi.org/10.1371/journal.pone.0198211

Woolnough, P. S., & MacLeod, M. D. (2001). Watching the birdie watching you: eyewitness memory for actions using CCTV recordings of actual crimes. *Applied Cognitive Psychology*, *15*(4), 395–411. https://doi.org/10.1002/acp.717

Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2001). A field study of own-race bias in South Africa and England. *Psychology, Public Policy, and Law*, *7*(1), 119–133. https://doi.org/10.1037/1076-8971.7.1.119

Yarmey, A. D. (1993). Adult Age and Gender Differences in Eyewitness Recall in Field

    Settings1. *Journal of Applied Social Psychology*, *23*(23), 1921–1932.

    https://doi.org/10.1111/j.1559-1816.1993.tb01073.x

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of

    habit-formation. *Journal of Comparative Neurology and Psychology*, *18*(5), 459–482.

    https://doi.org/10.1002/cne.920180503

Young, S. G., Hugenberg, K., Bernstein, M. J., & Sacco, D. F. (2011). Perception and

    Motivation in Face Recognition. *Personality and Social Psychology Review*, *16*(2),

    116–142. https://doi.org/10.1177/1088868311418987

Yuille, J. C. (1990). Adult ''Step-Wise ''Assault Interview protocol. Unpublished

    manuscript. Vancouver, BC, Canada: University of British Columbia.

Yuille, J. C., & Cutshall, J. L. (1986). A case study of eyewitness memory of a crime.

    Journal of Applied Psychology, *71*(2), 291–301. https://doi.org/10.1037/0021-

    9010.71.2.291

Yuille, J. C., & Cutshall, J. L. (1989). Analysis of the statements of victims, witnesses, and

    suspects. In Yuille, J. C. (Ed.), *Credibility assessment* (pp.175-191). Kluwer

    Academic.

Yuille, J. C., & Kim, C. K. (1987). A field study of the forensic use of hypnosis. Canadian

    Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement,

    19(4), 418–429. https://doi.org/10.1037/h0080001

Yuille, J. C., & Tollestrup, P. A. (1990). Some effects of alcohol on eyewitness memory.

    Journal of Applied Psychology, *75*(3), 268–273. https://doi.org/10.1037//0021-

    9010.75.3.268

Zeller, R. A. & Carmines, E. G. (1980). *Measurement in the Social Sciences.* University

    Press.

Appendix A

Search record

		Web of Science (Core collection)	Ovid (Psychinfo)	ProQuest (Social Sciences Premium Collection)
1	(Eyewitness* or Bystander* or Witness* or Observer*) near/2 (Memor* or Recall or Remember* or Recollect* or Testimon* or evidence*)	2,809	3,040	10,291

2	(Field or Natural*) near/2 (stud* or experiment* or research)	410,259	302,91	867,514
3	1 AND 2	51	77	156

Appendix B

Quality Assessment Tool (Protogerou and Hagger, 2020)

1	Introduction	Was the problem or phenomenon under investigation defined, described, and justified?
2	Introduction	Was the population under investigation defined, described, and justified?
3	Introduction	Were specific research questions or hypotheses stated?
4	Introduction	Were operational definitions of all study variables provided?
5	Participants	Were participant inclusion criteria stated?
6	Participants	Was the participant recruitment strategy described?
7	Participants	Was a justification/rationale for the sample size provided?
8	Data	Was the attrition rate provided? (Applies to cross-sectional and prospective studies)
9	Data	Was a method of treating attrition provided? (Applies to cross-sectional and prospective studies)
10	Data	Were the data analysis techniques justified (i.e., was the link between hypotheses/aims/research questions and data analyses explained)?

11	Data	Were the measures provided in the report (or in a supplement) in full?
12	Data	Was evidence provided for the validity of all the measures (or instrument) used?
13	Data	Was information provided about the person(s) who collected the data (e.g., training, expertise, other demographic characteristics)?
14	Data	Was information provided about the context (e.g., place) of data collection?
15	Data	Was information provided about the duration (or start and end date) of data collection?
16	Data	Was the study sample described in terms of key demographic characteristics?
17	Data	Was discussion of findings confined to the population from which the sample was drawn?
18	Ethics	Were participants asked to provide (informed) consent or assent?
19	Ethics	Were participants debriefed at the end of data collection?
20	Ethics	Were funding sources or conflicts of interest disclosed?

Note questions 18 and 19 were not used

Appendix C

Quality Assessment

Quality Assessment	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	20	Why omitted
Brónnimann, Herlihy, Müller & Ehlert (2013)	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	N	N	Y	Y	Y	Y	N	
Wagstaff et al. (2003)	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	N	N	Y	N	Y	Y	N	
Christianson & Hübinette (1993)	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	N	N	Y	Y	Y	Y	N	
Woolnough & MacLeod (2001)	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y	Y	Y	
Fisher, Geiselman & Amador (1989)	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	N	Y	Y	N	N	Y	N	
Hope, Gabbert & Fisher (2011)	Y	Y	Y	Y	Y	Y	Y	N	N	N	Y	N	N	Y	N	N	Y	Y	Review of current research into the self-administered interview with a case study of use in a real-life case but with limited information analysis

Study																			Notes
Newlands, George, Towell, Kemp & Clifford (1999)	Y	Y	Y	Y	Y	N	N	N	N	N	Y	N	Y	N	N	N	Y	N	Used data from another study with not enough informati on the data
Vredeveldt et al. (2015)	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	N	Y	N	Y	Y	Y	
Yuille & Kim (1987)	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	N	N	Y	Y	N	Y	N	
Yuille & Cutshall (1986)	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	N	Y	N	N	Y	N	
Fahsing, Ask & Granhag (2004)	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	N	N	Y	Y	Y	Y	N	
Tollestrup, Turtle & Yuille (1994)	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	N	N	Y	Y	Y	Y	N	
Cooper, Kennedy, Hervé & Yuille (2002)	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	N	Y	Y	Y	
Kuehn (1974)	Y	Y	Y	N	Y	Y	Y	N	N	N	Y	N	N	Y	Y	N	Y	N	No informati on method or measures
Macleod & Shepherd (1986)	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	N	N	Y	Y	N	Y	N	
Colomb, Ginet, Wright, Demarchi & Sadler (2013)	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	N	N	Y	N	Y	Y	N	Mostly focusing on officers' views and use of the Cognitive Interview

																			rather than the effect on memory of witnesses
van Koppen & Lochun (1997)	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	N	Y	Y	N	Y	N	
Granhag, Ask, Rebelius, Öhman & Mac Giolla (2013)	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	N	Y	Y	Y	N	Y	N	
Odinot, Wolters & van Koppen (2009)	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y	Y	N	
Clifford & George (1996)	Y	Y	Y	Y	Y	N	Y	N	N	Y	Y	N	N	N	N	N	Y	N	Focus on the effect of before and after training in the Cognitive Interview in terms of questioning rather than the effect on memory of witnesses
Dahl, Granér, Fransson, Bertilsson & Fredriksson (2018)	Y	Y	N	N	Y	Y	Y	N	N	N	Y	N	N	Y	N	N	Y	Y	Not enough detail on hypothesis, methods, or analysis

Ashkenazi & Fisher (2022)	Y	Y	Y	Y	N	Y	N	N	N	Y	Y	N	Y	Y	Y	Y	Y	N	

Preliminary analysis for multiple linear regression

A preliminary analysis was performed on the data to test whether assumptions for multiple regression were satisfied. Tabachnick and Fidell (2019) state that to determine the sample size required for multiple regression, the equation is: $N \geq 50 + 8m$ (where m is the number of IVs). In the case of the current data set, which has 13 IVs, this would mean that the sample size would need to be equal to or larger than 154. The current data had some missing data and hence variables with large amounts of missing data were excluded to make sure that there was an appropriate amount of data for the analysis.

Other criteria that need to be met for multiple regression include the assumptions of normality, homoscedasticity, linearity, and the absence of multicollinearity (Tabachnick & Fidell, 2019). The assumption of normality states that the residuals (the difference between the observed value of the dependent variable and the predictor variable) of the regression should follow a normal distribution. This can be tested by examining a normal Predicted Probability (P-P) plot and see if the residuals conform to the diagonal normality line. The assumption of homoscedasticity states that the residuals should be equally distributed. This can be tested by plotting the predicted values and residuals on a scatter plot and checking for equal distribution. The assumption of linearity states the predictor variables in the regression will have a straight-line relationship with the dependent variable. This assumption will be met if residuals are normally distributed and homoscedastic. Lastly, the absence of multicollinearity means that the predictor variables are not highly correlated with each other. This assumption can be tested by checking that variance inflation factor (VIF) values are

below 10.00. All the assumptions were met in the current dataset and hence multiple

regressions were appropriate to use.