VERBAL CREDIBILITY ASSESSMENT OF INCARCERATED VIOLENT

OFFENDERS' MEMORY REPORTS


by

MARGUERITE TERNES

B.A., St. Francis Xavier University, 2001

M.A., The University of British Columbia, 2003


THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY


in


THE FACULTY OF GRADUATE STUDIES

(Psychology)


THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)


March 2009

**Abstract**

This study investigated the verbal credibility of incarcerated offenders' memory reports of perpetrated violent crime through the use of Criteria Based Content Analysis (CBCA) and Reality Monitoring (RM). For a previous study, which examined memory in violent offenders, 150 male incarcerated violent crime perpetrators recalled up to five different types of memories: an act of perpetrated instrumental violence, an act of perpetrated reactive violence, a subjectively disturbing (traumatic) event, a positively valenced event, and a perpetrated act of violence for which the offender had poor memory (Cooper, 2005). The present study applied CBCA and RM to each of the memory reports of violence. In particular, this study compared the credibility of reports of instrumental violence to reports of reactive violence, compared credibility ratings with consistency with file information (whether details from memory reports were corroborated by correctional file information), compared the utility of CBCA and RM as credibility assessment tools, and compared the credibility ratings of psychopaths to nonpsychopaths.

The results showed that the vast majority (96%) of the memories were judged to be credible according to CBCA. CBCA proved to be somewhat effective at discriminating memories based on consistency with file information, as every memory partially confirmed by correctional file information was found credible according to CBCA, and every memory found noncredible according to CBCA was not confirmed or was partially contradicted by correctional file information. RM, as measured by the Memory Characteristics Questionnaire, showed limited effectiveness in discriminating memories based on consistency with file information. Further, CBCA and RM were not

similarly effective at discriminating memories based on the known truth of the memories. RM differentiated memories according to type of violence, but CBCA did not, suggesting that while memories for acts of instrumental violence are more vivid and detailed than memories for acts of reactive violence, they are not more likely to appear credible. Neither RM nor CBCA differentiated memory reports according to whether the participant met the diagnostic criteria for psychopathy. These results are discussed in terms of how they support existing research and theory.  Their implications to the criminal justice system are discussed.

**Table of Contents**

**Acknowledgements**

There are many people who have helped to make this dissertation possible. First, I would like to thank my supervisor, Dr. John Yuille, for giving me the opportunity to do this research, for sharing his wisdom and ideas, and for offering his support and encouragement along the way. Thanks to Dr. Don Read, Dr. Eric Eich, Dr. Peter Graf, Dr. Wolfgang Linden, Dr. Jonathan Schooler, and Dr. Ralph Hakstian for their valuable feedback and advice. All of you have gone above and beyond what I ever could have expected in helping me through the many challenges I faced in conducting this research.

Thanks also to Barry Cooper, for sharing his project with me and for advising and encouraging me throughout the completion of this project, and to Dorothee Griesel, for help with training and coding, as well as for her friendship and support.

I would also like to thank my family for their unconditional love and support throughout my many years of education, especially my sister Marianne who has always been by my side, and my husband Dan who has been there for me throughout my grad school studies. My parents and my brothers, Matt, Martin, and Will, have always cheered me on as I completed my academic pursuits. I would also like to express some appreciation to my Vancouver family, the Ternes', for their love and support.

The present research could not have been conducted without the tremendous effort of many individuals who served in various capacities (i.e., research assistants, coders). Special thanks to Willi Ternes, Benjamin Cheung, Alana Cook, Simone Viljoen, Justin Wareham, Judy Yang, LiLi Liu, Stephanie Fohring, Jessica Turpin, Laura Buchanan, Evelyne Martel, Joy Mainer, and Aksinia Shier. I would also like to thank the

## Dedication

This thesis is dedicated to my family. Thank you for your unconditional and never-ending love and support.

**Verbal Credibility Assessment of Incarcerated Violent Offenders' Memory Reports**

Although it is known that eyewitnesses sometimes lie and individuals suspected

of a criminal act often lie (DePaulo & Kirkendol, 1989; Ekman, 2002) little is known

about the lying behaviour of offenders who have already been convicted for their violent

crimes. It is known, however, that individuals lie on average about twice a day (DePaulo,

Kashy, Kirkendol, Wyer, & Epstein, 1996), mostly about matters that will have little

consequence. Although most everyday lies do not carry heavy consequences, deception

in the interrogation room, courtroom, or parole hearing holds enormous consequences at

both the individual and societal level. Knowing how violent offenders talk about their

crimes in an environment where there is little motivation to lie, will help those in the

criminal justice system (e.g., police, triers of fact, those conducting risk assessments, etc.)

ensure that the reports given by offenders are considered appropriately.

The present study examined the verbal credibility of incarcerated offenders'

memory reports of their violent crimes. Expanding on a previous study, in which 150

male incarcerated violent crime perpetrators recalled up to five different types of

memories: an act of perpetrated instrumental violence, an act of perpetrated reactive

violence, a subjectively disturbing (traumatic) event, a positively valenced event, and a

perpetrated act of violence for which the offender had poor memory (Cooper, 2005), the

present study used Criteria Based Content Analysis (CBCA), an important aspect of

Statement Validity Analysis (SVA; see Horowitz, 1991, for review), and Reality

Monitoring (RM; Johnson, 1988) to assess some of these memories for credibility. The

present study also examined the participants' correctional files to assess the consistency

of the participants' memories with correctional file information. Cooper (2005) examined

the questionnaire data related to the memories, not the content of the memories themselves.  The present study will be the first examination of the memory details provided by the participants, as well as the first examination of the participants' correctional files.

In the following sections, literature related to the present research is discussed. First, lies are defined, and types of lies are discussed. Second, types of credibility assessment are summarized, including physiological, nonverbal, and verbal approaches. The emphasis is on verbal credibility assessment, the focus of the present study. One of the specific verbal credibility assessment approaches to be used in the present study, SVA, and its three components, is highlighted and summarized. Then, research that has examined the utility of SVA is summarized and discussed, with a focus on research that has looked at verbal credibility in the criminal population. Next, the other verbal credibility assessment approach used in the present study, RM, is highlighted and reviewed. Third, unique issues regarding the use of offenders as research participants are discussed, with an emphasis on factors such as type of violence and motivation to lie. Although the research questions are introduced throughout these reviews, a systematic overview is provided in the fourth section of this dissertation. Fifth, the method of the study is described. Sixth, the results of this research are presented separately for each research question. The seventh and final section of this dissertation is a discussion of the findings, including a discussion of the implications of the results, and an outline of the limitations and strengths of this dissertation.

**Lies**

In eyewitness memory research, eyewitness statements can be (1) generally correct (but may contain some errors), (2) deceptive attempts, or (3) historically wrong but subjectively true narratives (i.e., an erroneous account provided without intention to mislead). The third type of statement is a mistaken memory and may be created through suggestive interviewing techniques in therapeutic settings (e.g., Lindsay & Read, 1995; Loftus, 1993; Porter, Birt, Yuille, & Hervé, 2001; Porter & Campbell, 2002).

For a statement to be considered a lie, the target of the lie must be deliberately misled and the target must not be notified of the liar's intention to deceive (Ekman, 1996). It is an intentional, interpersonal, and conscious act. Generally, lies are expressed as falsifications (i.e., untrue statements) or as concealments (i.e., hiding true statements), although Ekman (1996; 1997) has identified three other ways of lying. The first of these is telling the truth falsely. That is, speaking the truth, but making it seem like the opposite is actually meant. The second is telling a half-truth. That is, speaking the truth, but leaving out crucial information to mislead the target of the lie. The third is the *incorrect inference dodge*, a term coined by Ekman. An example of this technique would be a situation in which a person does not like a piece of art but does not want to express his or her dislike to the artist. When asked to comment, this person might say, "Incredible! I can't believe it!"

There are several types of lies that people may tell. An emotional lie occurs when a person displays or describes an emotion that is different from the emotion he or she is actually feeling (e.g., Ekman, 1997; Ekman & Friesen, 1974, 1982; Masip, Garrido, & Herrero, 2004; Seniuk & Yuille, 1996; Waxer, 1983). A person is telling an opinion lie

when he or she expresses an opinion or belief that is different from the one he or she actually holds (e.g., Frank & Ekman, 1997; Seniuk & Yuille, 1996). Finally, a factual lie may be a false denial of having experienced an event or a false assertion of having experienced an event. This type of lie constitutes the most common and most important type of lie that occurs in the court context (Seniuk & Yuille, 1996).

**Underlying Processes of Deception**

Ekman (1997) suggested two reasons why lies fail: cognitive load and emotional load. Theoretically, cognitive load results in clues primarily detectable in the liar's speech, such as errors, latencies, and pauses. Alternatively, emotional load is thought to be revealed in facial expressions, tone of voice, and/or body language. Ekman (2001) coined the term *leakage* to refer to behaviours that betray a lie. Emotions may betray a lie in two ways: a false presentation of an emotion or an emotional consequence of the lie (e.g., fear of being caught, guilt about lying). The resulting stress reaction manifests itself with the activation of the sympathetic nervous system. This activation drives a number of potential innate verbal, nonverbal, and paralinguistic behavioural responses designed instinctually as a component of the "fight or flight" system typically activated in times of stress (deTurck & Miller, 1990). Physiological measures, such as galvanic skin response, polygraph, or blood pressure, comparing sympathetic arousal during deceptive tasks have consistently supported these assumptions (Vrij, 1994).

During sympathetic arousal, the cognitive component of the deception-behaviour framework is activated (Vrij & Winkel, 1993). The control hypothesis states that an individual who is sympathetically aroused is aware of his or her nervousness and will therefore attempt to limit and control any behavioural cues he or she perceives as being

created by the stressor (i.e., the lie; Vrij, 1994). Although many potential behavioural

cues are salient to the individual, others are either perceived as less relevant or are not

consciously perceived at all. Thus, the individual will tend to focus their cognitive

processes on the more salient factors, while often neglecting the others (DePaulo, Lanier,

& Davis, 1983). Additionally, the relative conscious controllability of the salient

behavioural cues is variable. The verbal channel is an example of a behavioural channel

with high controllability because it has an elevated sending capacity (i.e., the number of

distinct messages that can be transmitted by a channel) and a heightened degree of

internal feedback (i.e., the extent to which an individual is aware of the channel's output

and their ability to change this information; DePaulo et al., 1983). Even though the verbal

channel is one of the most salient and controllable channels, there is still the potential of

self-monitoring and behavioural manipulation backfiring. If an individual attempts to

control too many channels concurrently, a cognitive overload may occur, and subsequent

attempted behavioural manipulations will fail.

**Lie Detection Techniques**

Various tools or techniques have been developed to detect deception or credibility. Psychophysiological, behavioural (nonverbal), and verbal techniques have been developed to evaluate statements according to the manner in which lies are presented and may betray themselves (e.g., heart rate, speech style, body language, facial expression, and speech content). Most of these techniques are based on the assumption that the intention to deceive creates cognitive and emotional consequences that are detectable. Since it has been established that no behaviour is diagnostic of lying (Ekman, 2001), all credibility assessment techniques pick up on behavioural changes rather than lies. The reason for a behavioural change may be deception, but it can also be due to other states (e.g., nervousness, embarrassment). Ekman (2001) refers to potentially meaningful changes in a person's speech, behavior, or physiology as "hot spots." Hot spots are signs of leakage, whether a lie or other feelings and/or thoughts are their source. Hot spots vary from person to person, thus, there are no universal signs of lying. That is, there is a large variety of possible behavioural changes that might occur and a large variety of reasons for behavioural changes. For example, while under stress, one person may tap his foot, another may restrict her movements so that her body shifts less than usual, and a third may show no difference in body movements, but does show physiological changes. Any of these changes may occur for a number of reasons, including nervousness, embarrassment, guilt, fatigue, or deception.

**Physiological Approaches to Credibility Assessment**

The most prevalent example of a physiological approach to credibility assessment is the polygraph. It is a physiological technique used to detect behavioural changes during

investigations of crime suspects. Generally, the physiological measures recorded in a polygraph examination include skin conductance, heart rate, blood pressure, and respiration. The polygraph detects the consequences of stress, rather than deception. Although some researchers have found very high accuracy rates using the polygraph to detect credibility (e.g., Raskin, 1987; Raskin, Kircher, Horowitz, & Honts, 1989), many researchers have expressed concerns about using the polygraph in that way (e.g., Iacono & Lykken, 1997). Criticisms of the polygraph include the high rate of positive decisions it yields (e.g., Gale, 1988; Honts & Perry, 1992) and that it can be "fooled" by guilty people (Honts, Raskin, & Kircher, 1994). While the polygraph is widely used in criminal investigations in North America and Israel, for the most part, it is not permitted as evidence in federal courts (Ben-Shakkar, Bar-Hillel, & Kremnitzer, 2002; Saxe & Ben-Shakkar, 1999). Additionally, in Germany, the polygraph was precluded as evidence from any legal procedures (Fiedler, Schmidt, & Stahl, 2002).

**Nonverbal Approaches to Credibility Assessment**

As mentioned, no behaviour is indicative of lying, but nonverbal behaviours can be clues to deception (Ekman, 2001). Behavioural-based lie detection consists of two steps: (1) recognizing a change in behaviour and (2) accurately interpreting the reason for the change (Frank & Ekman, 1997). This aspect of lie detection has been the subject of much discussion (e.g., for reviews see Ekman, 1992; DePaulo, 1992). Various researchers have found that facial expressions are valuable cues in the detection of lies (e.g., DePaulo et al., 2003; Etcoff, Ekman, Magee, & Frank, 2000; Porter, Campbell, Stapelton, & Birt, 2002). In particular, the *Facial Action Coding System* (*FACS*; Ekman & Friesen, 1978), a method developed to measure facial expressions of emotions, has

been useful in detecting truly felt emotions despite efforts to conceal these emotions (Ekman, Friesen, & O'Sullivan, 1997).

Researchers have also studied other behavioural cues to deception. Some investigators have found evidence of deceptive nonverbal behaviour, such as less eye contact, fewer illustrators (movements to modify or supplement speech), and increased adaptors (movements in which one's hand makes contact with another body part; e.g., DePaulo et al., 2003; Ekman & Friesen, 1972). However, these results have not been consistent (e.g., Mann, Vrij, & Bull, 2002; Porter, Doucette, Woodworth, Earle, & MacNeil, 2008; Vrij & Graham, 1997). These inconsistencies may be caused by attempts to control one's behaviour, individual differences, and situational differences. Overall, few behavioural cues discriminated between truths and lies in DePaulo et al.'s (2003) meta-analysis of deception cues. No significant differences were found for body activity, posture shifts, hand-, leg-, foot-, arm-, and hand-movements. That is, not behaviours, but behavioural changes, which may differ from one individual to another, must be interpreted in context to determine whether they are clues to deception. Behaviours, in isolation, cannot be used as reliable indicators of lying (Ekman, 2001). As the interviews used in the present study were not video recorded, behavioural cues to deception could not be assessed.

**Verbal Approaches to Credibility Assessment**

Compared to nonverbal behaviours and physiological associates of lying, the verbal content of lies is the easiest for a liar to control. As with nonverbal behaviour, there are no verbal expressions or behaviours reliably indicative of lying. Cues indicating verbal deception arise for the same reasons as nonverbal indications of deception:

emotion, content complexity, and attempted control. These factors lead people to provide over-generalized or indirect answers (Vrij, 2000). A lie's content complexity causes people to provide implausible or overly short statements. In their literature review, Vrij, Edward, Roberts, and Bull (2000) showed an association between lying and certain verbal behaviours, such as latency period, speech hesitations, quantity of details, reproduction of conversation, and auditory, visual, spatial, and temporal details. Moreover, Sporer and Schwandt's (2006) meta-analysis of paraverbal indicators of deception showed that voice pitch, response latency, and speech errors were positively associated with deception, while message duration was negatively related to deception.

Statement Validity Analysis (SVA) is the most popular technique in the field of verbal credibility assessment. The SVA technique is a comprehensive system developed in Germany to evaluate the veracity of written statements (for a discussion of SVA, see Raskin & Esplin, 1991). SVA was developed in the context of child abuse allegations. Such allegations are often made a considerable amount of time after the event took place, leaving little physical evidence, and most abuse cases happen in private settings, which results in few witnesses apart from the victim and perpetrator, so the child victim's statement is often the only evidence (cf., Steller & Koehnken, 1989; Yuille, 1988). Although some allegations prove to be false, most allegations are valid (Yuille, Tymofievish, & Marxsen, 1995). Despite this, children are often considered to be less credible than adults (e.g., Ceci & Bruck, 1993; Ceci, Ross, & Toglia, 1987), so assessing the credibility of children's allegations became necessary.

SVA began with the work of Undeutsch in Germany in the 1950s (Steller & Koehnken, 1989; Undeutsch, 1989). Although it was developed for use with children, the

procedure came to be applied to the statements of adults as well, particularly to the context of adults alleging a sexual assault or sexual abuse as a child (Seniuk & Yuille, 1996). SVA is made up of three components: (1) a properly conducted interview (e.g., the *Cognitive Interview*, Fisher & Geiselman, 1992; the *Step Wise Interview Protocol*, Yuille, Hunter, Joffe, & Zaparniuk, 1993), (2) Criteria Based Content Analysis (CBCA), and (3) the Validity Checklist to analyze the CBCA outcome (e.g., Lamb, Sternberg, Esplin, Hershkowitz, & Orbach, 1997; Vrij, 2000). The SVA procedure is based on the *Undeutsch hypothesis*: Statements based on genuine experiences differ in quality and content from invented or coached statements (Steller, 1989; Vrij, 2000). Two assumptions, related to a person's motivation and cognition, underlie this hypothesis (Koehnken, Schimossek, Aschermann, & Hoefer, 1995; Ruby & Brigham, 1997). First, this hypothesis assumes that a deceptive person's motivation to appear honest will cause the statement to be expressed differently than if the story was actually true (e.g., a liar will attempt to avoid stereotypical lying behaviours). Second, this hypothesis assumes that every person's cognitive and verbal abilities are limited when it comes to fabricating and maintaining a lie. In general, it is assumed that a person telling a true story will be able to provide more details, more contextual information, and more subjective experiences.

**Component 1: Interview**

The utility of verbal clues in detecting deception depends enormously on the quality of interviewing. If the interview is leading or suggestive these clues cannot be used because one does not know if the information provided comes from the subject or the interviewer. For example, the phrasing of the questions may suggest the answer that

the interviewer is expecting (e.g., Ceci & Bruck, 1993; Cohen & Harnick, 1980; Gudjonsson & Clark, 1986; Schooler & Loftus, 1986). Indeed, Vrij, Mann, Kristen, and Fisher (2007) found that using accusatory interviews, which consisted of suggestive accusations and statements, did not result in verbal cues to deceit and these interviews were not effective at eliciting the verbal clues to credibility necessary for CBCA to effectively discriminate between true and fabricated narratives. Several effective interview techniques have been established, including the Cognitive Interview and the Step Wise Interview.

**The Cognitive Interview.** The Cognitive Interview was established to maximize the amount and accuracy of information gathered in the interview with the victim, witness, or suspect in an eyewitness situation. The focus is on facilitating recollection at the time of the police interview as opposed to at the time of encoding, as the typical crime scenario is usually rapid, emotionally charged, and individuals are unlikely to use conscious memory strategies for remembering (Fisher & Geiselman, 1992; Fisher & Schreiber, 2007; Geiselman, Fisher, Cohen, Holland, & Surtes, 1986). These circumstances may interfere with the encoding process. Since it is not possible to improve encoding in these situations, researchers have concentrated on improving recall through interviewing techniques.

The theory underlying the cognitive interview is based on two regularly accepted principles of memory: (1) the memory trace is composed of many elements, and there must be feature overlap between the encoded event and the retrieval cue, and (2) the same piece of information may be accessible through various retrieval cues, as there may be several paths to the encoded event (Geiselman, Fisher, MacKinnon, & Holland, 1985).

The core elements of the Cognitive Interview are organized in terms of three basic psychological processes: memory/general cognition, social dynamics, and communication (Fisher & Schreiber, 2007). In terms of memory/general cognition, first, individuals are asked to reinstate the context surrounding the target event (e.g., think about how they felt, what the weather was like). Second, acknowledging that people have limited mental resources to process information, interviewers ask fewer but more open-ended questions. Third, interviewers tailor their questions to the mental representation and perceptual skills of each particular witness instead of asking all witnesses the same set of questions. Fourth, the witnesses are asked to think about the event in various ways, such as recalling the event in different orders or describing the event in terms of its semantic properties and its sensory properties. Fifth, witnesses are instructed to indicate when they are uncertain about the answer to a question, rather than guess at an answer. Sixth, interviewers warn witnesses not to expose themselves to the media, other witnesses, or any other sources of information that might contaminate the witnesses' memory. In terms of social dynamics, interviewers invest time at the beginning of the interview to develop rapport with the witnesses and throughout the interview, interviewers encourage active witness participation by refraining from interrupting witnesses during their narrative responses and constructing a social dynamic so that witnesses perceive themselves to be the dominant person in the conversation. In terms of communication, interviewers promote extensive, detailed responses by instructing witnesses to tell everything they remember no matter how trivial it may seem. To promote communication, the response formats can be compatible with the witness's

mental representation of the event, such as allowing the witness to draw a sketch of the layout of a room, rather than try to describe the layout in words.

The Cognitive Interview has been used to elicit statements that were examined for credibility in a number of studies (e.g., Hernandez-Fernaud & Alonso-Quecuty, 1997; Kohnken et al., 1995). Indeed, Vrij et al. (2008) found that interviewees who were asked to recall an event in reverse order, one aspect of the cognitive interview, provided more cues to deceit than interviewees who recalled the event naturally. These researchers have established that the Cognitive Interview is a useful credibility assessment tool.

**The Step Wise Interview.** The Step Wise Semi-Structured Interview was developed to help interview children involved in child sexual abuse cases by maximizing recall while minimizing contamination, by avoiding suggestive questions (Yuille, Hunter, & Harvey, 1990; Yuille et al., 1993). This interview method focuses on an uninterrupted free narrative and a higher proportion of open-ended questions than specific/close-ended questions. The main principle of the Step Wise Interview is a funnel approach to questioning: The interviewer begins by asking general questions, then proceeds to specific questions as required, until the subject's memory has been exhausted.

Although the Step Wise Interview was developed to interview children, it has been used successfully with adults, as well (e.g., Cooper, 2005; Porter & Yuille, 1996; Zaparniuk, Yuille, & Taylor, 1995). It has been empirically demonstrated that the Step Wise Interview elicits more unbiased accounts than traditional interview methods (Porter, Yuille, & Bent, 1995). As with the Cognitive Interview, the Step Wise Interview has been established as a successful credibility assessment tool by researchers who have used it to

elicit statements to be assessed for credibility (e.g., Porter & Yuille, 1996; Zaparniuk et al., 1995).

**Component 2: Criteria Based Content Analysis**

The most highly researched aspect of the SVA system is its core component, CBCA, which is reported to be the most widely used veracity assessment technique world-wide (Vrij, 2005). CBCA is the qualitative analysis of a statement's contents. On the basis of the work of Undeutsch (1984) and Arntzen (1983; as cited in Koehnken et al., 1995), Koehnken and Steller (1988; Steller, 1989; Steller & Koehnken, 1989) compiled a list of 19 criteria that had been found to contribute to a statement's credibility. This list included criteria such as logical structure, appropriate amount of details, spontaneous reproduction, contextual embedding, unusual details, unexpected complications, and interactions. Steller and Kohenken (1989) stressed that it is important that statements, but not individuals, are analyzed with regards to their credibility. On some occasions, a generally sincere person may tell a lie, and under certain circumstances a rather deceptive person may tell the truth. The theory behind CBCA is that criteria are more likely to occur in genuine, as opposed to fabricated statements, because the statement provider may not be creative enough or may lack the knowledge to include additional details.

SVA should not be considered a verbal lie detector. The underlying strategy of SVA is a search for signs of truthfulness (Koehnken et al., 1995). The presence of each criterion adds to a statement's credibility, yet their absence does not imply deceptiveness (Lamb, Sternberg, Esplin, Hershkowitz, & Orbach, 1997; Steller, 1989; Yuille, 1989). However, the absence of most of the criteria would raise serious doubts about the credibility of a statement (Seniuk & Yuille, 1996).

When applying CBCA to a statement, raters give a "0" if a criterion is absent in a given statement, "1" if present, and "2" if strongly present (Honts, 1994). According to Vrij (2005), for CBCA scores, an interrater reliability score of .50 or higher could be considered an adequate reliability, scores between .60 and .75 could be considered good, and scores over .75 could be considered excellent. Despite the scoring, CBCA is not a standardized test with set norms suggesting certain interpretations. There are no commonly agreed-upon "decision rules" to determine whether a statement is credible (e.g., Tye, Amato, Honts, Devitt, & Peters, 1999). Thus, CBCA is a complex qualitative procedure. If a memory account is found to be credible, this does not necessarily mean that it was based on true events. Similarly, an account containing few CBCA criteria may have been based on a genuine experience. A composite or total score would be hard to interpret, as certain criteria should be given more weight than others. For example 'logical structure' and an appropriate 'amount of details' are, according to some practitioners, mandatory for credible accounts, whereas 'unusual details' or 'unexpected complications' are optional (cf., Steller & Koehnken, 1989). A similar problem emerges if one attempted to establish a required minimum total score for credible statements. Especially in cases involving children, even a very low CBCA total score could be meaningless depending on age, personality, verbal, and intellectual abilities. Thus, CBCA should be viewed as a semi-standardized approach. Unfortunately, much research has disregarded these fundamental considerations behind this approach and has wrongfully used CBCA in quantitative ways (e.g., Parker & Brown, 2000). The present research examined CBCA in a qualitative manner, using an adapted version of the CBCA criteria, which lists 24 criteria rather than the traditional 19 (Yuille, 2005; see Table 1).

Table 1: CBCA Criteria

| General Characteristics of the Statement |
| --- |
| 1. Coherence |
| 2. Spontaneous Reproduction |
| 3. Appropriate Detail |
| **Specific Contents of the Statement** |
| 4. Contextual Embedding |
| 5. Descriptions of Interactions |
| 6. Reproduction of Conversation |
| 7. Unexpected Complications |
| **Peculiarities of Content** |
| 8. Unusual Details |
| 9. Peripheral Details |
| 10. Accurately Reported Details Misunderstood |
| 11. Related External Associations |
| 12. Accounts of Subjective Mental State |
| 13. Attributions of Another Person's Mental State |
| 14. Details Characteristic of a Particular Act |
| **Motivation Related Contents** |
| 15. Spontaneous Corrections |
| 16. Expressing Insecurities |
| 17. Admitting Lack of Memory |
| 18. Providing Reasons for Lack of Memory |
| 19. Raising Doubts About One's Own Testimony |
| 20. Self-Deprecation |
| 21. Pardoning the Perpetrator |
| **Stylistic Features** |
| 22. Long-winded vs. Issue Oriented |
| 23. Theme Related Changes |
| 24. Rigid Repetition |

This adapted version of the CBCA criteria was chosen to provide a more complete analysis of the participants' statements. The definitions of the criteria were also adapted so that they were more relevant to the present study. That is, the criteria were developed to be used with children in child sexual abuse cases, so some definitions did not apply to adult memories for non-abusive events. For example, Criterion 13, "Attribution of Perpetrator's Mental State" was changed to "Attribution of Another Person's Mental State" because in many of the memories, the participant was not a victim.

**Definitions of the Criteria.** As displayed in Table 1, the CBCA criteria are

divided into five different categories. In the first category, "General Characteristics of the

Statement," a statement possesses Criterion 1, "Coherence," if it makes sense and is

physically possible. Criterion 2, "Spontaneous Reproduction," is present in a statement

that is unstructured, lacks rigidity, and does not have a strict chronology. A statement has

Criterion 3, "Appropriate Detail," if it contains specific descriptions of places, people,

times, and events. Generally, the greater amount of detail, the more credible the

statement. However, in some circumstances, a large number of details may suggest a

noncredible statement. For example, fewer details would be expected in a child's

statement, or in a statement about a mundane event that occurred many years ago and had

been recalled few times since (Yuille, 2005).

Under the second category, "Specific Contents of the Statement," Criterion 4,

"Contextual Embedding," is present in a narrative that is embedded in the matrix of the

person's life. The recall of the event should include information about where and when

the event took place. Criterion 5, "Descriptions of Interactions," refers to reports of

actions and reactions, including non-verbatim reports of conversations. Criterion 6,

"Reproduction of Conversation," is present in a statement that has spontaneous reports of

verbatim dialogue spoken during the incident. A statement contains Criterion 7,

"Unexpected Complications," when the person has spontaneously reported at least one

occurrence of something unexpected during the incident, such as an interruption by a

third party or a complication in the event itself (Yuille, 2005).

In the third category, "Peculiarities of Content," Criterion 8, "Unusual Details,"

refers to spontaneously provided details concerning the event that are realistic but seem

unusual to the analyzer. To meet Criterion 9, "Peripheral Details," the person being interviewed must provide details that are related to the incident in question, but are not a central part of the incident. Criterion 10, "Accurately Reported Details Misunderstood," is present in a statement where a person has attempted to interpret an unfamiliar event or object in terms of his or her own experience, and, while the description is accurate, the interpretation is incorrect. This criterion is unlikely to be present in adults' statements. A narrative possesses Criterion 11, "Related External Associations," when it contains a spontaneous reference to an event related to the account of the incident being examined, but that did not occur within the incident. To meet Criterion 12, "Accounts of Subjective Mental State," the person must have spontaneously reported his or her thoughts, feelings, or physical sensations during the event. Criterion 13, "Attributions of Another Person's Mental State," is present in a statement where a person has spontaneously referred to his or her perceptions of the thoughts, intentions, or emotions of another person during the event. Criterion 14, "Details Characteristic of a Particular Act," refers to the spontaneous description of characteristics that are typical of a certain kind of incident, particularly when the details are not common knowledge (Yuille, 2005).

While Criteria 4 through 14 refer to specific details that may occur in eyewitness accounts of an event, the next seven criteria, in the "Motivation Related Contents" category, are used to draw inferences about the motives of the witness. It is assumed that one may betray a motive through the content of a report (Seniuk & Yuille, 1996). Criterion 15, "Spontaneous Corrections," is present in a narrative when a person has spontaneously offered corrections or additional information to material concerning the event that has previously been provided in the interview. Such corrections suggest a

motive of trying to be honest about the allegation, and, if the correction leads to an improvement in the quality of the account, according to other criteria, the correction may improve the assessed credibility of the allegation. A statement contains Criterion 16, "Expressing Insecurities," when the person expresses some uncertainty about some aspect of the event. To fulfill Criterion 17, "Admitting Lack of Memory," the person must have spontaneously reported a lack of memory or knowledge about some aspect of the event. Criterion 18, "Providing Reasons for Lack of Memory," refers to spontaneous explanations for lack of memory. Criterion 19, "Raising Doubts About One's Own Testimony," is present when the person has spontaneously expressed a concern that all or part of his or her allegation might seem unbelievable. A person giving a false account will be trying to appear credible, and is not likely to raise doubts about his or her statement. A narrative contains Criterion 20, "Self Deprecation," when the person spontaneously describes some aspect of his or her behaviour related to the event as wrong or inappropriate. To meet Criterion 21, "Pardoning the Perpetrator," the person must attempt to excuse the perpetrator. This criterion assumes that the perpetrator is someone other than the person giving the account (Yuille, 2005).

Under the last category, "Stylistic Features," for Criterion 22, "Reporting Style," a long-winded style is found when the person describes aspects of the event not related to the question asked or provides more detail about irrelevant matters than central details. An issue oriented style is found when most of the content relates to the key issues of the account. Generally, an issue oriented style is associated with credibility. Criterion 23, "Theme Related Changes," involves a change in the use of pauses, filled pauses, or clichés, which occur in relation to a shift of topic or theme in the narrative. These pauses

or expressions may serve to provide an opportunity to organize thoughts or avoid answers. They are associated with lack of credibility if they are not characteristic of the speaker's typical speech pattern and if the question asked should not be difficult to answer. Criterion 24, "Rigid Repetition," occurs when the person repeats some elements of the story without any additional details. Repetitions such as this are related to lack of credibility.

As mentioned, there are no established decision rules associated with CBCA to determine whether a statement is credible. However, as a general rule, the first five criteria should be present in any credible statement (Seniuk & Yuille, 1996). That is, the statement should describe a coherent event, be delivered in a spontaneous manner, with an appropriate amount of details, including details about context and the actions and reactions that constitute the core of the statement. None of the following 16 criteria need to be present to support credibility, although, as mentioned previously, the more that are present, the easier it is to provide such support. The presence of the final three criteria, the "Stylistic Features," in a statement, may reduce the credibility of that statement, depending on the circumstances surrounding the presence of these criteria.

### Component 3: The Validity Checklist

The Validity Checklist consists of sets of items referring to the witnesses' characteristics, the interview, and other evidence (for a detailed list, see Yuille, 1988). The statement-related factors include psychological characteristics (e.g., appropriateness of language, appropriateness of knowledge, susceptibility to suggestion), interview characteristics (e.g., adequacy of the interview), and motivational characteristics (e.g., pressure to report, secondary gain). The validity checklist also considers physical

evidence, statements made by other witnesses, the behaviour of the witness, and any other relevant information. All of this information is used to make a final judgment as to whether the account is based on a genuine personal experience. According to Honts (1994), the Validity Checklist is used to consider alternative interpretations to prevent premature conclusions in favour of the statement's credibility. However, there has not been any systematic empirical research on the Validity Checklist.

**Research on SVA**

During the past 15 years, SVA, especially CBCA, has been the focus of a considerable amount of research. There are two common research paradigms: laboratory and field research (cf., Ruby & Brigham, 1997). In most laboratory studies, the participants are asked to describe what they witnessed immediately following a staged event, a film, or a slide show. Alternatively, in most field studies, real life witnesses report crimes, often after a considerable delay. These procedural disparities mean that different storage and retrieval processes are likely to have taken place for accounts elicited in the laboratory and the field.

Both field and laboratory studies indicate that CBCA can differentiate truthful and deceptive statements better than chance. This was found to be true both for children's (e.g., Hershkowitz, 1999; Joffe, 1992; Lamb, Sternberg, Esplin, Hershkowitz, Orbach, & Hovav, 1997; Steller, 1989; Vrij, Akehurst, Soukara, & Bull, 2002; Yuille, 1988) and for adults' (e.g., Koehnken et al., 1995; Landry & Brigham, 1992; Vrij, Edward, & Bull, 2001; Zaparniuk et al., 1995) statements, with accuracy rating ranging from 55% to 90% for trained CBCA coders, and with most studies revealing a truth bias. That is, the procedure usually produces more false positive than false negative errors (e.g., Landry &

Brigham, 1992). Furthermore, CBCA does not have the potential to distinguish unintentional lies and truthful accounts because the former would not have the cognitive and motivational characteristics of a deceptive statement (e.g., Ost, Vrij, Costall, & Bull, 2002). CBCA judgements have been found to be affected by age and verbal ability, irrespective of the truthfulness of the statements (Santtila, Roppola, Runtti, & Niemi, 2000). Lastly, research has shown that training matters: although some practitioners suggest that several days of intense training are necessary to become a reliable CBCA coder (e.g., Koehnken, 2004; Yuille, 1988), even short training sessions have been shown to improve raters' accuracy, as defined as ability to differentiate true statements from false statements (e.g., Steller, 1989; Landry & Brigham, 1992).

Although a growing body of forensic psychology studies has indicated systematic differences between truthful and deceptive accounts of witnesses or alleged victims of crimes, few studies have examined verbal clues to deception in offenders or suspects of crime, partly because it is difficult to gain access to the statements of offenders or suspects. Motivated by the over-reliance on nonverbal deception detection in the suspect context, Porter and Yuille (1996) conducted a simulation study analyzing verbal cues in suspects' statements in an interrogation context. The participants committed a theft, under the pretext of testing the effectiveness of a newly introduced security system in the building. After witnessing the removal of a folder from an office, the participants were assigned to one of four experimental groups: truthful alibi, partial deception, false alibi, and truthful confession. The investigators used the Step Wise Interview Protocol to interrogate the mock suspects and used SVA to evaluate the credibility of the mock suspects' statements. The results revealed that the accounts differed only in terms of three

CBCA criteria: the amount of detail, coherence, and admitting lack of memory. A discriminant analysis correctly classified 78% of both truthful and deceptive accounts. Although CBCA was somewhat successful in this study, the support for SVA in the suspect context is limited, as the majority of CBCA items failed to discriminate between the types of accounts. However, a more recent study that applied CBCA to transcripts of statements referring to a simulated theft of money found more encouraging results. Gödert, Gamer, Rill, and Vossel (2005) found that total CBCA scores of truthful accounts were significantly higher than CBCA scores of fabricated accounts. In the Gödert et al. study, the difference between truthful and fabricated accounts was mainly due to three CBCA criteria: quantity of details, reproduction of conversation, and peripheral details. Three other criteria were also more frequently present in truthful accounts, as compared to fabricated accounts: unusual details, spontaneous corrections, and admitting lack of memory. Interestingly, Criterion 4, Contextual Details, was more frequently present in the fabricated accounts, as compared to the truthful accounts.

Lee, Klaver, and Hart (2008) applied CBCA to the memory reports of a sample of 45 adult male offenders. The participants each gave two narratives: one regarding the circumstances of a past conviction and another about a theft he did not commit. Although no information is provided on the type of interview conducted, participants were instructed that narratives should be no longer than five minutes in length, which is not in line with the interview methods shown to be effective in SVA. Four criteria, contextual embedding, reproduction of conversation, unexpected complications, and attribution of another's mental state, were judged to occur more often in truthful narratives than false

narratives, but one criterion, spontaneous corrections, was judged to occur more often in false narratives than truthful narratives.

Colwell, Hiscock, and Memon (2002) assessed the credibility of statements made by offenders within a maximum-security prison in Texas. The participants viewed a staged theft, and then were assigned to either tell the truth about what they had witnessed, or to lie about it. Trained interviewers interviewed each of the mock witnesses, using a structured interview (based on the Step Wise Interview Protocol), the Cognitive Interview, or the Inferential Interview (the Cognitive Interview, with inferential questions inserted, such as 'Was this an act of violence?'). Although these researchers did not use SVA to examine the credibility of the participants' statements, they did find that deceptive statements were systematically different from honest statements. In particular, they found that honest statements were more coherent, had longer responses, and contained less verbal hedging. This provides support for the Undeutsch hypothesis, as well as for Criterion 1 of CBCA.

DePaulo et al. (2003) conducted a meta-analysis on studies comparing truths and lies told by various types of participants, including witnesses. Included in their sample were three studies that considered deception in criminal suspects. They found significant effect sizes ranging from $d = |.14|$ to $d = |.55|$ for deception cues that were similar to those of SVA: details, talking time, plausibility, logical structure, discrepancy/ambivalence, verbal and vocal involvement, verbal and vocal immediacy, verbal and vocal uncertainty, word and phrase repetitions, spontaneous corrections, admitted lack of memory, and related external associations. Thus, several SVA cues have proven their validity in the

distinction of lies and truths in research conducted with victim, bystander, and crime suspect, and offender populations over the past 35 years.

**Reality Monitoring**

Reality Monitoring (RM) is an alternative verbal credibility assessment technique that has been the subject of many recent investigations (e.g., Barnier, Sharman, McKay, & Sporer, 2005; Bond & Lee, 2005; Granhag, Strömwall, & Landström, 2006; Masip, Sporer, Garrido, & Harrero, 2005; Ost et al., 2002; Santtila, Roppola, & Niemi, 1999; Smeets et al., 2006; Sporer, 1997, 2004; Sporer & Sharman, 2006; Strömwall, Bengtsson, Leander, & Granhag, 2004; Vrij, 2000; Vrij et al., 2001, 2004, 2007). RM is the mental process by which people discriminate whether their memory for an event is real or imagined. Johnson and Raye (1981) proposed that the origin of someone's memories may be known based on the characteristics of those memories. They differentiated between two possible origins of memories: an external origin, based on perceptual processes (i.e., memories of experienced events), and an internal origin, based on reasoning, imagination and thought processes.

Originally developed to distinguish true and confabulated memories (Johnson, 1988), the approach has also been applied to deception detection, as the information people use to determine the veracity of their own memories can also be used to evaluate other people's memories. In the RM approach, the truth is the recollection of something physically experienced and a lie is an internally generated memory. Johnson and Raye (1988) suggested that memories based on truly experienced events would contain more perceptual, contextual, and affective information and less cognitive-based information than memories based on imagined events. For example, memories for perceived events

contain sensory details such as smells and colours or information on spatial arrangement, while memories for imagined events are derived from internal sources, thus containing more cognitive operations such as intentions, planning, and drawing conclusions. Although Johnson's work is primarily concerned with the way people remember an event, not how they describe it (Vrij, 2000), RM provides a theoretical and extensive empirical basis that might be useful in deception detection. Truthful accounts should be more clear, realistic, and vivid, whereas deceptive accounts should contain more cognitive processes and inferences (DePaulo et al., 2003; Vrij, 2000).

Alonso-Quecuty (1992) was the first to apply the RM approach to the detection of deception. She found that truthful statements, told immediately after an event, contained more sensory and contextual information than fabricated statements.

Sporer (1997) published the first formalized list of criteria based on RM. These criteria, based on the Memory Characteristics Questionnaire (MCQ; Johnson, Foley, Suengas, & Raye, 1988), are made up of eight subscales: Clarity, Sensory Experiences, Spatial Information, Time Information, Emotions and Feelings, Reconstructability of the story, Realism, and Cognitive Operations. Other researchers have found similar subscales (McGinnis & Roberts, 1996; Sporer & Sharman, 2006; Suengas & Johnson, 1988). More recently, Takahashi and Shimizu (2007) examined the factor structure of a Japanese translation of the MCQ and also obtained eight factors: Clarity, Sensory Experiences, Time Information, Spatial Information, Retrospective Recollection, Events Before and After, Bizarreness, and Overall Impression. With 1183 participants, Cronbach's alpha was high ($\alpha = .93$), showing good internal consistency within the MCQ. These eight factors map well onto Sporer's eight factors, with Retrospective Recollection being

equivalent to Cognitive Operations, Events Before and After being equivalent to Reconstructability, Bizarreness being somewhat equivalent to Realism, and Overall Impression being equivalent to Emotions and Feelings. It should be noted that, while the MCQ was designed to measure RM, it has also been used to measure metamemory, or knowledge about memory (e.g., Gardiner & Richardson-Klavehn, 2000; Roediger & Marsh, 2003; Roediger, Marsh, & Lee, 2001). Metamemory includes factual knowledge about memory tasks and processes, memory monitoring, memory self-efficacy, and memory-related affect, topics which are covered by the MCQ (Hertzog, 1992).

In the past 10 years, researchers have examined RM's ability to distinguish truthful accounts from deceptive accounts. For example, Ost et al. (2002) found that RM did not distinguish participants who had reported that they had seen a non-existent film of the car crash in which Princess Diana was killed from those who had been asked to imagine having seen the film. Similarly, Porter and Yuille (1996) found that none of the RM criteria they used (frequency of verbal hedges, number of self-references, number of words, frequency of filled pauses) correctly differentiated truthful and deceptive accounts. Alternatively, Bond and Lee (2005) found that RM correctly classified 71% of the true and false statements.

Several researchers have compared CBCA and RM's abilities to distinguish truthful accounts from fabricated accounts (e.g., Granhag et al., 2006; Sporer, 1997, 2004; Strömwall et al., 2004; Vrij et al., 2007; 2004). Sporer (1997) found that the correct classification rate for CBCA was 65%, while for the RM criteria, 71% of all statements were correctly classified. A discriminant analysis including both CBCA and RM criteria correctly classified 79% of the statements. Similarly, Vrij et al. (2004) found overall

correct classification rates of 61% for CBCA and 69% for RM. For repeated events, Granhag et al. (2006) found correct classification rates of 58% for CBCA and 63% for RM, while Strömwall et al. (2004) found correct classification rates of 54% for CBCA, 63% for RM, and 69% for CBCA and RM together.

Some authors have questioned RM's assumption that internally generated memories are equivalent to intentionally distorted reports (e.g., Masip et al., 2005). Other authors (e.g., Masip et al., 2005; Ost et al., 2002; Porter & Yuille, 1996) have noted that the RM predictions concerning the discrimination between truthful statements and lies have not always been supported by the data and appear to be influenced by many variables, such as mode of presentation, preparation and delay, repeated statements, age of witnesses, and individual differences.

In a review of studies that have examined reality monitoring, Masip et al. (2005) concluded that while some RM criteria do not discriminate between truthful and deceptive accounts (e.g., internal information or cognitive operations), others seem quite promising (e.g., contextual, spatial and time information, and realism). These authors called for research assessing the validity of the procedure in field situations where real statements by real witnesses are analyzed.

It is unclear whether CBCA or RM is the preferable approach to verbal credibility assessment. A standard method of scoring or a standard set of criteria to be used has yet to be established for RM, since a variety of methods of scoring RM have been used across studies. This limits its usefulness and makes it difficult to compare results across studies. Despite this limitation, Vrij (2000) has suggested that RM might be more effective in adults because children tend to have difficulties with distinguishing fact from

fantasy. RM has other advantages over CBCA: RM has a solid theoretical foundation while CBCA has stemmed from clinical practice (Masip et al., 2005; Vrij, 2000); RM scoring is easier to teach than CBCA scoring; and RM contains criteria indicative of lying while CBCA is focused on detection of truth (Vrij et al., 2004). Regarding the last point, Vrij (2000) suggested that the RM criterion "cognitive operations" could be useful in differentiating between narrative truths but historical wrongs and truly experienced events, a distinction CBCA cannot perform. However, Sporer and Sharman (2006; Sporer, 2004) have acknowledged that there is little empirical evidence in studies involving deception that any of the RM criteria on the MCQ are more likely to be found in accounts of invented events than self-experienced events, and thus referred to the MCQ scales as 'truth criteria'. Sporer (1997) found five factor-analyzed dimensions underlying both RM and CBCA, suggesting an integrative approach would be possible.

**Investigating Memory Reports in Offenders**

As highlighted by Cooper (2005), utilizing incarcerated offenders as research participants presents several unique opportunities that would not likely exist if any other population was examined. This section summarizes these opportunities, including the opportunity to study different types of violence and the opportunity to study psychopathy, and discusses the offenders' motivation to lie in the present study.

**Types of Violence**

The distinction between instrumental and reactive violence has received much attention from researchers (e.g., Cooper, 2005; Cooper & Yuille, 2007; Cornell et al., 1996; Porter & Woodworth, 2007) in areas of sexual violence, general violence (e.g., Brown & Forth, 1997; Serin, 1991), and domestic violence (e.g., Chase, O'Leary, & Heyman, 2001; Tweed & Dutton, 1998). These dichotomous variables have received a variety of labels, including reactive-proactive (e.g., Hubbard, Dodge, Cillessen, Coie, & Schwartz, 2001; Poulin & Boivin, 2000), hostile-instrumental, impulsive-premeditated, hot blooded-cold blooded (e.g., Bushman & Anderson, 2001), impulsive-instrumental (e.g., Tweed & Dutton, 1998), and expressive-instrumental (e.g., Salfati, 2000), but the underlying meanings are fundamentally the same. Instrumental violence (also referred to as proactive, premeditated, or cold blooded violence) requires planning and is essentially a means to an end (Porter & Woodworth, 2007; Woodworth & Porter, 2002). For example, an individual may use instrumental violence to acquire goods, money, and/or sexual gratification. Conversely, reactive violence (also referred to as hostile, impulsive, hot blooded, or expressive violence) requires real or perceived provocation and typically

occurs in the context of emotional arousal (Berkowitz, 1983, 1990; Cornell et al., 1996; Dutton, 1995; Porter & Woodworth, 2007).

Researchers who have considered how the reactive-instrumental division would impact memory have found that this division is theoretically associated with differential eyewitness recall (e.g., Cooper, 2005; Cooper & Yuille, 2007; Hervé, Cooper, & Yuille, 2002, 2007; Hervé, Cooper, Yuille, & Daylen, 2003; Porter et al., 2001; Swihart, Yuille, & Porter, 1999; Taylor & Kopelman, 1984). Herve et al.'s (2007) biopsychosocial theory of eyewitness memory emphasizes the influence of individual differences in predisposing, precipitating, and perpetuating factors in determining how well an eyewitness recalls an event. That is, memory for perpetrated violence is dependent on the type of violence used and the affect associated with the violence. This theory proposes that, because the motivation for reactive violence is, by definition, internal (e.g., rage, anger) as opposed to the external motivation of instrumental violence (e.g., financial), the affect associated with reactive violence is likely to result from internal (e.g., subjective) as opposed to external (e.g., event-related) sources. Therefore, if an individual was focused on the source of affect during a reactive act of violence (e.g., an internal source such as rage), the individual would likely have relatively poorer memory for the details of the event, in comparison to if the event was instrumental in nature and the source of affect was associated with the event itself. In the latter case, one would expect high quality recall. Moreover, Porter et al. (2001) and Swihart et al. (1999) have suggested that instrumental violence should lead to higher quality recall, as compared to reactive violence, because instrumentally motivated offenders are likely to have fantasized (i.e., rehearsed) about the violence prior to engaging in it.

Cooper (2005; Cooper & Yuille, 2007) directly examined the memory consequences of committing such divergent types of violence. He found that acts of instrumental violence were recalled significantly better in comparison to acts of reactive violence. Participants rated memories for instrumentally violent acts higher for overall memory and vividness, compared to memories for reactively violent acts. Additionally, he found that participants' violent acts for which they had a lack of memory were more likely to be reactively than instrumentally motivated. Considering valence, he found that the reactive acts of violence were experienced with significantly higher levels of negative valence than the instrumental acts of violence, a finding congruent with Hervé et al.'s (2007) theory, which predicted that negative valence could lead to poor memory for violence.

Considering these findings, as well as the biopsychosocial theory, the relationship between the types of violence employed (e.g., reactive vs. instrumental) was investigated in the present research in terms of credibility. Based on the research and theory reviewed above, it was hypothesized that because the participants claimed that their memories for instrumental acts of violence were more vivid and more detailed than their memories for reactive acts of violence (Cooper, 2005), memory reports for instrumental acts of violence would be more likely to be found credible than memory reports for reactive acts of violence, regardless of whether the account was true or fabricated, because more detailed accounts would likely contain more details associated with credibility according to CBCA and RM. The biopsychosocial theory (Hervé et al., 2007) may have specific implications for RM credibility in memory for instrumental violence as compared to memory for reactive violence. That is, according to the biopsychosocial theory, during an

act of instrumental violence, the perpetrator's focus is external, on his or her

surroundings and the victim; during an act of reactive violence, the perpetrator's focus is

internal, on his or her thoughts and feelings. As mentioned, according to RM theory,

credible accounts should contain details related to perceptual processes, such as sounds

and colours, while noncredible accounts should contain details related to thought

processes, such as motivations. If focus during an instrumentally violent event is external,

memories for these events are likely to contain more details related to perceptual

processes than memories for reactively violent events. If focus during a reactively violent

event is internal, memories for these events are likely to contain more details related to

thought processes than memories for instrumentally violent events. These factors, not

related to the credibility of the memory accounts, may make accounts of perpetrated

instrumental violence appear to be more credible than accounts of reactive violence,

according to RM measures.

**Psychopathy**

The construct of psychopathy was operationalized by Hare with the Psychopathy

Checklist (PCL; Hare, 1980) and its revised version, the PCL-R (Hare, 1991, 2003).

Psychopathy is a specific form of personality disorder with a distinctive pattern of

interpersonal (e.g., glib, grandiose, callous, deceitful, manipulative), affective (e.g., lack

of empathy, shallow affect, short-tempered) and behavioral (e.g., parasitic, irresponsible,

impulsive) characteristics (Hart & Dempster, 1997; Hart & Hare, 1997; Hart, Hare, &

Harpur, 1992). Psychopathy, as defined by the PCL-R (Hare, 1991, 2003), forms two

distinct but correlated factors. Factor one describes interpersonal and affective traits, and

factor two describes the socially deviant lifestyle features of the disorder (Hare et al.,

1990; Harpur, Hakstian, & Hare, 1988). Much research has established that psychopaths commit a disproportionate amount of crime, including violent crime (Brown & Forth, 1997; Cornell et al., 1996; Porter & Woodworth, 2006; Serin, 1991; Williamson, Hare, & Wong, 1987; Woodworth & Porter, 2002; for review, see Hart & Dempster, 1997). Psychopaths are over-represented in jails and penitentiaries, compared to other criminals (Hemphill, Templeman, Wong, & Hare, 1998).

Although researchers have estimated that psychopaths make up only 1% of the general population, they make up a significant proportion of the prison population (Blaney & Davis, 1999). Generally, in North American prisons, 15-20% of the populations meet the criteria for psychopathy (Hare, 1991; 2003). The prevalence of psychopathy in Cooper's (2005) study was relatively high: of the participants for whom PCL-R information could be obtained (135 out of 150), 39.3 % met the diagnostic criteria for psychopathy.

Given that one of the defining features of psychopathy is deceitfulness, it is clear why this trait would be of interest in a study looking at credibility. Indeed, the PCL-R (Hare, 1991, 2003) contains two items that consider deception: "pathological lying" and "conning/manipulative." Several theoretical frameworks have been proposed to explain psychopathic deception. For example, Lykken (1995) suggested that affective factors that might inhibit deception in most people, such as anxiety and guilt, may be largely missing in a psychopathic individual. Lykken further suggested that because psychopaths experience less anxiety than nonpsychopaths, they are able to deceive more skilfully than others. Taking this a step further, some researchers have suggested that psychopaths may experience "duping delight" from successfully deceiving others (Ekman, 2002; Porter,

Birt, & Boer, 2001; Raskin & Hare, 1978). Theorists coming from an evolutionary standpoint have proposed that psychopaths may use a reproductive strategy intended to maximize their number of reproductive partners, but invest few resources into caring for the offspring (MacMillan & Kofoed, 1984; Mealey, 1995; Wiebe, 2004). For such a strategy to be effective, an immense amount of deception (i.e., infidelity, denial, false promises) may be required. Mealey (1995) refers to this as a "cheater strategy."

Although these various theories suggest that psychopaths should be better at telling lies than nonpsychopaths, the few empirical studies that have addressed deception in psychopaths have not confirmed this. For example, Raskin and Hare (1978) found that psychopaths were no better than nonpsychopaths at lying successfully during a polygraph examination about a mock crime. The authors attributed physiological arousal in psychopaths to duping delight rather than anxiety. Further, Poythress, Edens, and Watkins (2001) found that individuals with higher levels of psychopathic traits were no better at malingering than individuals with lower levels of psychopathic traits. Lee et al. (2008) found that psychopathic offenders were seven times less likely to be judged credible than nonpsychopathic offenders when telling the truth. Hare, Forth, and Hart (1989) suggested that psychopaths seem to not consider their deceptive verbal behaviour to be inconsistent with the truth and they seem unable or unwilling to monitor what they are saying, with the result that they will sometimes make a series of logically inconsistent statements with aplomb and with seeming unawareness of what they have done. Hare et al. (1989) described a psychopathic inmate who, when asked if he had ever committed a violent crime, replied, "No. But I once had to kill someone." Considering these research and anecdotal findings, Hare (2003) has suggested that while psychopaths may be no

more adept at deception than other offenders, they are more likely to use deception than are nonpsychopathic offenders. The present study is one of the first to consider deception in psychopaths by examining their verbal behaviour. It was expected that more credible accounts will be found among the nonpsychopathic participants, as compared to the psychopathic participants.

As mentioned above, other defining characteristics of psychopathy include callousness, lack of empathy, lack of remorse or guilt, and shallow affect. Psychopaths tend to be unconcerned about the pain and suffering of victims, and display a lack of feelings toward people in general. Hansen (1998) suggested that psychopaths have no emotional understanding of other people. They have no understanding of the fact that other people have their own subjective inner lives. This would suggest that psychopaths would be unlikely to discuss another person's mental state; thus, it was expected that there will be fewer incidences of CBCA Criterion 13, Attribution of Another Person's Mental State, in the reports of psychopaths, as compared to the reports of nonpsychopaths. Psychopaths tend to have low levels of arousal and perform better than nonpsychopaths under stressful conditions. They also tend to find it difficult to anticipate emotional distress as a result of their actions because of an inability to acquire conditioned emotional responses (Hare, 1970). Because psychopaths are less likely to feel aroused or emotional, it was expected that they would be less likely than nonpsychopaths to mention their own emotional states in their memory reports, which is covered by RM's Overall Impression Construct and is part of CBCA's Criterion 12, Accounts of Subjective Mental State.

**Motivation to Lie**

As mentioned in the limitations section of the Cooper (2005) study, although there was no a priori reason to expect that most of the participants would deliberately distort their memories, the fact remains that deception and manipulation are basic components of a criminal lifestyle and, thus, the participants themselves (Doren, 1987; Granhag & Strömwall, 2004). Indeed, one participant remarked to the author following his interview that he did not see the point of the study since it was likely that all of the participants had lied about their experiences. In line with this, Porter and Woodworth (2007) suggested that criminals have greater practice in deception than the average individual, as many have maintained lies for years or decades. As mentioned, psychopaths, in particular, are known for their tendency towards deceitfulness and they may have consciously lied about some of their autobiographical experiences for no other reason than mere duping delight (Ekman, 1992; Peticlerc, Hervé, Hare, & Spidel, 2000). Indeed, Spidel, Hervé, Greaves, Cooper, and Hare (2003) found that psychopaths lie significantly more often than nonpsychopaths and Porter and Woodworth (2007) found that psychopathic murderers were twice as likely as nonpsychopathic murderers to change their version of their homicide offences over the course of incarceration. Similarly, clinical researchers have reported that psychopathy is positively correlated with deception across domains such as sexual deception, general deception, and socially desirable responding (Book, Holden, Starzyk, Wasylkiw, & Edwards, 2006; Seto, Khattar, Lalumière, & Quinsey, 1997). Other researchers have also found that psychopathic individuals frequently engage in lying, deception, and manipulative behaviour (Hare et al., 1989; Kropp & Rogers, 1993; Rogers & Cruise, 2000).

Conversely, Klaver, Lee, and Hart (2007) found that psychopathy was unrelated to self-reported frequency of lying behaviour in incarcerated offenders, although the authors admitted that the self-report nature of the question leaves open the possibility that participants were deceptive about their own lying behaviour.

Another motivation for participants in the present study to lie concerns the interviewers. Most of the interviews (i.e., greater than 90%) were conducted by female undergraduate and graduate students. It is possible that some deception occurred due to the participants' attempts to impress the interviewers, to whom they may have been attracted.

Although it is likely that at least some of the participants deliberately distorted their memories, Cooper (2005) had hoped that the voluntary and confidential nature of his investigation provided a context in which the participants could discuss their past experiences truthfully. It was expected that a proportion of the memory reports in the present study would be deceptive, but that the majority would be judged to be credible.

## Hypotheses

The research questions and hypotheses for the proposed study have been introduced throughout the above literature review. This section provides an overview of the concrete research questions regarding credibility in violent offenders' memory reports.

### Hypothesis 1: CBCA and Consistency with File Information

Based on previously mentioned research that has examined how well CBCA measures credibility, it was expected that memory reports that were rated as credible according to CBCA would be more likely to be consistent with the participants' correctional file information (i.e., details confirmed and partially confirmed in correctional file), as compared to memory reports that were not rated as credible. It was hypothesized that memory reports that were rated as noncredible would be more likely to contain details that are contradictory to information in the offender's file. Related to this, it was expected that the CBCA criteria associated with credibility (i.e., Criteria 1-21) would have higher mean scores for memories that were consistent with file information than for memories that were contradicted by file information. This comparison was between-groups.

### Hypothesis 2: RM and Consistency with File Information

Based on previously mentioned research that has examined how well RM measures credibility, it was expected that memory reports that achieved high RM scores would be more likely to be consistent with the participants' correctional file information (i.e., details confirmed and partially confirmed in correctional file), as compared to memory reports that did not achieve high RM scores. It was hypothesized that memory

reports that achieved low RM scores would be more likely to contain details that were contradictory to information in the offender's file. This comparison was between-groups.

**Hypothesis 3: RM and CBCA**

Based on previously mentioned research that has compared RM and CBCA, it was expected that both credibility assessment approaches would be similarly effective at distinguishing credible accounts from noncredible accounts. In particular, it was expected that RM and CBCA would find a comparable number of credible accounts, and that memory reports judged to be credible according to CBCA would achieve higher RM scores than memory reports judged to be noncredible according to CBCA.

**Hypothesis 4: Type of Violence**

Based on previously mentioned memory research and theory that has shown that IV and RV events lead to differential recall, it was hypothesized that memories for instrumental violence would contain more details, especially details concerning external surroundings, than memories for reactive violence. A greater number of CBCA and RM criteria were expected in a more detailed account, which would lead to a higher likelihood that a memory report of an instrumental act would be found to be credible. That is, it was expected that more reports of instrumental violence would be found to be credible as compared to reports of reactive violence, without considering the truthfulness of these memories. This comparison was within-subjects.

**Hypothesis 5: Psychopathy and Credibility**

Based on previously mentioned research on psychopathy, it was hypothesized that the memory reports of psychopaths would be less likely to be consistent with file information than the memory reports of nonpsychopaths. It was also hypothesized that

fewer credible accounts, according to CBCA and RM, would be found among the participants in the proposed study who had been identified as psychopaths, as compared to the participants who had not been identified as psychopaths. It was further hypothesized that memory reports provided by psychopaths would contain fewer instances of CBCA's Criterion 12, Accounts of Subjective Mental States, especially instances related to affect, and would be scored lower for RM's Overall Impression Construct, as compared to memory reports provided by nonpsychopaths. In line with this, it was hypothesized that memory reports provided by psychopaths would contain fewer instances of CBCA's Criterion 13, Attribution of Another Person's Mental State, as compared to memory reports provided by nonpsychopaths. These comparisons were between-groups.

**Method**

**Participants**

The sample for this study consisted of 150 male incarcerated violent offenders who participated in the Cooper (2005) study. This section summarizes the participants' demographic information, which Cooper collected, and the method used by Cooper to collect the data which were analyzed for the present dissertation.

The participants' mean age was 34.93 (*SD* = 10.58; range: 19-77). Sixty-five percent were Caucasian, 17% were Aboriginal and approximately 18% reported a mixture of ethnic backgrounds. The participants reported an average of 11.25 years of education (*SD* = 2.13; range: 4.5-18) and indicated being incarcerated for a mean of 6.23 (*SD* = 5.88; range: 0.08-27.00) years for their index offences at the time of their interview.

The participants were interviewed at either Mountain Institution (58%) or Kent Institution (42%). Both institutions are federal penitentiaries situated around the Fraser Valley in British Columbia. In Canada, all offenders serving time for at least 2 years are incarcerated in federal penitentiaries governed by the Correctional Service of Canada (CSC). Mountain Institution is a medium-security protective custody institution housing over 500 federal offenders. Kent institution is a maximum-security institution containing approximately 300 offenders who are separated into five distinct populations. Participants were recruited through signs posted throughout the institutions or through word of mouth. To be eligible for participation, participants must have had at least one conviction for a violent or a sexual offence. They were also required to read and comprehend English. Interested participants either contacted the psychology department at their respective

institutions via a written request or approached the researchers in person to schedule an interview session. Participants received a $10 honorarium for their participation. The interviews took place in either a private office in the psychology department or in a private office in the participants' living units. On average, each interview took approximately 5 hours to complete (including the completion of the questionnaires). The interviews took place between 2001 and 2004.

**Memories**

As mentioned, each participant in the Cooper (2005) study was asked to provide memories for five events: an instrumentally violent event (IV), a reactively violent event (RV), a violent event for which he had poor memory (LM), a subjectively disturbing event, and a positive event. The present study did not consider the positive memories or the memories for subjectively disturbing events, but did consider each of the other types of memories. Altogether, 129 IV memories, 139 RV memories, and 84 LM memories were collected. Of these memories, several in each category were lost due to problems with recording, leaving 118, 130, and 80, respectively. All of these memory reports have been transcribed verbatim.

Cooper (2005) grouped the IV memories that the participants provided into the following categories: assaults/fights (54.1%); robberies (19.7%); stabbings/shootings/murders (9.0%); sexual assaults (9.0%); and break and enters and home invasions involving IV (7.4%). Approximately 1% of the memories for acts of IV could not be grouped into these categories. The RV memories the participants provided were classified into the following categories: assaults/fights (79%); stabbings/shootings/murders (15.9%); and break and enters and robberies involving RV

(1.4%). Approximately 4% of the memories for acts of RV could not be grouped into these categories. The LM memories the participants provided were grouped into the following categories: assaults/fights (67.1%); stabbings/shootings/murders (15.9%); sexual assaults (8.5%); and robberies involving violence (2.4%). Approximately 6% of the LM experiences could not be classified into these categories (Cooper, 2005).

**CBCA Training**

Sixteen coders, trained in CBCA, coded the memories for this project. Training included a one-day workshop in CBCA conducted by either a forensic psychologist and two doctoral students, or by two doctoral students. All trainers had been trained in CBCA by Dr. John Yuille, an expert in CBCA. The workshop involved an overview on SVA, including a detailed review of interviewing and CBCA, and practice transcripts, which everyone coded together. Following the workshop, coders were each given nine interview transcripts to code. They individually received feedback from one of the workshop presenters on the first three of these interview transcripts. After they had completed coding the remaining five transcripts, a workshop presenter compared their coding results with the coding results of the workshop presenters. Finally, coders met with a workshop presenter a final time to go over any discrepancies between the coding results, and to answer any questions. The entire process generally took several weeks. If the coder's interrater reliability was adequate (i.e., Pearson's correlation was significant at the .05 level for all practice memories), he or she was given the opportunity to code for this project. All coders who coded the additional nine interview transcripts had reached this level of reliability upon completion of the nine transcripts. Upon completion of training,

the interrater reliability of the raters' coding with practice transcripts ranged from .64 to

.93, with a mean reliability of .78 ($SD = .14$).

**CBCA Coding**

It was proposed that each memory report would be randomly assigned to at least

one coder, but that for 25% of the memory reports (approximately 85 memory reports),

two coders would be assigned to code these memories for CBCA. If interrater reliability,

according to Pearson's correlations, between the two coders was high enough (i.e.,

greater than .85), the remaining memories would have been coded by just one coder.

Unfortunately, interrater reliability was not found to be higher than .85 on all of the

memory reports, so all of the memories were coded by two or three coders. If any

memory had interrrater reliability, according to Pearson correlation, lower than .65, the

two independent coders were to meet to discuss their ratings and to come to an agreement

on any discrepant ratings. Fortunately, none of the memory reports had an interrater

reliability lower than .65. On several occasions, interrater reliability between two coders

on a particular memory report was adequate, but the coders did not agree on the

credibility of the memory report. In these instances, the coders met to discuss the

rationale behind their credibility decision, and were able to quickly come to an

agreement.

All coders were provided with coding sheets that included a list of the 24 CBCA

criteria (see Appendix B). The coders used the procedure described above, scoring each

of the criteria either 2 (strongly present), 1 (present), or 0 (absent). For all of the criteria,

except for Criteria 1 (Coherence), 2 (Spontaneous Reproduction), 3 (Appropriate Detail),

22 (Long-winded vs. Issue Oriented), 23 (Theme Related Changes), and 24 (Rigid

Repetition), the coders indicated on the coding sheet the line numbers (if any) where this criteria was displayed in the transcript. The remaining criteria are global judgments of the entire memory report, so including line numbers would not be appropriate. The line numbers were tallied to come up with a total frequency score for Criteria 4-21 for each memory. Based on the coders' evaluation of the CBCA criterion scores the coders decided whether the memory report was credible or noncredible.

Coders were instructed not to code details that were offered by the participant in response to a direct question. For example, if a participant mentioned a season or a date in response to the question, "Do you remember *when* this happened?" this was not coded as Criterion 4 (Contextual Embedding) because providing a response to a direct question in this way does not add to the statement's credibility. However, if information about the time of the event was provided in the free narrative portion of the interview or in response to a question that did not directly ask about the time of the event, this was coded as Criterion 4. In acknowledgement that most of the questions in the Step Wise Interview method are phrased in the form of "Do you remember X?" coders were instructed not to code any responses of "No" as Criterion 17 (Admitting Lack of Memory) for a similar rationale; although the participant who replies in this way is admitting lack of memory, it is not a spontaneous response.

Related to the hypotheses concerning psychopathy, each instance of Criteria 12 (Subjective Mental State) and 13 (Another Person's Mental State) in each memory report were coded as affective (e.g., "I was anxious"; "He was scared"), cognitive (e.g., "I decided I could take care of him later"; "She thought I had left already"), or physical (e.g., "I was tired"; "She was drunk").

**Reality Monitoring**

Reality Monitoring was assessed with the MCQ (Johnson et al., 1988; see Appendix C). The MCQ is made up of 39 statements that tap a broad range of memory characteristics. When the MCQ has been used in previous research to determine credibility (e.g., Mather, Henkel, & Johnson, 1997; McGinnis & Roberts, 1996; Santtila et al., 1999; Smeets et al., 2006; Sporer, 1997; Sporer & Sharman, 2006; Suengas & Johnson, 1988), the researchers have known the ground truth of all of the statements being evaluated, and a number of statements were known to be untrue. In previous studies, MCQ ratings were compared for true and false reports. Johnson et al. (1988) compared true and fabricated memories on each of the MCQ items, while most other researchers have used subscores based on factor analyses of the MCQ (McGinnis & Roberts, 1997; Santtila et al., 1999; Sporer, 1997; Sporer & Sharman, 2006; Smeets et al., 2006; Suengas & Johnson, 1988). The factor analyses or reliability analyses performed on the MCQ in the latter studies were conducted with sample sizes of 60 to 240 participants. Although a sample size of 200 would be adequate to conduct a factor analysis on a measure with this number of items, for a more stable factor pattern, at least 300 participants should have been used, and a sample size smaller than 200 would be considered too small (DeVellis, 2003). The fact that several studies from various laboratories have found a similar factor structure suggests that the factor structure of the MCQ is relatively stable. The stability of the factor structure is further confirmed by Takahashi and Shimizu's (2007) study to determine the factor structure of the MCQ. This study included over 1000 participants and identified eight factors, which, as mentioned,

were similar to the factors found in other studies that examined the factor structure of the MCQ.

All of the participants in the present study were asked to complete the MCQ in reference to each of their memories as part of Cooper's (2005) study. All participants were not able to complete the MCQ in relation to each of their memories, leaving 123 MCQ scores for IV memories, 138 MCQ scores for RV memories, and 81 MCQ scores for LM memories. The 38 items of the MCQ were considered individually and as part of established factors.

The present study is the first where there is an unknown number of untrue stories. Since there is not an established method of assessing RM credibility using the MCQ, two primary methods were employed in the present study. The first method considered the eight factors or constructs that make up the MCQ. Although the constructs that have been found have all been somewhat different from study to study (in regards to title and the specific items from the MCQ that compose each construct), there have been enough common elements that they can be summarized as: Clarity/Vividness, Sensory Components, Contextual/Spatial Attributes, Time Information, Valence/Overall Impression, Thoughts and Feelings/Cognitive Operations/Retrospective Recollection, Events Before and After, and Bizarreness. According to RM theory, all of these constructs should correlate positively with credibility, except for Thoughts and Feelings or Retrospective Recollection. The constructs defined by Takahashi and Shimizu (2007) were used for the first scoring method. The average scores for each construct that theoretically should correlate positively with credibility (Clarity: items 1-4, 8-11, 33, and 36; Time Information: items 17-22; Overall Impression: items 23, 24, and 28; Sensory

Experiences: items 5-7; Spatial Information: items 15 and 16; Bizarreness: 12-14; and Events Before and After: items 34 and 35) and the average for the reverse-scores of the construct that theoretically should correlate negatively with credibility (Retrospective Recollection: items 25-27, 29-32, 37, and 38) were totalled to form an RM score for each participant (see Appendix C for MCQ items).

The second scoring method also used Takahashi and Shimizu's (2007) constructs, but instead of using the reverse-scores of Retrospective Recollection, the average scores for each construct were summed to form an RM score for each participant. The rationale for this was based on Sporer and Sharman's (2006; Sporer, 2004) observation that there is little empirical evidence in studies involving deception that any of the RM criteria on the MCQ are more likely to be found in accounts of invented events than self-experienced events.

Two further methods of scoring the MCQ data were explored. The scoring method and analyses surrounding these scoring methods can be found in Appendix F.

**Establishing Consistency with File Information**

For all of the participants in this project, an effort was made to establish the consistency of their memories by reviewing their correctional files. By reviewing correctional files, it was possible to establish, or at least infer, whether the offenders' memories that are on record (i.e., any memory that occurred within a correctional institution or any memory that described an offence for which the offender had been charged) happened in the way the offenders described. The purpose of establishing the consistency of the participants' memories was to determine whether the participants

described the events accurately or whether they lied about or misremembered some of the events.

The crime information was coded directly from the official Criminal Profile Report (CPR) included in the participant's institutional file. This document is considered to be an objective and detailed account of what occurred during a serious offence. The CPR is based on the official police reports, crime scene information, autopsy report, witness reports, and court transcripts. The CPR was available for 27 of the participants. For incidents that occurred within a Canadian correctional facility, information was obtained from the Reports of Automated Data Applied to Reintegration (RADAR). The RADAR is a suite of reports that provide information on offenders and their milestones within the correctional planning process, including information on their offences, sentences, programs, and any incidents within an institution that they have been involved in. The RADAR was available for 94 of the participants, including all of the participants for which a CPR was available.

Four trained coders compiled a list of the actions described by the offender in his memory report. All coders were undergraduate volunteers. Training included a half-day workshop that involved familiarization with the CPR and the RADAR, as well as the coding sheet, which was prepared for the purposes of the present study (see Appendix D). Coders were ready to code once they demonstrated an understanding of the procedure through practice exercises. During coding, graduate students were available if any assistance was needed. Although some coders coded for both CBCA and consistency with file information, no coder was assigned to code the same memory for both credibility and consistency with file information, to ensure that knowledge of a memory

report's consistency with correctional file information did not influence credibility judgement. Using the CPR and the RADAR, the coders indicated on the coding sheet whether each detail was confirmed (detail described in the research interview was fully confirmed in the file review), partially confirmed (detail described in the research interview was partially confirmed in the file review), contradictory (detail described in the research interview was not confirmed by file information and information is present but contradicts the offender's version), or not confirmed (detail described in the research interview is not mentioned in the file review). If a participant was unsure about a detail, the coder did not consider whether that detail was consistent with file information to prevent an artificial inflation of the number of details contradicted by file information. For example, if a participant said that the event occurred in May or June, but according the information in the correctional file, the event actually occurred in July, this was not counted as a contradiction because this error would seem to be due to a memory failure rather than a deceptive attempt. Following this assessment, the author came to a conclusion about whether the memory report as a whole was confirmed (most of the memory details described in the research interview were confirmed or partially confirmed in the file review), partially confirmed (a proportion of the details described in the research interview were confirmed or partially confirmed in the file review), partially confirmed with some contradictions (a proportion of the details described in the research interview were confirmed or partially confirmed in the file review, but some critical details from the research interview were contradicted in the file review), contradicted (most details from the research interview were contradicted in the file review), or not confirmed (the event described in the research interview was not present in the file

review). Coding for consistency with file information occurred between April 2006 and February 2007.

**Assessment of Psychopathy**

The PCL-R (Hare, 1991, 2003) is the gold standard in the assessment of psychopathy (Fulero, 1995; Stone, 1995). Psychopathy, as measured by the PCL-R, is characterized by 20 criteria that measure the interpersonal (e.g., manipulative, superficially glib), affective (e.g., callousness, lack of remorse), and behavioral features (e.g., impulsive, criminally versatile) of psychopathy. Factor analysis indicates the PCL-R forms two distinct yet related factors (Hare et al., 1990; Harpur et al., 1988). Factor one assesses interpersonal and affective traits, while factor two measures socially deviant lifestyle features. The PCL-R was designed for the assessment of psychopathy in criminal populations and is usually scored via a review of the offender's correctional files and an interview with the offender. However, if the offender refuses an interview, a file-based assessment has been deemed sufficient to score the PCL-R, although researchers maintain that such a technique may slightly underestimate an individual's score (Hare, 1991, 2003; Harris, Rice, & Quinsey, 1993; Wong, 1988). Individual items are scored on a 3-point scale (i.e., 0, 1, 2) and are summed to yield a total score, ranging from 0 to 40 (Hart et al., 1992; Seto & Barbaree, 1999). The total score represents the degree to which an individual resembles the prototypical psychopath (Hart & Hare, 1997). Although the PCL-R can be used as a dimensional measure, a score of 30 or greater has been demonstrated to categorize or diagnose a psychopath (Hart & Dempster, 1997). That is, although the dimensions of psychopathy vary across individuals, at a certain level of severity, psychopathy can be construed as a discrete taxon or trait (Cooke, 1998). The

PCL-R score is highly reliable over time and has a high level of validity, according to several indices (Fulero, 1995; Hare, 2003; Stone, 1995).

In federal penitentiaries in the Pacific region of Canada, the PCL-R is part of a standard battery of psychological risk assessment inventories utilized on intake (Leis, Motiuk, & Ogloff, 1995). For the majority of participants in the present research (i.e., 135 out of 150), a PCL-R assessment was located in their correctional files.

**Statistical Analysis**

All statistical analyses were conducted in SPSS for Windows 16.0. All assumptions were met for each analysis. An alpha level of .05 was used for all statistical tests, unless otherwise mentioned.

**Interrater Reliability**

As mentioned previously, at least two coders coded each memory for the presence of CBCA criteria. Pearson's correlations were calculated on the CBCA criterion scores for each memory report. In the present study, all of the interrater correlations on the CBCA criterion scores were significant at the .01 level, with correlations ranging from .65 to 1.00. To calculate the mean correlation, the correlations were transformed into Fisher's *z*-scores. The mean of the Fisher's *z*-score was calculated and transformed into a correlation. This resulted in a mean correlation of .86 (standard deviation of .30).

Two coders coded 15% of the Criteria 12 and 13 frequencies for type of mental state (affective, cognitive, and physical). Pearson's correlations were calculated on these scores. All of the interrater correlations were significant at the .01 level, with correlations ranging from .72 to 1.00. To calculate the mean correlation, the correlations were transformed into Fisher's *z*-scores. The mean of the Fisher's *z*-score was calculated and transformed into a correlation. This resulted in a mean correlation of .93 (standard deviation of .08).

**CBCA General Findings: Credible vs. Noncredible Memories**

All of the criteria were present in at least one memory, except for Criterion 10 (Accurately Reported Details Misunderstood). Therefore, Criterion 10 will not be

considered in any of the analyses. An overview of the frequency of each criterion score,

as well as the mean frequency of occurrence of the CBCA criteria scored 1 or 2 is

provided in Table 2.

Table 2: Frequency of each criterion score across all coders and all types of memories
and mean frequency of occurrence of CBCA criteria receiving scores of 1 or 2 (standard
deviations in parentheses).

| Number of memory reports: 328 | | | | |
|---|---|---|---|---|
| | Score | | | Mean Frequency of Occurrence |
| | 0 | 1 | 2 | |
| Criterion 1 | 1 | 28 | 299 | n/a |
| Criterion 2 | 2 | 41 | 285 | n/a |
| Criterion 3 | 2 | 43 | 283 | n/a |
| Criterion 4 | 3 | 69 | 256 | 7.70 (6.63) |
| Criterion 5 | 1 | 30 | 297 | 18.29 (20.11) |
| Criterion 6 | 77 | 124 | 127 | 6.26 (6.82) |
| Criterion 7 | 280 | 39 | 9 | 1.68 (0.96) |
| Criterion 8 | 277 | 42 | 9 | 1.69 (1.46) |
| Criterion 9 | 186 | 115 | 27 | 2.85 (2.50) |
| Criterion 11 | 233 | 68 | 27 | 2.39 (1.80) |
| Criterion 12 | 10 | 100 | 218 | 9.41 (8.18) |
| Criterion 13 | 63 | 161 | 104 | 4.17 (3.39) |
| Criterion 14 | 297 | 25 | 6 | 1.65 (0.92) |
| Criterion 15 | 177 | 135 | 16 | 1.70 (1.11) |
| Criterion 16 | 32 | 127 | 169 | 7.57 (8.68) |
| Criterion 17 | 52 | 137 | 139 | 5.31 (5.27) |
| Criterion 18 | 158 | 149 | 21 | 1.87 (1.30) |
| Criterion 19 | 321 | 7 | 0 | 1.43 (0.54) |
| Criterion 20 | 298 | 29 | 1 | 1.30 (0.47) |
| Criterion 21 | 320 | 5 | 3 | 2.12 (1.64) |
| Criterion 22 | 256 | 25 | 47 | n/a |
| Criterion 23 | 308 | 20 | 0 | n/a |
| Criterion 24 | 321 | 4 | 3 | n/a |

n/a = not applicable because frequency of occurrence data was not collected for Criteria
1, 2, 3, 22, 23, and 24.

Of the 118 IV memories, 116 were found to be credible and 2 were found to be

noncredible. Of the 130 RV memories, 123 were found to be credible and 7 were found

to be noncredible. Of the 80 LM memories, 77 were found to be credible and 3 were

found to be noncredible. No participant had more than one memory that was found to be

noncredible. That is, each of the 12 noncredible memories was told by a different participant.

Circumstances did not allow for the memories to be evenly distributed among the coders. Each coder coded from 2 to 243 memories ($M = 34.6$, $SD = 65.7$), including the coding used for interrater reliability. The 12 noncredible memories were coded by seven of the coders. For these seven coders, one coder judged 6 out of 243 memories to be noncredible (2.5%), one judged 8 out of 217 memories to be noncredible (3.7%), one judged 2 out of 39 memories to be noncredible (5.1%), one judged 3 out of 33 memories to be noncredible (9.1%), one judged 1 out of 25 memories to be noncredible (4.0%), one judged 1 out of 16 to be noncredible (6.3%), and one judged 1 out of 14 memories to be noncredible (7.1%). One of the noncredible memories had been used in training and was coded by a team of graduate students and clinical psychologists.

Since the memory distribution among coders was disproportionate, the distribution of the frequency of the scoring for each CBCA criterion was examined, comparing the ratings of the coder who coded the greatest number of memory reports to the ratings of the coders of the remaining memory reports. These frequencies are presented in Table 3. An examination of these frequencies revealed no problems. That is, this coder was not responsible for a disproportionate number of 0 or 2 ratings for any of the CBCA criteria.

Table 3: Frequency of each CBCA criterion score according to coder, comparing the coder who coded the most transcripts to the other coders.

| | Coder 1 | | | Remaining Coders | | |
|---|---|---|---|---|---|---|
| Number of memory reports | 243 | | | 413 | | |
| Score | 0 | 1 | 2 | 0 | 1 | 2 |
| Criterion 1 | 0 | 17 | 226 | 0 | 33 | 380 |

| | Coder 1 | | | Remaining Coders | | |
|---|---|---|---|---|---|---|
| Score | 0 | 1 | 2 | 0 | 1 | 2 |
| Criterion 2 | 0 | 34 | 209 | 2 | 53 | 358 |
| Criterion 3 | 0 | 31 | 212 | 2 | 58 | 353 |
| Criterion 4 | 0 | 65 | 178 | 3 | 70 | 340 |
| Criterion 5 | 0 | 17 | 226 | 4 | 37 | 372 |
| Criterion 6 | 60 | 99 | 84 | 96 | 142 | 175 |
| Criterion 7 | 212 | 27 | 4 | 345 | 54 | 14 |
| Criterion 8 | 212 | 28 | 3 | 350 | 49 | 14 |
| Criterion 9 | 116 | 110 | 17 | 252 | 132 | 59 |
| Criterion 11 | 193 | 35 | 15 | 294 | 92 | 27 |
| Criterion 12 | 13 | 70 | 160 | 15 | 137 | 261 |
| Criterion 13 | 49 | 131 | 63 | 86 | 207 | 120 |
| Criterion 14 | 220 | 22 | 1 | 371 | 42 | 7 |
| Criterion 15 | 136 | 96 | 11 | 276 | 154 | 23 |
| Criterion 16 | 17 | 59 | 167 | 44 | 162 | 207 |
| Criterion 17 | 44 | 92 | 107 | 72 | 235 | 201 |
| Criterion 18 | 107 | 118 | 18 | 225 | 159 | 27 |
| Criterion 19 | 241 | 1 | 1 | 401 | 11 | 1 |
| Criterion 20 | 226 | 16 | 1 | 371 | 41 | 1 |
| Criterion 21 | 240 | 2 | 1 | 398 | 10 | 5 |
| Criterion 22 | 202 | 15 | 26 | 342 | 29 | 39 |
| Criterion 23 | 240 | 2 | 1 | 393 | 16 | 4 |
| Criterion 24 | 239 | 2 | 2 | 401 | 8 | 4 |

**Memory Length, Age, Rehearsal, and Intoxication**

A variety of methods were used to compare credible and noncredible memories (as judged by CBCA) to determine whether they differed on a number of variables, including length of memory, as judged by number of words in the portion of the interview related to that memory; age of memory in years; the participants' admitted intoxication during the event being recalled (1 = no, 2 = yes); and amount of rehearsal, as judged by the number of times the participant had told others about the event being recalled. First, paired-sample *t*-tests were conducted to compare the CBCA-judged noncredible memories to the credible memories, including only those 12 participants who provided a memory that had been judged to be noncredible. If a participant had more than

one credible memory, the mean of each of these variables for that participant's credible

memories was considered. Thus, the credible memories included four IV memories, three

RV memories, one LM memory, one mean of IV and RV memories, one mean of IV and

LM memories, and two means of RV and LM memories. Information on these variables

was not available for all participants or for all memories. It was decided to keep alpha at

.05 for these analyses despite the possibility of Type I error. If alpha had been lowered, it

would have been less likely that significant results would be found. In this situation,

nonsignificant results would be desirable, so that it could be claimed that these variables

do not affect CBCA-judged credibility. Keeping the alpha at .05 was considered to be a

reasonable trade-off between escalation of Type I error and making it too easy to achieve

a desirable result. All relevant means and information about the number of memories

included in each group can be found in Table 4.

According to the paired-sample $t$-tests, memories judged to be credible according

to CBCA were rehearsed significantly more often than memories judged to be

noncredible according to CBCA, $t(10) = 2.32$, $p < .05$. This difference resulted in a

medium effect size, $d = 0.67$. CBCA-judged credible and noncredible memories did not

significantly differ according to length of memory, age of memory, or intoxication.

Results of the $t$-tests are displayed in Appendix E, Table 35.

Table 4: Mean length of memory (in words), age of memory (in years), intoxication (1 = no, 2 = yes), and number of times rehearsed for CBCA-judged credible vs. noncredible IV, RV, and LM memories (standard deviations in parentheses) for the 12 participants who provided one memory that was judged to be noncredible according to CBCA.

|  | Credible | Noncredible |
|---|---|---|
| Length of memory | N = 12 | |
|  | 4744.46 (3934.64) | 5385.50 (5894.71) |

|  | Credible | Noncredible |
|---|---|---|
| Age of memory | N = 11 | |
|  | 16.83 (13.02) | 12.45 (7.58) |
| Intoxication | N = 10 | |
|  | 1.40 (0.52) | 1.40 (0.52) |
| Rehearsal* | N = 11 | |
|  | **5.91 (6.50)** | **2.41 (3.43)** |

*p < .05

Second, independent samples *t*-tests were used to compare the CBCA-judged credible and noncredible memories to determine whether they differed on the same variables: length of memory, age of memory in years, and amount of rehearsal, considering all of the memories. *Z*-tests of proportions were used to compare the CBCA-judged credible and noncredible memories to determine whether they differed in the participants' admitted intoxication during the event being recalled (1 = no, 2 = yes). As before, information on these variables was not available for all participants or for all memories. All relevant means and information about the number of memories included in each group can be found in Table 5. As before, it was decided to keep alpha at .05 for these analyses despite the possibility of Type I error.

The *t*-tests and *z*-test revealed no significant differences between memories judged credible according to CBCA or memories judged noncredible according to CBCA for any of these variables. Results of these tests can be found in Appendix E, Table 36.

Table 5: Mean length of memory (in words), age of memory (in years), intoxication (1 = no, 2 = yes), and number of times rehearsed for CBCA-judged credible vs. noncredible IV, RV, and LM memories (standard deviations in parentheses), considering all memories.

|  | IV | | RV | | LM | |
|---|---|---|---|---|---|---|
|  | Credible | Noncredible | Credible | Noncredible | Credible | Noncredible |
| Length of memory | N = 115 | N = 2 | N = 121 | N = 7 | N = 75 | N = 3 |
|  | 7882.37 | 12219.00 | 6821.06 | 4877.43 | 6570.16 | 2015.33 |
|  | (6258.53) | (13313.41) | (6438.67) | (3535.98) | (4834.28) | (584.53) |
| Age of memory | N = 110 | N = 2 | N =118 | N = 7 | N = 73 | N = 2 |
|  | 11.51 | 4.75 | 11.86 | 13.13 | 11.27 | 17.75 |
|  | (9.05) | (1.77) | (9.67) | (7.87) | (8.79) | (4.60) |
| Intoxication | N = 97 | N = 2 | N = 109 | N = 7 | N = 66 | N = 2 |
|  | 1.40 | 1.50 | 1.40 | 1.29 | 1.74 | 1.50 |
|  | (0.49) | (0.71) | (0.49) | (0.49) | (0.44) | (0.71) |
| Rehearsal | N = 111 | N = 2 | N = 119 | N = 7 | N = 74 | N = 2 |
|  | 11.40 | 3.25 | 27.06 | 2.71 | 18.41 | 0.50 |
|  | (24.10) | (0.35) | (107.81) | (4.23) | (37.16) | (0.71) |

*Note:* All findings were nonsignificant.

### CBCA Criterion Scores

To determine whether the criterion scores differed according to whether the memories were found to be credible or noncredible according to CBCA, a number of types of comparisons were considered. First, paired-sample *t*-tests were conducted to compare the CBCA-judged noncredible memories to the credible memories, including only those 12 participants who provided a memory that had been judged to be noncredible. If a participant had more than one credible memory, the mean of each of these variables was considered. As before, the credible memories included four IV memories, three RV memories, one LM memory, one mean of IV and RV memories, one mean of IV and LM memories, and two means of RV and LM memories. Criteria 19, 21, and 23 were not present in any of the memories considered in these comparisons, so these criteria were not considered in these comparisons. Alpha was set to .0025 in an effort to control for Type I error (alpha of .05 divided by 20 comparisons). All relevant means and information about the number of memories included in each group can be found in Table 6.

Criterion 3 (Appropriate Detail) was scored significantly higher in CBCA-judged credible memories (M = 1.67) than in noncredible memories (M = 1.08), $t(11) = 3.924$, $p < .0025$. The effect size for this difference was large, $d = 1.01$. The mean criterion scores for credible and noncredible memories did not differ for any of the other CBCA criteria. Results of the tests can be found in Appendix E, Table 37.

Table 6: Mean CBCA criterion scores for credible vs. noncredible memories (standard deviations in parentheses), considering only memories from the 12 participants who had a noncredible memory.

| Criteria | Credible | Noncredible |
|---|---|---|
| 1 | 2.00 (0.00) | 1.50 (0.67) |
| 2 | 1.75 (0.45) | 1.08 (0.67) |
| 3* | **1.67 (0.49)** | **1.08 (0.67)** |
| 4 | 1.58 (0.51) | 1.33 (0.65) |
| 5 | 1.96 (0.14) | 1.42 (0.67) |
| 6 | 1.08 (0.67) | 0.50 (0.67) |
| 7 | 0.17 (0.33) | 0.00 (0.00) |
| 8 | 0.13 (0.31) | 0.17 (0.58) |
| 9 | 0.71 (0.62) | 0.50 (0.67) |
| 11 | 0.25 (0.45) | 0.00 (0.00) |
| 12 | 1.33 (0.62) | 1.17 (0.58) |
| 13 | 0.79 (0.75) | 0.67 (0.49) |
| 14 | 0.04 (0.14) | 0.00 (0.00) |
| 15 | 0.29 (0.45) | 0.50 (0.67) |
| 16 | 1.13 (0.83) | 1.00 (0.74) |
| 17 | 1.00 (0.48) | 1.00 (0.60) |
| 18 | 0.54 (0.50) | 0.50 (0.67) |
| 20 | 0.13 (0.31) | 0.00 (0.00) |
| 22 | 0.00 (0.00) | 0.17 (0.58) |
| 24 | 0.00 (0.00) | 0.25 (0.62) |

\* $p < .0025$

Second, Welch's $t$-tests were conducted to determine whether the criterion scores differed according to whether the memories were found to be credible or noncredible according to CBCA, including all of the memories, divided according to type of memory to ensure that each data point was independent. The Welch test was chosen because the group sizes were unequal and almost half of the variances for the comparisons were

heterogeneous. A number of the mean criterion scores had standard deviations of zero. This means that the ratio of the nonzero variance to the zero variance was undefined, resulting in extremely heterogeneous variances. As of yet, no researchers have conducted Monte Carlo tests to explore the impact of a zero variance for one group on the error rates for Welch's $t$-tests. It would be difficult to make any conclusions about the findings of inferential tests under these conditions. For that reason, no comparisons were made for mean criterion scores where one of the mean scores had a variance of zero. Thus, for IV memories, Criteria 7, 8, 9, 11, 12, 13, 14, 15, 17, 18, 19, 20, 21, 23, and 24 were taken out of the analyses. For RV memories, Criteria 7, 11, 14, 19, 20, 21, 22, and 23 were taken out of the analyses. For LM memories, Criteria 1, 6, 7, 8, 9, 11, 12, 14, 15, 17 18, 19, 20, 21, 22, 23, and 24 were taken out of the analyses. An alpha of .0017 was chosen in an effort to reduce Type I error (an alpha of .05 divided by 29 comparisons). All relevant means are presented in Table 7.

Generally, criterion scores tended to be higher for CBCA-judged credible memories than for noncredible memories, although these differences were not significant. Results of the Welch's $t$-tests for all of the criteria for the IV, RV, and LM memories are included in Appendix E, Table 38.

Table 7: Mean CBCA criterion scores for credible vs. noncredible IV, RV, and LM memories (standard deviations in parentheses).

|  | IV | | RV | | LM | |
|---|---|---|---|---|---|---|
|  | Credible | Noncredible | Credible | Noncredible | Credible | Noncredible |
| N | 116 | 2 | 123 | 7 | 77 | 3 |
| Criterion 1 | 1.96 | 1.50 | 1.93 | 1.29 | 1.86 | 2.00[a] |
|  | (0.20) | (0.71) | (0.25) | (0.76) | (0.35) | (0.00) |
| Criterion 2 | 1.91 | 1.50 | 1.89 | 1.14 | 1.87 | 0.67 |
|  | (0.28) | (0.71) | (0.32) | (0.69) | (0.34) | (0.58) |
| Criterion 3 | 1.92 | 1.50 | 1.86 | 1.14 | 1.87 | 0.67 |
|  | (0.27) | (0.71) | (0.35) | (0.69) | (0.34) | (0.58) |

|  | IV | | RV | | LM | |
|---|---|---|---|---|---|---|
|  | Credible | Noncredible | Credible | Noncredible | Credible | Noncredible |
| Criterion 4 | 1.82 (0.39) | 1.50 (0.71) | 1.76 (0.47) | 1.57 (0.54) | 1.79 (0.41) | 0.67 (0.58) |
| Criterion 5 | 1.93 (0.25) | 1.50 (0.71) | 1.93 (0.25) | 1.57 (0.54) | 1.88 (0.32) | 1.00 (1.00) |
| Criterion 6 | 1.26 (0.76) | 1.50 (0.71) | 1.25 (0.74) | 0.43 (0.54) | 0.94 (0.78) | 0.00[a] (0.00) |
| Criterion 7 | 0.25 (0.51) | 0.00[a] (0.00) | 0.11 (0.38) | 0.00[a] (0.00) | 0.19 (0.46) | 0.00[a] (0.00) |
| Criterion 8 | 0.21 (0.49) | 0.00[a] (0.00) | 0.20 (0.46) | 0.29 (0.76) | 0.12 (0.36) | 0.00[a] (0.00) |
| Criterion 9 | 0.53 (0.65) | 0.00[a] (0.00) | 0.56 (0.64) | 0.86 (0.69) | 0.43 (0.64) | 0.00[a] (0.00) |
| Criterion 11 | 0.44 (0.69) | 0.00[a] (0.00) | 0.39 (0.61) | 0.00[a] (0.00) | 0.30 (0.61) | 0.00[a] (0.00) |
| Criterion 12 | 1.74 (0.48) | 1.00[a] (0.00) | 1.69 (0.48) | 1.29 (0.76) | 1.45 (0.64) | 1.00[a] (0.00) |
| Criterion 13 | 1.19 (0.71) | 1.00[a] (0.00) | 1.20 (0.71) | 0.71 (0.49) | 0.94 (0.70) | 0.33 (0.58) |
| Criterion 14 | 0.19 (0.49) | 0.00[a] (0.00) | 0.09 (0.29) | 0.00[a] (0.00) | 0.05 (0.28) | 0.00[a] (0.00) |
| Criterion 15 | 0.57 (0.61) | 1.00[a] (0.00) | 0.48 (0.61) | 0.57 (0.79) | 0.47 (0.53) | 0.00[a] (0.00) |
| Criterion 16 | 1.45 (0.65) | 1.50 (0.71) | 1.37 (0.69) | 1.14 (0.69) | 1.51 (0.60) | 0.33 0.58) |
| Criterion 17 | 1.21 (0.72) | 1.00[a] (0.00) | 1.19 (0.75) | 1.00 (0.82) | 1.52 (0.62) | 1.00[a] (0.00) |
| Criterion 18 | 0.57 (0.61) | 1.00[a] (0.00) | 0.51 (0.59) | 0.57 (0.79) | 0.73 (0.62) | 0.00[a] (0.00) |
| Criterion 19 | 0.02 (0.13) | 0.00[a] (0.00) | 0.03 (0.18) | 0.00[a] (0.00) | 0.01 (0.11) | 0.00[a] (0.00) |
| Criterion 20 | 0.13 (0.36) | 0.00[a] (0.00) | 0.09 (0.29) | 0.00[a] (0.00) | 0.06 (0.25) | 0.00[a] (0.00) |
| Criterion 21 | 0.06 (0.33) | 0.00[a] (0.00) | 0.02 (0.16) | 0.00[a] (0.00) | 0.01 (0.11) | 0.00[a] (0.00) |
| Criterion 22 | 0.33 (0.71) | 1.00[a] (1.41) | 0.40 (0.74) | 0.00[a] (0.00) | 0.39 (0.73) | 0.00[a] (0.00) |
| Criterion 23 | 0.07 (0.25) | 0.00[a] (0.00) | 0.06 (0.23) | 0.00[a] (0.00) | 0.06 (0.25) | 0.00[a] (0.00) |
| Criterion 24 | 0.03 (0.21) | 0.00[a] (0.00) | 0.02 (0.20) | 0.43 (0.79) | 0.01 (0.11) | 0.00[a] (0.00) |

*Note:* All differences nonsignificant.

[a] indicates that comparisons could not be made for this pair because at least one of the standard deviations was equal to 0.00.

Third, Welch's *t*-tests were conducted to compare CBCA criterion scores of the 12 CBCA-judged noncredible memories to the criterion scores of credible memories of participants who told only memories that were judged to be credible according to CBCA (N = 136). All memories could not be coded for CBCA due to recording problems. If a participant told more than one credible memory, the mean criterion scores were used. Thus, the credible memories included 10 IV memories, 7 RV memories, 4 LM memories, 43 means of IV and RV memories, 4 means of IV and LM memories, 17 means of RV and LM memories, and 51 means of IV, RV, and LM memories. Welch's *t*-tests were used to conduct these comparisons because approximately half of the variances of the two groups being compared were heterogeneous. Whenever a mean criterion score had a variance of zero, that criterion was taken out of the analysis, for the same reasons as mentioned above. Thus, Criteria 7, 11, 14, 19, 20, 21, and 23 were taken out of the analyses. An alpha of .003 was chosen in an effort to reduce Type I error (alpha of .05 divided by 16 comparisons). All relevant means are presented in Table 8.

Two of the CBCA criteria, Criteria 2 (Spontaneous Reproduction) and 3 (Appropriate Detail), had significantly higher scores in credible memories than noncredible memories. Specifically, this test showed significant differences for Criterion 2 ($t(11) = 4.15$, $p < .003$) and Criterion 3 ($t(11) = 4.09$, $p < .003$). Effect sizes for these differences were large: $d = 1.59$ for Criterion 2 and $d = 1.57$ for Criterion 3. The results of the Welch's *t*-tests are presented in Appendix E, Table 39.

Table 8: Mean CBCA criterion scores for memories from credible vs. noncredible groups (standard deviations in parentheses), considering all participants, averaging criterion scores for credible memories across type of memory.

| Criteria | Credible | Noncredible |
|---|---|---|
| n | 136 | 12 |
| 1 | 1.91 (0.25) | 1.50 (0.67) |
| 2* | **1.89 (0.27)** | **1.08 (0.67)** |
| 3* | **1.88 (0.27)** | **1.08 (0.67)** |
| 4 | 1.81 (0.29) | 1.33 (0.65) |
| 5 | 1.92 (0.20) | 1.42 (0.67) |
| 6 | 1.17 (0.61) | 0.50 (0.67) |
| 7 | 0.18 (0.35) | 0.00 (0.00)[a] |
| 8 | 0.19 (0.35) | 0.17 (0.58) |
| 9 | 0.54 (0.55) | 0.50 (0.67) |
| 11 | 0.41 (0.49) | 0.00 (0.00)[a] |
| 12 | 1.67 (0.40) | 1.17 (0.58) |
| 13 | 1.12 (0.55) | 0.67 (0.49) |
| 14 | 0.13 (0.30) | 0.00 (0.00)[a] |
| 15 | 0.50 (0.44) | 0.50 (0.67) |
| 16 | 1.42 (0.54) | 1.00 (0.74) |
| 17 | 1.25 (0.62) | 1.00 (0.60) |
| 18 | 0.61 (0.45) | 0.50 (0.67) |
| 19 | 0.02 (0.12) | 0.00 (0.00)[a] |
| 20 | 0.12 (0.26) | 0.00 (0.00)[a] |
| 21 | 0.04 (0.23) | 0.00 (0.00)[a] |
| 22 | 0.43 (0.70) | 0.17 (0.58) |
| 23 | 0.06 (0.22) | 0.00 (0.00)[a] |
| 24 | 0.03 (0.20) | 0.25 (0.62) |

*$p < .003$

[a] indicates that comparisons could not be made for this pair because at least one of the standard deviations was equal to 0.00.

Criterion scores of credible memories of participants who told CBCA-judged noncredible memories (N = 12) were compared to criterion scores of credible memories of participants who told only memories that were judged to be credible according to CBCA (N = 136) to determine whether there were any differences between credible memories told by participants who told a noncredible memory and credible memories told by participants who told all credible memories. For both groups, if a participant told more than one credible memory, the mean criterion scores were used. Welch's *t*-tests were used to conduct these comparisons because approximately half of the variances of the two groups being compared were heterogeneous. Whenever a mean criterion score

had a variance of zero, that criterion was taken out of the analysis, for the same reasons

mentioned above. Thus, Criteria 1, 19, 21, 22, 23, and 24 were taken out of the analyses.

An alpha of .003 was chosen in an effort to reduce Type I error (alpha of .05 divided by

17 comparisons). All relevant means are presented in Table 9.

There was no general tendency for the criterion scores to be higher in either

group, and none of the differences were significant. The results of the tests are presented

in Appendix E, Table 40.

Table 9: Mean CBCA criterion scores for credible memories from credible vs.
noncredible groups (standard deviations in parentheses).

| Criteria | Group 1 | Group 2 |
|---|---|---|
| n | 136 | 12 |
| 1 | 1.91 (0.25) | 2.00 (0.00)[a] |
| 2 | 1.89 (0.27) | 1.75 (0.45) |
| 3 | 1.88 (0.27) | 1.67 (0.49) |
| 4 | 1.81 (0.29) | 1.58 (0.51) |
| 5 | 1.92 (0.20) | 1.96 (0.14) |
| 6 | 1.17 (0.61) | 1.08 (0.67) |
| 7 | 0.18 (0.35) | 0.17 (0.33) |
| 8 | 0.19 (0.35) | 0.13 (0.31) |
| 9 | 0.54 (0.55) | 0.71 (0.62) |
| 11 | 0.41 (0.49) | 0.25 (0.45) |
| 12 | 1.67 (0.40) | 1.33 (0.62) |
| 13 | 1.12 (0.55) | 0.79 (0.75) |
| 14 | 0.13 (0.30) | 0.04 (0.14) |
| 15 | 0.50 (0.44) | 0.29 (0.45) |
| 16 | 1.42 (0.54) | 1.13 (0.83) |
| 17 | 1.25 (0.62) | 1.00 (0.48) |
| 18 | 0.61 (0.45) | 0.54 (0.50) |
| 19 | 0.02 (0.12) | 0.00 (0.00)[a] |
| 20 | 0.12 (0.26) | 0.13 (0.31) |
| 21 | 0.04 (0.23) | 0.00 (0.00)[a] |
| 22 | 0.43 (0.70) | 0.00 (0.00)[a] |
| 23 | 0.06 (0.22) | 0.00 (0.00)[a] |
| 24 | 0.03 (0.20) | 0.00 (0.00)[a] |

*Note:* All differences nonsignificant.
Group 1 = participants who told only memories judged to be credible; Group 2 =
participants who told one memory judged to be noncredible.

[a] indicates that comparisons could not be made for this pair because at least one of the standard deviations was equal to 0.00.

### Summary for Credible vs. Noncredible Memories Results

CBCA-judged credible and noncredible IV, RV, and LM memories did not differ in terms of length of memory, age of memory, or level of intoxication at time of event. When memories were compared within-participants, memories judged to be credible had been rehearsed more than memories judged to be noncredible. There was no difference between number of times rehearsed for credible and noncredible memories when memories were compared between groups.

When memories were compared within-participants, memories judged to be credible had significantly higher scores for Criterion 3 (Appropriate Detail) than memories judged to be noncredible. When memories were compared between groups, two of the CBCA criteria, Criteria 2 (Spontaneous Reproduction) and 3 (Appropriate Detail), were more strongly present in the credible, as compared to noncredible, violent memories. Criterion scores of the credible memories participants who told one noncredible memory did not differ from the criterion scores of the credible memories of participants who had all memories judged to be credible according to CBCA.

### Hypothesis 1: CBCA and Consistency with File Information

Correctional file information was available for 94 participants (62.7%). Only memories from participants for whom some file information was available were considered in the analyses related to consistency with file information. None of the memories were completely confirmed by file information (i.e., all or most details provided in the interview confirmed or partially confirmed in correctional file) and none of the memories could be completely contradicted by file information because if a

memory could be located in a correctional file, some of the details (i.e., at least the type

of offence and the time of year) had to match. Of the 196 memories for which a file

review was possible, 133 of the memories (67.9%) could not be confirmed by a file

review (i.e., the event described by the participant was not present in the participant's

correctional file), 55 of the memories (28.1%) were partially confirmed (i.e., some details

were confirmed or partially confirmed by file information, with no details contradicted; at

a minimum details related to the type of offence and the time of year had to be partially

confirmed), and 8 of the memories (4.1%) were partially confirmed with some

contradictions (i.e., some details were confirmed or partially confirmed by file

information, but at least one detail was contradicted by file information). For the

memories for which a file review was possible, 190 of the memories were judged to be

credible according to CBCA, and 6 of the memories were judged to be noncredible. All

of the memories that were partially confirmed with a file review were found to be

credible according to CBCA and no confirmation was found for the memories that were

found to be noncredible according to CBCA. The frequency totals are presented in Table

10.

Table 10: Frequencies of memories for CBCA credibility rating (according to coders'
decision) and consistency with file information.

|  | Not Confirmed | Partially Confirmed | Partially Confirmed with Some Contradictions | Totals |
|---|---|---|---|---|
| IV |  |  |  |  |
| Credible | 41 | 27 | 2 | 70 |
| Noncredible | 0 | 0 | 1 | 1 |
| RV |  |  |  |  |
| Credible | 58 | 12 | 4 | 74 |
| Noncredible | 3 | 0 | 0 | 3 |
|  |  |  |  |  |

| | Not Confirmed | Partially Confirmed | Partially Confirmed with Some Contradictions | Totals |
|---|---|---|---|---|
| LM | | | | |
| Credible | 29 | 16 | 1 | 46 |
| Noncredible | 2 | 0 | 0 | 2 |
| Totals | 133 | 55 | 8 | 196 |

In addition to relying on the coders' decision regarding the credibility of each memory report, credibility was also determined using the decision rules proposed by Zaparniuk et al. (1995). That is, according to the first decision rule, an account was considered credible if it contained the first five criteria plus any other two; according to the second decision rule, an account was considered credible if it contained the first three criteria plus any other two; and according to the third decision rule, an account was considered credible if it contained Criterion 2, Spontaneous Reproduction. Using one of the decision rules rather than the coders' decision to determine credibility had very little impact on the totals for CBCA credibility and consistency with file information, with the exception of the first decision rule, where one LM memory that was partially confirmed by file information did not meet the first decision rule's criteria for credibility. The totals for each decision rule are presented in Tables 11, 12, and 13.

Table 11: Frequencies of memories for CBCA credibility rating (according to decision rule 1: Criteria 1-5, plus any other 2) and consistency with file information.

| | Not Confirmed | Partially Confirmed | Partially Confirmed with Some Contradictions | Totals |
|---|---|---|---|---|
| IV | | | | |
| Credible | 41 | 27 | 3 | 71 |
| Noncredible | 0 | 0 | 0 | 0 |
| RV | | | | |
| Credible | 57 | 12 | 4 | 73 |
| Noncredible | 4 | 0 | 0 | 4 |

| | Not Confirmed | Partially Confirmed | Partially Confirmed with Some Contradictions | Totals |
|---|---|---|---|---|
| LM | | | | |
| Credible | 29 | 15 | 1 | 45 |
| Noncredible | 2 | 1 | 0 | 3 |
| Totals | 133 | 55 | 8 | 196 |

Table 12: Frequencies of memories for CBCA credibility rating (according to decision rule 2: Criteria 1-3, plus any other 2) and consistency with file information.

| | Not Confirmed | Partially Confirmed | Partially Confirmed with Some Contradictions | Totals |
|---|---|---|---|---|
| IV | | | | |
| Credible | 41 | 27 | 3 | 71 |
| Noncredible | 0 | 0 | 0 | 0 |
| RV | | | | |
| Credible | 59 | 12 | 4 | 75 |
| Noncredible | 2 | 0 | 0 | 2 |
| LM | | | | |
| Credible | 29 | 16 | 1 | 46 |
| Noncredible | 2 | 0 | 0 | 2 |
| Totals | 133 | 55 | 8 | 196 |

Table 13: Frequencies of memories for CBCA credibility rating (according to decision rule 3: Criterion 2) and consistency with file information.

| | Not Confirmed | Partially Confirmed | Partially Confirmed with Some Contradictions | Totals |
|---|---|---|---|---|
| IV | | | | |
| Credible | 41 | 27 | 3 | 71 |
| Noncredible | 0 | 0 | 0 | 0 |
| RV | | | | |
| Credible | 60 | 12 | 4 | 76 |
| Noncredible | 1 | 0 | 0 | 1 |
| LM | | | | |
| Credible | 29 | 16 | 1 | 46 |
| Noncredible | 2 | 0 | 0 | 2 |
| Totals | 133 | 55 | 8 | 196 |

**CBCA Criterion Scores and Consistency with File Information**

Two methods were used to determine whether the criterion scores differed according to the memories' consistency with correctional file information. First, Welch's *t*-tests were conducted to compare the CBCA criterion scores for partially confirmed memories and memories that were partially confirmed with some contradictions within each memory type. This test was chosen because the group sizes were unequal and almost half of the variances for the comparisons were heterogeneous. Whenever a mean criterion score had a variance of zero, that criterion was taken out of the analysis, for the same reasons as mentioned above. Thus, comparisons could not be made for most of the criteria. For IV memories, Criteria 1, 7, 8, 12, 14, 15, 18, 19, 20, 21, 23, and 24 were taken out of the analyses. For RV memories, Criteria 1, 2, 3, 4 5, 7, 12, 14, 18, 19, 20, 21, 22, 23, and 24 were taken out of the analyses. Comparisons could not be made for LM memories because only one LM memory was partially confirmed with some contradictions. An alpha of .0026 was chosen in an effort to reduce Type I error (i.e., an alpha of .05 divided by 19 comparisons). Specifically, the criterion scores for memories that were partially confirmed by file information were compared to memories that were partially confirmed with contradictions for IV and RV memories.

Although criterion scores tended to be higher for memories that were partially confirmed than for memories that were partially confirmed with some contradictions for both IV and RV memories, these differences were not significant. Mean criterion scores according to consistency with file information are presented on Table 14. Results of the

Welch's *t*-tests to compare the criterion scores across consistency with file information

for all of the criteria for the IV and RV memories are presented in Appendix E, Table 41.

Table 14: Mean CBCA criterion scores for IV, RV, and LM memories that were partially confirmed and partially confirmed with some contradictions (standard deviations in parentheses).

| | IV | | RV | | LM | |
|---|---|---|---|---|---|---|
| | PC | PCwC | PC | PCwC | PC | PCwC |
| N | 27 | 3 | 12 | 4 | 16 | 1[b] |
| Criterion 1 | 1.96 | 2.00[a] | 2.00 | 2.00[a] | 1.69 | 2.00 |
| | (0.19) | (0.00) | (0.00) | (0.00) | (0.48) | |
| Criterion 2 | 1.93 | 1.67 | 2.00 | 2.00[a] | 1.87 | 2.00 |
| | (0.27) | (0.58) | (0.00) | (0.00) | (0.34) | |
| Criterion 3 | 1.93 | 1.67 | 1.92 | 2.00[a] | 1.94 | 2.00 |
| | (0.27) | (0.58) | (0.29) | (0.00) | (0.25) | |
| Criterion 4 | 1.89 | 1.67 | 2.00 | 1.75[a] | 1.87 | 1.00 |
| | (0.32) | (0.58) | (0.00) | (0.50) | (0.34) | |
| Criterion 5 | 1.96 | 1.67 | 2.00 | 2.00[a] | 1.87 | 2.00 |
| | (0.19) | (0.58) | (0.00) | (0.00) | (0.34) | |
| Criterion 6 | 1.44 | 1.67 | 1.58 | 1.25 | 0.81 | 1.00 |
| | (0.75) | (0.58) | (0.67) | (0.50) | (0.75) | |
| Criterion 7 | 0.44 | 0.00[a] | 0.17 | 0.00[a] | 0.25 | 1.00 |
| | (0.58) | (0.00) | (0.39) | (0.00) | (0.58) | |
| Criterion 8 | 0.33 | 0.00[a] | 0.25 | 0.50 | 0.06 | 0.00 |
| | (0.56) | (0.00) | (0.45) | (1.00) | (0.25) | |
| Criterion 9 | 0.89 | 0.33 | 0.58 | 0.50 | 0.63 | 2.00 |
| | (0.80) | (0.58) | (0.67) | (0.58) | (0.81) | |
| Criterion 11 | 0.56 | 0.33 | 0.42 | 0.50 | 0.75 | 0.00 |
| | (0.85) | (0.58) | (0.67) | (0.58) | (0.93) | |
| Criterion 12 | 1.96 | 1.00[a] | 1.83 | 2.00[a] | 1.50 | 2.00 |
| | (0.19) | (0.00) | (0.39) | (0.00) | (0.63) | |
| Criterion 13 | 1.48 | 0.33 | 1.25 | 0.75 | 1.00 | 1.00 |
| | (0.64) | (0.58) | (0.75) | (0.96) | (0.82) | |
| Criterion 14 | 0.33 | 0.00[a] | 0.08 | 0.00[a] | 0.00 | 0.00 |
| | (0.62) | (0.00) | (0.29) | (0.00) | (0.00) | |
| Criterion 15 | 0.78 | 1.00[a] | 0.67 | 0.25 | 0.56 | 1.00 |
| | (0.64) | (0.00) | (0.65) | (0.50) | (0.51) | |
| Criterion 16 | 1.44 | 1.67 | 1.25 | 1.00 | 1.69 | 2.00 |
| | (0.58) | (0.58) | (0.75) | (0.82) | (0.60) | |
| Criterion 17 | 1.37 | 1.67 | 1.25 | 1.25 | 1.62 | 2.00 |
| | (0.74) | (0.58) | (0.75) | (0.50) | (0.62) | |
| Criterion 18 | 0.70 | 1.00[a] | 0.42 | 0.00[a] | 0.75 | 1.00 |
| | (0.67) | (0.00) | (0.52) | (0.00) | (0.68) | |

| | IV | | RV | | LM | |
|---|---|---|---|---|---|---|
| | PC | PCwC | PC | PCwC | PC | PCwC |
| Criterion 19 | 0.00 (0.00) | 0.00[a] (0.00) | 0.08 (0.29) | 0.00[a] (0.00) | 0.00 (0.00) | 0.00 |
| Criterion 20 | 0.22 (0.51) | 0.00[a] (0.00) | 0.17 (0.39) | 0.00[a] (0.00) | 0.13 (0.34) | 0.00 |
| Criterion 21 | 0.07 (0.39) | 0.00[a] (0.00) | 0.00 (0.00) | 0.00[a] (0.00) | 0.00 (0.00) | 0.00 |
| Criterion 22 | 0.15 (0.53) | 0.67 (1.16) | 0.83 (0.84) | 0.00[a] (0.00) | 0.88 (0.89) | 0.00 |
| Criterion 23 | 0.07 (0.27) | 0.00[a] (0.00) | 0.08 (0.29) | 0.00[a] (0.00) | 0.13 (0.34) | 0.00 |
| Criterion 24 | 0.00 (0.00) | 0.33[a] (0.00) | 0.00 (0.00) | 0.00[a] (0.00) | 0.06 (0.25) | 0.00 |

PC = Partially Confirmed; PCwC = Partially Confirmed with Some Contradictions
[a] indicates that comparisons could not be made for this pair because at least one of the standard deviations was equal to 0.00.
[b] Only one memory fell in this position, so standard deviation was not applicable.
*Note:* All comparisons nonsignificant.

Second, Welch's *t*-tests were conducted to compare the CBCA criterion scores for partially confirmed memories and memories that were partially confirmed with some contradictions, considering all of the memories together, regardless of memory type. This comparison was conducted in an effort to increase the power of the comparisons by increasing the number of participants in each group. This comparison was possible because each memory that was partially confirmed with some contradictions had come from a different participant. That is, no participant had more than one memory that was partially confirmed with some contradictions. Three participants had one memory that had been partially confirmed and another that had been partially confirmed with some contradictions. For these participants, the memories that had been partially confirmed were dropped out of the analysis to keep the groups independent. Four participants had more than one memory that had been partially confirmed. For these participants, the mean criterion scores for each of the partially confirmed memories were considered.

Altogether, the partially confirmed memories included 25 IVs, 8 RVs, 14 LMs, 2 means of IV and RVs, 1 mean of IV and LMs, and 1 mean of IV, RV, and LMs. Welch's *t*-tests were used because a number of the comparisons involved groups that had significantly different variances according to Levene's test for equality of variances. Whenever a mean criterion score had a variance of zero, that criterion was taken out of the analysis, for the reasons mentioned above. This resulted in Criteria 1, 14, 19, 20, 21, 22, 23, and 24 being taken out of the analyses. An alpha of .0033 (i.e., an alpha of .05 divided by 15 comparisons) was used in an effort to control for Type I error.

Although there was a general tendency for most of the criterion scores to be higher for memories that were partially confirmed than for memories that were partially confirmed with some contradictions, none of these differences were significant. Means related to these comparisons are presented in Table 15. Results of the Welch's *t*-tests are presented in Appendix E, Table 42.

Table 15: Mean CBCA criterion scores for memories that were partially confirmed and memories that were partially confirmed with some contradictions (standard deviations in parentheses), considering all types of memories together.

|  | Partially Confirmed | Partially Confirmed, Some Contradictions |
|---|---|---|
| n | 48 | 8 |
| Criterion 1 | 1.89 (0.31) | 2.00 (0.00)[a] |
| Criterion 2 | 1.95 (0.21) | 1.88 (0.35) |
| Criterion 3 | 1.95 (0.21) | 1.88 (0.35) |
| Criterion 4 | 1.91 (0.29) | 1.63 (0.52) |
| Criterion 5 | 1.95 (0.21) | 1.88 (0.35) |
| Criterion 6 | 1.26 (0.75) | 1.38 (0.52) |
| Criterion 7 | 0.30 (0.54) | 0.13 (0.35) |
| Criterion 8 | 0.24 (0.47) | 0.25 (0.71) |
| Criterion 9 | 0.69 (0.77) | 0.63 (0.74) |
| Criterion 11 | 0.59 (0.82) | 0.38 (0.52) |
| Criterion 12 | 1.81 (0.41) | 1.63 (0.52) |
| Criterion 13 | 1.32 (0.73) | 0.63 (0.74) |
| Criterion 14 | 0.19 (0.45) | 0.00 (0.00)[a] |

| | Partially Confirmed | Partially Confirmed, Some Contradictions |
|---|---|---|
| Criterion 15 | 0.71 (0.62) | 0.63 (0.52) |
| Criterion 16 | 1.49 (0.58) | 1.38 (0.74) |
| Criterion 17 | 1.38 (0.73) | 1.50 (0.53) |
| Criterion 18 | 0.64 (0.55) | 0.50 (0.53) |
| Criterion 19 | 0.02 (0.14) | 0.00 (0.00)[a] |
| Criterion 20 | 0.18 (0.42) | 0.00 (0.00)[a] |
| Criterion 21 | 0.04 (0.29) | 0.00 (0.00)[a] |
| Criterion 22 | 0.51 (0.80) | 0.00 (0.00)[a] |
| Criterion 23 | 0.10 (0.31) | 0.00 (0.00)[a] |
| Criterion 24 | 0.00 (0.00) | 0.13 (0.35)[a] |

*Note:* All differences nonsignificant.
[a] indicates that comparisons could not be made for this pair because at least one of the standard deviations was equal to 0.00.

The distribution of the frequency of the scoring for each CBCA criterion was examined, comparing the ratings of memory reports that were partially confirmed to the ratings of memory reports that were partially confirmed with some contradictions. These frequencies and the related proportions for each criterion are presented in Table 16.

Table 16: Frequencies and proportions of each CBCA criterion score according to consistency with file information.

| | Partially Confirmed | | | | | | Partially Confirmed, Some Contradictions | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of memory reports | 55 | | | | | | 8 | | | | | |
| Score | 0 | | 1 | | 2 | | 0 | | 1 | | 2 | |
| | *f* | *p* | *f* | *p* | *f* | *p* | *f* | *p* | *f* | *p* | *f* | *p* |
| Criterion 1 | 0 | .00 | 6 | .11 | 49 | .89 | 0 | .00 | 0 | .00 | 8 | 1.00 |
| Criterion 2 | 0 | .00 | 4 | .07 | 51 | .93 | 0 | .00 | 1 | .12 | 7 | .88 |
| Criterion 3 | 0 | .00 | 4 | .07 | 51 | .93 | 0 | .00 | 1 | .12 | 7 | .88 |
| Criterion 4 | 0 | .00 | 5 | .09 | 50 | .91 | 0 | .00 | 3 | .38 | 5 | .63 |
| Criterion 5 | 0 | .00 | 3 | .05 | 52 | .95 | 0 | .00 | 1 | .12 | 7 | .88 |
| Criterion 6 | 11 | .20 | 17 | .31 | 27 | .49 | 0 | .00 | 5 | .62 | 3 | .38 |
| Criterion 7 | 39 | .71 | 14 | .25 | 2 | .04 | 7 | .88 | 1 | .12 | 0 | .00 |
| Criterion 8 | 43 | .78 | 11 | .20 | 1 | .02 | 7 | .88 | 0 | .00 | 1 | .12 |
| Criterion 9 | 25 | .45 | 19 | .35 | 11 | .20 | 4 | .50 | 3 | .38 | 1 | .12 |
| Criterion 11 | 35 | .64 | 8 | .15 | 12 | .21 | 5 | .62 | 3 | .38 | 0 | .00 |
| Criterion 12 | 1 | .02 | 9 | .16 | 45 | .82 | 0 | .00 | 3 | .38 | 5 | .62 |
| Criterion 13 | 9 | .16 | 21 | .38 | 25 | .45 | 4 | .50 | 3 | .38 | 1 | .12 |
| Criterion 14 | 47 | .85 | 6 | .11 | 2 | .04 | 8 | 1.00 | 0 | .00 | 0 | .00 |

| Score | Partially Confirmed | | | | | | Partially Confirmed, Some Contradictions | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | | 1 | | 2 | | 0 | | 1 | | 2 | |
| | *f* | *p* | *f* | *p* | *f* | *p* | *f* | *p* | *f* | *p* | *f* | *p* |
| Criterion 15 | 21 | .38 | 30 | .55 | 4 | .07 | 3 | .38 | 5 | .62 | 0 | .00 |
| Criterion 16 | 4 | .07 | 21 | .38 | 30 | .55 | 1 | .12 | 3 | .38 | 4 | .50 |
| Criterion 17 | 7 | .13 | 18 | .33 | 30 | .55 | 0 | .00 | 4 | .50 | 4 | .50 |
| Criterion 18 | 24 | .44 | 26 | .47 | 5 | .09 | 4 | .50 | 4 | .50 | 0 | .00 |
| Criterion 19 | 54 | .98 | 1 | .02 | 0 | .00 | 8 | 1.00 | 0 | .00 | 0 | .00 |
| Criterion 20 | 46 | .84 | 8 | .14 | 1 | .02 | 8 | 1.00 | 0 | .00 | 0 | .00 |
| Criterion 21 | 54 | .98 | 0 | .00 | 1 | .02 | 8 | 1.00 | 0 | .00 | 0 | .00 |
| Criterion 22 | 37 | .67 | 8 | .15 | 10 | .18 | 7 | .88 | 0 | .00 | 1 | .12 |
| Criterion 23 | 50 | .91 | 5 | .09 | 0 | .00 | 8 | 1.00 | 0 | .00 | 0 | .00 |
| Criterion 24 | 54 | .98 | 1 | .02 | 0 | .00 | 7 | .88 | 1 | .12 | 0 | .00 |

**Summary for CBCA and Consistency with File Information Results**

All memories that were partially confirmed were found to be credible when the CBCA coder's decision was used and no confirmation was found for the memories that were found to be noncredible according to CBCA. Using one of the recommended decision rules to determine credibility produced similar results as when the coder's decision was used, except that one LM memory that was partially confirmed by file information did not meet the first decision rule's criteria for credibility. None of the criterion scores significantly differed according to consistency with file information.

**Hypothesis 2: RM and Consistency with File Information**

As mentioned, several methods were used in the present study to score for RM. In the first method, the average scores for each of Takahashi and Shimizu's (2007) constructs that theoretically should correlate positively with credibility (Clarity, Time Information, Overall Impression, Sensory Experiences, Spatial Information, Bizarreness, and Events Before and After) and the average of the reverse-scores of the construct that theoretically should correlate negatively with credibility (Retrospective Recollection)

were totalled to form an RM score for each participant. For the second RM scoring method, the average scores of each of Takahashi and Shimizu's constructs were summed to form an RM score for each participant, based on Sporer and Sharman's (2006) observation that there is little empirical evidence in studies involving deception that any of the RM criteria on the MCQ are more likely to be found in fabricated accounts, as compared to accounts based on experienced events.

The total scores for each method of scoring for RM were sorted into ascending order and divided into thirds. For one of the RV memories that was partially confirmed by file information, there was no MCQ data available. Due to problems with the audio-recording, there was not enough information for CBCA coding to be possible for one of the IV memories, but there was enough information about the memory for it to be partially confirmed by file information. Although there was a tendency in both scoring methods for memories that were partially confirmed to be sorted into the highest third of the scores, when considering the totals for the bottom and top thirds for each memory type, this difference was not significant according to a $z$-test of proportions. Specifically, for the first scoring method, no significant differences were found between memories in the top third of RM scores and the bottom third of RM scores for the IV ($z = 0.62$, $p > .05$ for partially confirmed memories; $z = 1.43$, $p > .05$ for memories partially confirmed with some contradictions), RV ($z = 1.02$, $p > .05$ for partially confirmed memories; $z = 1.02$, $p > .05$ for memories partially confirmed with some contradictions), or LM ($z = 0.37$, $p > .05$ for partially confirmed memories; $z = 1.01$, $p > .05$ for memories partially confirmed with some contradictions) memories. For the second scoring method, no significant differences were found between memories in the top third and bottom third of RM scores

for the IV ($z = 0.00$, $p > .05$ for partially confirmed memories; $z = 1.43$, $p > .05$ for

memories partially confirmed with some contradictions), RV ($z = 0.79$, $p > .05$ for

partially confirmed memories; $z = 1.02$, $p > .05$ for memories partially confirmed with

some contradictions), or LM ($z = 1.01$, $p > .05$ for partially confirmed memories; $z =$

1.01, $p > .05$ for memories partially confirmed with some contradictions) memories. The

two methods resulted in very similar totals with regards to consistency with file

information. These totals can be found on Tables 17 and 18.

Table 17: Frequencies of memories for consistency with file information and RM score
using the first RM scoring method (sum of mean scores of the constructs, reverse scoring
constructs theoretically believed to be negatively correlated with credible reports).

|  | Lowest 1/3 | | Middle 1/3 | | Highest 1/3 | |
|---|---|---|---|---|---|---|
|  | Partially Confirmed | Some Contradictions | Partially Confirmed | Some Contradictions | Partially Confirmed | Some Contradictions |
| IV | 9 | 2 | 10 | 1 | 9 | 0 |
| RV | 1 | 1 | 7 | 0 | 3 | 3 |
| LM | 4 | 0 | 7 | 0 | 5 | 1 |
| Total | 14 | 3 | 24 | 1 | 17 | 4 |

*Note:* All findings nonsignificant.

Table 18: Frequencies of memories for consistency with file information and RM score
using the second RM scoring method (sum of average scores of the constructs).

|  | Lowest 1/3 | | Middle | | Highest 1/3 | |
|---|---|---|---|---|---|---|
|  | Partially Confirmed | Some Contradictions | Partially Confirmed | Some Contradictions | Partially Confirmed | Some Contradictions |
| IV | 7 | 2 | 12 | 1 | 9 | 0 |
| RV | 2 | 1 | 4 | 0 | 5 | 3 |
| LM | 7 | 0 | 5 | 0 | 4 | 1 |
| Total | 16 | 3 | 21 | 1 | 18 | 4 |

*Note:* All findings nonsignificant.

      As mentioned, two further methods of scoring the MCQ data were explored. The

analyses surrounding these scoring methods can be found in Appendix F.

      A number of comparisons were conducted to determine whether RM total score

differed according to the memories' consistency with file information. First, Welch's *t*-

tests were used to compare RM scores in memories that were partially confirmed to memories that were partially confirmed with some contradictions, considering each memory type separately. Analyses could not be conducted for LM memories because only one LM memory was partially confirmed with some contradictions. Alpha was set to .00625 in an effort to control for Type I error (i.e., alpha of .05 divided by eight comparisons over the four scoring methods). Welch's $t$-tests were used because the group sizes were unequal and a number of the variances were heterogeneous. No significant differences were found for either of the first two scoring methods. Specifically, for the first scoring method (sum of average scores of the constructs, reverse scoring constructs theoretically believed to be negatively correlated with credible reports), no difference was found for the IV memories ($t(2) = 1.38$, $p > .05$) or the RV memories ($t(3) = 0.10$, $p > .05$). For the second scoring method (sum of average scores of the constructs), no difference was found for the IV memories ($t(2) = 1.72$, $p > .05$) or the RV memories ($t(3) = -0.16$, $p > .05$). Relevant means can be found on Table 19.

Table 19: Mean RM scores for the first and second scoring methods (standard deviations in parentheses) for IV, RV, and LM memories according to consistency with file information.

| | | Partially Confirmed | Partially Confirmed, Some Contradictions |
|---|---|---|---|
| IV | n | 28 | 3 |
| | RM Scoring Method 1 | 37.30 (5.09) | 27.10 (12.71) |
| | RM Scoring Method 2 | 39.85 (5.73) | 28.06 (11.71) |
| RV | n | 11 | 4 |
| | RM Scoring Method 1 | 37.22 (5.06) | 36.59 (12.40) |
| | RM Scoring Method 2 | 38.88 (5.53) | 40.09 (14.82) |
| LM | n | 16 | 1 |
| | RM Scoring Method 1 | 26.97 (6.63) | 33.92[a] |
| | RM Scoring Method 2 | 27.50 (6.69) | 36.81[a] |

[a] Only one memory fell in this position, so standard deviation was not applicable.
*Note:* All differences nonsignificant.

Second, Welch's *t*-tests were conducted to compare the RM scores for partially confirmed memories and memories that were partially confirmed with some contradictions, considering all of the memories together, regardless of memory type. As with the CBCA criterion scores, this comparison was conducted in an effort to increase the power of the comparisons by increasing the number of participants in each group. For the three participants who had one memory that had been partially confirmed and another that had been partially confirmed with some contradictions, the memories that had been partially confirmed were dropped out of the analysis to keep the groups independent. For the four participants who had more than one memory that had been partially confirmed, the mean RM scores for each of the partially confirmed memories were considered. As before, altogether, the partially confirmed memories included 25 IVs, 8 RVs, 14 LMs, 2 means of IV and RVs, 1 mean of IV and LMs, and 1 mean of IV, RV, and LMs. Welch's *t*-tests were used because a number of the comparisons involved groups that had significantly different variances according to Levene's test for equality of variances. An alpha of .013 (i.e., an alpha of .05 divided by four comparisons) was used in an effort to control for Type I error.

No significant difference was found between memories that were partially confirmed ($M = 34.23$) and memories that were partially confirmed with some contradictions ($M = 32.70$) for the first scoring method, $t(8) = 0.36$, $p > .05$. No significant difference was found between memories that were partially confirmed ($M = 38.17$) and memories that were partially confirmed with some contradictions ($M = 35.17$) for the second scoring method, $t(12) = 0.57$, $p > .05$. Findings related to the third and fourth scoring methods can be found in Appendix F.

A number of tests were conducted to determine whether RM construct scores differed according to the memories' consistency with file information. First, Welch's *t*-tests were conducted to compare RM construct scores in memories that were partially confirmed to memories that were partially confirmed with some contradictions, considering each memory type separately. Analyses could not be conducted for LM memories because only one LM memory was partially confirmed with some contradictions. Welch's *t*-tests were used because the group sizes were unequal and the variances were heterogeneous. An alpha of .003 (i.e., an alpha of .05 divided by 16 comparisons) was used in an effort to control for Type I error.

For IV and RV memories, none of the construct scores significantly differed according to consistency with file information. The means related to this are presented in Table 20. The results of the Welch's *t*-tests are presented in Appendix E, Table 43.

Table 20: Mean scores for each RM construct (standard deviations in parentheses), across categories of consistency with file information, considering each memory type separately.

|  | Construct | Partially Confirmed | Partially Confirmed, Some Contradictions |
|---|---|---|---|
| IV | n | 28 | 3 |
|  | Clarity | 5.45 (0.85) | 3.47 (1.79) |
|  | Sensory Experiences | 2.81 (1.37) | 1.89 (1.54) |
|  | Spatial Information | 5.64 (1.03) | 3.83 (2.25) |
|  | Time Information | 5.46 (1.01) | 3.83 (1.32) |
|  | Overall Impression | 4.35 (1.12) | 3.56 (1.64) |
|  | Retrospective Recollection | 5.28 (1.06) | 4.48 (0.50) |
|  | Events Before and After | 5.02 (1.52) | 3.33 (2.08) |
|  | Bizarreness | 5.85 (0.93) | 3.67 (2.31) |
| RV | n | 11 | 4 |
|  | Clarity | 5.34 (0.87) | 5.55 (2.12) |
|  | Sensory Experiences | 3.36 (1.79) | 3.75 (2.38) |
|  | Spatial Information | 5.64 (1.00) | 4.88 (2.10) |
|  | Time Information | 4.97 (0.94) | 5.25 (1.77) |
|  | Overall Impression | 3.97 (0.84) | 4.33 (1.98) |

| | Construct | Partially Confirmed | Partially Confirmed, Some Contradictions |
|---|---|---|---|
| RV | Retrospective Recollection | 4.83 (0.67) | 5.75 (1.24) |
| | Events Before and After | 4.68 (2.02) | 5.50 (2.38) |
| | Bizarreness | 6.09 (0.70) | 5.08 (2.08) |
| LM | n | 16 | 1 |
| | Clarity | 3.08 (1.10) | 4.20 [a] |
| | Sensory Experiences | 2.15 (1.12) | 3.33 [a] |
| | Spatial Information | 2.94 (1.41) | 4.00 [a] |
| | Time Information | 4.03 (0.89) | 5.50 [a] |
| | Overall Impression | 3.25 (0.59) | 4.00 [a] |
| | Retrospective Recollection | 4.26 (0.68) | 5.44 [a] |
| | Events Before and After | 3.81 (1.71) | 5.00 [a] |
| | Bizarreness | 3.98 (2.05) | 5.33 [a] |

[a] Only one memory fell in this position, so standard deviation was not applicable.
*Note:* All differences nonsignificant.

Second, independent samples t-tests were conducted to compare the RM construct scores for partially confirmed memories and memories that were partially confirmed with some contradictions, considering all of the memories together, regardless of memory type. Levene's test for equality of variances showed that all variances were homogeneous. For the three participants who had one memory that had been partially confirmed and another that had been partially confirmed with some contradictions, the memories that had been partially confirmed were dropped out of the analysis to keep the groups independent. For the four participants who had more than one memory that had been partially confirmed, the mean RM construct scores for each of the partially confirmed memories were considered. An alpha of .006 (i.e., an alpha of .05 divided by eight comparisons) was used in an effort to control for Type I error.

There were no significant differences between memories that were partially confirmed and memories that were partially confirmed with some contradictions for any of the RM construct scores. Means related to these comparisons are presented in Table

21. Results of the independent samples *t*-tests for these comparisons are presented in

Appendix E, Table 44.

Table 21: Mean scores for each RM construct (standard deviations in parentheses), across categories of consistency with file information, considering different memory types together.

|  | Partially Confirmed | Partially Confirmed, Some Contradictions |
| --- | --- | --- |
| n | 48 | 8 |
| Clarity | 4.79 (1.29) | 4.60 (1.98) |
| Sensory Experiences | 2.63 (1.15) | 3.00 (1.99) |
| Spatial Information | 4.90 (1.53) | 4.38 (1.90) |
| Time Information | 4.98 (1.13) | 4.75 (1.56) |
| Overall Impression | 3.86 (0.99) | 4.00 (1.61) |
| Retrospective Recollection | 4.99 (1.00) | 5.24 (1.07) |
| Events Before and After | 4.80 (1.48) | 4.63 (2.20) |
| Bizarreness | 5.44 (1.44) | 4.58 (1.99) |

*Note:* All comparisons nonsignificant.

**Summary for RM and Consistency with File Information Results**

RM total score did not differ according to the known truth of the memory no

matter which RM scoring method was used and no matter which method of comparison

was used. None of the RM constructs differed according to the memories' consistency

with file information.

**Hypothesis 3: RM and CBCA**

A variety of methods were used to determine whether RM scores differed

according to whether the memory was found to be credible or noncredible according to

CBCA. First, paired-sample *t*-tests were conducted to compare the RM scores of the 12

memories judged to be noncredible according to CBCA to the RM scores to the credible

memories of the same participants. If any of these participants had more than one

credible memory, the mean RM score of the credible memories was used. As before, the

credible memories included four IV memories, three RV memories, one LM memory, one mean of the IV and RV memories, one mean of the IV and LM memories, and two means of the RV and LM memories. An alpha of .0125 was used for these comparisons in an effort to control for Type I error (i.e., an alpha of .05 divided by four comparisons for the four scoring methods). RM scores did not differ according to CBCA credibility for either the first or second scoring method. Specifically, for the first scoring method, there was no significant difference between RM scores for memories that were found credible and memories that were found noncredible according to CBCA, $t(11) = 2.12$, $p > .013$. For the second RM scoring method, there was no difference between RM scores for memories that were found credible and memories that were found noncredible according to CBCA, $t(11) = 2.35$, $p > .013$. Means related to these findings are presented in Table 22.

Table 22: Mean RM scores for the first and second scoring methods (standard deviations in parentheses) according to CBCA credibility, considering only the 12 participants who had one memory judged to be noncredible.

|  | Credible | Noncredible |
|---|---|---|
| RM Scoring Method 1 | 37.73 (6.98) | 32.00 (8.85) |
| RM Scoring Method 2 | 38.97 (8.01) | 31.60 (11.00) |

*Note:* All comparisons nonsignificant.

Second, *t*-tests were conducted to compare RM scores of CBCA-judged credible and noncredible memories within each memory type. Levene's test for equality of variances showed that all variances were homogeneous. An alpha of .004 was used for these comparisons in an effort to control for Type I error (i.e., an alpha of .05 divided by 12 comparisons over the four scoring methods). RM scores did not differ according to CBCA credibility for any of the memory types. Specifically, for the first scoring method, there was no significant difference between RM scores for memories that were found

credible and memories that were found noncredible according to CBCA, for IV memories

($t$(114) = -0.90, $p$ = .370), RV memories ($t$(126) = 0.44, $p$ = .662), or LM memories ($t$(75)

= 2.06, $p$ = .043). For the second RM scoring method, there was no difference between

RM scores for memories that were found credible and memories that were found

noncredible according to CBCA, for IV memories ($t$(114) = -0.43, $p$ = .666), RV

memories ($t$(126) = 0.64, $p$ = .522), or LM memories ($t$(75) = 2.66, $p$ = .010).  Means

related to these findings are presented in Table 23. Information related to the third and

fourth RM scoring methods can be found in Appendix F.

Table 23: Mean RM scores for the first and second scoring methods (standard deviations in parentheses) for IV, RV, and LM memories according to CBCA credibility.

|  |  | Credible | Noncredible |
|---|---|---|---|
|  | n | 114 | 2 |
| IV | RM Scoring Method 1 | 36.83 (6.13) | 40.75 (2.95) |
|  | RM Scoring Method 2 | 38.56 (7.11) | 40.75 (2.95) |
|  | n | 121 | 7 |
| RV | RM Scoring Method 1 | 35.59 (6.29) | 34.51 (6.88) |
|  | RM Scoring Method 2 | 36.75 (7.73) | 34.80 (8.99) |
|  | n | 74 | 3 |
| LM | RM Scoring Method 1 | 29.23 (6.45) | 21.45 (5.64) |
|  | RM Scoring Method 2 | 29.53 (7.39) | 18.04 (4.80) |

*Note:* All comparisons nonsignificant.

A variety of methods were used to determine whether RM construct scores

differed according to whether the memory was found to be credible or noncredible

according to CBCA. First, paired-sample $t$-tests were conducted to compare the RM

construct scores of the 12 memories judged to be noncredible according to CBCA to the

RM construct scores of the credible memories of the same participants. If any of these

participants had more than one credible memory, the mean RM scores were used. As

before, the credible memories included four IV memories, three RV memories, one LM

memory, one mean of IV and RV memories, one mean of IV and LM memories, and two

means of RV and LM memories. An alpha of .006 was used for these comparisons in an effort to control for Type I error (i.e., an alpha of .05 divided by eight comparisons). RM construct scores did not differ according to CBCA credibility judgment. Means related to these findings are presented in Table 24. The results of the paired-sample $t$-tests are presented in Appendix E, Table 45.

Table 24: Mean scores for each RM construct (standard deviations in parentheses), according to CBCA credibility, considering only the 12 participants who provided one memory that was judged to be noncredible.

| RM Construct | Credible | Noncredible |
|---|---|---|
| Clarity | 5.13 (1.21) | 3.92 (1.81) |
| Sensory Experiences | 3.90 (1.82) | 2.78 (1.25) |
| Spatial Information | 5.67 (1.10) | 4.33 (2.27) |
| Time Information | 4.76 (1.58) | 4.08 (1.91) |
| Overall Impression | 4.38 (0.97) | 3.89 (0.84) |
| Retrospective Recollection | 4.48 (0.95) | 3.66 (1.46) |
| Events Before and After | 3.94 (2.15) | 3.67 (2.02) |
| Bizarreness | 5.97 (1.07) | 5.28 (1.66) |

*Note:* All comparisons nonsignificant.

Second, Welch's $t$-tests were conducted to determine whether RM construct scores differed according to whether the memory was found to be credible or noncredible according to CBCA, within memory type. This test was used because Levene's test for equality of variances showed that a number of the variances were heterogeneous. Whenever a mean criterion score had a variance of zero, that criterion was taken out of the analysis, for the same reasons as mentioned above. Thus, for IV memories, the Retrospective Recollection Construct score was removed from the analyses. In an effort to control for Type I error, an alpha of .002 was used (i.e., an alpha of .05 divided by 23 comparisons).

These comparisons resulted in no significant differences for RM construct scores across CBCA-judged credibility. For RV and LM memories, CBCA-judged credible

memories generally tended to have higher RM construct scores than CBCA-judged

noncredible memories. For IV memories, the means of the RM construct scores did not

follow a clear pattern for CBCA-judged credibility. The means related to this are

presented in Table 25. The results of the Welch's *t*-tests are presented in Appendix E,

Table 46.

Table 25: Mean scores for each RM construct (standard deviations in parentheses), according to CBCA credibility, considering all memories.

|  | Construct | Credible | Noncredible |
|---|---|---|---|
| IV | n | 114 | 2 |
|  | Clarity | 5.28 (1.13) | 5.50 (0.71) |
|  | Sensory Experiences | 2.97 (1.63) | 4.50 (1.18) |
|  | Spatial Information | 5.49 (1.29) | 5.75 (0.35) |
|  | Time Information | 4.93 (1.26) | 5.00 (0.47) |
|  | Overall Impression | 4.30 (1.18) | 4.33 (0.47) |
|  | Retrospective Recollection | 4.87 (1.09) | 4.00 (0.00)[a] |
|  | Events Before and After | 4.89 (1.53) | 6.00 (1.41) |
|  | Bizarreness | 5.84 (1.07) | 5.67 (0.94) |
| RV | n | 123 | 7 |
|  | Clarity | 4.92 (1.18) | 4.51 (1.27) |
|  | Sensory Experiences | 3.05 (1.56) | 2.95 (0.59) |
|  | Spatial Information | 5.21 (1.44) | 4.93 (2.24) |
|  | Time Information | 4.48 (1.32) | 4.93 (1.42) |
|  | Overall Impression | 4.23 (1.22) | 3.86 (0.92) |
|  | Retrospective Recollection | 4.58 (1.13) | 4.14 (1.61) |
|  | Events Before and After | 4.48 (1.91) | 3.71 (1.82) |
|  | Bizarreness | 5.78 (1.04) | 5.76 (1.58) |
| LM | n | 76 | 3 |
|  | Clarity | 3.38 (1.27) | 1.47 (0.40) |
|  | Sensory Experiences | 2.26 (1.31) | 1.22 (0.38) |
|  | Spatial Information | 3.69 (1.70) | 2.00 (1.32) |
|  | Time Information | 3.96 (1.35) | 1.50 (0.73) |
|  | Overall Impression | 3.66 (0.96) | 3.67 (1.00) |
|  | Retrospective Recollection | 4.15 (1.09) | 2.30 (0.42) |
|  | Events Before and After | 3.80 (1.65) | 2.00 (1.32) |
|  | Bizarreness | 4.62 (1.60) | 3.89 (1.84) |

*Note:* All comparisons nonsignificant.
[a] indicates that comparisons could not be made for this pair because at least one of the standard deviations was equal to 0.00.

**Summary for RM and CBCA Results**

RM scores were not significantly different for memories that were found to be credible or noncredible according to CBCA whether the comparisons were within-participants or between groups. RM construct scores did not significantly differ according to CBCA credibility judgment, no matter whether the comparisons were within-participants or between groups for each type of memory.

**Hypothesis 4: Credibility and Type of Violence**

**Type of Violence and Consistency with File Information**

Of the 71 IV memories for which some file information for those participants was available, 41 were not confirmed, 27 were partially confirmed, and 3 were partially confirmed with some contradictions. Of the 77 RV memories for which some file information was available, 61 were not confirmed, 12 were partially confirmed, and 4 were partially confirmed with some contradictions. McNemar's test for correlated proportions was conducted to determine whether the proportion of partially confirmed IV memories differed from the proportion of partially confirmed RV memories, including only memories that had been partially confirmed or not confirmed, including only participants who had provided both an IV and an RV memory, and including only participants for which file consistency information was available for more than one memory. This test showed that IV memories were significantly more likely to be partially confirmed than RV memories, $\chi^2(1) = 7.54$, $p < .01$. This relationship remained significant when only participants whose sole source of file information was the RADAR, $\chi^2(1) = 5.26$, $p < .05$. These frequencies are presented in Table 26.

Table 26: Frequencies of memories for consistency with file information and type of violence, considering both participants whose sources of information was the CPR and RADAR, as well as participants whose only source of file information was the RADAR.

*p < .05

|  |  | CPR & RADAR | | | RADAR only | | |
|---|---|---|---|---|---|---|---|
|  |  | RV | |  | RV | |  |
|  |  | Not Confirmed | Partially Confirmed | Totals | Not Confirmed | Partially Confirmed | Totals |
| IV | Not Confirmed | 28 | 6 | 34 | 19 | 6 | 25 |
|  | Partially Confirmed* | 20 | 3 | 23 | 17 | 2 | 19 |

**CBCA and Type of Violence**

Overall, two out of 120 IV memories and seven out of 134 RV memories were found to be not credible according to CBCA. Considering only participants who had both an IV and an RV memory, 97 participants had IV and RV memories judged to be credible according to CBCA, 5 participants had IV memories that were judged to be credible and RV memories that were judged to be noncredible according to CBCA, and 2 participants had IV memories that were judged to be noncredible and RV memories that were judged to be credible according to CBCA. According to a McNemar test, this difference was not significant, $\chi^2 (1) = 1.29, p > .05$.

Paired sample *t*-tests were conducted on the CBCA criterion scores to determine whether there were any differences in criterion scores for IV and RV memories. An alpha of .002 was used in an effort to control for Type I error (i.e., an alpha of .05 divided by 23 comparisons). Although criterion scores for IV memories tended to be higher than criterion scores for RV memories, none of the differences were significant. Relevant means can be found on Table 27. Results of the paired sample *t*-tests are presented in Appendix E, Table 47.

Table 27: Mean score and frequency of occurrence for each CBCA criterion for IV vs. RV memories (standard deviations in parentheses).

|  | Mean Criterion Scores | |
|---|---|---|
|  | IV | RV |
| Criterion 1 | 1.95 (0.22) | 1.90 (0.33) |
| Criterion 2 | 1.91 (0.29) | 1.85 (0.38) |
| Criterion 3 | 1.92 (0.28) | 1.82 (0.40) |
| Criterion 4 | 1.81 (0.39) | 1.75 (0.47) |
| Criterion 5 | 1.92 (0.27) | 1.92 (0.28) |
| Criterion 6 | 1.26 (0.76) | 1.21 (0.75) |
| Criterion 7 | 0.25 (0.51) | 0.10 (0.37) |
| Criterion 8 | 0.20 (0.48) | 0.21 (0.48) |
| Criterion 9 | 0.52 (0.65) | 0.58 (0.65) |
| Criterion 11 | 0.43 (0.69) | 0.37 (0.60) |
| Criterion 12 | 1.73 (0.48) | 1.67 (0.50) |
| Criterion 13 | 1.19 (0.70) | 1.18 (0.71) |
| Criterion 14 | 0.19 (0.49) | 0.08 (0.28) |
| Criterion 15 | 0.58 (0.61) | 0.48 (0.61) |
| Criterion 16 | 1.45 (0.65) | 1.36 (0.69) |
| Criterion 17 | 1.20 (0.71) | 1.18 (0.75) |
| Criterion 18 | 0.58 (0.61) | 0.52 (0.60) |
| Criterion 19 | 0.02 (0.13) | 0.03 (0.17) |
| Criterion 20 | 0.13 (0.36) | 0.08 (0.28) |
| Criterion 21 | 0.06 (0.33) | 0.02 (0.15) |
| Criterion 22 | 0.34 (0.72) | 0.38 (0.73) |
| Criterion 23 | 0.07 (0.25) | 0.05 (0.23) |
| Criterion 24 | 0.03 (0.21) | 0.05 (0.28) |

*Note:* All differences nonsignificant.

### RM and Type of Violence

Paired Sample *t*-tests were conducted on the RM scores, considering both scoring methods, for the IV and RV memories to determine whether the two types of memories differed in terms of RM scores. Alpha was set to .013 in an effort to control for Type I error (i.e., an alpha of .05 divided by four comparisons). For the first scoring method (sum of average scores of the constructs, reverse scoring constructs theoretically believed to be negatively correlated with credible reports) IV memories had significantly higher RM scores ($M = 37.34$) as compared to RV memories ($M = 35.99$), $t(115) = 2.55$, $p <$

.013. The effect size for this difference was small, $d = 0.22$. For the second scoring

method (sum of average scores of the constructs), IV memories had significantly higher

RM scores ($M = 39.02$) as compared to RV memories ($M = 37.19$), $t(115) = 2.86$, $p <$

.013. The effect size for this difference was small, $d = 0.25$. Results related to the third

and fourth RM scoring methods are presented in Appendix F.

Paired Sample $t$-tests were conducted to determine whether RM construct scores

differed according to type of violence. Alpha was set to .006 in an effort to control for

Type I error (i.e., an alpha of .05 divided by eight comparisons). The means of the Time

Information construct ($t(115) = 2.983$, $p < .006$), and the Events Before and After

construct ($t(115) = 2.804$, $p < .006$) were significantly higher for IV memories than for

RV memories. The effect sizes for this difference were small: $d = 0.33$ for the time

information construct and $d = 0.26$ for the events before and after construct. The means

related to this are presented in Table 28. The results of the $t$-tests are presented in

Appendix E, Table 48.

Table 28: Mean scores for each RM construct (standard deviations in parentheses), for IV
and RV memories.

| Construct | IV | RV |
|---|---|---|
| Clarity | 5.32 (1.08) | 5.02 (1.14) |
| Sensory Experiences | 3.11 (1.66) | 3.20 (1.57) |
| Spatial Information | 5.56 (1.22) | 5.31 (1.39) |
| Time Information* | **4.98 (1.26)** | **4.55 (1.32)** |
| Overall Impression | 4.34 (1.19) | 4.18 (1.23) |
| Retrospective Recollection | 4.84 (1.05) | 4.60 (1.14) |
| Events Before and After* | **4.97 (1.49)** | **4.53 (1.87)** |
| Bizarreness | 5.91 (0.95) | 5.80 (1.06) |

*$p < 0.006$

**Summary for Credibility and Type of Violence Results**

IV memories were more likely to be partially confirmed and less likely to be not confirmed, as compared to RV memories. IV and RV memories did not differ for number of credible vs. noncredible memories. None of the mean CBCA criterion scores differed significantly in IV, as compared to the RV memories.

IV memories had significantly higher RM scores, as compared to RV memories for the first, second, and fourth RM scoring methods. Two RM constructs (i.e., Time Information and Events Before and After) had significantly higher mean scores in IV, as compared to RV memories.

**Hypothesis 5: Verbal Credibility and Psychopathy**

**Psychopathy and Consistency with File Information**

For the following analyses, unless otherwise mentioned, a psychopath is defined as a participant who scored 30 or above on the PCL-R and a nonpsychopath is defined as a participant who scored less than 30 on the PCL-R (Hare, 2003). Altogether, 18 of the memories of psychopaths and 35 of the memories of nonpsychopaths were partially confirmed by correctional file information; 5 of the memories of psychopaths and 2 of the memories of nonpsychopaths were partially confirmed with some contradictions. These frequencies were compared separately for IV, RV, and LM memories. Alpha was set to .017 in an effort to control for Type I error (i.e., an alpha of .05 divided by three comparisons). None of the differences were significant: $\chi^2(1) = 0.002$, p > .017 for IV memories, with Yates correction; $\chi^2(1) = 2.026$, p > .017 for RV memories, with Yates correction; and $\chi^2(1) = 0.071$, p > .017 for LM memories with Yates correction. Relevant frequencies are presented in Table 29.

Table 29: Frequency of memory reports determined to be not confirmed, partially confirmed, and partially confirmed with some contradictions for psychopaths and nonpsychopaths for IV, RV, and LM memories.

|  | Psychopath? | NC | PC | PCwSC | Totals[a] |
|---|---|---|---|---|---|
| IV | Yes | 13 | 9 | 1 | 23 |
|  | No | 26 | 17 | 2 | 45 |
| RV | Yes | 23 | 4 | 3 | 30 |
|  | No | 35 | 8 | 0 | 43 |
| LM | Yes | 11 | 5 | 1 | 17 |
|  | No | 18 | 10 | 0 | 28 |

NC = Not Confirmed, PC = Partially Confirmed, PCwSC = Partially Confirmed with Some Contradictions
[a]Totals refers to total number of memory reports for psychopaths and nonpsychopaths in each memory type who had provided that type of memory, and for whom PCL-R and file information data was available.
Note: All comparisons nonsignificant.

## Psychopathy and CBCA

Altogether, 106 of the memories of psychopaths and 184 of the memories of nonpsychopaths were judged to be credible according to CBCA; 8 of the memories of psychopaths and 4 of the memories of nonpsychopaths were judged to be noncredible according to CBCA. These frequencies were compared separately for IV, RV, and LM memories. Alpha was set to .017 in an effort to control for Type I error (i.e., an alpha of .05 divided by three comparisons). None of the differences were significant: $\chi^2(1) = 0.10$, $p > .017$ for IV memories, with Yates correction; $\chi^2(1) = 4.52$, $p > .017$ for RV memories, with Yates correction; and $\chi^2(1) = 0.17$, $p > .017$ for LM memories with Yates correction. Relevant frequencies are presented in Table 30.

Table 30: Frequency of memory reports determined to be credible and noncredible according to CBCA for psychopaths and nonpsychopaths for IV, RV, and LM memories.

|  | Psychopath? | Credible | Noncredible |
|---|---|---|---|
| IV | Yes | 39 | 0 |
|  | No | 69 | 2 |
| RV | Yes | 42 | 6 |
|  | No | 70 | 1 |

|     | Psychopath? | Credible | Noncredible |
| --- | --- | --- | --- |
| LM | Yes | 25 | 2 |
|     | No | 45 | 1 |

Note: All comparisons nonsignificant.

For the analyses related to the CBCA criterion scores and psychopathy, a psychopath was defined as any participant who scored 30 or above on the PCL-R and a nonpsychopath was defined as any participant who scored 20 or below on the PCL-R. This revised distinction was implemented to make the two groups of participants more distinct from each other, in relation to the psychopathy construct. Analyses were conducted on only the memories found to be credible according to CBCA to reduce the possibility of extraneous variables related to credibility affecting the comparisons. Welch's *t*-tests were conducted to compare the CBCA criterion scores of psychopaths and nonpsychopaths, separately for IV, RV, and LM memories. This test was used because almost half of the comparisons involved heterogeneous variances, as assessed by Levene's test for equality of variances. Whenever a mean criterion score had a variance of zero, that criterion was taken out of the analysis, for the reasons mentioned above. This resulted in Criteria 1, 3, 19, and 24 being taken out of the analyses for IV memories; Criteria 1, 21, and 24 being taken out of the analyses for RV memories; and Criteria 1, 2, 3, 14, 19, 21, and 24 being taken out of the analyses for LM memories. An alpha of .001 was used in an effort to control for Type I error. None of the criterion scores significantly differed according to psychopathy for IV, RV, or LM memories. Means related to these comparisons are presented in Table 31. The results of the Welch's *t*-tests are presented in Appendix E, Table 49.

Table 31: Mean CBCA criterion scores for psychopaths and nonpsychopaths for IV, RV, and LM memories (standard deviations in parentheses).

| Criterion | IV | | RV | | LM | |
|---|---|---|---|---|---|---|
| | Non-psychopath (≤ 20 on PCL-R) | Psychopath (≥ 30 on PCL-R) | Non-psychopath (≤ 20 on PCL-R) | Psychopath (≥ 30 on PCL-R) | Non-psychopath (≤ 20 on PCL-R) | Psychopath (≥ 30 on PCL-R) |
| N | 19 | 39 | 18 | 42 | 9 | 25 |
| 1 | 2.00 (0.00) | 1.95 (0.22)[a] | 2.00 (0.00) | 1.88 (0.33)[a] | 2.00 (0.00) | 1.76 (0.44)[a] |
| 2 | 1.89 (0.32) | 1.95 (0.22) | 1.89 (0.32) | 1.95 (0.22) | 2.00 (0.00) | 1.88 (0.33)[a] |
| 3 | 1.84 (0.38) | 2.00 (0.00)[a] | 1.89 (0.32) | 1.90 (0.30) | 2.00 (0.00) | 1.84 (0.37)[a] |
| 4 | 1.79 (0.42) | 1.85 (0.37) | 1.83 (0.38) | 1.76 (0.53) | 1.89 (0.33) | 1.72 (0.46) |
| 5 | 1.95 (0.23) | 1.95 (0.22) | 1.94 (0.24) | 1.95 (0.22) | 1.89 (0.33) | 1.96 (0.20) |
| 6 | 1.00 (0.82) | 1.44 (0.68) | 1.17 (0.86) | 1.50 (0.55) | 0.78 (0.67) | 1.00 (0.76) |
| 7 | 0.21 (0.42) | 0.21 (0.41) | 0.11 (0.32) | 0.07 (0.34) | 0.22 (0.44) | 0.32 (0.56) |
| 8 | 0.16 (0.50) | 0.21 (0.47) | 0.11 (0.32) | 0.29 (0.55) | 0.22 (0.44) | 0.24 (0.52) |
| 9 | 0.53 (0.77) | 0.49 (0.56) | 0.50 (0.51) | 0.74 (0.70) | 0.56 (0.88) | 0.52 (0.71) |
| 11 | 0.42 (0.69) | 0.44 (0.68) | 0.28 (0.58) | 0.60 (0.70) | 0.44 (0.73) | 0.36 (0.70) |
| 12 | 1.89 (0.32) | 1.59 (0.60) | 1.72 (0.46) | 1.77 (0.47) | 1.56 (0.73) | 1.41 (0.57) |
| 13 | 1.26 (0.65) | 1.21 (0.73) | 1.28 (0.67) | 1.19 (0.70) | 1.33 (0.87) | 0.81 (0.68) |
| 14 | 0.32 (0.58) | 0.15 (0.43) | 0.11 (0.32) | 0.10 (0.30) | 0.00 (0.00) | 0.00 (0.00)[a] |
| 15 | 0.37 (0.50) | 0.62 (0.59) | 0.39 (0.50) | 0.67 (0.69) | 0.44 (0.53) | 0.60 (0.50) |
| 16 | 1.63 (0.60) | 1.44 (0.68) | 1.56 (0.71) | 1.24 (0.79) | 1.78 (0.44) | 1.56 (0.65) |
| 17 | 1.11 (0.81) | 1.13 (0.77) | 1.28 (0.83) | 1.21 (0.84) | 1.33 (0.87) | 1.56 (0.65) |
| 18 | 0.58 (0.69) | 0.54 (0.68) | 0.44 (0.62) | 0.55 (0.63) | 0.44 (0.53) | 0.80 (0.58) |
| 19 | 0.00 (0.00) | 0.03 (0.16)[a] | 0.06 (0.24) | 0.02 (0.15) | 0.00 (0.00) | 0.04 (0.20)[a] |
| 20 | 0.21 (0.42) | 0.08 (0.27) | 0.06 (0.24) | 0.14 (0.35) | 0.22 (0.44) | 0.04 (0.20) |
| 21 | 0.16 (0.50) | 0.05 (0.32) | 0.00 (0.00) | 0.05 (0.22)[a] | 0.00 (0.00) | 0.00 (0.00)[a] |
| 22 | 0.26 (0.65) | 0.36 (0.74) | 0.33 (0.69) | 0.62 (0.88) | 0.33 (0.71) | 0.48 (0.82) |
| 23 | 0.05 (0.23) | 0.03 (0.16) | 0.06 (0.24) | 0.07 (0.26) | 0.33 (0.11) | 0.08 (0.28) |
| 24 | 0.00 (0.00) | 0.08 (0.35)[a] | 0.00 (0.00) | 0.05 (0.31)[a] | 0.00 (0.00) | 0.04 (0.20)[a] |

Note: all comparisons nonsignificant.
[a] indicates that comparisons could not be made for this pair because at least one of the standard deviations was equal to 0.00.

To determine whether psychopaths and nonpsychopaths differed in the frequency in which they discussed their own and other people's mental states, *t*-tests were used to compare frequencies for CBCA Criteria 12 and 13 affect, cognitive, and physical. Alpha was set to .006 in an attempt to control for Type I error (alpha of .05 divided by eight comparisons). None of the frequencies significantly differed according to psychopathy

for IV, RV, or LM memories. Means related to these comparisons are presented in Table

32. The results of the *t*-tests are presented in Appendix E, Table 50.

Table 32: Mean frequencies for CBCA Criteria 12 and 13 Affect, Cognition, and Physical details for IV, RV, and LM memories (standard deviations in parentheses) for psychopaths and nonpsychopaths.

| Criterion Frequency | IV | | RV | | LM | |
|---|---|---|---|---|---|---|
| | Non-psychopath (≤ 20 on PCL-R) | Psychopath (≥ 30 on PCL-R) | Non-psychopath (≤ 20 on PCL-R) | Psychopath (≥ 30 on PCL-R) | Non-psychopath (≤ 20 on PCL-R) | Psychopath (≥ 30 on PCL-R) |
| N | 19 | 39 | 10 | 42 | 9 | 25 |
| 12 Total | 12.00 (7.54) | 9.33 (8.00) | 9.33 (8.44) | 10.98 (9.24) | 12.78 (12.80) | 7.76 (9.89) |
| 12 Affect | 5.00 (2.79) | 4.03 (3.92) | 4.00 (2.77) | 4.62 (3.58) | 3.78 (3.63) | 2.88 (4.50) |
| 12 Cognition | 6.53 (5.72) | 4.62 (3.92) | 4.50 (6.29) | 5.93 (6.63) | 7.22 (8.29) | 3.68 (5.70) |
| 12 Physical | 0.47 (0.70) | 0.69 (1.40) | 0.83 (0.99) | 0.43 (0.86) | 1.78 (2.49) | 1.32 (1.38) |
| 13 Total | 4.21 (4.26) | 3.72 (2.96) | 4.11 (4.40) | 3.67 (3.46) | 4.11 (3.33) | 2.20 (3.23) |
| 13 Affect | 2.00 (2.96) | 1.97 (2.11) | 2.22 (1.80) | 2.00 (2.38) | 1.78 (1.72) | 1.28 (2.03) |
| 13 Cognition | 1.79 (2.23) | 1.51 (1.48) | 1.78 (2.56) | 1.55 (1.64) | 2.11 (2.03) | 0.48 (0.92) |
| 13 Physical | 0.37 (0.50) | 0.23 (0.54) | 0.17 (0.51) | 0.17 (0.44) | 0.22 (0.67) | 0.44 (1.00) |

Note: All comparisons nonsignificant.

Comparisons of CBCA criterion scores and frequencies for Criteria 12 and 13

affect, cognitive, and physical for the more traditional definition of psychopathy (i.e.,

those scoring below 30 on the PCL-R are nonpsychopaths; those scoring 30 or above on

the PCL-R are psychopaths) were also conducted. The results were very similar to the

above results. Tables related to these findings are presented in Appendix G, Tables 59 -

62.

**Psychopathy and RM**

For the analyses related to the RM scores and psychopathy, a psychopath was

defined as any participant who scored 30 or above on the PCL-R and a nonpsychopath

was defined as any participant who scored 20 or below on the PCL-R. As before, this

revised distinction was implemented to make the two groups of participants more distinct

from each other, in relation to the psychopathy construct. Analyses were conducted on only the memories found to be credible according to CBCA to reduce the possibility of extraneous variables related to credibility affecting the comparisons.

$T$-tests were conducted to compare RM scores for the first two scoring methods for psychopaths and nonpsychopaths. Alpha was set to .004 in an effort to control for Type I error (alpha of .05 divided by twelve comparisons). RM scores did not differ according to psychopathy for either the first or the second scoring method for IV, RV, or LM memories. Means related to these tests are presented on Table 33. Results of the $t$-tests are presented in Appendix E, Table 51.

Table 33: Mean RM scores for the first and second scoring methods (standard deviations in parentheses) for IV, RV, and LM memories according to psychopathy.

|  |  | Nonpsychopath ($\leq$ 20 on PCL-R) | Psychopath ($\geq$ 30 on PCL-R) |
|---|---|---|---|
| IV | n | 21 | 42 |
|  | RM Scoring Method 1 | 37.28 (6.22) | 37.70 (6.95) |
|  | RM Scoring Method 2 | 38.20 (7.32) | 39.57 (7.88) |
| RV | n | 20 | 52 |
|  | RM Scoring Method 1 | 34.87 (6.90) | 36.52 (6.00) |
|  | RM Scoring Method 2 | 35.22 (8.07) | 38.08 (7.31) |
| LM | n | 10 | 26 |
|  | RM Scoring Method 1 | 29.47 (6.67) | 28.52 (6.12) |
|  | RM Scoring Method 2 | 29.20 (6.06) | 28.92 (7.72) |

*Note:* All comparisons nonsignificant.

Similar comparisons were conducted for the third and fourth scoring method. These results are presented in Appendix F.

$T$-tests were conducted to compare RM construct scores for psychopaths and nonpsychopaths for IV, RV, and LM memories. Alpha was set to .002 in an effort to control for Type I error (alpha of .05 divided by 24 comparisons). RM scores did not differ according to psychopathy for either the first or the second scoring method for IV,

RV, or LM memories. Means related to these tests are presented on Table 34. Results of

the *t*-tests are presented in Appendix E, Table 52.

Table 34: Mean RM construct scores (standard deviations in parentheses) for IV, RV, and
LM memories according to psychopathy.

|  |  | Nonpsychopath ($\leq$ 20 on PCL-R) | Psychopath ($\geq$ 30 on PCL-R) |
|---|---|---|---|
| IV | n | 21 | 42 |
|  | Clarity | 5.46 (0.97) | 5.40 (1.14) |
|  | Sensory Experiences | 3.19 (1.83) | 3.29 (1.68) |
|  | Spatial Information | 5.52 (1.23) | 5.58 (1.56) |
|  | Time Information | 4.77 (1.45) | 5.02 (1.23) |
|  | Overall Impression | 4.06 (0.95) | 4.53 (1.27) |
|  | Retrospective Recollection | 4.46 (1.19) | 4.93 (1.08) |
|  | Events Before and After | 4.71 (1.69) | 4.93 (1.59) |
|  | Bizarreness | 6.02 (0.85) | 5.88 (1.17) |
| RV | n | 20 | 52 |
|  | Clarity | 4.99 (1.22) | 5.08 (1.19) |
|  | Sensory Experiences | 2.98 (1.64) | 3.23 (1.46) |
|  | Spatial Information | 4.98 (1.59) | 5.46 (1.47) |
|  | Time Information | 4.38 (1.34) | 4.97 (1.19) |
|  | Overall Impression | 4.15 (1.13) | 4.14 (1.16) |
|  | Retrospective Recollection | 4.18 (1.18) | 4.78 (1.06) |
|  | Events Before and After | 3.73 (1.92) | 4.73 (1.85) |
|  | Bizarreness | 5.85 (1.08) | 5.69 (1.12) |
| LM | n | 10 | 27 |
|  | Clarity | 3.20 (1.47) | 3.50 (1.26) |
|  | Sensory Experiences | 2.20 (1.44) | 2.33 (1.36) |
|  | Spatial Information | 4.05 (1.99) | 3.61 (1.34) |
|  | Time Information | 3.52 (1.24) | 4.22 (1.48) |
|  | Overall Impression | 3.40 (0.87) | 3.49 (0.95) |
|  | Retrospective Recollection | 3.87 (0.80) | 4.20 (1.20) |
|  | Events Before and After | 4.70 (1.62) | 3.07 (1.46) |
|  | Bizarreness | 4.27 (2.11) | 4.60 (1.46) |

*Note:* All comparisons nonsignificant.

Comparisons of all four RM scores and RM construct scores for the more

traditional definition of psychopathy (i.e., those scoring below 30 on the PCL-R are not

psychopaths; those scoring 30 or above on the PCL-R are psychopaths) were also

conducted. The results were very similar to the above results. Tables related to these findings are presented in Appendix G, Tables 63 - 66.

**Summary for Psychopathy and Credibility Results**

Although there was tendency for psychopaths to be more likely to have memories partially confirmed with some contradictions than nonpsychopaths, especially for RV memories, this difference was not significant. Similarly, there was tendency for psychopaths to be more likely to have memories judged to be noncredible according to CBCA than nonpsychopaths, especially for RV memories, but this difference was not significant. CBCA criterion scores, CBCA criterion frequencies concerning one's own and another person's mental state, RM total scores, and RM construct scores did not differ according to psychopathy.

**Discussion**

The present investigation was the first thorough qualitative study on verbal credibility assessment in violent offenders. In the following sections, the results are discussed in terms of how the present research supports the Undeutsch hypothesis, as well as other research in this area. First, findings related to each hypothesis are discussed. Second, the limitations of the present investigation are reviewed and suggestions for future research are offered. Third, implications of the research for the criminal justice system are proposed. Finally, the summary and conclusions of the present investigation are presented.

**CBCA-Judged Credible vs. Noncredible Memories**

The vast majority (approximately 96%) of the violent memories examined in the present study were found to be credible according to CBCA. It appears that these results were mostly independent of effects of memory age, level of intoxication during the event, memory length, and amount of rehearsal. However, when a within-participants comparison was made, considering only participants who had provided a memory that had been judged to be noncredible according to CBCA, credible memories had been rehearsed significantly more often than noncredible memories.

Although the results regarding credibility were not completely unexpected, as it had been hypothesized that most memories would be found to be credible, a larger proportion of the memories were expected to be noncredible based on what is known about incarcerated offenders, especially those diagnosed as psychopaths. The previously mentioned conversation with the participant, as well as research findings that suggest that psychopaths lie more frequently than nonpsychopaths due to duping delight (e.g., Ekman,

1992; Hare et al., 1989; Porter & Woodworth, 2007; Spidel et al., 2003) led us to expect that a significant proportion of the participants would lie during these interviews. In line with this, one researcher has suggested that all statements of psychopaths should not be believed without outside verification, since psychopaths enjoy lying just for the sake of lying (Doren, 1987).

There are several likely explanations to account for the large proportion of credible memories in the present study. First, as mentioned previously, there was little motivation for the participants to tell a lie. The participants were assured that anything they shared during their interviews would remain confidential, and would have no effect on their sentences. Thus, it was expected that most memories would be based on true stories, and true stories were expected to be found credible. The fact that a higher proportion of credible accounts was found than was expected implies that incarcerated offenders are less likely to lie in this type of research context than clinical experience or anecdotal evidence would suggest. The present study is the first to demonstrate that incarcerated offenders generally provide credible accounts when they have little motivation to lie.

Second, as mentioned, CBCA is known to have a truth bias (e.g., Landry & Brigham, 1992; Steller & Koehnken, 1989; Yuille, 1988). That is, when the analysis results in an error, it tends to be that a false statement is judged credible. This bias suggests that some memories that were untrue may have been found credible. Although no statistical evidence was found to support this suggestion, anecdotal evidence suggests that this may have happened for at least one account. One participant described an act of perpetrated reactive violence where he seriously stabbed a fellow inmate while in prison.

He further claimed to have been sent to segregation as a result of this event. This event should have been reported in his CPR or RADAR, but could not be confirmed by either report. Interestingly, when an interviewer asked him about this event in a follow-up interview examining memory consistency, he claimed that such an event had never happened. This memory was found to be credible by two independent CBCA coders. While it is possible that this event actually happened in the way the participant described, but that the participant was not caught or charged with this crime, clearly at least some aspects of this memory were fabricated, as the participant claimed to have been caught and charged when he originally shared this memory.

Further evidence of the truth bias came from several memories that gave CBCA coders the impression of being false, but since these memories had a large number of CBCA criteria, they were judged to be true. For example, one memory involved a very detailed description of torturing a victim to death over a number of days. This event could not be confirmed by file information, but since only the RADAR was available for this participant, it could not be concluded unequivocally that the participant had invented this event (see below for a discussion on the limitations in determining consistency with file information in the present study). The CBCA coders felt that the memory seemed false because the participant seemed to be enjoying the telling of the gory details in a manner that would be consistent with self-serving exaggeration or fabrication. Alternatively, telling the gory details in this manner would also be consistent with a sadistic personality, but coders did not consider personality factors in their judgments. In any case, like the story about the big fish that got away, this story just did not ring true. However, the story was judged to be coherent, spontaneously reproduced, to have an appropriate amount of

detail, with contextual embedding, with descriptions of interactions, and six other CBCA criteria, so it was judged to be credible. Although the verbal style of the storyteller alerted CBCA evaluators to examine this memory closely, there was nothing in the CBCA criteria that showed that this story was false. This example highlights that CBCA is just one component of SVA and is not meant to be used independently of the other two components. The present study included the first two components: an appropriate interview and CBCA. However, these two components are meant to be used with the third component, the validity checklist, for successful verbal credibility assessment. If the validity checklist had been used in the present study, perhaps the affect of the storyteller in the above example would have been judged as inappropriate and the memory would have been judged to be false.

This tentative evidence for a truth bias highlights one of CBCA's weaknesses (Rassin, 2000). CBCA uses an inductivistic approach, searching for signs of truthfulness. Deceptive statements can be interpreted in ways that make them appear truthful. One single sign of deceit would outweigh several signs of truthfulness. Historically, CBCA was intended to prove that children can indeed be credible witnesses; therefore, the focus was on the search for truthfulness. CBCA is problematic if the purpose is to prove someone was lying, since none of the criteria are signs of deception, with the possible exception of the final three criteria, which will be discussed in more detail below. This demonstrates why SVA is meant to be a guideline, rather than a diagnostic measure.

Depending on how credible and noncredible memories were compared, a couple of the CBCA criteria, Criteria 2 (Spontaneous Reproduction) and 3 (Appropriate Amount of Detail) were significantly more strongly present in the credible, as compared to the

noncredible, violent memories. It was not surprising that these CBCA criteria differed in noncredible and credible memories. For Criteria 1-5, the presence of these criteria in a statement adds to credibility, and the absence of any of these criteria may suggest a lack of credibility (Seniuk & Yuille, 1996). Most of the memories, whether they were judged to be credible or noncredible, contained Criteria 1-5. If one of these criteria was missing from a statement, or not strongly present in a statement, it would stand out, increasing the likelihood that the statement would be judged to be noncredible.

Two of the three comparisons of CBCA criterion scores in credible vs. noncredible memories resulted in similar criteria differing according to whether an account was judged to be credible or noncredible. That is, when credible and noncredible memories were compared within-participants, only Criterion 3 was scored significantly higher in credible memories, as compared to noncredible memories, and when all of the noncredible memories were compared to the credible memories of the participants who had told only memories that had been judged to be credible, only Criteria 2 and 3 were scored significantly higher in credible memories, as compared to noncredible memories. Overall, no matter how the memories were divided according to CBCA-judged credibility, memories that were judged to be credible tended to have higher CBCA criterion scores than memories that were judged to be noncredible. This demonstrates that the coders applied the criteria as they had been trained. The fact that more of these differences were not significant is likely due to the small number of memories judged to be noncredible according to CBCA, which resulted in a lack of power and an inability to make several statistical comparisons because of lack of variance. This was especially true for the comparison of noncredible and credible memories within memory types where

only two IV memories were found to be noncredible. It is likely that if a greater number of statements had been judged to be noncredible, therefore increasing the power of the tests, a greater number of the CBCA criteria would have been found to significantly differ according to CBCA credibility.

The within-participants comparison and the comparison of the noncredible memories to the credible memories of the participants who had provided only credible memories were further limited by the fact that memories for different types of events were combined. Although tests related to the fourth hypothesis indicated that IV and RV memories do not differ in terms of CBCA criterion scores, the three memory types are theoretically different from each other and it had been hoped that they could be examined separately. The memory types were combined only to strengthen the power of the comparisons.

The fact that no clear pattern of CBCA criteria that were more likely to be present in credible than noncredible accounts was found may also be partly due to the fact that according to CBCA theory, there is no particular pattern of criteria that indicates credibility. As mentioned, individual differences in people and events cause a variety of criteria to be present in any credible account. The present results regarding CBCA credibility highlight the importance of not using CBCA as a scale. Critical thinking is crucial even within CBCA. Even if a narrative has a large variety of CBCA criteria, this does not necessarily mean that it is true. For example, one participant shared a detailed account about a time he attempted to exact revenge on another person. The participant and the victim in the story had been living together in a group home. One night, while the victim was sleeping, the participant constructed a trap, a web of fishing line and fishing

hooks, and used nails to attach the trap to the victim's bedroom door. The participant threw eggs at the sleeping victim to provoke the victim to wake up and leave his bedroom. While the participant ran to his own bedroom and hid in his bed, the victim got caught in the trap, and a fish hook was embedded in his eye. According to the participant, the victim ended up losing his eye as a result. This story had many CBCA criteria: it was told spontaneously, had an appropriate amount of detail, contextual embedding, descriptions of interactions, and unusual details. However, upon careful reading of the transcript, it was clear that the story was not coherent. Among the holes in this story are the facts that it is not likely that the trap could have been nailed to the doorframe without waking up the victim, and a trap stuck to the doorframe by anything weaker than nails would not likely hold up to a person running through it. Further, a trap constructed the way the participant had described would not be strong enough to embed a fish hook into a person's eye. If the criteria in this memory had been added up, they likely would have met most decision rule thresholds, and the memory would have been judged credible. However, without Criterion 1, Coherence, the memory should not be and was not found credible. This anecdotal evidence underscores the reasons why CBCA should be considered qualitatively rather than used as a quantitative tool.

The above example also highlights a limitation of CBCA. In some situations, CBCA coders may not have enough background knowledge to evaluate the coherence of an account. For example, finding that the above example lacked coherence required CBCA coders to have some knowledge about fishhooks. That is, the design of a fishhook makes it difficult to be embedded in an eye; to do so would require a lot of force. A

CBCA coder without this knowledge may have found this story to be coherent and, therefore, credible.

Criterion 24 (Rigid Repetition) was one of the few CBCA criteria to be scored higher in CBCA-judged noncredible accounts than in credible accounts, no matter how the memories were compared according to CBCA-judged credibility, although this difference was not significant. As this criterion, along with Criterion 22 (Long-winded vs. Issue Oriented) and Criterion 23 (Theme-Related Changes) is one of the few whose presence may indicate a lack of credibility, this finding was not unexpected. Criteria 22-24 are among the criteria not generally considered in most CBCA research (e.g., Akehurst et al., 2001; Buck, Warren, Betman, & Brigham, 2002; Gödert et al., 2005; Granhag et al., 2006; Koehnken et al., 1995; Landry & Brigham, 1992; Orbach & Lamb, 1999; Parker & Brown, 2000; Porter & Yuille, 1996; Ruby & Brigham, 1997; Strömwall et al., 2004; Vrij et al., 2000, 2002; Vrij & Mann, 2006; Zaparniuk et al., 1995). As mentioned, an adapted version of the CBCA criteria, containing 24 criteria rather than the traditional 19, was used to provide a more complete analysis of these memories.

Interestingly, Criteria 22 and 23 tended to be more strongly present in the memories judged to be credible according to CBCA than in the memories judged to be noncredible according to CBCA, although these differences were not significant. This means that memories that were judged to be credible were found to be more long-winded and to have more theme-related changes than memories that were judged to be noncredible according to CBCA. Since these criteria are thought to be indicative of lack of credibility, this was contrary to expectations. It is also contrary to Koehnken et al.'s (1995) finding that an issue-oriented reporting style was associated with credibility.

Although these findings were based on a small number of data points, they suggest that these criteria may be present in accounts that appear to be credible.

Criteria 22-24 are examples of Ekman's (2001) aforementioned hot spots. The presence of any of these criteria in a memory would not indicate lack of credibility, but their presence would suggest that the CBCA coder should consider the credibility of this memory carefully. For example, one participant shared a memory of an instrumentally violent incident where he entered a woman's house and sexually assaulted her. For most of the interview, he used pronouns normally, but when he spoke about preparing for the assault and committing the assault, he vacillated between using the personal pronoun "I" as he described his actions and dropping the personal pronoun:

I decided at probably two o'clock in the morning and I finally decided that ok, I'm going to do this. Um, had to been the direction of that apartment. Um, broke into the apartment. Um, got in without anybody noticing anything, um, without anybody seeing breaking and entering, without waking the women that lived there. Um, I went to the kitchen and got a knife. I don't know, maybe it was on the counter or in the drawer. I don't remember exactly I got the knife from, but it was in the kitchen somewhere. Um, went from the kitchen into the bathroom after that. Um, got a towel, I believe off the towel rack or somewhere in the bathroom. Um, draped it over my head so I couldn't be recognized. Um, I believe I turned on the light quickly in the bathroom to, I guess to be certain that the woman was alone. Just very quickly on and off to see, so I could see if the women was in there alone. Um, once I was sure that she was home alone I went into the bedroom. Um, snuck up to the bed where she was sleeping. Um, put my hand over her mouth.

Held the knife to her throat. I don't recall exactly, I think I said to the effect of, uh, "keep your mouth shut and do what I tell you to do, and you won't get hurt." Um, I recall that she seemed very scared. Um, I think, I think at that point, um, I think she told me "don't worry I won't scream". Um, that point, um, I pulled the covers down, removed the clothing that she had on. I forced sex to the period, I think about, um, half an hour.

This is an example of a theme related change, rather than just a style of speaking, because he used pronouns in this way only when he spoke about the specific details related to his offence. A stylistic change such as this might indicate that the participant is hiding something, or it might indicate that he is feeling guilt or remorse and wants to detach himself from his actions. According to Ekman (1997), an individual may show changes in emotional arousal, which may result in verbal and nonverbal behavioural changes, for many different reasons, only one of which is because he or she is lying. Since many other criteria were present in this participant's account, this memory was found to be credible. Likewise, if a statement was long-winded, but otherwise appeared credible (i.e., contained other criteria indicating credibility), it would likely be judged to be credible.

In contrast, one participant's reactively violent memory was coherent (Criterion 1), but received low CBCA scores for Spontaneous Reproduction, Appropriate Detail, Contextual Embedding, and Descriptions of Interactions (Criteria 2-5, respectively), and a high score for Criterion 24, Rigid Repetition. Whenever the interviewer asked him for more detail, he repeated the same assertion each time. On page 16 of this participant's transcript he stated, "And they had seen this, as they got closer, the guy stopped and

grabbed me by the throat and tried to get my money and drugs from me." On the following page he stated, "The details is what I just mentioned, the guy grabbed me by the throat – tried to get my money and I killed him." A page after that, he repeated, "I was about to leave when one of the guys grabbed me by the throat. And tried to rob me of my money and my drugs." Finally, the interviewer asked for more detail one last time and the participant stated, "They grabbed me by the throat and tried to take my money out of me." It is possible that this participant was nervous about the interview, which caused him to withhold details about this memory. However, this is unlikely, as the other memories this participant provided had an appropriate amount of details. Alternatively, it is possible that this participant did not remember this event very well, which led to a few number of details. This explanation is also improbable, since the participant gave his memory for this memory a seven out of ten for vividness and an eight out of ten for amount of details, suggesting that he felt he remembered this event well. Since this memory was rigidly repetitive, lacked most of the criteria that should be present in a credible memory, and alternative explanations to account for these criteria or lack of criteria could not be found, it was judged to be noncredible.

As mentioned briefly above, researchers have pointed out that one of the weaknesses of CBCA is that it is focused on the detection of truth and does not contain criteria indicative of lying (Vrij et al., 2004). This criticism is based on the list of 19 CBCA criteria, rather than the list of 24 criteria used in the present study. Including all 24 criteria in verbal credibility assessment might improve CBCA's effectiveness as a lie detection tool. While none of the CBCA criteria are indicative of lying, and the present results do not suggest that any of the criteria should be interpreted as being indicative of

110

lying, the final three criteria may alert the evaluator to a hot spot, which would motivate him or her to take a closer look at possible reasons for this change in verbal style.

The credible statements of the participants who had shared one memory that had been judged to be noncredible according to CBCA were compared to the credible statements of the participants who had all memories judged to be credible according to CBCA. This comparison was made to determine whether there was something about the statements of participants who had memories judged to be noncredible, or something about the way these participants shared their memories, that made them less credible than the statements of participants who had all memories judged to be credible according to CBCA. None of the CBCA criteria significantly differed for these two groups of statements. There was no general pattern of results. That is, there was no tendency for the criterion scores of one group to be higher than the criterion scores of the other group, suggesting that neither group of memories appeared to be more credible than the other. This suggests that it was the statements being judged, rather than the participants, further demonstrating that CBCA was applied correctly in the present study.

Altogether, the results regarding CBCA credibility suggest that coders effectually used the CBCA criteria to come to a decision about a memory's credibility. The majority of the statements of these participants contained a number of CBCA criteria and were consequently judged to be credible. These results further suggest that verbal credibility assessment is possible with memory reports of this length and about these types of events. The effectiveness of CBCA at determining the veracity of a memory is discussed in the next section.

**CBCA and Consistency with File Information**

Unfortunately, the majority of the memories considered in the present study could not be confirmed (80.8%), even when some file information was available (i.e., 67.9% of memories for which file information was available could not be confirmed). If a memory for an event was not confirmed, this does not necessarily mean that it did not happen. It means that this event was not in any official record that could be accessed. The memories that were not confirmed were often of less serious violent offences (e.g., grabbed glasses off another guy's face, fight in elementary school), were not likely to be reported for some reason (e.g., a fight with a family member), or were about events that had happened in another country. That is, most of the memories that were not confirmed would not be expected to be in the participants' correctional reports.

None of the memories were completely confirmed. This is likely due to the fact that the interviewers elicited very detailed memories through the use of the Step Wise Interview. The memories were so detailed that the information in a correctional file could never confirm all, or even most, of the details. None of the memories were completely contradicted. If a memory could be located in a participant's correctional files, there would have to be information present in the file that matched the information about the event provided by the participant. If there was enough information in the file for a memory to be identified, it would have to be at least partly confirmed.

All of the memories for which some consistency with file information was found (i.e., those in the partially confirmed category) were coded as credible. This supports the first hypothesis, which predicted that memories found to be credible according to CBCA would be more likely to be consistent with correctional file information. This finding also

provides support for the usefulness of CBCA as a credibility assessment tool. This translates into a 100% hit rate for partially confirmed memories, with no false negatives.

For eight of the memories, some of the details were confirmed in the correctional file, but some of the details were disconfirmed or important details were left out of either the summaries in the correctional file or the participant's statement about the event. For example, one participant claimed to have been hunting and fishing the morning before an assault, but according to information in his correctional file, he was actually hosting a party at his house, at which the victim was a guest. Seven of these memories were coded as credible and one was coded as noncredible. This translates into 87.5 % of the memories that were partially contradicted being falsely judged to be true (i.e., false alarm), and 12.5 % being correctly judged to be noncredible. This is understandable since there were true and not true aspects of each of these memory reports. Other researchers have noted that some of the most difficult types of lies to determine are those that are based on partial truths (DePaulo et al., 2003; Porter & Yuille, 1996). Since CBCA has a known truth bias, as discussed earlier, it is not surprising that these memories tended to be coded as credible rather than noncredible. Although this finding is based on just eight memory reports, it highlights some of the limitations and difficulties in doing field research versus laboratory research. In laboratory research, a story is either the truth or a lie and ground truth is easy to determine. In field research, truth and deception are more ambiguous. This finding also highlights some of the limitations of CBCA. If a lie is embedded in a true story, or if an untruth is told unintentionally (e.g., a storyteller got two days confused), the factors thought to produce CBCA criteria are still going to be present (i.e., the cognitive and motivational components of the Undeutsch hypothesis). For this

reason, Köhnken (2004) suggested that statements be partitioned into individual descriptions of single events or components and CBCA be applied to each segment separately.

Six of the memories for which file information was available were found to be not credible according to CBCA. One of these memories was partially confirmed with some contradictions and five were not confirmed. Although it could not be established that these five memories were fabrications, anecdotal evidence suggests that at least one of these memories was correctly found to be not credible. For one of the memories that was found to be not credible and could not be confirmed, when the participant was asked about the event in a follow-up interview several years after he had first shared his memory for this event, he refused to talk about it. He claimed that if he shared this memory, the interviewer would have to break the limits of confidentiality. That is, he had never been charged for this particular assault and, having just read the limits to confidentiality he felt that it was such a severe assault that the interviewer would be ethically obligated to report him to legal authorities. When the interviewer reminded him that he had shared the story before, and that the interviewer was familiar with the memory and did not feel that the assault was so serious that it would need to be reported, the participant said that he must not have told the whole story in the previous interview. While the participant  implied that he lied by omission when he told the story the first time, it is also possible that he was using this as an excuse so that he did not have to repeat a story that he made up the first time and would now have difficulty piecing together for a second interview. In any case, it is clear that the participant was admitting that this story was a lie. This anecdotal evidence supports the second part of the first

hypothesis, which predicted that memory reports that were rated as noncredible would be more likely to contain details that are contradictory to information in the offender's file.

Using decision rules rather than simply the CBCA coders' credibility decision had little effect on the totals for CBCA credibility and consistency with file information. This result is not surprising since the coders were informed of the decision rules as part of their training. For the first decision rule (Criteria 1-5, plus any two other criteria), the most stringent decision rule, one memory that had been partially confirmed was found noncredible. This suggests that this decision rule might be too strict, as it resulted in at least one false negative. The second (Criteria 1-3, plus any two other criteria) and third (Criterion 2) resulted in fewer noncredible memories than when the decision was left to the coder. Since both of these decision rules resulted in the aforementioned noncredible memory being found credible (i.e., the memory the participant refused to discuss in a follow-up interview), these decision rules may be too lenient, leading to an increased number of false positives. Although no conclusions can be made on such a small number of findings, these results suggest that the decision rules recommended by Zaparniuk et al. (1995) would be better used as guidelines for CBCA coders to keep in mind, rather than as absolute decision rules.

### CBCA Criteria and Consistency with File Information

None of the CBCA criteria were significantly more strongly present in memories that were partly confirmed, as compared to memories that were partly confirmed with some contradictions, over two different comparisons of memories that were partially confirmed and memories that were partially confirmed with some contradictions. When the frequency of each criterion score was examined for memories that were partially

confirmed and memories that were partially confirmed with some contradictions, no clear patterns emerged.

Although not true for all criteria, for all types of memories, the general pattern for the criterion scores was for scores to be higher for memories that were partially confirmed than for memories that were partially confirmed with some contradictions. This is the pattern that would be expected based on the Undeutsch hypothesis, since memories that were partially confirmed should be found credible and memories that were partially confirmed with some contradictions were presumably based partly on deceit, and thus should be found noncredible.

It is likely that none of the CBCA criteria differed according to consistency with file information because of lack of power. Although $t$-tests, which are known to be powerful tests, were used to compare partially confirmed memories and memories that were partially confirmed with some contradictions, the small number of memories that could be located in the participants' correctional files lowered the power of these comparisons. It also often led to a lack of variance, which meant that no comparisons could be made. This was especially true for memories that were partially confirmed with some contradictions. Only one memory fell into this category for LM memories, which meant that no comparisons could be made for LM memories across consistency with file information.

Altogether, only marginal support was found for the first hypothesis. There was clearly a tendency for memories that were partially confirmed to be judged credible according to CBCA, although memories that were partially confirmed with some

contradictions were also more likely to be judged credible. Too few memories were judged to be noncredible to make any firm conclusions about this hypothesis.

**RM and Consistency with File Information**

Inconsistent with the second hypothesis, RM total score did not differ according to the memories' consistency with file information, even though several RM scoring methods were attempted and comparisons were conducted considering the various memory types separately and together. Further, none of the RM construct scores differed according to the memories' consistency with file information, either when comparisons were conducted considering the various memory types separately or together. Moreover, generally, RM total scores or construct scores did not tend to be higher, even nonsignificantly higher, for memories that were partially confirmed than for memories that were partially confirmed with some contradictions. These findings were not consistent with previous research that has used the MCQ to distinguish truth and lies (e.g., Johnson, 1988; Johnson et al., 1988; Masip et al., 2005; Sporer, 1997; Suengas & Johnson, 1988).

The findings related to the comparison of RM scores when memory types were combined should be considered cautiously since RM was shown to differ according to memory type in the comparison of IV and RV memories. Since IV memories were more likely to be partially confirmed than RV memories, and IV memories had significantly higher RM scores than RV memories, this could influence this comparison. This is not a concern in this case since there were no differences between scores according to consistency with file information.

As in the previous section, the findings related to RM scores and consistency with file information are limited by the low number of accounts that were confirmed to be untrue. That is, because so few memories were known to be partly contradicted, it is difficult to find patterns or trends in the RM scores that differentiate memories that had some contradictions from memories that were partially confirmed.

The findings of the present study suggest that RM, as assessed through the MCQ, is not very effective at distinguishing truthful accounts from untruthful accounts in these types of memories. Although it is not clear why RM did not work in these cases, but did work in previous studies, there are several possible explanations for these discrepant results. These explanations will be discussed in the following section, as RM and CBCA are compared.

**RM and CBCA**

Contrary to the third hypothesis, RM and CBCA were not similarly effective at distinguishing credible accounts from noncredible accounts. Memories that were found to be credible according to CBCA were not more likely to have a higher total RM score, no matter whether the comparison was within-participants or between groups for each type of memory.

There were no differences in RM construct scores for CBCA judged-credible and noncredible memories whether the comparison was within-participants or between groups. For the within-groups comparison, the RM construct scores of credible memories tended to be higher than noncredible memories. This was especially true for the Retrospective Recollection and the Bizarreness constructs. For these two constructs, no matter how the memories were compared according to credibility, the construct scores

118

were higher for credible memories than for noncredible memories. These differences may have been significant if more participants could have been included in this test, since this would have increased the power of the test. However, because these comparisons considered memories of different types of events together, the results of these comparisons should be interpreted cautiously. As mentioned, an IV and an RV event theoretically lead to differential memory encoding, which was shown to affect RM scores in the present study. The within-participants analysis considered all of the noncredible and credible memories of the 12 participants together, regardless of type of memory. This may have affected the RM scores and RM construct scores independent of the effects of CBCA-judged credibility.

It was not expected that the Retrospective Recollection construct would consistently have higher scores for memories that were judged to be credible than for memories that were judged to be noncredible. This is not consistent with RM theory or with research that has used the MCQ to distinguish truth and lies (e.g., DePaulo et al., 2003; Johnson, 1988; Johnson et al., 1988; Masip et al., 2005; Sporer, 1997; Suengas & Johnson, 1988; Vrij, 2000). The Retrospective Recollection construct is expected to be more likely present in noncredible or fabricated accounts, according to RM theory. These findings do, however, support Masip et al.'s (2005) observation that the cognitive operations (i.e., Retrospective Recollection) criterion is not effective at discriminating between truthful and fabricated accounts, as well as Sporer and Sharman's (2006) note that there is little empirical evidence in studies involving deception that cognitive operations are found more frequently in accounts of invented events than self-experienced events.

Altogether, the results comparing CBCA and RM suggest that the two verbal credibility assessment tools did not similarly distinguish credible and noncredible memories, despite previous research that has shown that CBCA and RM, as measured by the MCQ, are related (Sporer, 1997, 2004). Although these findings were not expected, based on the literature, they are in line with the present results regarding CBCA and RM and consistency with file information. That is, CBCA seemed to be somewhat effective at discriminating memories that were known to be based on real experiences from memories that were known to be at least partially fabricated, while RM showed limited effectiveness at this type of discrimination, so it is not surprising that CBCA and RM did not perform similarly. There are several possible explanations for these discrepant performances.

First, as mentioned, RM is based on the theory that memories based on truly experienced events would contain more perceptual, contextual, and affective information and less cognitive-based information than memories based on imagined events. The MCQ was developed, based on this theory, to distinguish true and imagined events (Johnson et al., 1988). Although the MCQ has effectively discriminated truthful and fabricated accounts in a number of studies (e.g., Mather et al., 1997; McGinnis & Roberts, 1996; Smeets et al., 2006; Sporer, 1997; Sporer & Sharman, 2006), most of these studies have used a modified version of the MCQ, which was adapted for the judgment of accounts by an external observer. Since research has shown that people tend to be ineffective at determining the source of their own memories (Garry, Manning, Loftus, & Sherman, 1996; Johnson, Hastroudi, & Lindsay, 1993), perhaps RM did not work in this context because the original self-report version of the MCQ was used. It

could be that people are better at judging RM in others' accounts than at judging RM in their own accounts, even when they are not explicitly asked to make a source monitoring judgment. Another issue related to the self-report aspect of the MCQ, as used in the present study, is the possibility of response distortion. That is, participants may have overstated or downplayed some of their responses on the MCQ due to factors such as demand characteristics. These factors could also account for why CBCA seemed to perform more effectively than RM when consistency of file information was considered. That is, the CBCA judgement involved the consideration of another's account, while RM was assessed through self-report. If people tend to be better at judging others' memories than at judging their own memories, and if people tend to base their responses not just on their memory, but also factors such as what they believe the experimenter expects, it would be expected that CBCA would perform better than RM in the way each was assessed in the present study.

Second, although RM and CBCA are based on the same basic theory, that accounts based on real events are qualitatively different from accounts based on imagined or fabricated events, different criteria are used by each method to determine credibility. While the criteria from each method overlap somewhat, there are criteria in each that the other method does not consider. For example, CBCA's Criteria 2 (Spontaneous Reproduction) and 3 (Appropriate Amount of Details) were effective at discriminating between credible and noncredible memories, and the MCQ does not have any items that matched up with these criteria. CBCA Criterion 12 was also found to be consistently scored higher in memories that were judged to be credible according to CBCA, as compared to memories judged to be noncredible, as well as in memories that were

partially confirmed by file information, as compared to memories that were partially confirmed with some contradictions. This criterion is similar to RM's Retrospective Recollection Construct. Both Criterion 12 and the Retrospective Recollection Construct are concerned with thoughts and feelings, however, the Retrospective Recollection Construct does not distinguish event-relevant thoughts and feelings (i.e., thoughts and feelings one remembers experiencing during an event) and cognitive and emotional inventions (i.e., thoughts and feelings related to thinking about the event). In contrast, Criterion 12 is related only to event-relevant thoughts and feelings. Thus, since CBCA had more criteria that contributed to distinguishing memories for events that had actually been experienced from memories for events that had not been experienced, as compared to RM, this could play a role in their differing performances.

Third, there may be some problems with the way that RM was measured in the present study, even though a number of scoring methods were attempted. As mentioned, there is not an established method of assessing RM credibility using the MCQ, since previous studies that have used the MCQ to determine credibility in autobiographical memories have evaluated statements with known ground truth, so MCQ ratings could be compared for true and false reports (e.g., McGinnis & Roberts, 1996; Santtila et al., 1999; Smeets et al., 2006; Sporer, 1997; Sporer & Sharman, 2006; Suengas & Johnson, 1988). None of the researchers who have used the MCQ to measure RM have used a total score, there are no established cut-off points, and there are no recommended interpretations of MCQ scores with regards to credibility. The only credibility judgment associated with RM as assessed via the MCQ is when Sporer (1997; Sporer & Sharman, 2006) had coders make a global rating of credibility after applying a version of the MCQ to another's

memory report. The lack of guidelines for using the MCQ to judge RM credibility limits its usefulness as a verbal credibility assessment tool. The lack of clear guidelines for scoring the MCQ according to credibility may contribute to RM's relative failure to differentiate memories according to CBCA credibility or consistency with file information in the present study.

Finally, the flaws in the data (e.g., the small number of memories judged to be noncredible) may partially account for the discrepant performances of CBCA and RM in the present study. As mentioned previously, these flaws resulted in low power and a lack of variance which adversely affected the analyses of the present study.

**Credibility and Type of Violence**

This study represents the first time that verbal credibility was compared across types of violence. Mixed support was found for the hypothesis that IV memories would appear to be more credible than RV memories.

**Type of Violence and Consistency with File Information**

Of the memories for which some file information was available, approximately 56% of the IV memories and 80% of the RV memories could not be confirmed; 39% of the IV memories and 14% of the RV memories were partially confirmed; and 4% of the IV memories and 5% of the RV memories were partially confirmed with some contradictions. File information that confirmed some of the details of the memories was more likely to be found for IV memories than for RV memories, and it was more likely that no file information would be found for RV memories than for IV memories. This difference was unexpected; IV memories were expected to appear more credible than RV memories, but no difference was expected for the actual truthfulness of IV and RV

memories. Although it had not been expected, it can be explained through a consideration of the types of IV and RV memories that were recalled for the Cooper (2005) study.

Cooper (2005) categorized the violent memories according to the type of violence or crime committed. Those categories were sexual assault or sexual assault with a weapon; stabbings, shootings, or murders; violence in a break and enter, robbery, or home invasion; robbery; and fights or assaults. Most of the memories that were not likely to be included in a participant's correctional file (e.g., fight in elementary school, fight with a family member) would fall into the latter category, fights or assaults. In contrast, memories that fell into the first four categories would be more likely to be included in a participant's correctional file. Indeed, if a memory that fell into one of the first two categories was recalled, but could not be found in an offender's correctional file, this could imply lack of credibility.

For the IV memories, 54.1% were in the fights or assaults category and 45.1% were in the other four categories. For the RV memories, 79% were in the fights or assaults category, and 17.9% were in the other four categories. Thus, IV memories were more likely to be the type of memory that was likely to be found in a participant's correctional file, while RV memories were more likely to be the type of memory that was unlikely to be found in a participant's correctional file. RV memories were not more likely than IV memories to be partially contradicted by correctional file information. This means that there was no evidence that RV memories were more likely to be untruthful according to consistency with file information truth than IV memories. This provides further evidence that the differences regarding memories not confirmed and memories partially confirmed is likely due to the specific types of violence recalled and their

likelihood of being part of an inmate's correctional file, rather than the truthfulness of these memories.

### Type of Violence and CBCA

Approximately 98% of the IV memories and 95% of the RV memories were found to be credible according to CBCA. This difference was not significant and none of the CBCA criteria differed according to type of violence. Thus, little support was found for the fourth hypothesis when CBCA credibility was examined across types of violence.

### Type of Violence and RM

In support of the fourth hypothesis, IV memories had significantly higher RM scores, as compared to RV memories when the first, second, or fourth RM scoring method was used. Further, two of the RM constructs (i.e., Time Information and Events Before and After) had significantly higher mean scores in IV, as compared to RV memories, while five of the other six RM constructs (i.e., Clarity, Spatial Information, Overall Impression, Retrospective Recollection, and Bizarreness) had nonsigificantly higher mean scores in IV, as compared to RV memories.

Thus, considering RM on its own, IV memories do appear more credible than RV memories. This was expected since Cooper (2005; Cooper & Yuille, 2007) found that acts of instrumental violence were recalled significantly more vividly and with greater detail in comparison to acts of reactive violence. It was expected that a more vivid and detailed account would appear to be more credible. This was also in line with other research in this area (Porter et al., 2001; Swihart et al., 1999; Taylor & Kopelman, 1989) and with Hervé et al.'s (2007) biopsychosocial theory of eyewitness memory, which suggests that because the motivation for reactive violence is internal (e.g., anger), and the

motivation for instrumental violence may be internal or external (e.g., financial gain, revenge), accounts of instrumentally violent memories would likely contain more details about the event that are external to the individual and accounts of reactively violent memories would likely contain more details about the event that are internal to the individual (e.g., thoughts and feelings) and fewer external details. This would have the effect of making statements about IV events appear to be more credible than statements about RV events, since the latter statements should theoretically contain more external details, which is associated with credibility, according to RM theory.

**Verbal Credibility and Type of Violence**

The results regarding verbal credibility are qualified by the finding that IV memories were more likely to be consistent with file information than RV memories. It could be then, that IV memories had stronger indicators of verbal credibility than RV memories because IV memories were actually more likely to be true than RV memories. There would be no reason to expect this, and it is unlikely to be the case, since there was no evidence that RV memories were more likely than IV memories to be contradicted by file information, but it is a possibility that cannot be completely discarded without empirical evidence to the contrary.

Considering the results regarding type of violence and RM and CBCA together, it seems that IV memories were not found to be more credible than RV memories. No support was found for the premise that IV memories would appear to be more credible than RV memories, according to CBCA. Indeed, the CBCA criteria differentiated between memories better according to CBCA credibility and consistency with file information than according to type of violence. In contrast, RM, as judged by the MCQ,

126

differentiated between IV and RV memories better than it differentiated between

memories according to CBCA credibility or consistency with file information. When

discriminating memories based on consistency with file information or CBCA credibility,

total RM scores were not very useful, but IV memories had higher RM scores than RV

memories, for three of the four RM scoring methods. In addition, none of the RM

constructs differentiated between memories according to consistency with file

information or according to CBCA credibility, but two of the RM constructs were scored

significantly higher in IV memories than in RV memories. This suggests that IV

memories contained more RM criteria indicative of credibility than RV memories beyond

what could be expected by differences in consistency with file information.

Since CBCA showed a tendency to differentiate memories according to

consistency with file information, but did not differentiate memories according to type of

violence and RM did not differentiate memories according to consistency with file

information, but did differentiate memories according to type of violence, there is little

support for the fourth hypothesis. It would appear that in the present study, RM, as

assessed by the MCQ, is measuring something other than verbal credibility. For example,

as mentioned, the MCQ has been used to measure metamemory, awareness of one's

memory processes (e.g., Roediger & Marsh, 2003). Thus, it could be that IV memories

were scored higher than RV memories for metamemory, rather than verbal credibility,

which is still in line with Hervé et al.'s (2007) biopsychosocial theory of eyewitness

memory. Indeed, Cooper (2005; Cooper & Yuille, 2007) concluded that memories for

instrumentally violent acts more vivid and more detailed than memories for reactively

violent acts, measuring memory through the use of several items on the MCQ (Items 8, 9,

and 33). The reasons that memory for instrumentally violent acts has been found to be more detailed and vivid than memory for reactively violent acts could also account for metamemory scores being higher for IV than for RV. Although individuals may not be aware of whether their focus during an event was internal or external, or of any of the other predisposing, precipitating, or perpetuating factors that affect memory, they are likely aware of how well they remember details about the event and how clear memory for an event is. In addition, the experience of planning and fantasizing about IV (i.e., rehearsal) likely has the effect of not only improving memory for this type of event, in terms of amount of detail and level of vividness, but also increasing metamemory. That is, research has shown that rehearsal improves memory quality and that individuals are aware of the effects of rehearsal on memory (Huet & Mariné, 2002).

**Credibility and Psychopathy**

### Psychopathy and Consistency with File Information

Based on many research studies that have suggested that psychopathic individuals tend to deceive more frequently than nonpsychopathic individuals (e.g., Book et al., 2006; Hare et al., 1989; Kropp & Rogers, 1993; Rogers & Cruise, 2000; Seto et al., 1997; Spidel et al., 2003; Woodworth & Porter, 2007), it had been expected that the participants identified as psychopaths in the present study would be more likely than the nonpsychopathic participants to provide memories partially confirmed with some contradictions. Although there was a tendency for psychopathic participants to be more likely to provide a memory that was partially confirmed with some contradictions, especially for RV memories, this tendency was not significant. It is likely that the low number of memories found to be partially confirmed with some contradictions resulted in

128

low power, which contributed to the lack of significance for this comparison. This lack of power limits the capacity for the present results to support the fifth hypothesis.

### Psychopathy and CBCA Credibility

As mentioned previously, it had been expected that participants identified as psychopaths would be more likely than nonpsychopaths to lie about the perpetrated violent events they had experienced. Since it was expected that fabricated memories would be judged to be noncredible according to CBCA, it was expected that psychopaths would be more likely to provide CBCA-judged noncredible memories than nonpsychopaths. Although there was a tendency for psychopathic participants to be more likely to provide a memory that was judged to be noncredible according to CBCA, particularly for RV memories, this tendency was not significant. Similar to the findings related to psychopathy and consistency with file information, it is likely that the low number of memories judged to be noncredible according to CBCA resulted in low power, which contributed to the lack of significance for this comparison. Although this lack of power limits the capacity for the present results to support the fifth hypothesis, the fact that psychopathy was similarly related to consistency with file information and CBCA-judged credibility (i.e., nonsignificant tendencies for psychopaths to be more likely to provide memories that were partially contradicted by file information and judged to be noncredible, especially for RV memories) lends some marginal support to the first hypothesis.

There was no clear tendency for the CBCA criterion scores of nonpsychopaths to be higher or lower than those of psychopaths. This finding was not entirely surprising, as only memory reports judged to be credible according to CBCA were considered in these

analyses. Lee et al. (2008) also found few differences between psychopaths and nonpsychopaths for CBCA criterion scores for credible memories. It was unfortunate that the present study did not find enough memories to be judged noncredible to allow for an adequate analysis of the interactions between psychopathy, credibility, and consistency with file information.

Nonpsychopaths were no more likely than psychopaths to discuss their own emotional experiences (i.e., CBCA Criterion 12 Affect) or another person's mental state (i.e., CBCA Criterion 13). These findings are strengthened by the fact that they were consistent for two approaches for discriminating psychopaths and nonpsychopaths. These findings were surprising as psychopaths are known to have deficiencies in affective experiences (e.g., Hare, 1980, 1991, 2003; Herba et al., 2007). The PCL-R includes items related to this, such as lack of remorse or guilt, shallow affect, callous/lack of empathy, and failure to take responsibility for one's own actions (Hare, 2003). Psychopaths typically display low levels of emotion, as compared to nonpsychopaths, so it was expected that they would be less likely to discuss their emotional states than nonpsychopaths. Psychopaths also typically show little concern or awareness of other people's thoughts and feelings, so it was expected that they would be less likely to discuss another person's mental state than nonpsychopaths.

There are no clear explanations to account for the frequency with which psychopaths discussed their own and other's emotional states in the present study. The most likely explanation is that the psychopathic participants in the present study learned to discuss emotions in therapy or other sessions designed for rehabilitation. It is standard for inmates to undergo psychological assessment upon intake (Leis, Motiuk, & Ogloff,

1995). Most offenders undergo therapy while incarcerated (Morgan, 2002). Hare (1999) has noted that psychopaths in prison often learn to use the correctional facilities to their own advantage to help shape a positive image of themselves, usually for the benefit of the parole board. Thus, psychopathic inmates might take classes, enroll in programs for drug and alcohol abuse, and join religious groups in an effort to appear as if they are rehabilitating themselves. Through these rehabilitative efforts, rather than being rehabilitated, psychopaths generally learn the language of emotionally healthy individuals. For example, Hare (1999) describes psychopathic individuals who have claimed to experience strong emotions but were unable to describe the subtleties of various affective states, equating love with sexual arousal and anger with irritability. Thus, it is possible that the psychopathic participants in the present study were describing emotions that they believed they ought to have experienced, rather than emotions they remembered experiencing. Similarly, it is also possible that the psychopathic participants in the present study described the mental states of other people because they have learned that what others describe as a given emotional state is reflected in a distinct pattern of verbal cues and body language. Psychopaths can use this information to infer a mental state that they do not understand (Hare, 1999).

Related to the above points, it is possible that a selection bias influenced the results related to the likelihood of the psychopathic participants to discuss their own emotional experiences. It is probable that the inmates who were more likely to attend therapy and discuss the events they had experienced were also more likely to volunteer to participate in this type of research. Thus, the psychopathic inmates who would be more likely to discuss their feelings were more likely to participate in the present research.

Another possible explanation for the propensity of psychopaths in the present study to discuss their own emotional experiences is related to the concept of "proto-emotions," somewhat primitive responses associated with one's own needs and experiences. Psychopaths are thought to be likely to experience proto-emotions (Hare, 1999; Patrick, 2007). Thus, a psychopathic individual's displays of anger, hostility, envy, and frustration are likely to be much more intense and genuine than feelings of empathy, love, shame, and sorrow, all of which are deeper emotions. It is possible that many of the emotional experiences described by the psychopathic participants in the present study were proto-emotions. References to experiences of either a proto-emotion or a more complex emotion would add to a statement's credibility.

**Psychopathy and RM Credibility**

Psychopaths and nonpsychopaths did not differ in RM total scores in the present study. This was not surprising, since RM seemed to be related to metamemory rather than credibility in the present study, and research and theory suggest that psychopaths are likely to remember events with a similar amount of quality and quantity as nonpsychopaths (Christianson et al., 1996; Hervé et al., 2007).

None of the RM construct scores differed according to the psychopathy of the participants in the present study. In general, this was not surprising since none of the RM constructs seemed to be related to credibility in the present study. However, it had been expected that the Overall Impression Construct, which is concerned with the emotional experiences related to the event, would be scored lower in psychopaths than in nonpsychopaths. It is likely that this expected difference did not emerge for the same reasons that the Criterion 12 and 13 frequencies for affect did not differ according to

psychopathy. That is, the psychopathic participants have likely learned how to discuss their own and others' emotional states and answered the MCQ questions related to this appropriately in an effort to appear normal.

**Psychopathy and Verbal Credibility**

Altogether, there was little support for the fifth hypothesis. There was a nonsignificant tendency for the psychopathic participants to be more likely than the nonpsychopathic participants to provide an account that was partially confirmed with some contradictions or judged to be noncredible according to CBCA, especially for RV memories. It is expected that if a greater number of memory reports were partially confirmed with some contradictions or judged to be noncredible according to CBCA that the power of the comparisons would have increased to the point that these trends would have been significant.

**Limitations and Suggestions for Future Research**

There were a number of factors that limit the results of the present study. The most obvious drawback was the fact that consistency with file information data could not be attained for all of the memories. This was mainly due to the fact that the interviews were conducted between 2001 and 2004 and the consistency with file information data was collected between 2006 and 2007. In the years between, many of the participants had been released or had moved on to other institutions where it was no longer possible to access their correctional file information. As a result, a complete file review could be conducted for only 27 of the participants. For a further 67 participants, only a RADAR was available. Although the RADAR did contain some information on offences both inside and outside of correctional institutions, it did not contain detailed accounts of any

offences. This means that for some of the memories that were partially confirmed, this confirmation came from one line about the offences on the RADAR, and for others, there were several pages of information about the offences in the participants' CPRs. A memory for which only RADAR information was available could never be partially confirmed with some contradictions because there would not be enough information available on the RADAR for any contradicting facts. Also, it was not uncommon for critical events to be present in a participant's CPR but missing from the participant's RADAR. It is unfortunate that circumstances did not allow for the participants' CPRs to be reviewed at the time of their interviews so that file consistency could be determined for more of the memories.

A further limitation of the way consistency with file information was attained in the present study concerns the memories that were partially confirmed with some contradictions. When a number of details provided by a participant were contradicted in the participant's file, the participant was not telling a true and accurate story. However, it cannot be certain that the participant was lying. The participant could have been misremembering, which does not meet Ekman's (1996) definition of a lie (i.e., for a statement to be a lie, the person making the statement must know that he or she is not telling the truth). If the participant was misremembering, his statement would have the motivational and cognitive components of a true statement, which are thought to cause the CBCA criteria to appear in a statement. It should be noted that this should not be a problem with RM, as theoretically RM should distinguish between memories based on true events and memories based on events that did not actually happen (Johnson et al., 1988). The contradicted details in some of these memories suggest that they are

fabricated stories rather than misremembered stories. For example, one participant described being severely injured by the victim and a subsequent hospital visit after a reactively violent event. The police report in this participant's CPR indicated that he had not been injured and, rather than visiting the hospital after the assault, he was arrested and immediately taken to the police station. The discrepancies between this participant's memory report and the description of this event in his correctional file are consistent with overstating the reactivity of the event to give the impression that the participant was less guilty, in line with Porter and Woodworth's (2007) findings that offenders, especially psychopathic offenders, tend to exaggerate the reactivity of their homicides.

This highlights one of the limitations of field research in credibility assessment: Ground truth is often difficult, if not impossible, to ascertain. That consistency with file information was achieved for a proportion of the memories in the present study strengthens the present findings, as the utility of verbal credibility assessment tools could be measured against known truth in an applied setting. Future research examining credibility in violent offenders should ensure that correctional files are reviewed at the time of the research interview to increase the chances of determining the ground truth of the memories. The likelihood of determining the ground truth of the memories could also be raised if the interviewers suggested that participants share only memories for which they had been charged. Increasing the number of memory accounts for which consistency with file information could be established would increase the internal validity of field studies like the present study.

The present investigation was also limited by the fact that no other clues to credibility were examined. Research suggests that combining various clues to deception

improves the level of accuracy in distinguishing truth from lies (e.g., Vrij et al., 2000). While using video cameras to record the interviews was not feasible for the present project, future research should explore the combination of verbal clues to deception with physiological or nonverbal clues.

The results of the present investigation were also limited by the lack of motivation to lie or tell the truth. Although the present field study had clear forensic relevance because the participants were violent offenders and they were discussing violent crimes they had actually committed, the participants were aware that anything they talked about in the interview would have no impact on their sentencing. If the participants had believed that the apparent credibility of their memory accounts would affect their correctional sentences, they may have shared different memories or expressed their memories in different ways. This limits the ecological validity of the present results, since any time verbal credibility assessment might be used in a forensic context (e.g., psychological assessment, risk assessment, court testimony), the offender being assessed would likely be highly motivated to ensure that his story was believed. The lack of motivation to lie has specific limitations for the use of CBCA, since, as mentioned, the Undeutsch hypothesis is based on the assumption that a deceptive person's motivation to appear honest will cause the statement to be expressed differently than if the story was actually true. Theoretically, low motivation should make CBCA more difficult. This may contribute to the low number of memories in the present study that were found noncredible according to CBCA. The fact that some memories were found noncredible suggests that the lack of motivation in the present study did not negate the results.

However, CBCA likely would have been more effective at discriminating truthful and deceptive accounts if the participants had been more motivated to appear credible.

As mentioned previously, many of the present analyses were limited by a very sample size. Although the total sample size of 150 participants is large for a field study, so few memories were found noncredible according to CBCA and so few memories were partially confirmed or partially confirmed with some contradictions by file information that the power of the comparisons was greatly reduced. This means that it is likely that in some cases these tests did not detect differences that exist.

Finally, it should be mentioned that the Type-I error rate is potentially elevated due to the large number of analyses conducted. However, this is not a concern for a few reasons. First, most of the comparisons were hypothesized a priori. Second, the familywise error rate was taken into consideration and the alpha was reduced for most of the comparisons so that most significant differences in the present study were significant at the .001 level.

**Implications**

The previously mentioned weaknesses in this data set make it difficult to make many conclusions about this population. Despite these limitations, the results of the present investigation may have a number of implications for the criminal justice system, as well as for those interested in studying memory in this population.

Probably the most striking finding of the present study was that the vast majority of accounts of violent memories were judged to be credible. This knowledge may be valuable for any researcher interested in studying memory in offenders. It is a commonly held assumption that incarcerated offenders, especially those who have been diagnosed as

137

psychopaths are likely to lie about their experiences, even without much motivation (e.g.,

Doren, 1987; Granhag & Strömwall, 2004; Kropp & Rogers, 1993; Porter & Woodworth,

2007; Rogers & Cruise, 2000; Spidel et al., 2003). This has led some researchers to

conclude that the memory reports of offenders are not worthy of consideration in a

memory context, since it is unlikely that they are truthful (B. S. Cooper, 2008, personal

communication). The present study demonstrated that offenders can, and usually do,

generate credible accounts of events they have experienced. This suggests that there is no

reason that the memory accounts of this population should not be given the same

attention, with regards to memory research and theory, as the memory accounts of other

populations.

This finding also has implications for the criminal justice system, especially for

those who work with this population, or with a similar population, on a daily basis. That

is, professionals such as correctional officers, prison officials, probation and parole

officers, and mental health workers have no reason to expect that most offenders will lie

about their experiences with little or no motivation to lie. It would be helpful for these

professionals to know that offenders can provide truthful accounts, even when discussing

their offences.

The small number of memories judged to be noncredible and found to be partially

confirmed with some contradictions makes it difficult to make any conclusions about the

utility of RM or CBCA as credibility assessment tools. Although no firm conclusions can

be made due to the flaws in the data set, there was a tendency for CBCA to effectively

discriminate between accounts based on truthfulness, and no such tendency was evident

in the findings related to RM. Based on this tendency, it can be speculated that, given a

data set that included more noncredible memories or more file consistency data, CBCA would have demonstrated itself to be an effective verbal credibility assessment tool. If future research with incarcerated offenders does indeed demonstrate CBCA's utility in this context, this would have many implications for a variety of professionals in the criminal justice system, including mental health professionals within the prison context, mental health professionals working in other areas of the criminal justice system, law enforcement, triers of fact, probation and parole officers, customs and immigration officers, and security personnel. Credibility assessment could be applied to a variety of tasks where it would be important to know if potentially violent individuals are being truthful, such as risk assessments, security screening, and expert testimony.

The present results suggest that the self-report version of the MCQ should not be used as a credibility assessment tool, since RM showed no tendency to differentiate memories according to truthfulness. Credibility assessment is a complex task with many variables to consider and so it makes sense that a simple self-report questionnaire would not be effective at distinguishing truthful accounts from fabricated accounts.

As the present results regarding RM seemed to apply more to metamemory than to verbal credibility, the present results may also have implications for memory in violent offenders. The results of the present study supplement the results of Cooper's (2005; Cooper & Yuille, 2007) findings regarding memory for instrumental acts of violence compared to memory for reactive acts of violence. These findings can be applied to a variety of tasks performed by a forensic psychologist, including risk assessments and providing expert testimony. For example, based on the present results, in combination with theories on memory and some of the past research on the topic, experts could

discuss the fact that memories for instrumental acts of violence are generally more detailed, more vivid, and contain more information related to time information and events before and after than memories for reactive acts of violence. Further, during a risk assessment interview, if an offender provided a detailed description of a reactive murder but claimed to have little memory for an instrumental physical assault, there may be reason to be concerned about the credibility of this offender's statements.

The present results regarding psychopathy do not provide much evidence that psychopaths are more likely to be deceptive than nonpsychopaths, inconsistent with previous research (e.g., Kropp & Rogers, 1993; Porter & Woodworth, 2007; Rogers & Cruise, 2000; Spidel et al., 2003). Moreover, it seems that the credible accounts of psychopaths do not differ qualitatively, in terms of the CBCA criteria, from the credible accounts of nonpsychopaths. A psychopath is as likely to include any of the CBCA criteria, including descriptions of their own thoughts and feelings and the thoughts of feelings of others, in their credible accounts as a nonpsychopath. These findings suggest that psychopathic offenders do not have unique indicators of deception, independent of the ground truth of their statements. Additionally, the findings regarding RM and psychopathy suggest that psychopaths remember events with a similar amount of detail and vividness to nonpsychopaths. These findings can be applied to a variety of tasks performed by a forensic psychologist, such as risk assessments and providing expert testimony. For example, experts could discuss the fact that, although psychopaths may not feel complex emotions, they may report experiencing a variety of emotions.

Although the results of the present research may be useful in informing triers of fact about issues regarding verbal credibility, the present results offer no evidence that

either CBCA or RM is ready to be used as scientific evidence in legal systems. In the past, it has been argued that SVA should be used as forensic evidence in North American courts (Honts, 1994). Honts believed that the four Daubert criteria (*Daubert v. Merrell Dow Pharmaceuticals,* 1993), the set of guidelines for admitting expert evidence in US federal courts, were met by SVA: (1) It is testable and has been tested, (2) the error rate is known, (3) the procedure has been subject to peer review, and (4) the procedure is generally accepted in the scientific community (Honts, 1994). Others have argued that CBCA is not precise enough to be permitted in the courtroom (e.g., Lamb et al., 1997; Vrij, 2005). The present results are consistent with the latter viewpoint: Several issues remain unresolved for both RM and CBCA and should be taken into account before admitting either procedure in court. These issues include the lack of credibility decision or scoring rules for RM, the fact that theory and research results do not match for RM (i.e., the Retrospective Recollection Construct does not seem to be indicative of lack of credibility), the lack of precision for CBCA credibility decisions (i.e., although all of the memories that were partially confirmed were found to be credible according to CBCA, seven out of the eight memories that had some contradictions were also found to be credible), and the small number of memories found to be contradicted by file information did not allow for conclusive evidence that either CBCA or RM are reliable at distinguishing true accounts from fabricated accounts. Before it can be concluded that CBCA or RM is reliable enough to be admitted in court as evidence of verbal credibility, a number of ecologically valid studies should show that using each assessment tool can consistently improve assessors' abilities to distinguish truthful accounts from fabricated accounts to a level significantly better than chance. Further, for the results of a CBCA or

RM assessment to be maximally informative to triers of fact, the error rates for these techniques should be known, an issue that has not yet been resolved by researchers.

**Summary and Conclusions**

As mentioned, the present study was the first to provide a thorough qualitative examination of verbal credibility in violent offenders. The current results suggest that incarcerated offenders' descriptions of their perpetrated violent events are generally quite credible. These results show that CBCA had a tendency to be effective at distinguishing memory accounts of acts of perpetrated violence according to truthfulness. There was some evidence that CBCA seemed to be limited by a truth bias and appeared to be less effective when deceptive accounts were based on some truth, but more research needs to be conducted before any firm conclusions can be made.

The present results also demonstrated that RM, as measured by the MCQ, did not seem to be effective at discriminating accounts according to truthfulness. Future research should look into the reasons for RM's ineffectiveness in this context. Specifically, various methods of measuring RM in a statement should be compared to determine which methods are most effective at credibility assessment. Further, research should look into which RM constructs are most effective at distinguishing truthful accounts and fabricated accounts. If future research continues to find that the Retrospective Recollection Construct is not associated with lack of credibility, perhaps the recommendations surrounding this construct should be altered to account for these findings.

RM discriminated between types of violent memory, but CBCA did not. This suggests that, although IV memories were remembered better than RV memories, there

was no evidence to suggest that they are more likely to appear credible than RV memories.

The present results also demonstrated that the credible accounts of psychopaths do not differ qualitatively from the credible accounts of nonpsychopaths. Additionally, psychopaths remembered the events they had experienced as well as nonpsychopaths. Hopefully future research will explore the interactions between psychopathy, credibility ratings, and ground truth.

Although the present results were weakened by flaws in the data set, they are strengthened by the efforts to increase internal validity through the assessment of consistency with file information. The results of the present study are further strengthened by the field approach of the present investigation. No other approach would have allowed for the consideration of these types of participants and these types of memories. Thus, the ecological validity of this study is much stronger than that of laboratory studies. This is especially true for the way CBCA was examined in the present study, since two of the three SVA components were used. Most field research considers transcripts of interviews by legal professionals, interviews that often contain many closed questions (e.g., Buck et al., 2002), which limits the effectiveness of CBCA. There is a very small body of field research on verbal credibility assessment (e.g., Lamb et al., 1997; Orbach & Lamb, 1999), especially field research on offenders or suspects (Lee et al., 2008; Mann, Vrij, & Bull, 2004; Vrij & Mann, 2001). It is hoped that future research will build on this body of research, to further investigate how credibility can be determined in the reports of violent offenders. In particular, a study that examined CBCA and RM in witness or offender statements that were attained in a forensic context (e.g.,

police interview), which would presumably result in the statement providers being highly motivated to have their account appear truthful; used a non-leading interview method, such as the cognitive interview or the Step Wise Interview; and for which other information about the events in question was available so that ground truth could be established, would be an ideal method to test the efficacy of these credibility assessment methods. This method would be improved even further if the validity checklist was examined, as well.

# References

Akehurst, L., Koehnken, G., Hoefer, E. (2001). Content credibility of accounts derived from live and video presentations. *Legal and Criminological Psychology, 6,* 65-83.

Alonso-Quecuty, M. L. (1992). Deception detection and reality monitoring: A new answer to an old question? In F. Lösel, D. Bender, & T. Bliesener (Eds.), *Psychology and Law: International Perspectives* (pp. 328-332). Berlin: Walter de Gruyter.

Barnier, A. J., Sharman, S. J., McKay, L., & Sporer, S. L. (2005). Discriminating adults' genuine, imagined, and deceptive accounts of positive and negative childhood events. *Applied Cognitive Psychology, 19,* 985-1001.

Ben-Shakkar, G., Bar-Hillel, M., & Kremnitzer, M. (2002). Trial by polygraph: Reconsidering the use of the guilty knowledge technique in court. *Law and Human Behavior, 26,* 527-541.

Berkowitz, L. (1983). The experience of anger as a parallel process in the display of impulsive, "angry" aggression. In R. G. Green & E. I. Donnerstein (Eds.), *Aggression:  Theoretical and empirical views* (Vol. 1, pp. 103-134). New York: Academic Press.

Berkowitz, L. (1990). On the formation and regulation of anger and aggression: A cognitive-neoassociationistic analysis. *American Psychologist, 45*, 494-503.

Blaney, P. H., & Davis, R. D. (1999). *Oxford Textbook of Psychopathology.* New York : Oxford University Press.

Bond, G. D., & Lee, A. Y. (2005). Language of lies in prison: Linguistic classification of

    prisoners' truthful and deceptive natural language. *Applied Cognitive Psychology,*

    *19,* 313-329.

Book, A. S., Holden, R. R., Starzyk, K. B., Wasylkiw, L., & Edwards, M. J. (2006).

    Psychopathic traits and experimentally induced deception in self-report

    assessment. *Personality and Individual Differences, 41,* 601-608.

Brown, S.L. & Forth, A.E. (1997). Psychopathy and sexual assault: Static factors,

    emotional precursors, and rapist subtypes. *Journal of Consulting and Clinical*

    *Psychology, 65*, 848-857.

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.

    *Applied Developmental Psychology, 23,* 267-283.

Bushman, B. J. & Anderson, C. A. (2001). Is it time to pull the plug on the hostile versus

    instrumental aggression dichotomy? *Psychological Review, 108*, 273-279.

Ceci, S. J., & Bruck, M. (1993). The suggestibility of the child witness: A historical

    review and synthesis. *Psychological Bulletin, 113*, 403-439.

Ceci, S. J., Ross, D. F., & Toglia, M.P. (1987). Suggestibility of children's memory:

    Psycholegal implications. *Journal of Experimental Psychology: General, 116,* 38-

    49.

Chase, K. A., O'Leary, K. D., Heyman, R. E. (2001). Categorizing partner-violent men

    within the reactive-proactive typology model. *Journal of Consulting and Clinical*

    *Psychology , 69*, 567-572.

Christianson, S-A., Forth, A. E., Hare, R. D., Strachan, C., Lidberg, L., & Lars-Hakan, T. (1996). Remembering details of emotional events: A comparison between psychopathic and non-psychopathic offenders. *Personality and Individual Differences, 20*, 437-443.

Clark, C. R. (1997). Sociopathy, malingering, and defensiveness. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (2nd ed., pp. 68-84), New York: Guilford Press.

Cohen, R. L., & Harnick, M. A. (1980). The susceptibility of child witnesses to suggestion: An empirical study. *Law and Human Behavior, 4*, 201-210.

Colwell, K., Hiscock, C. K., & Memon, A. (2002). Interviewing techniques and the assessment of statement credibility. *Applied Cognitive Psychology, 16,* 287-300.

Cooke, D. J. (1998). Cross-cultural aspects of psychopathy. In T. Millon, E. Simonsen, M. Bicket-Smith, & R. D. Davis (Eds.), *Psychopathy: Antisocial, criminal, and violent behaviour,* (pp. 260-276). New York: The Guilford Press.

Cooper, B. S. (2005). *Memory for mayhem.* Unpublished Doctoral Dissertation. The University of British Columbia, Vancouver, British Columbia, Canada.

Cooper, B. S., Griesel, D., & Yuille, J. C. (2007). Clinical-forensic risk assessment: The past and current state of affairs. *Journal of Forensic Psychology Practice, 7,* 1-64.

Cooper, B. S., & Yuille, J. C. (2007). An investigation of violent offenders' memories for instrumental and reactive violence. In S. A. Christianson (Ed.) *Offenders' Memories of Violent Crimes* (pp. 75-97). West Sussex, England: John Wiley & Sons.

Cooper, B. S., & Yuille, J. C. (2007). Psychopathy and deception. In H. Hervé & J. C.

 Yuille (Eds.), *The Psychopath: Theory, Research, and Practice* (pp. 487-503).

 Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Cornell, P. G., Warren, J., Hawk, G., Stafford, E., Oram, G., & Pine, D. (1996).

 Psychopathy in instrumental and reactive violent offenders. *Journal of Consulting*

 *and Clinical Psychology, 64*, 783-790.

Daubert v. Merrell Dow Pharmaceuticals (92-102), 509 U.S. 579 (1993).

DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychological*

 *Bulletin, 111,* 203-243.

DePaulo, B. M., Kashy, D. A, Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996).

 Lying in everyday life. *Journal of Personality and Social Psychology, 74,* 63-79.

DePaulo, B. M., & Kirkendol, S. E. (1989). The motivational impairment effect in the

 communication of deception. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 51-

 70). Boston: Kluwer Academic Publishers.

DePaulo, B. M., Lanier, K., & Davis, T. (1983). Detecting the deceit of the motivated

 liar. *Journal of Personality and Social Psychology, 45,* 1096-1103.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper,

 H. (2003). Cues to deception. *Psychological Bulletin, 129,* 74-118.

deTurck, M. A., & Miller, G. R. (1990). Training observers to detect deception: Effects

 of self-monitoring and rehearsal. *Human Communication Research, 16,* 603-620.

DeVellis, R. F. *Scale Development: Theory and Applications* (2nd Ed.). Thousand Oaks,

 CA: Sage Publications.

Doren, D. M. (1987). *Understanding and treating the psychopath.* New York: Wiley.

Dutton, D. G. (1995). *The batterer*. New York: Basic Books.

Ekman, P. (2002). *Telling lies: Clues to deceit in the marketplace, politics, and marriage* (3rd ed.). New York: Norton.

Ekman, P. (2001). *Telling lies.* New York: W.W. Norton & Company.

Ekman, P. (1997). Lying and deception. In N. L. Stein, P. A. Ornstein, B. Tversky, & C. Brainerd (Eds.), *Memory for everyday and emotional events.* Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Ekman, P. (1996). Why don't we catch liars? *Social Research, 63,* 801-817.

Ekman, P. (1992). Facial expressions of emotion: New findings, new questions. *Psychological Science, 3,* 34-38.

Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior, 6,* 238-252.

Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System.* Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., & Friesen, W. V. (1974). Detecting deception from the body or face. *Journal of Personality and Social Psychology, 29,* 288-298.

Ekman, P., & Friesen, W. V. (1972). Hand movements. *The Journal of Communication, 22,* 353-374.

Ekman, P. Friesen, W. V., & O'Sullivan, M. (1997). Smiles when lying. In P. Ekman & E. L. Rosenberg (Eds.), *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)* (pp. 201-214). New York: Oxford University Press.

Etcoff, N. L., Ekman, P., Magee, J. J., & Frank, M. G. (2000). Lie detection and language comprehension. *Nature, 405,* 139-141.

Fiedler, K., Schmidt, J., & Stahl, T. (2002). What is the current truth about polygraph lie detection? *Basic and Applied Social Psychology, 24,* 313-324.

Fisher, R. P., & Geiselman, R. E. (1992). *Memory-enhancing techniques for investigative interviewing: The Cognitive Interview.* Springfield, IL: Charles C. Thomas.

Fisher, R. P., & Shreiber, N. (2007). Interview protocols for improving eyewitness memory. In M. P. Toglia, J. D. Read, D. F. Ross, & R. C. L. Lindsay (Eds.), *The Handbook of Eyewitness Psychology: Volume I: Memory for Events* (pp. 53-80). Mahwah, NJ: Erlbaum Associates.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7,* 286-299.

Frank, M. G., & Ekman, P. (1997). The ability to detect deceit generalizes across different types of high-stake lies. *Journal of Personality and Social Psychology, 72,* 1429-1439.

Fulero, S. M. (1995). Review of the Hare Psychopathy Checklist-Revised. In J. C.Conoley & J. C. Impara (Eds.), *Twelfth Mental Measurements Yearbook* (pp. 453-454). Lincoln, NE: Buros Institute.

Gale, A. (Ed.). (1988). *The polygraph test: Lies, truth, and science.* London: Sage.

Gardiner, J. M., & Richardson-Klavehn, A. (2000). Remembering and knowing. In E. Tulving & F.I.M. Craik (Eds.), *Handbook of memory* (pp. 229-244). New York: Oxford University Press.

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

Geiselman, E. R., Fisher, R. P., Cohen, G., Holland, H., Surtes, L. (1986). Eyewitness responses to leading and misleading questions under the cognitive interview. *Journal of Police Science and Administration, 14*, 31-39.

Geiselman, R.E., Fisher, R.P., Firstenberg, I., Hutton, L.A., Sullivan, S., Avetissian, I., & Prosk, A. (1984). Enhancement of eyewitness memory: An empirical evaluation of the cognitive interview. *Journal of Police Science and Administration, 12,* 74-80.

Geisleman, 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, 2*, 401-412.

Glass, G. V., & Hopkins, K. D. (1996). *Statistical methods in education and psychology.* Needham Heights, MA: Allyn & Bacon.

Gödert, H. W., Gamer, M., Rill, H. -G., & Vossel, G. (2005). Statement validity assessment: Inter-rater reliability of criteria-based content analysis in the mock-crime paradigm. *Legal and Criminological Psychology, 10,* 225-245.

Granhag, P. A., & Strömwall, L. A. (2004). Research on deception detection: Past and present. In P. A. Granhag & L. A. Strömwall (Eds.). *The Detection of Deception in Forensic Contexts* (pp. 3-12), Cambridge, UK: Cambridge University Press.

Granhag, P. A., Strömwall, L. A., & Landström, L. A. (2006). Children recalling an event repeatedly: Effects on RM and CBCA scores. *Legal and Criminological*

*Psychology, 11,* 81-98.

Gudjosson, G. H., & Clark, N. K. (1986). Suggestibility in police interrogation: A social

psychological model. *Social Behavior, 1*, 83-104.

Hansen, H. (1998). Treating the "untreatable" in Denmark: Past and present. In T.

Millon, E. Simonsen, M. Birket-Smith, & R. D. Davis (Eds.), *Psychopathy:*

*Antisocial, criminal, and violent* (pp. 458-462). New York: The Guilford Press.

Hare, R. D. (2003). *The Hare Psychopathy Checklist-Revised, Second Edition.* Toronto,

Canada: Multi-Health Systems.

Hare, R. D. (1991). *The Hare Psychopathy Checklist-Revised.* Toronto, Canada: Multi-

Health Systems.

Hare, R. D. (1980). A research scale for the assessment of psychopathy in criminal

populations. *Personality and Individual Differences, 1,* 111-117.

Hare, R. D. (1970). *Psychopathy: Theory and Research.* Toronto: John Wiley & Sons,

Inc.

Hare, R. D. (1999). *Without Conscience: The Disturbing World of the Psychopaths*

*Among Us.* New York: Guilford Press.

Hare, R. D., Forth, A. E., & Hart, S. D. (1989). The psychopath as prototype for

pathological lying and deception. In J. C. Yuille (Ed.), *Credibility Assessment* (pp.

25-49)*.*  Boston: Kluwer Academic Publishers.

Hare, R. D., Harpur, T. J., Hakstian, A. R., Forth, A. E., Hart, S. D., & Newman, J. P.

(1990). Revised Psychopathy Checklist: Reliability and factor structure.

*Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2*,

338-341.

Harpur, T. J., Hakstian, A. R., & Hare, R. D. (1988). Factor structure of the Psychopathy Checklist. *Journal of Consulting and Clinical Psychology, 56*, 741-747.

Harris, G. T., Rice, M., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior, 20,* 315-335.

Hart, S. D., & Dempster, R. J. (1997). Impulsivity and psychopathy. In C. D. Webster & M. A. Jackson (Eds.), *Impulsivity; Theory, assessment, and treatment*, (pp. 212-232). New York: The Guilford Press.

Hart, S. D. & Hare, R. D. (1997). Psychopathy: Assessment and association with criminal conduct. In D. M. Stoff, J. Breiling, & J. D. Maser (Eds.), *Handbook of antisocial behavior* (pp. 22-35), New York: Wiley.

Hart, S. D., Hare, R. D., & Harpur, T. J. (1992). The Psychopathy Checklist-Revised (PCL-R): An overview for researchers and clinicians. In J.C. Rosen & P. McReynolds (Eds.), *Advances in psychological assessment, Volume 8* (pp. 103-130), New York: Plenum Press.

Hemphill, J. F., Templeman, R., Wong, S., & Hare, R. D. (1998). Psychopathy and crime: Recidivism and criminal careers. In D.J. Cooke et al. (Eds.), *Psychopathy: Theory, research and implications for society* (375-399), Kluwer Academic Publishers. Printed in the Netherlands.

Hernandez-Fernaud, E. & Alonso-Quecuty, M. (1997). The Cognitive Interview and lie detection: a new magnifying glass for Sherlock Holmes? *Applied Cognitive Psychology, 11,* 55-68.

Hershkowitz, I. (1999). The dynamics of interview involving plausible and implausible

    allegations of child sexual abuse. *Applied Developmental Science, 3,* 86-91.

Hertzog, C. (1992). Improving memory: The possible roles of metamemory. In D.

    Herrmann, H. Weingartner, A. Searleman, & C. McEvoy (Eds.). *Memory*

    *Improvement: Implications for Memory Theory* (pp. 61-78). New York: Springer-

    Verlag.

Hervé, H.F., Cooper, B.S., & Yuille, J.C. (June, 2002). *A biopsychosocial model of*

    *eyewitness memory*. Paper presented at the Canadian Psychological Association's

    Annual Convention, Vancouver, British Columbia.

Hervé, H.F., Cooper, B.S., & Yuille, J.C. (2007). Memory formation in offenders:

    Perspectives from a biopsychosocial model of eyewitness memory. In S. A.

    Christianson (Ed.) *Offenders' Memories of Violent Crimes* (pp. 38-74). West

    Sussex, England: John Wiley & Sons.

Hervé, H.F., Cooper, B.S., Yuille, J.C., & Daylen (July, 2003). *The psychopathic*

    *eyewitness: Perspectives from a biopsychosocial model of eyewitness memory*.

    Paper presented at the Society for Applied Research on Memory and Cognition's

    (SARMAC) conference, Aberdeen, Scotland.

Hess, U., & Kleck, R. E. (1997). Differentiating emotions elicited and

    deliberateemotional facial expressions. In P. Ekman & E. Rosenberg (Eds.), *What*

    *the face reveals: Basic and applied studies of spontaneous expression using the*

    *Facial Action Coding System (FACS)* (pp. 271-288). New York: Oxford

    University Press.

Honts, C.R. (1994). Assessing children's credibility: Scientific and legal issues in 1994. *North Dakota Law Review, 70,* 879-903.

Honts, C. R., & Perry, M. V. (1992). Polygraph admissibility: Changes and challenges. *Law and Human Behavior, 16,* 357-379.

Honts, C. R., Raskin, D. C., & Kircher, J. C. (1994). Mental and physical countermeasures reduce the accuracy of polygraph tests. *Journal of Applied Psychology, 17,* 252-259.

Horowitz, S.W. (1991). Empirical support for statement validity assessment. *Behavioral Assessment,13*, 293-313.

Hubbard, J. A., Dodge, K. A., Cillessen, A. H. N., Coie, J. D., & Schwartz, D. (2001). The dyadic nature of social information processing in boy's reactive and proactive aggression. *Journal of Personality and Social Psychology, 80*, 268-280.

Huet, N., & Mariné, C. (2002). Metamemory assessment and memory behaviour in a simulated memory professional task. *Contemporary Educational Psychology, 22,* 507-520.

Iacono, W. G., & Lykken, D. T. (1997). The validity of the lie detector: Two surveys of scientific opinion. *Journal of Applied Psychology, 82,* 426-433.

Inbau, F. E., Reid, J. E., Buckley, J. P., & Jayne, B. C. (2004). *Essentials of the Reid Technique: Criminal Interrogation and Confessions.* Sudbury, MA: Jones & Bartlett Publishers.

Joffe, R. D. (1992). Criteria-Based Content Analysis: An experimental investigation with children. Unpublished doctoral dissertation, Department of Psychology, University of British Columbia.

Johnson, M. K. (1988). Reality monitoring: An experimental phenomenological approach. *Journal of Experimental Psychology: General, 117,* 390-394.

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,* 371-376.

Johnson, M. K., Hastroudi, S., & Lindsay, D. S. (1993). Source monitoring. P*sychological Bulletin, 114,* 3-28.

Johnson, M. K., & Raye, C. L. (1988). False memory and confabulation. *Trends in Cognitive Sciences, 2,* 137-145.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88,* 67-85.

Kassin, S. M. (2004). True or false: 'I'd know a false confession if I saw one.' In P. A. Granhag & L. A. Strömwall (Eds.). *The Detection of Deception in Forensic Contexts* (pp. 172-194), Cambridge, UK: Cambridge University Press.

Kassin, S. M., & Gudjonsson, G. H. (2004). The psychology of confession evidence: A review of the literature and issues. *Psychological Science in the Public Interest, 5*, 31-67.

Klaver, J. R., Lee, Z., & Hart, S. D. (2007). Psychopathy and nonverbal indicators of deception in offenders. *Law and Human Behavior, 31,* 337-351.

Koehnken, G. (2004). Statement Validity Analysis and the 'detection of the truth.' In P. A. Granhag & L. A. Strömwall (Eds.). *The Detection of Deception in Forensic Contexts* (pp. 41-63), Cambridge, UK: Cambridge University Press.

Koehnken, G., Schimossek, E., Aschermann, E., & Hoefer, E. (1995). The Cognitive
    Interview and the assessment of the credibility of adults' statements. *Journal of
    Applied Psychology, 80,* 671-684.

Koehnken, G., & Steller, M. (1988). *Issues in criminological and legal psychology* (No.
    13). Leicester, England: British Psychological Society.

Kropp, P. R., & Rogers, R. (1993). Understanding malingering: Motivation, method, and
    detection. In M. Lewis & C. Saarni (Eds.), *Deception and lying in everyday life.*
    New York: Guilford.

Lamb, M.E., Sternberg, K.J., Esplin, P.W., Hershkowitz, I., & Orbach, Y. (1997).
    Assessing the credibility of children's allegations of sexual abuse: A survey of
    recent research. *Learning and Individual Differences, 9,* 175-194.

Lamb, M. E., Sternberg, K. J., Esplin, P. W., Hershkowitz, I., Orbach, Y., & Hovav, M.
    (1997). Criterion-Based Content Analysis: A field validation study. *Child Abuse
    & Neglect, 21,* 255-264.

Landry, K., & Brigham, J. (1992). The effect of training in criteria-based content analysis
    on the ability to detect deception in adults. *Law and Human Behavior, 16,* 663-
    676.

Lee, Z., Klaver, J. R., & Hart, S. D. (2008). Psychopathy and verbal indicators of
    deception in offenders. *Psychology, Crime & Law, 14,* 73-84.

Leis, T. A., Motiuk, L. L., & Ogloff, J. R. P. (1995). *Forensic psychology: Policy and
    practice in corrections.* Correctional Service of Canada: Minister of Supply and
    services Canada.

Lindsay, D. S. & Read, J. D. (1995). "Memory work" and recovered memories of

    childhood sexual abuse: Scientific evidence and public, professional, and personal

    issues. *Psychology, Public Police, and Law, 1,* 846-908.

Loftus, E. F. (1993). The reality of repressed memories. *American Psychologist, 48,* 518-

    537.

Lykken, D. T. (1995). *The antisocial personalities.* Hillsdale, NJ: Lawrence Erlbaum

    Associates.

MacMillan, J., & Kofoed, L. (1984). Sociobiology and antisocial personality. *The*

    *Journal of Nervous and Mental Disease, 172,* 701-706.

Mann, S., Vrij, A., & Bull, R. (2002). Suspects, lies and videotape: An analysis of

    authentic high-stakes liars. *Law and Human Behavior, 26,* 365-376.

Masip, J., Garrido, E., & Herrero, C. (2004). Defining deception. *Anales de psicologia,*

    *20,* 147-171.

Masip, J., Sporer, S. L., Garrido, E., & Herrero, C. (2005). The detection of deception

    with the reality monitoring approach: A review of the empirical evidence.

    *Psychology, Crime & Law, 11,* 99-122.

Mather, M., Henkel, L. A., & Johnson, M. K. (1997). Evaluating characteristics of false

    memories: Remember/know judgments and memory characteristics questionnaire

    compared. *Memory & Cognition, 25,* 826-837.

McGinnis, D., & Roberts, P. (1996). Qualitative characteristics of vivid memories

    attributed to real and imagined experiences. *The American Journal of Psychology,*

    *109,* 59-77.

Mealey, L. (1995). The sociobiology of sociopathy: An integrated evolutionary model. *Behavioral and Brain Sciences, 18,* 523-599.

Memon, A., Vrij, A., & Bull, R. (1998). *Psychology and law: Truthfulness, accuracy and credibility*. London: McGraw-Hill publishing company.

Morgan, R. D., & Flora, D. B. (2002). Group psychotherapy with incarcerated offenders: A research synthesis. *Group Dynamics: Theory, Research, and Practice, 6,* 203-218.

Orbach, Y., & Lamb, M. E. (1999). Assessing the accuracy of a child's account of sexual abuse: A case study. *Child Abuse & Neglect, 23,* 91-98.

Ost, J., Vrij, A., Costall, A., & Bull, R. (2002). Crashing memories and reality monitoring: Distinguishing between perceptions, imaginations, and 'false memories.' *Applied Cognitive Psychology, 16,* 125-134.

Parker, A. & Brown, J. (2000). Detection of deception: Statement Validity Analysis as a means of determining truthfulness or falsity of rape allegations. *Legal and Criminological Psychology, 5,* 237-259.

Patrick, C. J. (2007). Getting to the heart of psychopathy. In H. Hervé & J. C. Yuille (Eds.), *The Psychopath: Theory, Research, and Practice* (pp. 207-252). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Peticlerc, A.M., Hervé, H.F., Hare, R.D., & Spidel, A. (June, 2000). *Psychopaths' reasons to deceive*. Poster presented at the Canadian Psychological Association's 61st annual convention.

Porter, S., Birt, A. R., & Boer, D. P. (2001). Investigation of the criminal and conditional

  release profiles of Canadian federal offenders as a function of psychopathy and

  age. *Law and Human Behavior, 25,* 647-661.

Porter, S., Birt, A. R., Yuille, J. C., & Hervé, H. F. (2001). Memory for murder: A

  psychological perspective on dissociative amnesia in legal contexts. *International*

  *Journal of Law and Psychiatry, 24,* 23-41.

Porter, S. & Campbell, M. A. (2002). Pinpointing reality: How well can people judge true

  and mistaken emotional childhood memories? *Canadian Journal of Behavioural*

  *Science, 34,* 217-229.

Porter, S., Campbell, M. A., Stapelton, J., & Birt, A. R. (2002). The influence of judge,

  target, and stimulus characteristics on accuracy of detecting deceit. *Canadian*

  *Journal of Behavioural Science, 34,* 172-196.

Porter, S., Doucette, N. L., Woodworth, M., Earle, J., & MacNeil, B. (2008). Halfe the

  world knowes not how the other halfe lies: Investigation of verbal and non-verbal

  signs of deception exhibited by criminal offenders and non-offenders. *Legal and*

  *Criminological Psychology, 13,* 27-38.

Porter, S., & Woodworth, M. (2007). "I'm sorry I did it…but he started it": A

  comparison of the official and self-reported homicide descriptions of psychopaths

  and non-psychopaths. *Law and Human Behavior.31*, 91-107.

Porter, S., & Woodworth, M. (2006). Psychopathy and aggression. In C. Patrick (Ed.),

  *Handbook of psychopathy* (481-494). New York: Guilford.

Porter, S., & Yuille, J. C. (1996). The language of deceit: An investigation of the verbal cues to deception in the interrogation context. *Law and Human Behavior, 20,* 443-458.

Porter, S., Yuille, J. C., & Bent, A. (1995). A comparison of the eyewitness accounts of deaf and hearing children. *Child Abuse and Neglect, 19,* 51-61.

Poulin, F. & Boivin, M. (2000). Reactive and proactive aggression: Evidence of a two-factor model. *Psychological Assessment, 12*, 115-122.

Poythress, N. G., Edens, J. F., & Watkins, M. M. (2001). The relationship between psychopathic personality features and malingering symptoms of major mental illness. *Law and Human Behavior, 25,* 567-582.

Raskin, D. C. (1987). Methodological issues in estimating polygraph accuracy in field applications. *Canadian Journal of Behavioural Science, 19,* 389-404.

Raskin, D. C., & Esplin, P. W. (1991). Statement Validity Assessment: Interview procedures and content analysis of children's statements of sexual abuse. *Behavioral Assessment, 12,* 265-291.

Raskin, D. C., & Hare, R. D. (1978). Psychopathy and detection of deception in a prison population. *Psychophysiology, 15,* 126-136.

Raskin, D. C., Kircher, J. C., Horowitz, S. W., & Honts, C. R. (1989). Recent laboratory and field research on polygraph techniques. In J. C. Yuille (Ed.), *Credibility Assessment.* Dordrecht, Netherlands: Kluwer Academic Publishers.

Rassin, E. (2000). Criteria Based Content Analysis: The less scientific road to truth. *Expert Evidence, 7,* 265-278.

Roediger, H. L., & Marsh, E. L. (2003). Episodic and autobiographical memory. In I. B.

    Weiner (Series Ed.) & A. F. Healy & R. W. Proctor (Vol. Eds.), *Handbook of*

    *Psychology: Volume 4: Experimental Psychology* (pp. 475-498). Hoboken, NJ:

    Wiley.

Roediger, H. L., Marsh, E. J., & Lee, S. C. (2001). Varieties of memory. In D. L. Medin

    & H. Pashler (Eds.), *Steven's Handbook of Experimental Psychology, Third*

    *Edition, Volume 2: Memory and Cognitive Processes* (pp. 1-41). New York: John

    Wiley & Sons.

Rogers, R., & Cruise, K. R. (2000). Malingering and deception among psychopaths. In C.

    B. Gacono (Ed.), *The clinical and forensic assessment of psychopathy: A*

    *practitioner's guide* (pp. 269-284). Mahwah, NJ: Lawrence Erlbaum Associates.

Ruby, C. L., & Brigham, J. C. (1997). The usefulness of the Criteria-Based Content

    Analysis technique in distinguishing between truthful and fabricated allegations:

    A critical review. *Psychology, Public Policy, and Law, 3,* 705-737.

Salfati, C. G. (2000). The nature of expressiveness and instrumentality in homicide.

    *Homicide Studies, 4*, 265-293.

Santtila, P., Roppola, H., Runtti, M., & Niemi, P. (2000). Assessment of child witness

    statements using criteria-based content analysis (CBCA): The effects of age,

    verbal ability, and interviewer's emotional style. *Psychology, Crime & Law, 6,*

    159-179.

Santtila, P., Roppola, H., & Niemi, P. (1999). Assessing the truthfulness of witness

    statements made by children (aged 7-8, 10-11, and 13-14) employing scales

derived from Johnson and Raye's model of reality monitoring. *Expert Evidence, 6,* 273-289.

Saxe, L., & Ben-Shakkar, G. (1999). Admissibility of polygraph tests: The application of scientific standards post-Daubert. *Psychology, Public Policy, and Law, 5*, 203-223.

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

Seniuk, G. T. G., & Yuille, J. C. (1996). *Fact finding and the judiciary.* Saskatoon: Commonwealth of Learning.

Serin, R.C. (1991). Psychopathy and violence in criminals. *Journal of Interpersonal Violence, 6*, 423-431.

Seto, M. C., & Barbaree, H. E. (1999). Psychopathy, treatment behavior, and sex offender recidivism. *Journal of Interpersonal Violence, 14,* 1235-1248.

Seto, M. C., Khattar, N. A., Lalumière, M. L., & Quinsey, V. L. (1997). Deception and sexual strategy in psychopathy. *Personality and Individual Differences, 22,* 301-307.

Smeets, T., Jelicic, M., Peters, M. J. V., Candel, I., Horselenberg, R., & Merckelbach, H. (2006). 'Of course I remember seeing that film' – How ambiguous questions generate crashing memories. *Applied Cognitive Psychology, 20,* 779-789.

Spidel, A., Hervé, H. F., Greaves, C., Cooper, B. S., & Hare, R. D. (2003). Psychopathy and deceptive motivations in young offenders. In M. Vanderhallen, G. Vervaeke,

P. J. Van Koppen, & J. Goethals (Eds.), *Much ado about crime: Chapters on psychology and law* (pp. 265-276). Braxelles: Politeia.

Sporer, S. L. (2004). Reality monitoring and deception detection. In P. −A. Granhag and L. Strömwall (Eds.), *Deception Detection in Forensic Contexts* (pp. 64-102). Cambridge: Cambridge University Press.

Sporer, S. L. (1997). The less travelled road to truth: Verbal cues in deception detection in accounts of fabricated and self-experienced events. *Applied Cognitive Psychology, 11,* 373-397.

Sporer, S. L., & Schwandt, B. (2006). Paraverbal indicators of deception: A meta-analytic synthesis. *Applied Cognitive Psychology, 20,* 421-446.

Sporer, S. L., & Sharman, S. J. (2006). Should I believe this? Reality monitoring of accounts of self-experienced and invented recent and distant autobiographical events. *Applied Cognitive Psychology, 20,* 837-54.

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

Steller, M., & Koehnken, G. (1989). Statement Analysis: Credibility assessment of children's testimonies in sexual abuse cases. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence.* New York: Springer.

Stone, G. L. (1995). Review of the Hare Psychopathy Checklist-Revised. In J. C. Conoley & J. C. Impara (Eds.), *Twelfth Mental Measurements Yearbook* (pp. 454-455). Lincoln, NE : Buros Institute.

Strömwall, L. A., Bengtsson, L., Leander, L., & Granhag, P. A. (2004). Assessing children's statements: The impact of a repeated experience on CBCA and RM ratings. *Applied Cognitive Psychology, 18,* 653-668.

Suengas, A. G., & Johnson, M. K. (1988). Qualitative effects of rehearsal on memories for perceived and imagined complex events. *Journal of Experimental Psychology: General, 117,* 377-389.

Swihart, G., Yuille, J.C., & Porter, S. (1999). The role of state dependent memory in "redouts." *International Journal of Law and Psychiatry, 125*, 199-212.

Takahashi, M., & Shimizu, H. (2007). Do you remember the day of your graduation ceremony from junior high school?: A factor structure of the Memory Characteristics Questionnaire. *Japanese Psychological Research, 49,* 275-81.

Taylor, P. J., & Kopelman, M. D. (1984). Amnesia for criminal offences. *Psychological Medicine, 14,* 581-588.

Tweed, R. G., & Dutton, D. G. (1998). A comparison of impulsive and instrumental subgroups of batterers. *Violence & Victims, 13*, 217-230.

Tye, M. C., Amato, S. L., Honts, C. R., Devitt, M. K., & Peters, D. (1999). The willingness of children to lie and the assessment of credibility in an ecologically relevant laboratory setting. *Applied Developmental Science, 3,* 92-110.

Undeutsch, U. (1984). Courtroom evaluation of eyewitness testimony. *International Review of Applied Psychology, 33,* 51-67.

Undeutsch, U. (1989). The development of Statement Reality Analysis. In J. C. Yuille (Ed.), *Credibility Assessment* (pp. 101-120). Dordrecht, The Netherlands: Kluwer Academic Publisher.

Vrij, A. (2005). Criteria-Based Content Analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law,* 3-41.

Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and the implications for professional practice.* Chichester: Wiley.

Vrij, A. (1994). The impact of information and setting on detection of deception by police detectives. *Journal of Nonverbal Behavior, 18,* 117-136.

Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2004). Detecting deceit via analyses of verbal and nonverbal behaviour in children and adults. *Human Communication Research, 30, 8-41.*

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 Behavor, 26,* 261-283.

Vrij, A., Edward, K., & Bull, R. (2001). Stereotypical verbal and nonverbal responses while deceiving others. *Personality and Social Psychology Bulletin, 27,* 899-909.

Vrij, A., Edward, K., Roberts, K. P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior, 24,* 239-263.

Vrij, A., & Graham, S. (1997). Individual differences between liars and the ability to detect lies. *Expert Evidence, 5,* 144-148.

Vrij, A., & Mann, S. (2006). Criteria-Based Content Analysis: An empirical test of its underlying processes. *Psychology, Crime & Law, 12,* 337-349.

Vrij, A., & Mann, S. (2001). Telling and detecting lies in a high-stake situation: The case of a convicted murderer. *Applied Cognitive Psychology, 15,* 187-203.

Vrij, A., Mann, S. A., Fisher, R. P., Leal, S., Milne, R., & Bull, R. (2008). Increasing

cognitive load to facilitate lie detection: The benefit of recalling an event in

reverse order. *Law and Human Behavior, 32,* 253-265.

Vrij, A., Mann, S., Kristen, S., & Fisher, R. P. (2007). Cues to deception and ability to

detect lies as a function of police interview styles. *Law and Human Behavior, 31,*

499-518.

Vrij, A., & Winkel, F. W. (1993). Objective and subjective indicators of deception. In N.

K. Clark & G. M. Stephenson (Eds.), *Children, Evidence, and Procedure.*

Leicester, UK: The British Psychological Society.

Waxer, P. H. (1983). Emotional deceit: False words versus false action. *Motivation and*

*Emotion, 7,* 365-376.

Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). HCR-20: Assessing risk

for violence (version 2). Mental Health, Law, and Policy Institute, Simon Fraser

University, Burnaby, British Columbia, Canada.

Wiebe, R. P. (2004). Psychopathy and sexual coercion: A Darwinian analysis.

*Counselling and Clinical Psychology Journal, 1,* 23-41.

Williamson, S., Hare, R. D., & Wong, S. (1987). Violence: Criminal psychopaths and

their victims. *Canadian Journal of Behavioral Science, 19,* 454-462.

Wong, S. (1988). Is Hare's Psychopathy Checklist reliable without an interview?

*Psychological Reports, 62,* 931-934.

Woodworth, M., & Porter, S. (2002). In Cold blood: Characteristics of criminal

homicides as a function of psychopathy. *Journal of Abnormal Psychology, 111,*

436-445.

Yuille. J. C. (2005). Statement analysis: An overview of the content criteria: Applications to the study of adult victims' and perpetrators' accounts of violent crimes and other experiences (e.g., positive, traumatic). Unpublished Workshop manual.

Yuille, J. C. (1989). *Credibility Assessment.* Dordrecht, Boston, London: Kluwer Academic Publishers.

Yuille, J. C. (1988). The systematic assessment of children's testimony. *Canadian Psychology, 29,* 247-262.

Yuille, J. C., Hunter, R., & Harvey, W. (1990). A coordinated approach to interviewing in child sexual abuse investigations. *Canada's Mental Health, 38,* 14-17.

Yuille, J.C., Hunter, R., Joffe, R., & Zaparniuk, J. (1993) Interviewing children in sexual abuse cases. In G.S. Goodman & B.L. Bottoms (Eds.), *Child Victims, Child Witnesses: Understanding and Improving Children's Testimony.* New York: Guilford.

Yuille, J. C., Tymofievich, M., & Marxsen, D. (1995). The nature of allegations of child sexual abuse. In T. Ney (Ed.), *True and false allegations of child sexual abuse: Assessment and case management.* New York: Brunner/Mazel.

Zaparniuk, J., Yuille, J. C., & Taylor, S. (1995). Assessing the credibility of true and false statements. *International Journal of Law and Psychiatry, 18,* 343-352.

Zaparniuk, J., Yuille, J. C., & Taylor, S. (1995). Assessing the credibility of true and false statements. *International Journal of Law and Psychiatry, 18,* 343-352.

# Appendices

A) UBC Ethical Approval

B) CBCA Coding Sheet

C) Memory Characteristics Questionnaire

D) File Coding Sheet

E) Results of Statistical Tests

F) Additional RM Scoring Methods

G) Additional Psychopathy Analyses

A) UBC Ethical Approval

# CERTIFICATE OF APPROVAL- MINIMAL RISK RENEWAL

| PRINCIPAL INVESTIGATOR: John C. Yuille | DEPARTMENT: UBC/Arts/Psychology, Department of | UBC BREB NUMBER: H99-80469 |
|---|---|---|

| INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT: | |
|---|---|
| **Institution** | **Site** |
| N/A | N/A |
| **Other locations where the research will be conducted:** Mountain Institution, Correctional Services of Canada Kent Institution, Correctional Services of Canada | |

| CO-INVESTIGATOR(S): |
|---|
| Dorothee Griesel |
| Barry Cooper |
| Gayla Swihart |
| Marguerite Rodgers Ternes |

| SPONSORING AGENCIES: |
|---|
| Social Sciences and Humanities Research Council of Canada (SSHRC) - "Autobiographical Memories in Incarcerated Men" - "Memory for Mayhem" - "Autobiographical Memories in Incarcerated Persons" |

| PROJECT TITLE: |
|---|
| Autobiographical Memories in Incarcerated Men |
| Funding (a student research grant from the Canadian Psychological Association Foundation) has been requested for a sub-project within "Autobiographical Memories in Incarcerated Men". This sub-project is entitled the "Consistency of Male Violent Offenders' Autobiographical Memories". The procedures of this project have already been described in our original proposal for "Autobiographical Memories in Incarcerated Men". Please add this new study title to the ethics certificate so that the title matches the one of the CPA funding application. Please note that the overall project, "Autobiographical Memories in Incarerated Men" is still ongoing, i.e., this title should not be replaced. |

**EXPIRY DATE OF THIS APPROVAL:  January 25, 2009**

| APPROVAL DATE:  January 25, 2008 |
|---|

| The Annual Renewal for Study have been reviewed and the procedures were found to be acceptable on ethical grounds for research involving human subjects. |
|---|

*Approval is issued on behalf of the Behavioural Research Ethics Board*

B) CBCA Coding Sheet

Name of coder: _____

PP#:_____          ☐ Memory for Mayhem:  ☐ POS   ☐ IV   ☐ RV   ☐ LM   ☐ SD

## CBCA Coding Sheet

|  | Rating | Location (line #)/notes |
|---|---|---|
| **General Characteristics** | | |
| 1.  Coherence | | |
| 2.  Spontaneous reproduction | | |
| 3.  Appropriate Detail | | |
| **Specific Contents** | | |
| 4.  Contextual Embedding | | |
| 5.  Descriptions of Interactions | | |
| 6.  Reproduction of Conversation | | |
| 7.  Unexpected Complications | | |
| **Peculiarities of Content** | | |
| 8.  Unusual Details | | |
| 9.  Peripheral Details | | |
| 10. Accurately Reported Details Misunderstood | | |
| 11. Related External Associations | | |
| 12. Accounts of Subjective Mental State | | |

| | | |
|---|---|---|
| 13. Attribution of Perpetrator's Mental State | | |
| 14. Details Characteristic of a Particular Act | | |
| **Motivation Related Contents** | | |
| 15. Spontaneous Corrections | | |
| 16. Expressing Insecurities | | |
| 17. Admitting Lack of Memory | | |
| 18. Providing Reasons for Lack of Memory | | |
| 19. Raising Doubts about one's own Testimony | | |
| 20. Self-deprecation | | |
| 21. Pardoning the Perpetrator | | |
| **Stylistic Features** | | |
| 22. Long-winded vs. Issue Oriented | | |
| 23. Theme related Changes | | |
| 24. Rigid Repetition | | |

**DECISION:**      ☐ **Credible (i.e., based on genuine experience)**
                   ☐ **Noncredible (e.g., fabricated)**

C) Memory Characteristics Questionnaire (Johnson et al., 1988)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1. My memory for this event is: | 1<br>dim | 2 | 3 | 4 | 5 | 6 | 7<br>sharp/<br>clear |
| 2. My memory for this event is: | 1<br>black & white | 2 | 3 | 4 | 5 | 6 | 7<br>entirely color |
| My memory for this event involves: | | | | | | | |
| 3. Visual detail: | 1<br>little or none | 2 | 3 | 4 | 5 | 6 | 7<br>a lot |
| 4. Sound: | 1<br>little or none | 2 | 3 | 4 | 5 | 6 | 7<br>a lot |
| 5. Smell: | 1<br>little or none | 2 | 3 | 4 | 5 | 6 | 7<br>a lot |
| 6. Touch: | 1<br>little or none | 2 | 3 | 4 | 5 | 6 | 7<br>a lot |
| 7. Taste: | 1<br>little or none | 2 | 3 | 4 | 5 | 6 | 7<br>a lot |
| 8. Overall vividness is: | 1<br>vague | 2 | 3 | 4 | 5 | 6 | 7<br>very vivid |
| 9. My memory for this event is: | 1<br>sketchy | 2 | 3 | 4 | 5 | 6 | 7<br>very detailed |
| 10. Order of events is: | 1<br>confusing | 2 | 3 | 4 | 5 | 6 | 7<br>comprehensible |
| 11. Story line is: | 1<br>simple | 2 | 3 | 4 | 5 | 6 | 7<br>complex |
| 12. Story line is: | 1<br>bizarre | 2 | 3 | 4 | 5 | 6 | 7<br>realistic |
| 13. My memory for the location where the event takes place is: | 1<br>vague | 2 | 3 | 4 | 5 | 6 | 7<br>clear/distinct |
| 14. General setting is: | 1<br>unfamiliar | 2 | 3 | 4 | 5 | 6 | 7<br>familiar |
| 15. Relative spatial arrangement of objects in my memory for the event is: | 1<br>vague | 2 | 3 | 4 | 5 | 6 | 7<br>clear/<br>distinct |
| 16. Relative spatial arrangement of people in my memory for this event is: | 1<br>vague | 2 | 3 | 4 | 5 | 6 | 7<br>clear/<br>distinct |

My memory for the…

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 17. Time when the event takes place is: | 1<br>vague | 2 | 3 | 4 | 5 | 6 | 7<br>clear/<br>distinct |
| 18. Year when the event takes place is: | 1<br>vague | 2 | 3 | 4 | 5 | 6 | 7<br>clear/<br>distinct |
| 19. Season when the event takes place is: | 1<br>vague | 2 | 3 | 4 | 5 | 6 | 7<br>clear/distinct |
| 20. Day when the event takes place is: | 1<br>vague | 2 | 3 | 4 | 5 | 6 | 7<br>clear/distinct |
| 21. Hour when the event takes place is: | 1<br>vague | 2 | 3 | 4 | 5 | 6 | 7<br>clear/distinct |
| 22. The event seems: | 1<br>short | 2 | 3 | 4 | 5 | 6 | 7<br>long |
| 23. The overall tone of the memory is: | 1<br>negative | 2 | 3 | 4 | 5 | 6 | 7<br>positive |
| 24. In this event I was: | 1<br>a spectator | 2 | 3 | 4 | 5 | 6 | 7<br>a participant |
| 25. At the time, the event seemed like it would have a serious implication: | 1<br>not at all | 2 | 3 | 4 | 5 | 6 | 7<br>definitely |
| 26. Looking back, this event did have serious implications: | 1<br>not at all | 2 | 3 | 4 | 5 | 6 | 7<br>definitely |
| 27. I remember how I felt at the time when the event took place: | 1<br>not at all | 2 | 3 | 4 | 5 | 6 | 7<br>definitely |
| 28. Feelings at the time were: | 1<br>negative | 2 | 3 | 4 | 5 | 6 | 7<br>positive |
| 29. Feelings at the time were: | 1<br>not intense | 2 | 3 | 4 | 5 | 6 | 7<br>intense |
| 30. As I am remembering now, my feelings are: | 1<br>not intense | 2 | 3 | 4 | 5 | 6 | 7<br>intense |
| 31. I remember what I thought at the time: | 1<br>not at all | 2 | 3 | 4 | 5 | 6 | 7<br>clearly |
| 32. This memory reveals or says about me: | 1<br>not much | 2 | 3 | 4 | 5 | 6 | 7<br>a lot |

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 33. Overall, I remember this event: | hardly | | | | | | very well |
| 34. I remember events relating to this memory that took place in advance of the event: | 1 not at all | 2 | 3 | 4 | 5 | 6 | 7 clearly |
| 35. I remember events relating to this memory that took place after the event: | 1 not at all | 2 | 3 | 4 | 5 | 6 | 7 clearly |
| 36. Do you have any doubts about the accuracy of your memory for this event: | 1 a great deal of doubt | 2 | 3 | 4 | 5 | 6 | 7 no doubt whatsoever |
| 37. Since it happened, I have thought about this event: | 1 not at all | 2 | 3 | 4 | 5 | 6 | 7 many times |
| 38. Since it happened, I have talked about this event: | 1 not at all | 2 | 3 | 4 | 5 | 6 | 7 many times |
| 39. About when did this event happen?  Circle one: | Just today Yesterday Few days ago Last week Few weeks ago Last month Few months ago Last year or longer (indicate age: _____ ) | | | | | | |

D) File Coding Sheet

## **Memory for Mayhem: File Review**

## **Basic Information**

Participant #:

……………………………………………………………………………………….

Date of original interview:

……………………………………………………………………………………

Date of file review:

………………………………………………………………………………………

Location of file review:

……………………………………………………………………………………

Name of reviewer:

…………………………………………………………………………………….

---

**Definitions of Coding Scheme for Corroboration of Details in Narratives**

(1) *Confirmed* = Detail described in research interview fully confirmed in file review.
(2) *Partially Confirmed* = Detail described in research interview partially confirmed in file review.
(3) *Contradictory*= Detail in described in research interview is not confirmed by file information (i.e., information is present but contradicts the offender's version).
(4) *Not confirmed* = Detail described in research interview is not mentioned in file review.

---

**Documents in Case Management File that can be Used to Corroborate Details in Narratives**

- Criminal Profile / Criminal Profile Report
- Correctional Plan Progress Report
- Correctional Plan
- Intake Assessment
- Preliminary Assessment
- Report to Crown Council
- Police Report

Note: Others are possible. Only independent information (i.e., not reported by the offender himself) should be used to corroborate the details in the narratives!

## Memory:  ☐ LM    ☐ RV    ☐ IV

| **Summary from Research Interview** | **File Review** |
|---|---|
| Circle ⊘ if information is not present. | Is this memory in the file? |
| Circle ✓ if information is present and provide description. | ☐ No ☐ Yes |
| | If yes, continue: |

| Summary from Research Interview | File Review |
|---|---|
| ⊘ ✓City/town: …………………………………………………….. | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
| ⊘ ✓Area: ………………………………………………………..… | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
| ⊘ ✓Specific description of location of crime: …………………………………………… | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
| …………………………………………………………….. | |
| ⊘ ✓Year: …………………………………………………….. | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
| ⊘ ✓Season: ……………………………………………….. | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
| ⊘ ✓Month: …………………………………………….. | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
| ⊘ ✓Day/date: …………………………………………….. | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
| ⊘ ✓Time of the day: ………………………………………….. | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
| ⊘ ✓Weather: ……………...…………………………………… | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
| Description of self: | |
|      ⊘ ✓Clothing: ………………………………………… | |
|            ……………………………………………………… | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
| Description of victim: | |
|      ⊘ ✓Clothing: ………………………………………… | |
|            ……………………………………………………… | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
|      ⊘ ✓Hair ………………………………………………… | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
|      ⊘ ✓Height ……………………………………………… | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
|      ⊘ ✓Weight ……………………………………………… | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
| Description of other person: | |
|      Role of person in this event: ………………………………………… | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
|      ⊘ ✓Clothing: ……………………………………….. | |
|            ……………………………………………………… | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
|      ⊘ ✓Hair ………………………………………………… | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
|      ⊘ ✓Height ……………………………………………… | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
|      ⊘ ✓Weight ……………………………………………… | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
| (Use extra sheet if more person descriptions present.) | |
| Details of actions: | |
| - ……………………………………………………………… | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
| - ……………………………………………………………… | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
| - ……………………………………………………………….. | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
| - ……………………………………………………………… | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
| - ……………………………………………………………… | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
| - ……………………………………………………………… | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
| - ……………………………………………………………… | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
| - ……………………………………………………………… | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
| - ……………………………………………………………… | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
| - ……………………………………………………………… | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
| - ……………………………………………………………… | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
| - ……………………………………………………………… | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
| - ……………………………………………………………… | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
| - ……………………………………………………………… | ☐ (1)  ☐ (2)   ☐ (3)   ☐ (4) |
| (Use extra sheet if more actions described.) | |

Overall number of details omitted in the narrative but mentioned in the file: ……..

E) Results of Statistical Tests

Table 35: Results of paired-sample $t$-tests to compare CBCA-judged credible and noncredible memories in memory length (in words), memory age (in years), intoxication (1 = no, 2 = yes), and number of times rehearsed, considering only the 12 participants who provided noncredible memories.

|  | $t$ | df | $p$ |
|---|---|---|---|
| Memory length | 0.33 | 11 | .745 |
| Age of memory | -0.96 | 10 | .359 |
| Intoxication | 0.00 | 9 | 1.00 |
| Rehearsal | -2.32 | 10 | .043 |

α = .05

Table 36: Results of the $t$-tests and $z$-test of proportions to compare CBCA-judged credible and noncredible memories in memory length (in words), memory age (in years), intoxication (1 = no, 2 = yes), and number of times rehearsed, considering all memories, and considering the memory types separately.

|  | IV | | | RV | | | LM | | |
|---|---|---|---|---|---|---|---|---|---|
|  | Statistic | df | $p$ | Statistic | df | $p$ | Statistic | df | $p$ |
| Memory length | $t$ = -0.96 | 115 | .341 | $t$ = 0.82 | 126 | .412 | $t$ = 1.62 | 76 | .109 |
| Age of Memory | $t$ = 1.05 | 110 | .295 | $t$ = -0.34 | 123 | .735 | $t$ = -1.03 | 73 | .305 |
| Intoxication | $z$ = 0.28 | n/a | .390 | $z$ = 0.62 | n/a | .268 | $z$ = 0.77 | n/a | .221 |
| Rehearsal | $t$ = 0.48 | 111 | .635 | $t$ = 0.60 | 124 | .553 | $t$ = 0.68 | 74 | .500 |

α = .05

Table 37: Results of the paired-sample $t$-tests to compare CBCA scores of each criterion according to CBCA judgment of credible and noncredible for all of the CBCA criteria, considering only the 12 participants who provided a noncredible memory (df = 11).

| Criteria | $t$ | $p$ |
|---|---|---|
| 1 | 2.57 | .026 |
| 2 | 2.97 | .013 |
| 3 | 3.92 | <.0025 |
| 4 | 1.15 | .275 |
| 5 | 2.60 | .025 |
| 6 | 2.55 | .027 |
| 7 | 1.77 | .104 |
| 8 | -0.21 | .838 |
| 9 | 1.00 | .339 |
| 11 | 1.92 | .082 |
| 12 | 0.67 | .517 |
| 13 | 0.76 | .463 |
| 14 | 1.00 | .339 |
| 15 | -1.10 | .295 |
| 16 | 0.61 | .555 |

| | | |
|---|---|---|
| 17 | 0.00 | 1.00 |
| 18 | 0.15 | .884 |
| 20 | 1.39 | .191 |
| 22 | -1.00 | .339 |
| 24 | -1.39 | .191 |

α = .0025

Table 38: Results of the Welch's *t*-tests to compare CBCA score of each criterion according to CBCA judgment of credible and noncredible for all of the CBCA criteria for the IV, RV, and LM memories.

| CBCA Criteria | IV | | | RV | | | LM | | |
|---|---|---|---|---|---|---|---|---|---|
| | *t* | df | *p* | *t* | df | *p* | *t* | df | *p* |
| Criterion 1 | 0.91 | 1 | .529 | 2.27 | 6 | .064 | n/a | n/a | n/a |
| Criterion 2 | 0.83 | 1 | .560 | 2.83 | 6 | .029 | 3.59 | 2 | .067 |
| Criterion 3 | 0.84 | 1 | .553 | 2.74 | 6 | .033 | 3.59 | 2 | .067 |
| Criterion 4 | 0.64 | 1 | .253 | 0.90 | 7 | .403 | 3.34 | 2 | .075 |
| Criterion 5 | 0.86 | 1 | .547 | 1.79 | 6 | .123 | 1.53 | 2 | .265 |
| Criterion 6 | -0.48 | 1 | .714 | 3.87 | 7 | .006 | n/a | n/a | n/a |
| Criterion 8 | n/a | n/a | n/a | -0.29 | 6 | .784 | n/a | n/a | n/a |
| Criterion 9 | n/a | n/a | n/a | -1.11 | 7 | .306 | n/a | n/a | n/a |
| Criterion 12 | n/a | n/a | n/a | 1.40 | 6 | .208 | 6.24 | n/a | n/a |
| Criterion 13 | n/a | n/a | n/a | 2.50 | 8 | .038 | 1.76 | 2 | .208 |
| Criterion 15 | n/a | n/a | n/a | -0.30 | 6 | .771 | n/a | n/a | n/a |
| Criterion 16 | -0.10 | 1 | .934 | 0.86 | 7 | .419 | 3.45 | 2 | .067 |
| Criterion 17 | n/a | n/a | n/a | 0.59 | 7 | .574 | n/a | n/a | n/a |
| Criterion 18 | n/a | n/a | n/a | -0.20 | 6 | .851 | n/a | n/a | n/a |
| Criterion 22 | -0.67 | 1 | .623 | n/a | n/a | n/a | n/a | n/a | n/a |
| Criterion 24 | n/a | n/a | n/a | -1.36 | 6 | .223 | n/a | n/a | n/a |

n/a = Welch's *t*-tests were not performed because at least one variance was zero.

α = .0017

Table 39: Results of the Welch's *t*-tests to compare the CBCA scores of the 12 CBCA-judged noncredible memories to the mean CBCA scores of the credible memories of the participants who provided only memories that were judged to be credible.

| Criteria | *t* | df | *p* |
|---|---|---|---|
| 1 | 2.08 | 11 | .061 |
| 2 | 4.15 | 11 | .002 |
| 3 | 4.09 | 11 | .002 |
| 4 | 2.53 | 11 | .028 |
| 5 | 2.62 | 11 | .024 |
| 6 | 3.33 | 13 | .006 |
| 8 | 0.15 | 12 | .887 |

| | | | |
|---|---|---|---|
| 9 | 0.20 | 12 | .848 |
| 12 | 2.98 | 12 | .011 |
| 13 | 3.04 | 14 | .009 |
| 15 | 0.01 | 12 | .995 |
| 16 | 1.90 | 12 | .081 |
| 17 | 1.37 | 13 | .193 |
| 18 | 0.55 | 12 | .593 |
| 22 | 1.49 | 14 | .157 |
| 24 | -1.24 | 11 | .239 |

$\alpha = .003$

Table 40: Results of the Welch's *t*-tests to compare CBCA scores of credible memories of participants who provided only memories that were judged to be credible according to CBCA (Group 1) and credible memories of participants who provided one noncredible memory (Group 2).

| Criteria | *t* | df | *p* |
|---|---|---|---|
| 2 | 1.05 | 12 | .313 |
| 3 | 1.47 | 12 | .168 |
| 4 | 1.52 | 12 | .155 |
| 5 | -0.76 | 15 | .459 |
| 6 | 0.44 | 13 | .671 |
| 7 | 0.14 | 13 | .893 |
| 8 | 0.70 | 14 | .496 |
| 9 | -0.91 | 13 | .378 |
| 11 | 1.14 | 13 | .276 |
| 12 | 1.88 | 12 | .085 |
| 13 | 1.49 | 12 | .162 |
| 14 | 1.83 | 21 | .082 |
| 15 | 1.55 | 13 | .145 |
| 16 | 1.19 | 12 | .257 |
| 17 | 1.69 | 15 | .112 |
| 18 | 0.45 | 13 | .659 |
| 20 | -0.09 | 12 | .928 |

$\alpha = .003$

Table 41: Results of the Welch's *t*-tests to compare CBCA criterion scores across consistency with file information for all of the criteria for the IV and RV memories.

| CBCA Criteria | IV | | | RV | | |
|---|---|---|---|---|---|---|
| | *t* | df | *p* | *t* | df | *p* |
| Criterion 2 | 0.77 | 2 | .519 | n/a | n/a | n/a |
| Criterion 3 | 0.77 | 2 | .519 | n/a | n/a | n/a |
| Criterion 4 | 0.66 | 2 | .576 | n/a | n/a | n/a |
| Criterion 5 | 0.88 | 2 | .468 | n/a | n/a | n/a |

| | | | | | | |
|---|---|---|---|---|---|---|
| Criterion 6 | -0.61 | 3 | .587 | 1.06 | 7 | .326 |
| Criterion 8 | n/a | n/a | n/a | -0.48 | 3 | .658 |
| Criterion 9 | 1.51 | 3 | .229 | 0.24 | 6 | .818 |
| Criterion 11 | 0.60 | 3 | .591 | -0.24 | 6 | .818 |
| Criterion 13 | 3.23 | 3 | .059 | 0.95 | 4 | .392 |
| Criterion 15 | n/a | n/a | n/a | 1.33 | 7 | .226 |
| Criterion 16 | -0.63 | 2 | .581 | 0.54 | 5 | .613 |
| Criterion 17 | -0.82 | 3 | .478 | 0.00 | 8 | 1.00 |
| Criterion 22 | -0.77 | 2 | .519 | n/a | n/a | n/a |

n/a = Welch's *t*-tests were not performed because at least one variance was zero.

α = .0026

Table 42: Results of Welch's *t*-tests to compare CBCA criterion scores of memories that were partially confirmed to memories that were partially confirmed with some contradictions, considering all memory types together.

| CBCA Criteria | *t* | df | *p* |
|---|---|---|---|
| Criterion 2 | 0.57 | 8 | .587 |
| Criterion 3 | 0.57 | 8 | .587 |
| Criterion 4 | 1.50 | 8 | .173 |
| Criterion 5 | 0.57 | 8 | .587 |
| Criterion 6 | -0.54 | 13 | .599 |
| Criterion 7 | 1.20 | 13 | .251 |
| Criterion 8 | -0.40 | 8 | .969 |
| Criterion 9 | 0.23 | 10 | .822 |
| Criterion 11 | 1.01 | 14 | .332 |
| Criterion 12 | 0.98 | 9 | .356 |
| Criterion 13 | 2.47 | 9 | .035 |
| Criterion 15 | 0.41 | 11 | .690 |
| Criterion 16 | 0.42 | 8 | .688 |
| Criterion 17 | -0.58 | 12 | .575 |
| Criterion 18 | 0.66 | 10 | .530 |

α = .0033

Table 43: Results of Welch's *t*-tests for comparing each RM construct according to consistency with file information, considering each type of memory separately.

| | Construct | *t* | df | *p* |
|---|---|---|---|---|
| IV | Clarity | 1.90 | 2 | .192 |
| | Sensory Experiences | 0.99 | 2 | .411 |
| | Spatial Information | 1.38 | 2 | .298 |
| | Time Information | 2.07 | 2 | .160 |
| | Overall Impression | 0.81 | 2 | .495 |
| | Retrospective Recollection | 2.26 | 4 | .082 |
| | Events Before and After | 1.36 | 2 | .294 |

| | | Bizarreness | 1.62 | 2 | .243 |
|---|---|---|---|---|---|
| RV | | Clarity | -0.20 | 3 | .856 |
| | | Sensory Experiences | -0.30 | 4 | .781 |
| | | Spatial Information | 0.70 | 4 | .529 |
| | | Time Information | -0.30 | 4 | .780 |
| | | Overall Impression | -0.36 | 3 | .743 |
| | | Retrospective Recollection | -1.41 | 4 | .237 |
| | | Events Before and After | -0.61 | 5 | .569 |
| | | Bizarreness | 0.95 | 3 | .407 |

α = .003

Table 44: Results of the independent samples *t*-tests for comparing each RM construct according to consistency with file information, considering all memory types together (df = 54).

| Construct | *t* | *p* |
|---|---|---|
| Clarity | 0.36 | .718 |
| Sensory Experiences | -0.76 | .451 |
| Spatial Information | 0.88 | .386 |
| Time Information | 0.51 | .609 |
| Overall Impression | -0.34 | .733 |
| Retrospective Recollection | -0.63 | .535 |
| Events Before and After | 0.29 | .772 |
| Bizarreness | 1.47 | .148 |

α = .006

Table 45: Results of paired-sample *t*-tests for comparing each RM construct according to CBCA credibility judgment, considering only participants who provided a memory that was judged to be noncredible (df = 11).

| Construct | *t* | *p* |
|---|---|---|
| Clarity | 2.26 | .045 |
| Sensory Experiences | 2.11 | .058 |
| Spatial Information | 2.01 | .070 |
| Time Information | 1.22 | .248 |
| Overall Impression | 1.23 | .246 |
| Retrospective Recollection | 2.21 | .049 |
| Events Before and After | 0.53 | .610 |
| Bizarreness | 1.30 | .220 |

α = .006

Table 46: Results of Welch's *t*-tests for comparing each RM construct according to CBCA credibility, considering all memories.

| | Construct | *t* | df | *p* |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| IV | Clarity | -0.43 | 1 | .738 |
| | Sensory Experiences | -1.81 | 1 | .309 |
| | Spatial Information | -0.95 | 2 | .469 |
| | Time Information | -0.20 | 1 | .871 |
| | Overall Impression | -0.10 | 1 | .934 |
| | Events Before and After | -1.10 | 1 | .464 |
| | Bizarreness | 0.26 | 1 | .838 |
| RV | Clarity | 0.83 | 7 | .434 |
| | Sensory Experiences | 0.38 | 12 | .708 |
| | Spatial Information | 0.33 | 6 | .750 |
| | Time Information | -0.81 | 7 | .445 |
| | Overall Impression | 1.03 | 7 | .336 |
| | Retrospective Recollection | 0.70 | 6 | .506 |
| | Events Before and After | 1.08 | 7 | .318 |
| | Bizarreness | 0.03 | 6 | .978 |
| LM | Clarity | 6.98 | 4 | .003 |
| | Sensory Experiences | 3.87 | 4 | .016 |
| | Spatial Information | 2.14 | 2 | .151 |
| | Time Information | 5.53 | 3 | .017 |
| | Overall Impression | -.04 | 2 | .973 |
| | Retrospective Recollection | 6.77 | 3 | .005 |
| | Events Before and After | 2.23 | 2 | .141 |
| | Bizarreness | 0.67 | 2 | .568 |

$\alpha = .002$

Table 47: Results of Paired Sample *t*-Tests for CBCA criterion scores for IV vs. RV memories.

| | *t* | *p* |
|---|---|---|
| Criterion 1 | 1.39 | .167 |
| Criterion 2 | 0.90 | .368 |
| Criterion 3 | 1.51 | .134 |
| Criterion 4 | 0.90 | .372 |
| Criterion 5 | 0.28 | .783 |
| Criterion 6 | 0.00 | 1.00 |
| Criterion 7 | 2.03 | .045 |
| Criterion 8 | -0.30 | .765 |
| Criterion 9 | -0.13 | .895 |
| Criterion 11 | 0.99 | .327 |
| Criterion 12 | 0.80 | .426 |
| Criterion 13 | -0.12 | .907 |
| Criterion 14 | 2.08 | .040 |
| Criterion 15 | 0.48 | .635 |
| Criterion 16 | 0.24 | .812 |
| Criterion 17 | 0.13 | .895 |

| Criterion 18 | 0.47 | .640 |
|---|---|---|
| Criterion 19 | -1.42 | .158 |
| Criterion 20 | 0.47 | .640 |
| Criterion 21 | 1.14 | .259 |
| Criterion 22 | -0.82 | .417 |
| Criterion 23 | -0.45 | .657 |
| Criterion 24 | -1.42 | .158 |

α = .002

Table 48: Results of paired sample t-tests for each RM construct according to type of violence.

| Construct | $t$ | $p$ |
|---|---|---|
| Clarity | 2.74 | .007 |
| Sensory Experiences | -0.68 | .496 |
| Spatial Information | 1.82 | .071 |
| Time Information | 2.98 | .003 |
| Overall Impression | 1.44 | .153 |
| Retrospective Recollection | 2.04 | .043 |
| Events Before and After | 2.80 | .006 |
| Bizarreness | 0.94 | .349 |

α = .006

Table 49: Results of Welch's $t$-tests comparing Criterion scores for psychopaths (PCL-R score greater or equal to 30) and nonpsychopaths (PCL-R score equal to or less than 20) for IV, RV, and LM memories.

| | IV | | | RV | | | LM | | |
|---|---|---|---|---|---|---|---|---|---|
| Criterion | $t$ | df | $p$ | $t$ | df | $p$ | $t$ | df | $p$ |
| 1 | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| 2 | -0.67 | 27 | .509 | -0.76 | 24 | .453 | n/a | n/a | n/a |
| 3 | n/a | n/a | n/a | -0.18 | 30 | .860 | n/a | n/a | n/a |
| 4 | -0.50 | 32 | .618 | 0.59 | 44 | .562 | 1.17 | 20 | .255 |
| 5 | -0.02 | 35 | .983 | -0.12 | 30 | .903 | -0.60 | 10 | .560 |
| 6 | -2.01 | 31 | .053 | -1.52 | 23 | .142 | -0.82 | 16 | .422 |
| 7 | 0.05 | 35 | .963 | 0.43 | 34 | .671 | -0.53 | 18 | .602 |
| 8 | -0.34 | 34 | .733 | -1.53 | 52 | .133 | -0.10 | 17 | .923 |
| 9 | 0.20 | 27 | .845 | -1.47 | 43 | .150 | 0.11 | 12 | .915 |
| 11 | -0.08 | 35 | .939 | -1.83 | 39 | .075 | 0.30 | 14 | .767 |
| 12 | 2.55 | 55 | .014 | -0.90 | 27 | .375 | 0.43 | 12 | .675 |
| 13 | 0.31 | 40 | .762 | 0.21 | 35 | .839 | 1.54 | 12 | .149 |
| 14 | 1.08 | 28 | .291 | 0.18 | 30 | .860 | n/a | n/a | n/a |
| 15 | -1.67 | 42 | .102 | -1.75 | 44 | .087 | -0.77 | 14 | .455 |
| 16 | 1.12 | 40 | .270 | 1.54 | 36 | .132 | 1.11 | 21 | .280 |
| 17 | -0.10 | 34 | .919 | 0.27 | 33 | .788 | -0.72 | 11 | .488 |

| 18 | 0.21 | 35 | .835 | -0.59 | 33 | .559 | -1.70 | 15 | .111 |
| 19 | n/a | n/a | n/a | 0.53 | 23 | .604 | n/a | n/a | n/a |
| 20 | 1.27 | 26 | .216 | -1.12 | 47 | .268 | 1.20 | 9 | .261 |
| 21 | 0.85 | 25 | .405 | n/a | n/a | n/a | n/a | n/a | n/a |
| 22 | -0.50 | 40 | .619 | -1.35 | 41 | .184 | -0.51 | 16 | .617 |
| 23 | 0.46 | 27 | .648 | -0.23 | 35 | .818 | 0.25 | 12 | .806 |
| 24 | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a |

$\alpha = .001$

n/a indicates that Welch's *t*-tests could not be computed because at least one of the variances was equal to zero.

Table 50: Results of *t*-tests comparing frequencies for CBCA Criteria 12 and 13 affect, cognition, and physical for psychopaths (PCL-R score 30 or higher) and nonpsychopaths (PCL-R score 20 or lower) for IV, RV, and LM memories.

| Criterion | IV $df = 56$ | | RV $df = 58$ | | LM $df = 32$ | |
|---|---|---|---|---|---|---|
| | *t* | *p* | *t* | *p* | *t* | *p* |
| 12 Total | 1.21 | .230 | -0.65 | .520 | 0.68 | .499 |
| 12 Affect | 0.97 | .337 | -0.65 | .516 | 1.72 | .095 |
| 12 Cognition | 1.49 | .141 | -0.78 | .441 | 1.41 | .167 |
| 12 Physical | -0.64 | .524 | 1.60 | .115 | 0.68 | .499 |
| 13 Total | 0.51 | .610 | 0.42 | .676 | 1.51 | .141 |
| 13 Affect | 0.04 | .970 | 0.36 | .724 | 0.65 | .518 |
| 13 Cognition | 0.56 | .576 | 0.42 | .677 | 3.26 | .003 |
| 13 Physical | 0.94 | .351 | 0.00 | 1.00 | -0.60 | .551 |

$\alpha = .002$

Table 51: Results of *t*-tests comparing RM scores for psychopaths (PCL-R score equal to or greater than 30) and nonpsychopaths (PCL-R score equal to or less than 20).

| | IV $df = 61$ | | RV $df = 70$ | | LM $df = 35$ | |
|---|---|---|---|---|---|---|
| | *t* | *p* | *t* | *p* | *t* | *p* |
| Scoring Method 1 | -0.23 | .816 | -1.00 | .319 | 0.41 | .687 |
| Scoring Method 2 | -0.67 | .509 | -1.44 | .153 | 0.10 | .919 |
| Scoring Method 3 | -0.02 | .988 | -1.00 | .320 | 0.08 | .935 |
| Scoring Method 4 | -0.75 | .454 | -1.42 | .161 | -0.52 | .609 |

$\alpha = .004$

Table 52: Results of *t*-tests comparing RM construct scores of psychopaths (PCL-R score 30 or above) and nonpsychopaths (PCL-R score 20 or below) for IV, RV, and LM memories.

| | IV | RV | LM |
|---|---|---|---|

|  | df = 56 | | df = 58 | | df = 35 | |
|---|---|---|---|---|---|---|
| Construct | *t* | *p* | *t* | *p* | *t* | *p* |
| Clarity | 0.21 | .832 | -0.30 | .762 | -0.61 | .548 |
| Sensory Experiences | -0.22 | .824 | -0.62 | .536 | -0.26 | .796 |
| Spatial Information | -0.15 | .879 | -1.23 | .222 | 0.77 | .445 |
| Time Information | -0.69 | .492 | -1.83 | .071 | -1.34 | .190 |
| Overall Impression | -1.50 | .140 | 0.04 | .968 | -0.27 | .787 |
| Retrospective Recollection | -1.58 | .118 | -2.09 | .040 | -0.81 | .423 |
| Events Before and After | -0.49 | .623 | -2.04 | .045 | 2.92 | .006 |
| Bizarreness | 0.46 | .650 | 0.56 | .575 | -0.55 | .585 |

$\alpha = .002$

F) Additional RM Scoring Methods

**Additional Analyses Regarding Hypothesis 2: RM and Consistency with File Information**

Although the MCQ has been found to be made up of eight factors (e.g., Takahashi & Shimizu, 2007), which generally means that interpreting an item total score is inappropriate (Floyd & Widaman, 1995), two additional scoring methods were examined to consider each item on the MCQ separately. In the third scoring method, all MCQ items theoretically believed to correlate positively with credibility (items 1-24, 28, 33-36) were summed with the reverse-scores of all MCQ items theoretically believed to correlate negatively with credibility (items 25-27, 29-32, 37, and 38) to form an RM score for each participant. In the fourth scoring method, all MCQ items were summed to form an RM score for each participant. As with the second scoring method, the rationale for this was based on Sporer and Sharman's (2006) observation that there is little empirical evidence in studies involving deception that any of the RM criteria on the MCQ are more likely to be found in accounts of invented events than self-experienced events.

As with the first and second scoring methods, the total scores for each method were sorted into ascending order and divided into thirds. Although there was a tendency in both scoring methods for memories that were partially confirmed to be sorted into the highest third of the scores, when considering the totals for the bottom and top thirds for each memory type, this difference was not significant according to a $z$-test of proportions. Specifically, for the third scoring method, no significant differences were found between memories in the top third of RM scores and the bottom third of RM scores for the IV ($z = 0.26$, $p > .05$ for partially confirmed memories; $z = 1.43$, $p > .05$ for memories partially confirmed with some contradictions), RV ($z = 0.39$, $p > .05$ for partially confirmed

memories; $z = 1.24$, $p > .05$ for memories partially confirmed with some contradictions),

or LM ($z = 0.37$, $p > .05$ for partially confirmed memories; $z = 1.01$, $p > .05$ for memories

partially confirmed with some contradictions) memories. For the fourth scoring method,

no significant differences were found between memories in the top third and bottom third

of RM scores for the IV ($z = 0.27$, $p > .05$ for partially confirmed memories; $z = 1.43$, $p >$

.05 for memories partially confirmed with some contradictions), RV ($z = 0.42$, $p > .05$ for

partially confirmed memories; $z = 1.02$, $p > .05$ for memories partially confirmed with

some contradictions), or LM ($z = 0.65$, $p > .05$ for partially confirmed memories; $z =$

1.01, $p > .05$ for memories partially confirmed with some contradictions) memories. The

four scoring methods resulted in very similar totals with regards to consistency with file

information. These totals can be found on Tables 53 and 54.

Table 53: Frequencies of memories for consistency with file information and RM score
using the third scoring method (sum of all items on the MCQ, reverse scoring items
theoretically believed to be negatively correlated with credible reports).

|  | Lowest 1/3 | | Middle 1/3 | | Highest 1/3 | |
|---|---|---|---|---|---|---|
|  | Partially Confirmed | Some Contradictions | Partially Confirmed | Some Contradictions | Partially Confirmed | Some Contradictions |
| IV | 9 | 2 | 9 | 1 | 10 | 0 |
| RV | 3 | 1 | 4 | 0 | 4 | 3 |
| LM | 9 | 0 | 2 | 0 | 5 | 1 |
| Total | 21 | 3 | 15 | 1 | 20 | 4 |

*Note:* All findings nonsignificant.

Table 54: Frequencies of memories for consistency with file information and RM Score
using the fourth scoring method (sum of all items on the MCQ).

|  | Lowest 1/3 | | Middle | | Highest 1/3 | |
|---|---|---|---|---|---|---|
|  | Partially Confirmed | Some Contradictions | Partially Confirmed | Some Contradictions | Partially Confirmed | Some Contradictions |
| IV | 8 | 2 | 11 | 1 | 9 | 0 |
| RV | 2 | 1 | 5 | 0 | 4 | 3 |
| LM | 7 | 0 | 4 | 0 | 5 | 1 |
| Total | 17 | 3 | 20 | 1 | 18 | 4 |

*Note:* All findings nonsignificant.

A number of comparisons were conducted to determine whether RM total score differed according to the memories' consistency with file information. First, Welch's *t*-tests were used to compare RM scores for memories that were partially confirmed to memories that were partially confirmed with some contradictions, considering memories of each type separately. Welch's *t*-tests were used because the group sizes were unequal and some of the variances were heterogeneous, according to Levene's test for equality of variances. No significant differences were found for either the third or fourth scoring methods. Specifically, for the third scoring method (sum of MCQ item scores, reverse scoring items part of constructs theoretically believed to be negatively correlated with credible reports), no difference was found for the IV memories ($t(2) = 1.55$, $p > .05$) or the RV memories ($t(3) = 0.13$, $p > .05$). For the fourth scoring method (sum of MCQ item scores), no difference was found for the IV memories ($t(2) = 2.15$, $p > .05$) or the RV memories ($t(3) = -0.34$, $p > .05$). Analyses could not be conducted for LM memories because only one LM memory was partially confirmed with some contradictions. Relevant means can be found on Table 55.

Table 55: Mean RM scores for the third and fourth scoring methods according to consistency with file information (standard deviations in parentheses) for IV, RV, and LM memories.

|  | RM Scoring Method | Partially Confirmed | Partially Confirmed, Some Contradictions |
|---|---|---|---|
| | n | 27 | 3 |
| IV | RM Scoring Method 3 | 174.20 (20.51) | 134.33 (44.77) |
| | RM Scoring Method 4 | 195.13 (25.49) | 139.67 (43.84) |
| | n | 12 | 4 |
| RV | RM Scoring Method 3 | 175.91 (16.56) | 174.50 (54.69) |
| | RM Scoring Method 4 | 187.55 (22.94) | 199.00 (66.09) |
| | n | 16 | 1 |
| LM | RM Scoring Method 3 | 136.38 (24.20) | 166.00[a] |
| | RM Scoring Method 4 | 135.00 (27.79) | 180.00[a] |

*Note:* All findings nonsignificant.

Second, Welch's *t*-tests were conducted to compare the RM scores for partially confirmed memories and memories that were partially confirmed with some contradictions, considering all of the memories together, regardless of memory type. As with the CBCA criterion scores, this comparison was conducted in an effort to increase the power of the comparisons by increasing the number of participants in each group. Welch's *t*-tests were used because a number of the comparisons involved groups that had significantly different variances according to Levene's test for equality of variances. An alpha of .013 (i.e., an alpha of .05 divided by four comparisons) was used in an effort to control for Type I error.

No significant difference was found between memories that were partially confirmed (*M* = 163.74) and memories that were partially confirmed with some contradictions (*M* = 158.38) for the third scoring method, *t*(8) = 0.31, *p* > .05. No significant difference was found between memories that were partially confirmed (*M* = 175.87) and memories that were partially confirmed with some contradictions (*M* = 174.38) for the fourth scoring method, *t*(8) = 0.07, *p* > .05.

**Additional Analyses Regarding Hypothesis 3: RM and CBCA**

As mentioned, a variety of methods were used to determine whether RM scores differed according to whether the memory was found to be credible or noncredible according to CBCA. First, paired-sample *t*-tests were conducted to compare the RM scores of the 12 memories judged to be noncredible according to CBCA to the RM scores to the credible memories of the same participants. If any of these participants had more than one credible memory, the mean RM score was used. An alpha of .0125 was used for

these comparisons in an effort to control for Type I error (i.e., an alpha of .05 divided by four comparisons for the four scoring methods). RM scores did not differ according to CBCA credibility for either the third or fourth scoring method. Specifically, for the third scoring method, there was no significant difference between RM scores for memories that were found credible and memories that were found noncredible according to CBCA, $t(11) = 2.12, p > .013$. For the fourth RM scoring method, there was no difference between RM scores for memories that were found credible and memories that were found noncredible according to CBCA, $t(11) = 2.35, p > .013$. Means related to these findings are presented in Table 56.

Table 56: Mean RM scores for the third and fourth scoring methods (standard deviations in parentheses) according to CBCA credibility, considering only the 12 participants who had one memory judged to be noncredible.

|  | Credible | Noncredible |
| --- | --- | --- |
| RM Scoring Method 3 | 176.79 (30.09) | 152.42 (39.15) |
| RM Scoring Method 4 | 183.71 (38.35) | 148.42 (53.53) |

*Note:* All comparisons nonsignificant.

Second, *t*-tests were conducted to determine whether RM scores differed according to whether the memory was found to be credible or noncredible according to CBCA. Levene's test for equality of variances showed that all variances were homogeneous. An alpha of .002 was used for these comparisons in an effort to control for Type I error (i.e., an alpha of .05 divided by 24 comparisons over the four scoring methods). RM scores did not differ according to CBCA credibility for any of the memory types. Specifically, for the third scoring method, there was no significant difference between RM scores for memories that were found credible and memories that were found noncredible according to CBCA, for IV memories ($t(114) = -0.76, p > .05$), RV memories ($t(126) = -0.11, p > .05$), or LM memories ($t(75) = 3.11, p > .002$). For the fourth RM

scoring method, there was no difference between RM scores for memories that were found credible and memories that were found noncredible according to CBCA, for IV memories (t(114) = -0.07, $p$ > .05), RV memories ($t$(126) = 0.63, $p$ > .05), or LM memories ($t$(75) = 2.54, $p$ > .01). Means related to these findings are presented in Table 57.

Table 57: Mean RM scores for the third and fourth scoring methods (standard deviations in parentheses) for IV, RV, and LM memories according to CBCA credibility, considering all memories.

|  |  | Credible | Noncredible |
|---|---|---|---|
| IV | n | 114 | 2 |
|  | RM Scoring Method 3 | 170.38 (25.09) | 184.00 (11.31) |
|  | RM Scoring Method 4 | 186.28 (32.99) | 188.00 (11.31) |
| RV | n | 121 | 7 |
|  | RM Scoring Method 3 | 164.70 (26.56) | 165.86 (26.26) |
|  | RM Scoring Method 4 | 175.93 (36.42) | 167.00 (42.57) |
| LM | n | 74 | 3 |
|  | RM Scoring Method 3 | 139.18 (26.25) | 100.00 (23.39) |
|  | RM Scoring Method 4 | 141.56 (34.72) | 78.67 (13.65) |

*Note:* All comparisons nonsignificant.

**Additional Analyses Regarding Hypothesis 4: IV vs. RV**

Paired Sample *t*-tests were conducted on the RM scores, considering the third and fourth scoring methods for the IV and RV memories to determine whether the two types of memories differed in terms of RM scores. Alpha was set to .013 in an attempt to control for Type I error (alpha of .05 divided by four comparisons). For the third scoring method (sum of all items on the MCQ, reverse scoring items theoretically believed to be negatively correlated), there was no significant difference between RM scores for IV (*M* = 172.07) and RV memories (*M* = 166.65), $t$(115) = 2.29, $p$ > .013. For the fourth scoring method (sum of all items on the MCQ), IV memories had significantly higher scores (*M*

= 187.71) as compared to RV memories ($M = 178.13$), $t(115) = 3.02$, $p < .013$. The effect size for this difference was small, $d = 0.28$.

**Additional Analysis Regarding Hypothesis 5: Psychopathy and RM**

For the analyses related to the RM scores and psychopathy, a psychopath was defined as any participant who scored 30 or above on the PCL-R and a nonpsychopath was defined as any participant who scored 20 or below on the PCL-R. As before, this revised distinction was implemented to make the two groups of participants more distinct from each other, in relation to the psychopathy construct. Analyses were conducted on only the memories found to be credible according to CBCA to reduce the possibility of extraneous variables related to credibility affecting the comparisons.

*T*-tests were conducted to compare RM scores for the third and fourth scoring methods for psychopaths and nonpsychopaths. Alpha was set to .004 in an effort to control for Type I error (alpha of .05 divided by twelve comparisons). RM scores did not differ according to psychopathy for either the first or the second scoring method for IV, RV, or LM memories. Means related to these tests are presented on Table 58. Results of the *t*-tests are presented in Appendix F, Table 51.

Table 58: Mean RM scores for the third and fourth scoring methods (standard deviations in parentheses) for IV, RV, and LM memories according to psychopathy.

|  |  | Nonpsychopath ($\leq 20$ on PCL-R) | Psychopath ($\geq 30$ on PCL-R) |
|---|---|---|---|
| IV | n | 21 | 42 |
|  | RM Scoring Method 3 | 173.38 (25.57) | 173.49 (26.51) |
|  | RM Scoring Method 4 | 183.67 (34.15) | 190.63 (34.80) |
| RV | n | 20 | 52 |
|  | RM Scoring Method 3 | 162.65 (28.51) | 169.63 (25.68) |
|  | RM Scoring Method 4 | 170.05 (36.81) | 183.20 (34.65) |
| LM | n | 10 | 26 |
|  | RM Scoring Method 3 | 138.60 (27.61) | 137.77 (26.82) |
|  | RM Scoring Method 4 | 135.00 (28.08) | 141.92 (38.53) |

*Note:* All comparisons nonsignificant

G) Additional Psychopathy Analyses

Table 59: Mean CBCA criterion scores for psychopaths and nonpsychopaths for IV, RV, and LM memories (standard deviations in parentheses).

| | IV | | RV | | LM | |
|---|---|---|---|---|---|---|
| Criterion | Non-psychopath (PCL-R < 30) | Psychopath (PCL-R ≥ 30) | Non-psychopath (PCL-R < 30) | Psychopath (PCL-R ≥ 30) | Non-psychopath (PCL-R <30) | Psychopath (PCL-R ≥ 30) |
| N | 69 | 39 | 70 | 42 | 45 | 25 |
| 1 | 1.96 (0.21) | 1.95 (0.22) | 1.96 (0.20) | 1.88 (0.33) | 1.89 (0.32) | 1.76 (0.44) |
| 2 | 1.90 (0.30) | 1.95 (0.22) | 1.87 (0.34) | 1.95 (0.22) | 1.84 (0.37) | 1.88 (0.33) |
| 3 | 1.87 (0.33) | 2.00 (0.00)[a] | 1.84 (0.37) | 1.90 (0.30) | 1.87 (0.34) | 1.84 (0.37) |
| 4 | 1.80 (0.41) | 1.85 (0.37) | 1.79 (0.41) | 1.76 (0.53) | 1.82 (0.39) | 1.72 (0.46) |
| 5 | 1.91 (0.28) | 1.95 (0.22) | 1.91 (0.28) | 1.95 (0.22) | 1.84 (0.37) | 1.96 (0.20) |
| 6 | 1.16 (0.80) | 1.44 (0.68) | 1.19 (0.77) | 1.50 (0.55) | 0.89 (0.80) | 1.00 (0.76) |
| 7 | 0.23 (0.49) | 0.21 (0.41) | 0.14 (0.43) | 0.07 (0.34) | 0.16 (0.42) | 0.32 (0.56) |
| 8 | 0.19 (0.49) | 0.21 (0.47) | 0.17 (0.42) | 0.29 (0.55) | 0.07 (0.25) | 0.24 (0.52) |
| 9 | 0.58 (0.72) | 0.49 (0.56) | 0.46 (0.58) | 0.74 (0.70) | 0.42 (0.62) | 0.52 (0.71) |
| 11 | 0.43 (0.70) | 0.44 (0.68) | 0.29 (0.54) | 0.60 (0.70) | 0.31 (0.60) | 0.36 (0.70) |
| 12 | 1.77 (0.42) | 1.59 (0.60) | 1.62 (0.52) | 1.77 (0.47) | 1.48 (0.66) | 1.41 (0.57) |
| 13 | 1.23 (0.68) | 1.21 (0.73) | 1.18 (0.68) | 1.19 (0.70) | 0.98 (0.72) | 0.81 (0.68) |
| 14 | 0.22 (0.54) | 0.15 (0.43) | 0.09 (0.28) | 0.10 (0.30) | 0.07 (0.33) | 0.00 (0.00)[a] |
| 15 | 0.54 (0.61) | 0.62 (0.59) | 0.41 (0.55) | 0.67 (0.69) | 0.38 (0.54) | 0.60 (0.50) |
| 16 | 1.45 (0.63) | 1.44 (0.68) | 1.44 (0.65) | 1.24 (0.79) | 1.49 (0.59) | 1.56 (0.65) |
| 17 | 1.23 (0.71) | 1.13 (0.77) | 1.17 (0.74) | 1.21 (0.84) | 1.44 (0.62) | 1.56 (0.65) |
| 18 | 0.59 (0.58) | 0.54 (0.68) | 0.46 (0.53) | 0.55 (0.63) | 0.73 (0.65) | 0.80 (0.58) |
| 19 | 0.00 (0.00) | 0.03 (0.16)[a] | 0.03 (0.17) | 0.02 (0.15) | 0.00 (0.00) | 0.04 (0.20)[a] |
| 20 | 0.14 (0.39 | 0.08 (0.27) | 0.04 (0.20) | 0.14 (0.35) | 0.07 (0.25) | 0.04 (0.20) |
| 21 | 0.07 (0.36) | 0.05 (0.32) | 0.01 (0.12) | 0.05 (0.22) | 0.02 (0.15) | 0.00 (0.00)[a] |
| 22 | 0.28 (0.66) | 0.36 (0.74) | 0.33 (0.68) | 0.62 (0.88) | 0.36 (0.68) | 0.48 (0.82) |
| 23 | 0.09 (0.28) | 0.03 (0.16) | 0.04 (0.20) | 0.07 (0.26) | 0.07 (0.25) | 0.08 (0.28) |
| 24 | 0.00 (0.00) | 0.08 (0.35)[a] | 0.01 (0.12) | 0.05 (0.31) | 0.00 (0.00) | 0.04 (0.20)[a] |

Note: all comparisons nonsignificant.
[a] indicates that comparisons could not be performed for that pair because at least one standard deviation was equal to 0.00.

Table 60: Results of Welch's *t*-tests for comparing each CBCA criterion according to psychopathy (psychopaths score 30 or above on PCL-R; nonpsychopaths score less than 30) for IV, RV, and LM memories.

| | IV | | | RV | | | LM | | |
|---|---|---|---|---|---|---|---|---|---|
| Criterion | *t* | df | *p* | *t* | df | *p* | *t* | df | *p* |
| 1 | 0.18 | 74 | .858 | 1.36 | 60 | .180 | 1.30 | 38 | .202 |
| 2 | -0.98 | 99 | .330 | -1.55 | 110 | .124 | -0.41 | 54 | .681 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 3 | n/a | n/a | n/a | -0.98 | 100 | .331 | 0.29 | 46 | .770 |
| 4 | -0.64 | 86 | .521 | 0.25 | 71 | .805 | 0.94 | 43 | .350 |
| 5 | -0.72 | 95 | .473 | -0.81 | 104 | .423 | -1.71 | 68 | .092 |
| 6 | -1.90 | 90 | .060 | -2.51 | 106 | .014 | -0.57 | 52 | .570 |
| 7 | 0.30 | 91 | .762 | 0.97 | 101 | .333 | -1.28 | 40 | .206 |
| 8 | -0.18 | 82 | .862 | -1.16 | 69 | .252 | -1.56 | 30 | .129 |
| 9 | 0.75 | 96 | .457 | -2.19 | 74 | .032 | -0.57 | 44 | .569 |
| 11 | -0.01 | 81 | .994 | -2.46 | 70 | .017 | -0.30 | 43 | .770 |
| 12 | 1.72 | 59 | .091 | -1.65 | 107 | .102 | 0.48 | 61 | .631 |
| 13 | 0.14 | 74 | .887 | -0.03 | 99 | .973 | 0.97 | 57 | .335 |
| 14 | 0.67 | 94 | .504 | -0.17 | 83 | .867 | n/a | n/a | n/a |
| 15 | -0.66 | 81 | .510 | -2.02 | 72 | .047 | -1.74 | 53 | .088 |
| 16 | 0.10 | 74 | .920 | 1.42 | 74 | .161 | -0.45 | 46 | .653 |
| 17 | 0.69 | 74 | .491 | -0.27 | 78 | .786 | -0.72 | 48 | .473 |
| 18 | 0.43 | 69 | .668 | -0.78 | 75 | .439 | -0.44 | 55 | .661 |
| 19 | n/a | n/a | n/a | 0.15 | 92 | .879 | n/a | n/a | n/a |
| 20 | 1.06 | 102 | .292 | -1.67 | 58 | .100 | 0.49 | 60 | .629 |
| 21 | 0.32 | 86 | .752 | -0.92 | 56 | .361 | n/a | n/a | n/a |
| 22 | -0.58 | 72 | .561 | -1.84 | 70 | .071 | -0.64 | 42 | .523 |
| 23 | 1.44 | 106 | .154 | -0.61 | 71 | .545 | -0.20 | 46 | .843 |
| 24 | n/a | n/a | n/a | -0.67 | 48 | .506 | n/a | n/a | n/a |

α = .001

n/a = Welch's *t*-tests were not performed because at least one variance was zero.

Table 61: Mean frequencies for CBCA Criteria 12 and 13 affect, cognition, and physical for IV, RV, and LM memories (standard deviations in parentheses).

| | IV | | RV | | LM | |
|---|---|---|---|---|---|---|
| Criterion Frequency | Non-psychopath (PCL-R < 30) | Psychopath (PCL-R ≥ 30) | Non-psychopath (PCL-R < 30) | Psychopath (PCL-R ≥ 30) | Non-psychopath (PCL-R < 30) | Psychopath (PCL-R ≥ 30) |
| N | 69 | 39 | 70 | 42 | 45 | 25 |
| 12 Total | 10.80 (7.13) | 9.33 (8.00) | 8.46 (8.45) | 10.98 (9.24) | 8.33 (8.03) | 7.76 (9.89) |
| 12 Affect | 5.06 (4.34) | 4.03 (3.92) | 3.83 (3.28) | 4.62 (3.58) | 3.76 (4.92) | 2.88 (4.50) |
| 12 Cognition | 4.99 (4.45) | 4.62 (3.92) | 3.84 (5.46) | 5.93 (6.63) | 3.36 (4.83) | 3.68 (5.70) |
| 12 Physical | 0.69 (0.99) | 0.31 (0.60) | 0.61 (0.99) | 0.58 (1.37) | 1.20 (1.90) | 1.30 (1.35) |
| 13 Total | 4.16 (4.20) | 3.72 (2.96) | 3.51 (3.65) | 3.67 (3.46) | 2.87 (2.98) | 2.20 (3.23) |
| 13 Affect | 2.45 (3.36) | 1.97 (2.11) | 2.14 (2.47) | 2.00 (2.38) | 1.60 (1.94) | 1.28 (2.03) |

| | | | | | |
|---|---|---|---|---|---|
| 13 Cognition | 1.42 (1.89) | 1.51 (1.48) | 1.26 (1.76) | 1.55 (1.64) | 1.04 (1.40) | 0.48 (0.92) |
| 13 Physical | 0.31 (0.60) | 0.23 (0.54) | 0.11 (0.36) | 0.17 (0.43) | 0.22 (0.59) | 0.41 (0.97) |

Note: All comparisons nonsignificant.

Table 62: Results of *t*-tests for comparing frequencies for CBCA Criteria 12 and 13 affect, cognition, and physical according to psychopathy (PCL-R score of 30 cut-off) for IV, RV, and LM memories.

| | IV df = 108 | | RV df = 117 | | LM df = | |
|---|---|---|---|---|---|---|
| Criterion | *t* | *p* | *t* | *p* | *t* | *p* |
| 12 Total | 0.81 | .422 | -1.24 | .219 | 0.43 | .669 |
| 12 Affect | 1.08 | .283 | -0.78 | .435 | 0.82 | .416 |
| 12 Cognition | 0.29 | .775 | -1.56 | .121 | -0.13 | .900 |
| 12 Physical | -0.01 | .992 | 0.10 | .918 | -0.24 | .810 |
| 13 Total | 0.51 | .614 | 0.11 | .917 | 0.99 | .325 |
| 13 Affect | 0.74 | .462 | 0.53 | .595 | 0.73 | .469 |
| 13 Cognition | -0.30 | .765 | -0.50 | .621 | 1.93 | .057 |
| 13 Physical | 0.69 | .494 | -0.74 | .460 | -1.04 | .302 |

α = .002

Table 63: Mean RM total scores for psychopaths and nonpsychopaths for IV, RV, and LM memories for all RM scoring methods (standard deviations in parentheses).

| | IV | | RV | | LM | |
|---|---|---|---|---|---|---|
| | Non-psychopath (PCL-R < 30) | Psychopath (PCL-R ≥ 30) | Non-psychopath (PCL-R < 30) | Psychopath (PCL-R ≥ 30) | Non-psychopath (PCL-R < 30) | Psychopath (PCL-R ≥ 30) |
| N | 69 | 39 | 70 | 42 | 45 | 25 |
| Scoring Method 1 | 36.93 (6.00) | 37.70 (6.95) | 35.47 (6.08) | 36.52 (6.01) | 28.92 (6.71) | 28.52 (6.12) |
| Scoring Method 2 | 38.50 (7.05) | 39.57 (7.88) | 36.34 (7.52) | 38.08 (7.31) | 29.04 (7.58) | 29.92 (7.72) |
| Scoring Method 3 | 170.70 (25.62) | 173.49 (26.51) | 164.32 (24.82) | 169.63 (25.68) | 136.93 (26.62) | 137.77 (26.82) |
| Scoring Method 4 | 185.04 (33.27) | 190.63 (34.80) | 173.53 (34.34) | 183.20 (34.65) | 137.39 (35.00) | 141.92 (38.53) |

Note: All comparisons nonsignificant.

Table 64: Results of *t*-tests for comparing RM total scores for each RM scoring method for psychopaths and nonpsychopaths (PCL-R score of 30 cut-off) for IV, RV, and LM memories.

| | IV | RV | LM |
|---|---|---|---|
| | | | |

|  | df = 110 | | df = 124 | | df = 71 | |
|---|---|---|---|---|---|---|
|  | t | p | t | p | t | p |
| Scoring Method 1 | -0.62 | .536 | -0.95 | .342 | 0.25 | .804 |
| Scoring Method 2 | -0.75 | .458 | -1.29 | .198 | 0.07 | .948 |
| Scoring Method 3 | -0.55 | .583 | -1.17 | .246 | -0.13 | .897 |
| Scoring Method 4 | -0.85 | .399 | -1.55 | .124 | -0.51 | .611 |

α = .004

Table 65: Mean RM construct scores for psychopaths and nonpsychopaths for IV, RV, and LM memories (standard deviations in parentheses).

|  | IV | | RV | | LM | |
|---|---|---|---|---|---|---|
|  | Non-psychopath (PCL-R < 30) | Psychopath (PCL-R ≥ 30) | Non-psychopath (PCL-R < 30) | Psychopath (PCL-R ≥ 30) | Non-psychopath (PCL-R < 30) | Psychopath (PCL-R ≥ 30) |
| N | 69 | 39 | 70 | 42 | 45 | 25 |
| Clarity | 5.26 (1.11) | 5.40 (1.14) | 4.89 (1.06) | 5.08 (1.19) | 3.19 (1.29) | 3.50 (1.26) |
| Sensory Experiences | 3.01 (1.69) | 3.29 (1.68) | 3.13 (1.61) | 3.23 (1.46) | 2.17 (1.30) | 2.33 (1.36) |
| Spatial Information | 5.47 (1.11) | 5.58 (1.56) | 5.09 (1.46) | 5.46 (1.47) | 3.55 (1.86) | 3.61 (1.34) |
| Time Information | 4.82 (1.28) | 5.02 (1.23) | 4.36 (1.21) | 4.97 (1.19) | 3.66 (1.34) | 4.22 (1.48) |
| Overall Impression | 4.31 (1.17) | 4.53 (1.27) | 4.43 (1.13) | 4.78 (1.06) | 3.76 (0.89) | 3.49 (0.95) |
| Retrospective Recollection | 4.31 (1.17) | 4.53 (1.27) | 4.43 (1.13) | 4.78 (1.06) | 4.06 (1.03) | 4.20 (1.20) |
| Events Before and After | 4.96 (1.51) | 4.93 (1.59) | 4.32 (1.89) | 4.73 (1.85) | 3.98 (1.64) | 3.07 (1.46) |
| Bizarreness | 5.87 (0.96) | 5.88 (1.17) | 5.80 (1.06) | 5.69 (1.12) | 4.51 (1.67) | 4.60 (1.46) |

Note: All comparisons nonsignificant.

Table 66: Results of Welch's *t*-tests for comparing each RM construct according to psychopathy (PCL-R cut-off of 30), for IV, RV, and LM memories.

|  | Construct | t | df | p |
|---|---|---|---|---|
| IV | Clarity | -0.62 | 84 | .540 |
|  | Sensory Experiences | -0.85 | 87 | .397 |
|  | Spatial Information | -0.41 | 66 | .685 |

|  | | | | |
|---|---|---|---|---|
|  | Time Information | -0.82 | 89 | .414 |
|  | Overall Impression | -0.91 | 81 | .368 |
|  | Retrospective Recollection | -0.70 | 89 | .488 |
|  | Events Before and After | 0.12 | 83 | .907 |
|  | Bizarreness | -0.63 | 74 | .950 |
| RV | Clarity | -0.92 | 102 | .362 |
|  | Sensory Experiences | -0.36 | 116 | .723 |
|  | Spatial Information | -1.39 | 109 | .169 |
|  | Time Information | -2.81 | 111 | .006 |
|  | Overall Impression | 0.79 | 113 | .434 |
|  | Retrospective Recollection | -1.76 | 114 | .081 |
|  | Events Before and After | -1.22 | 111 | .225 |
|  | Bizarreness | 0.59 | 106 | .559 |
| LM | Clarity | -0.98 | 55 | .330 |
|  | Sensory Experiences | -0.50 | 52 | .622 |
|  | Spatial Information | -0.16 | 68 | .875 |
|  | Time Information | -1.61 | 50 | .113 |
|  | Overall Impression | 1.21 | 51 | .233 |
|  | Retrospective Recollection | -0.52 | 45 | .605 |
|  | Events Before and After | 2.46 | 59 | .017 |
|  | Bizarreness | -0.26 | 60 | .792 |

$\alpha = .002$