THE UTILITY OF THE COGNITIVE INTERVIEW AS A CREDIBILITY ASSESSMENT TOOL

by

GAYLA SWIHART

B.A., The University of British Columbia, 1996

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES

(>Department of Psychology, Forensic Area

We accept this thesis as conforming To the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

September 1998
© Gayla Swihart, 1998
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Psychology
The University of British Columbia
Vancouver, Canada

Date September 30, 1998

DE-6 (2/88)
ABSTRACT

The Cognitive Interview is currently being used by various law enforcement agencies as a credibility assessment tool. This research is designed to investigate whether there is utility in using the CI in this way. Seventy-eight subjects were interviewed, each telling 2 stories, one true and one false, and these stories were coded for various verbal content characteristics—number of words, number of details, number of new details generated by the CI, filled pauses, repetitions, subjective details, and references to the difficulty of the task. There was a significant difference in the number of overall new details between the true and false accounts. There were more new details (as a function of total number of words) for the true stories at least when the true stories were told first. There were several interactions involving version and order for word count, details, and new details, but no other significant findings for other coded clues to deception. The findings suggest that the CI may be useful as a credibility assessment tool, however caution is suggested before accepting its utility until further investigation is complete.
# Abstract

**List of Tables**

**List of Figures**

**Acknowledgments**

## 1. Introduction

### 1.a Physiological Clues to Deception

### 1.b Nonverbal Behavioral Clues to Deception

### 1.c Verbal Clues to Deception

### 1.d Cognitive Interview

#### 1.d.i The Cognitive Interview as a Credibility Assessment Tool

## 2. Method

### 2.a Participants

### 2.b Interviewer Training

### 2.c Procedure

### 2.d Evaluation of Interview Transcripts

## 3. Results

### 3.a Reliability Checks

### 3.b Final Analyses

### 3.c Details

#### 3.c.i New Details

#### 3.c.ii New Details Combined

### 3.d Word Count

### 3.e Manipulation Check

## 4. Discussion

## 5. Conclusion

**Tables**

**Figures**

**References**
LIST OF TABLES

Table 1. Details

Table 2. New Details : Word Count Ratios

Table 3. New Details Combined

Table 4. Word Count

Table 5. Reference to difficulty of task
LIST OF FIGURES

Figure 1. Number of details as a function of story version and story order. 34

Figure 2. Rates of new details to word count as a function of story version and story order. 35

Figure 3: Ratio of new details overall to word count of V2+V3 as a function of order. 36

Figure 4: Word count as a function of story version and story order. 37
ACKNOWLEDGMENTS

To my nephew
Ty Swihart

I am indebted to many, so my list is long....

A very special thank-you to Jenny Lang for your boundless energy, enthusiasm, organizational skills, & hard work. You are the greatest! Thank-you to my supervisor John Yuille, for your support, patience, and good advice; to Eric Eich for your support, encouragement, & statistical assistance; and to Bob Hare for agreeing to be the chairperson for my defense. Also, many thanks Don Dutton for your last minute assistance & for adding a jovial atmosphere to my defense.

A special thanks to Leora Stacee-Chun for your commitment and hard work, & to Rhea Del Vecchio, Patricia Vilaysne, & the other interviewers and volunteers who so kindly shared their time this year to assist me with this project. Also, thank-you to Tracy Lavin for getting me through stats class, creating my figures, assisting me with my presentation & for being such a good friend.

I am indebted to David Marxsen for your invaluable assistance in the creation and the completion of this project, and for your encouragement in between; and to Steve Porter for the good advice & for paving the way with your excellent research on verbal clues to deception. Also, thank-you to Mike Papsdorf & Tom Rutledge for your statistical advice, and to Diane Sirkia for the moral support.

To Gladys Payne (Mom): I am deeply grateful for your friendship, support, & unconditional love. Thank-you to Garret Swihart, Sr., Darcy Taylor, & David Retson for your support & encouragement. Also, to Rob Kercher- your patience & support when I was embarking on this educational path is much appreciated. To Tim Joynt, thank-you for your constant support & friendship; & to Louise Sicolo, for believing in me since the days of the Monte.... Also, to Anastasia Spalding, my gratitude for your love, friendship, & encouragement. Thanks to Garret & Marianne Swihart for your love & support. Eternal thanks to Rosaline Beekhan for the strength & courage you have shown since age 17 (at least)....

Finally, thank-you Lisa Furlan, for being my little ray of sunshine, my tireless & enthusiastic supporter & cheerleader; and thank-you to John DeHart for showing me that there is a better way, for inspiring me to try to find it, & for being living proof that grand dreams get grand results....
1. INTRODUCTION

Eyewitnesses sometimes lie, and suspects often do. The research presented here is designed to see whether it is possible to distinguish between individuals’ true and false statements. Credibility assessment, also called lie detection, has been widely explored, and the breadth of literature on the topic of lying is enormous. However, the focus here is on the circumstance in which someone is reporting on a personal event that is historically true (an event that did occur), and one that is historically false (it did not occur).

Credibility assessment can be divided into two broad categories. Verbal, which focuses on actual statement content and involves statement analysis, and nonverbal, which focuses on various other cues including physiological and behavioral clues.

1.1 Physiological Clues to Deception

There has been much research on physiological, behavioral, and verbal/content clues to deception, most of which compare samples of lying and truthful behavior, and the results have been conflicting or ambiguous (Hocking et al., 1979). The polygraph is a machine that is used to assess physiological responses in individuals answering questions. It monitors heart-rate, skin conductance, and respiration. The assumption underlying the polygraph is that individuals have emotional responses to lying and that these responses result in physiological changes. However, there are many problems with this type of test, which include the following: 1) individual differences in physiological responses make some people more difficult to detect; 2) the test produces both false positives and false negatives, although there is considerable debate about the rate of these errors (e.g. Raskin & Hare, 1978; Waid, Orne, & Orne, 1981); 3) there is no special pattern of physiological activity that occurs when one lies. Rather, it is changes in physical
activity as a result of being asked incriminating questions which is indicative of lying. Further, the polygraph can tell when there is a change in physiological response, but not what that change is (Gale, 1988); 4) mental and physical countermeasures may reduce the accuracy of polygraph tests and these counter measures require less than thirty minutes to learn (Honts, Raskin, & Kircher, 1994). Because of these problems, using physiological measures to detect deception is problematic, and one of the reasons that polygraph evidence is rarely admitted in court (Geiselman et al., 1985).

1.6 Nonverbal Behavioral Clues to Deception

Behavioral clues to deception have been extensively studied and include smiling, eye movements, hand and arm behavior, (Knapp, Hart, & Dennis, 1974), as well as posture shifts, qualifying an emotional expression, curtailing a response, modulating emotions, falsifying an emotional response (Ekman & Freisen, 1969), facial affect, and voice volume and speech rate (Mehrabian, 1971). Ekman (1992) has also researched illustrators (body or hand motions that illustrate speech), emblems (culturally established gestures with accepted meanings), and manipulators (nervous habits, such as grooming, scratching, picking). It is often assumed that people have less control over their nonverbal than verbal behavior, therefore, it should be easier to detect deception via nonverbal channels. However, a problem with nonverbal behavior is that it is fleeting, in most cases there is no proof that it occurred, and it can be easily denied.

In a courtroom context, the fact-finder may not be able to see the nonverbal behaviors of the witness. As well, there are individual as well as cultural differences in response to stress and/or lying (which cannot necessarily be differentiated), and as stated above, there is no lie response. Each individual is different from every other individual, and their behavior is not static.
across all situations. Also, some individuals have more practice at lying and may be better at it than others (Senuik & Yuille, 1995). Therefore, if there is no baseline, that is, if it is not known how the individual behaves when telling the truth, it will be difficult to assess changes in behavior when the individual is lying. It appears that nonverbal clues are particularly valuable when evaluating denials or minimal statements that involve brief verbal descriptions. For more extensive verbal descriptions, verbal clues, that is, methods based on style and content of speech, may come into play.

1.c Verbal Clues to Deception

Verbal clues to deceit have been extensively studied, but have not had much more success in detecting deceit than nonverbal clues. A study by Stiff & Miller (1986) found that none of the nonverbal clues and only one of the verbal clues they used as indicators of deception were related to actual deception. Nonverbal clues coded for included blink rate, smiling duration, hand gesture duration, posture shifts, speech error rate, and pause duration, among others. Verbal clues included statements of personal/mutual/other responsibility, self/mutual/other references, modifiers, number of words, and verbal content (how plausible, definite, clear, and concise the statement was). Only verbal content was related to actual deception. That is, the deceptive statements were less clear, consistent, definite, and plausible than the truthful statements. Ekman (1988) found that there was a reduction in the number of self-references in deceptive interviews, but that this was the only textual measure which showed any difference between truthful and untruthful statements. Ekman (1992) and Senuik & Yuille (1995) suggest that if there are nonverbal clues that alert one to the possibility that an individual is lying, the “behavioral clues are not evidence in themselves, but rather clues or warning signs to seek more information”
They further suggest that analyzing the content of the statement may be useful at this point.

Porter & Yuille (1996) describe four methods of using verbal clues to detect deceit. The first is Statement Validity Analysis (SVA), developed initially by Udo Undeutsch (1982). It consists of a validity checklist which looks at motives, physical evidence, and individual characteristics of a witness, as well as Criteria-Based Content Analysis (CBCA), which analyzes the actual verbal report. The latter consists of nineteen criteria (see Senuik & Yuille, 1995; Stellar & Kohnken, 1988) that are based on the “Undeutsch Hypothesis” which states that memories of events that were actually experienced will be quantitatively and qualitatively different than descriptions of event that were not directly experienced by the teller. That is, the story will be more rich in detail, and these will be more connected to other real-world events than false stories (Lamb, Sternberg, Hershkowitz, Orbach, & Hovav (1997b). It was initially developed to assess the credibility of children’s allegations of sexual abuse, but results of research using adults instead of children looks promising (Landry & Brigham, 1992). However, Lamb et al. (1997b) suggest that caution be exercised when using CBCA in a forensic setting as its precision is still too poor. Also, Yuille (personal communication) has noted that considerable training and experience is needed to reliably apply CBCA. Therefore, further research is necessary to investigate this and other possible methods of discriminating between true and false statements.

The second method described by Porter & Yuille (1996) is called “reality monitoring” (Johnson & Raye, 1981), which hypothesizes that descriptions of events that were actually occurred will have more references to external sensory information, and descriptions of event
that were invented will include more internally generated cognitive details and subjective information.

The third method, "scientific content analysis" or SCAN developed by Sapir (1987), is based on the theory that lengthier introductions, unnecessary connectors, and significant deviations in pronoun usage, among other things, are more likely to occur in untruthful versus truthful statements.

Lexical diversity, the fourth method, uses the "type-token ratio" (TTR), which "is obtained by dividing the number of distinct words (types) by the total number of words (tokens) in statements or segments of statements" (Porter & Yuille, 1996, p.446). It is assumed that a person who is trying to be credible may display low lexical diversity, as their language behavior tends to be more stereotyped. However, it can not yet be determined whether this method is useful as a credibility assessment tool in criminal cases (Porter & Yuille, 1995).

There are other verbal clues to deception that have been studied including verbal hedges (Schooler, Gerhard, & Loftus, 1986), number of words (Knapp et al., 1974), Information Manipulation Theory, which suggests that fabricated statements will violate general conversational expectations (McCornack, 1992), and motivational impairment effect, which posits that it is more difficult to lie when the consequences of being caught are great (DePaulo, Kirkendol, Tang, & O'Brien, 1988). These clues are described in more detail in Porter & Yuille (1995). Findings on verbal clues to deception are still inconclusive, and because of individual differences between liars it is necessary to use a within-subjects design in order to compare an individual’s true statement with his/her own false statement. The research presented here used this method.
1.4 Cognitive Interview

The utility of verbal clues in detecting deception depends enormously on the quality of interviewing. If the interviewing 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 (Ceci & Bruck, 1993; Gudjonsson & Clark, 1986; Schooler & Loftus, 1986; Cohen & Harnick, 1980). One established tool for effective interviewing is the Cognitive Interview (Geiselman, 1988).

Eyewitness memory is often the sole factor in determining whether a criminal case will be solved (Geiselman, Fisher, Firstenberg, Hutton, Sullivan, Avetissan, & Prosk, 1984). Therefore, it is important to maximize the amount and accuracy of information gathered in the interview with the victim, witness, or suspect, particularly in the early stages of an investigation, as an increase in amount of detail may provide the investigators with more leads to explore. Ronald Fisher and Edward Geiselman have developed an interviewing protocol called the Cognitive Interview, which uses accepted memory theory to enhance recall. The focus is on facilitating recollection at the time of the police interview as opposed to the time at encoding, as the typical crime scenario is usually rapid, emotionally charged, and individuals are unlikely to use conscious memory strategies for remembering (Geiselman, Fisher, Cohen, Holland, & Surtes, 1986).

The theory underlying the cognitive interview is based on two regularly accepted principles of memory. These are; 1) the memory trace is composed of many elements, and there must be feature overlap between the encoded event and retrieval cue, and 2) that 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 following four suggestions are the key elements to the Cognitive Interview:

1) Reconstruct the circumstances: The witness is told to reconstruct in his/her mind the context that surrounded the incident and to think about what the surrounding environment looked like at the scene, such as rooms, the lighting, the weather, any smells, any nearby people or objects. The witness is also asked to think about how he/she was feeling at the time and think about his/her reaction to the incident and to mentally put him/her-self back to the time that the incident occurred.

It is not always practical, nor necessarily desirable to revisit the scene of the crime, but the more similar the context is at retrieval as at encoding, the more details the witness is likely to generate. Thus, mentally reinstating the context should aid memory retrieval.

2) Be complete: The witness is advised that some people hold back information because they are not quite sure that the information is important. The witness is then asked to please not edit anything out of his/her report, even things he/she thinks may not be important.

Some individuals may not know what is important to the investigator, and sometimes the memory of an unimportant detail may lead to a memory of an important detail.

3) Recall in different orders: It is natural to go through the incident from beginning to end. However, the witness is asked to also try to go through the events in reverse order. Alternatively, he/she is asked to try starting with the thing that impressed him/her the most in the incident and then go from there, going both forward in time and backward.

When an individual recalls in a forward manner, he/she may tell the story by what he/she thinks MUST have happened, rather than what actually did happen, and asking the witness to
recall the event in another order may result in a more critical examination of the memory.

4) Change Perspective: The witness is told to try to recall the incident from different perspectives that he/she may have had, or to adopt the perspective of others who were present during the incident. For example, he/she is told to try to place him/her-self in the role of a prominent character in the incident and think about what he or she must have seen.

People usually have many perspectives on an incident, but only report one. By asking witnesses to report the perspectives of others it may be possible to obtain a more complete report (Geiselman, 1987).

Geiselman (1988) reports that each of these suggestions is effective when used alone, but are not as effective as when used altogether. He also reports that the cognitive interview is effective with college-educated students, as well as with lesser educated individuals, who are most likely to be the victims or witnesses of crime. A study by Fisher, Geiselman, & Amador (1989) found that detectives trained using the CI elicited 47% more information (details) after than before training, and 63% more information than untrained detectives. Kohnken, Schimossek, Ascherman, & Hofer (1995) found that the CI significantly increased the amount of correctly recalled details of a film showing a blood donation episode, as well as more incorrect details about the film, but accuracy rates were still the same as those of a standard or structured interview. This is important early in an investigation, when an increased number of details generated can significantly increase the possibility of solving the crime. Geiselman et al. (1986) did a series of experiments on the effect of leading questions when using the CI and found that individuals interviewed using the CI were significantly less affected by leading/misleading questions.

Further research compared the cognitive retrieval mnemonics used in the CI with
hypnosis, and with a standard police interview (Geiselman et al., 1985). Results showed that the number of incorrect items generated was the same between the three conditions, and that both CI and hypnosis interviews elicited a greater number of correct details than the standard interview, especially when there was high event density in the crime scenario. Although the CI and hypnosis appear to be equally effective at enhancing memory, evidence elicited via hypnosis is usually not admitted in court because of a heightened risk of susceptibility to suggestion. The CI, on the other hand, meets the standards of a long-standing test for the admissibility of scientific evidence in court in the United States (Frye, 1923). Another advantage of the CI over hypnosis is that it takes little time to train interviewers effectively (Kohnken et al., 1995), and less time to explain the cognitive techniques used to interviewees.

This research explores the possibility that the Cognitive Interview may also have credibility assessment value. That is, it may prove useful in discriminating between true and fabricated statements.

1.4.1 The Cognitive Interview as a Credibility Assessment Tool

Two studies have researched the use of the Cognitive Interview in credibility assessment, and these are only tangentially related to this research. Kohnken et al. (1995) researched the claim by Stellar and Wellerhaus (1992) that the CI has undesirable effects on criteria-based content analysis (CBCA), and so should not be used in cases where CBCA is used. Kohnken et al. (1995) argued that Stellar et al. (1995) did not clearly show that this was or was not the case, and further found that, in their research, the CI did not interfere with the potential of CBCA to distinguish between true and false statements.

One study has been completed to date which addresses the question of whether or not the
cognitive interview might successfully be used as a credibility assessment tool. Hernandez-Fernaud & Alonso-Quecuty (1997) used the CI and a standard Spanish police interview in a between-subjects design and examined true and false statements and found that true and false stories are qualitatively and quantitatively different from each other. Specifically, they found that truthful statements contained a larger number of accurate details than false statements, false statements had a greater number of false details than truthful statements, and the true statements had more contextual and sensorial information than the false statements. The interviewing method used does not affect the differences between the true and false accounts so the CI does not hamper the ability to detect differences between true and false accounts.

It is well established that the Cognitive Interview increases the amount of recall in people who are telling the truth, but it may not do the same when people are being deceptive. The first purpose of the present study was to determine whether the Cognitive Interview discriminates between true and false accounts by virtue of differential enhancement of number of details. That is, it was expected that there would be more details in the true account than the false account. That is, it was expected that the memory enhancing effects of the CI would only work for the true story and not the false story. The second hypothesis is that the various features of the Cognitive Interview, i.e. exhaustive recall, contextual reinstatement, recounting the event from another perspective, and recounting in another order (e.g. backwards), may be more difficult for someone who invents a story than for someone who is able to draw on an actual memory. It was expected that these people may have more inconsistencies within their stories, be less flexible, and have more filled pauses and more repetitions than when they were telling the truthful account.

It was expected that there would be differences between the standard account, that is, the
free recall version of the story, and the story when told using the Cognitive Interview.

Specifically, there should be more words, more details, and more new details in the story when told using the Cognitive Interview. It was also expected that there may be order effects. That is, if someone told the true story first, they would know what to expect and it may prepare them for telling the false story, so when they lie for the second story, they may be better able to do it convincingly. For example, they may give more details when the false story is told second, as they may have inferred from the CI instructions during the first story that giving more details was important in some way.
2. METHOD

2.a Participants

Adults (n =116, 36 males, 80 females) participated as interviewees. They were undergraduate students from the University of British Columbia and received course credit for their participation. It was requested that interviewees be fluent in English and they were informed that they would be audio- and video-taped during their interview.

2.b Interviewer training

Twelve interviewers were trained by the researcher, who was trained by a psychologist with extensive experience using the Cognitive Interview. Interviewers were given a brief history of the theory underlying the CI and its development, and then were shown an example of the interviewer using the CI with one of the trainees. A question and answer period followed, and trainees were encouraged to go over the CI instructions carefully before the next training meeting. The second training meeting began with a question and answer period, as the trainees had had a chance to practice using the CI protocol. Then the trainees broke up into pairs and were supervised by the researcher and two previously trained volunteers to assess the quality of their interviews. Also, videotapes of the completed interviews were checked randomly (one out of five) as the interviews progressed to check for interviewing quality. Twenty interviews were thrown out because of improperly conducted interviews.
2.3 Procedure

The participant was greeted by the interviewer and given an instruction booklet which described the experiment and their role in it. They were told that they would be required to tell the interviewer two stories, one true and one false, and that the interviewer was not to know which was which - i.e. the interviewer was blind to the veracity and order of the stories. The instruction booklets delineated to the participant which story to tell first, the true or the false, and this was randomized, so approximately half of the interviews begin with the true story, and approximately half begin with the false story.

The participant was encouraged to lie credibly, and although there was no external incentive, they were told that if they did not attempt to do this, the study would be of little use. The instruction booklet also stated that the false story should be completely false.

A checklist was attached to the instructions which gave the participant some suggestions regarding the type of event about which it would be appropriate to tell a story (e.g. a life threatening experience, being involved in a physical fight, getting caught breaking rules or laws, unpleasant adult experiences, an embarrassing event, witnessing an accident, accident with injuries, illnesses, operations, other), and they were asked to write on this checklist which of their stories was true and which was false (according to content, not order). If their story was not one of the suggested topics they were to briefly write the nature of the event and whether or not it was the true story. The participants each had 10–15 minutes to prepare their stories.

When the interviewer returned, the video- and audio-taping machines were turned on, and the interviewer asked the participant to begin with the first story. The participant was instructed to simply tell the interviewer what happened. This was the free recall condition, and is referred to as version 1. After the free recall, the interviewer asked the participant to retell the
story, this time using features of the CI: context reinstatement, exhaustive recall, and using another perspective. This is version 2 of the story. After this telling, the participant was then asked to tell the story a final time, backwards. This is referred to as version 3 of the story. This process was repeated for the second story. After the interview was complete, the participant was asked to fill out a debriefing form, in which he/she was asked the following questions: how much of the false story was true, where did the idea for the false story come from, did it happen to someone you know, did you see it on television or read it in a book? At this time they were given a short personality test (NEO-PIR short version), for use in future research with this data.

2.d Evaluation of interview transcripts

All of the interviews were transcribed onto computer diskette from the audio-tape, or the videotape when audio-tape quality was not ideal. They were transcribed word for word and included pauses and unintelligible words, coughs, laughter, etc. Three interviews had to be thrown out due to audio-visual equipment difficulties. A further 13 interviews were thrown out because the participant did not tell a story about an event about which one would/could assess credibility, for example, one’s feelings about their first year at university.

These transcriptions were then coded by 5 trained coders. Eight clues were coded for, and clues were divided up between coders. One coder coded for number of details in each version of each story, number of new details generated by the Cognitive Interview instructions (in versions 2 and 3 of each story), filled pauses (e.g. umm, ahh, coughing), repetitions, subjective details (e.g. I thought, I felt), and verbal references to the difficulty of the task. A second coder was trained in each of these clues for purposes of a reliability check. The coders for these clues received at least 14 hours of training. The training consisted of explanations and description of
the categories of clues to be coded, and examples were generated. Also, the coders and the
trainer went through several examples together in detail, and discussed questions and problems
throughout.

The remaining items—word count and subjective self-references (e.g. I, me, my, mine)—were coded using a computer search for the specific set of words associated with each item. Less training was required for these items and the task was divided between 4 coders.
3. RESULTS

3.a Reliability checks

There was one rater per clue for the items that were not rated via computer search, interrater reliability could not be assessed. Additionally, 15% of the interviews were coded by a second coder for all clues to assess interrater reliability. Three clues, contradictions, subjective details, and references to the difficulty of the task were excluded because of poor reliability. There were very few instances of contradictions, and a floor effect likely was responsible for the lack of reliability. It was difficult to train coders to assess for subjective details, and it is not surprising that this clue was not coded reliably. The remaining clues were all highly reliable ($r > .75$). One exception was for new details in the third version of the second story. Again, likely due to a floor effect, the reliability was $r = .68$. However, for all other versions of both the first and second story, reliability was $r > .84$, and averaged across all versions of both stories, $r = .85$. When all new details generated for a story by the Cognitive Interview were combined (i.e. version 2 and version 3), $r = .87$. As a result, new details as an item was included in the analyses.

3.b Final Analyses

The final analyses included 78 interviews (25 males, 53 females). A repeated measures 2 (true story told first/true story told second) $\times$ 2 (true story/false story) $\times$ 3 (version 1/version 2/version 3) ANOVA was run on each of the 5 clues; word count, details, new details : word count ratios, filled pauses : word count ratios, and repetitions : word count ratios.

There was a main effect of version for total number of words and for total number of details. The length of the versions varied systematically. That is, the first (free recall) and second (three combined CI components) versions were on average much longer that the third
(backwards) version (mean word count for version 1 = 468.21, mean word count for version 2 = 451.24, mean word count of version 3 = 229.62). In order to facilitate comparability between versions the remaining clues were transformed into ratios, using word count as the denominator. For example, the number of filled pauses for version 1 of the true story was divided by the number of words in version 1 of the true story. This transformation was also done for number of new details and number of repetitions. There was also a main effect of version for new details: word count ratios.

Several interaction effects were found and involved version as a factor, which was expected due to the different memory enhancement instructions of the versions. Many of the interaction effects were a result of the different lengths of the three versions. For example, there was a significant difference in the number of words between version 1 of the true story when the true story was told first, and the number of words in version 3 of the true story when the true story was told first. This is due to the fact that the length of the versions vary greatly (mean word count of version 1 = 404.91, mean word count of version 3 = 234.33). Because these results were expected and are not meaningful for the purposes of this study, they will not be explored further. Interactions that occurred within a version are relevant here because they discriminate between either true/false stories, order effects (true story being told first versus true story being told second), or both.

Many interactions also involved an effect of order, however this was not a focus of the present research. Tukey's HSD test, (Tukey, 1953-as found in Kirk, 1995) was used here to follow up on significant interactions of interest, as it is one of the most common post hoc procedure that controls the familywise type I error rate at or less than alpha for all a posteriori pairwise comparisons.
To test hypothesis 1, which proposed that there would be a differential enhancement in the number of details between true and false accounts when interviewed using the CI, the number of details and the number of new details generated by the Cognitive Interview were analyzed, and the length of the true versus the false stories were compared. The average overall length of the true (1160 words) and false (1142 words) stories did not differ significantly, nor did the average overall number of details (121 details in the true story and 114 details in the false story).

3. c Details

There was a significant interaction effect of order x truth x version for details, $F(2, 152) = 5.02$, $p < .01$ (see Table 1). Follow-up analyses showed that there was a significant difference in the number of details between the true and false accounts in the free recall version when the true story was told second, $q < .02$ (mean number details in free recall version of true story when true story was told second = 55.85, mean number details in free recall version of false story when true story was told second = 43.52).

3. c. i New Details

In the present study, 21% more details were generated by the Cognitive Interview for the true statements, and 20% more details were generated for the false statements. This increase was expected for the true accounts, but not for the false, so these results only partially conformed with expectations.

There was a significant interaction effect of order x truth x version for new details, using the new details: word count ratio $F(1,76) = 17.28$, $p < .0001$ (see Table 2). Follow-up analyses
showed that there was a significant difference in the number of new details: word count ratios between the true and false accounts for version 2 (three combined CI components) when the true story was told first, q < .001 (mean of new detail : word count ratios for version 2 of the true story when true story was told first = 0.054, mean of new detail : word count ratios for version 2 of the false story when true story was told first = 0.038). There was a significant difference in the number of new details : word count ratios between the true and false accounts for version 2 of the story when the true story was told second, q < .01 (mean of new detail : word count ratios for version 2 of the true story when true story was told second = 0.034, mean of new detail: word count ratios for version 2 of the false story when true story was told second = 0.048). There was also a significant difference in the number of new details : word count ratio between the true story is first condition versus the true story is second condition for version 2 of the true stories, q < .001 (mean of new details: word count ratios of true story in version 2 when the true story is told first = 0.054, mean of new details: word count ratios of true story in version 2 when the true story is told second = 0.034). There were no significant differences found in new details: word count ratios for version 3 (backwards version) of the stories.

3.c.ii New Details Combined

Analyses were also run on the overall number of new details within a story. That is, the number of new details in version 2 (details that were not included in version 1, the free recall condition) were added to the number of new details in version 3 (details that were not included in versions 1 and 2 of the story). There were two reasons for this. One, because of the lower reliability for the coding of new details for version 3 in the second story, which was alleviated when the new details were combined; and two, because the purpose of the Cognitive Interview is
to generate more new details overall, so combining these two is conceptually logical. The overall number of new details was transformed into a ratio using the word count for versions 2 and 3 as a denominator, and all analyses were run on this ratio.

There was a significant truth x order interaction for new details overall: word count v2 + v3 combined ratios, $F (1,76) = 19.02, p < .0001$ (see Table 3). Follow-up analyses showed that there was a significant difference in new details overall: word count v2 + v3 combined ratios between the true and false stories when the true story was told first, $q < .001$ (mean of new details : word count ratios for true stories in order 1 = 0.043, mean of new details: word count ratios for false stories in order 1 = 0.030). There was also a significant difference in new details overall: overall word count ratios between the true story when it was told first and the true story when it was told second, $q < .0001$, (mean of new details overall: word count overall for true story when it was told first = 0.043, mean of new details overall: word count overall for true story when it was told second = 0.028). The mean of the number of new details overall: word count v2 + v3 combined ratio for the true story was 0.036 and for the false story was 0.029. If one looks at the number of new details as a function of the total number of words used, then the true stories produce more new details than the false stories, at least when the true story was told first. There was approximately a 20% larger increase in the new details: word count ratios for the true accounts than in the new details: word count ratios for the false accounts. This is consistent with hypothesis 1 which proposed that the number of details would discriminate between the true and the false accounts, specifically, that more new details would be produced in the true story than the false story.
**3.d Word Count**

There was significant order x truth x version interaction for word count, $F(2, 152) = 4.21, p < .02$ (see Table 4). Follow-up analyses showed that there was a significant difference in word count for the true stories between the true story told first and the true story told second conditions for the free recall version of the true story only, $q < .0001$ (mean word count for version 1 for true story when the true story was told first $= 404.91$, mean word count for version 1 for true story when the true story was told second $= 559.82$). This was the only interaction effect within a version for word count.

Hypothesis #2 suggested that it would be more difficult to tell a fabricated story using the cognitive enhancement techniques of the cognitive interview, for example telling the story backwards. This was assessed using the following verbal measures: increases in the number of filled pauses, number of repetitions, and actual verbal references to the difficulty of the task. The number of filled pauses and number of repetitions were transformed into ratios using the word count as the denominator as was done with the details and new details. No significant effects were found for either clue.

Sufficient interrater agreement could not be reached on actual verbal references to the difficulty of the task. It appears, however, that only 19 of the 78 participants actually referred to the difficulty of the task. Of these 19, 7 occurred in the true story and 12 occurred in the false story (see Table 5).

It appears that the Cognitive Interview worked, as these findings are consistent with previous research showing that it functions to increase the number of details generated from the witness (Geiselman et al., 1988). The results partially support the hypotheses that the Cognitive Interview may function as a credibility assessment tool, as the true stories produced more new
details than the false stories (as a function of the total number of words used) at least when the true story was told first. There were some complex interactions involving the order in which the subjects gave true and false accounts, but these do not appear to qualify the preceding two conclusions.

3. Manipulation Check

Before being interviewed, participants were provided with an instruction booklet which asked them to tell the interviewer one true story and one completely false story. After their interview was complete, the participants were asked to read a debriefing form, and answer the questions that were included. This form asked them how much of their false story was true, where they got the idea for the false story, and whether or not it happened to someone they know. Most of the participants (86%) reported that 90% or more of their story was completely false. Four percent said that their false story was 75%-90% false, and 4% said that 50%-75% of their false story was false. Five percent reported that their false story was less than 50% false, and one individual (1%) did not answer this question on the debriefing form. Answers to this question were difficult to interpret, as some individuals reported a story that happened to someone else as a complete lie, because it did not happen to them, whereas others reported this to be mostly true, except for the fact that it did not happen to them. There was no clear pattern in the source of the idea for the lie, and indeed, many participants did not fill this question in. Eleven participants (14%) said their story or something similar to it had happened to someone they know.
4. DISCUSSION

The major purpose of this study was to investigate whether or not the Cognitive Interview enhances the ability to discriminate between true and false statements. Results showed that the CI does consistently produce an increase in the number of details, but it did so for both the true and the false accounts, counter to the expectation that this increase should only occur for true accounts. Results also showed that there are more new details in the true stories than the false, at least when the true story was told first, which conformed with expectations. Also, it was thought that it would be more difficult to do the various versions using the CI (in particular, the backwards version) when telling a fabricated story, and this hypothesis was not supported. Many of the differences found in this research between true and false accounts were complex and are likely not useful in a forensic setting. Most of the relevant significant differences were found within two- or three-way interactions, and were difficult to interpret, as no clear pattern was shown.

As a preliminary attempt to determine what clear differences existed between the true and false accounts the amount of detail and reported lack of memory were analyzed. These have been found in previous research to sometimes discriminate between true and false accounts (e.g. Jones & McGraw, 1987, found in Horowitz, 1991; Stellar et al., 1988, found in Raskin & Esplin, 1991; Lamb et al., 1997a; Porter et al., 1995; MacDonald & Michaud, 1987, as found in Porter et al., 1995). All found more details in plausible than implausible accounts. Hypothesis 1 suggested the Cognitive Interview may be able to discriminate between true and false accounts by virtue of differential enhancement of number of details. As reported above, there was no difference in amount of detail, although there were significant interaction effects. An inspection of Graph 1 showed that the greatest number of details were obtained in version 1 (free recall) when the true
story is told second, and that none of the other differences for versions 1 (free recall) and 2 (three combined CI components) were significant. It appears that the order \times truth \times version interaction effect for details was due primarily to the shorter length of the stories in version 3 – the backwards version- (both true and false stories were much shorter than those in versions 1 and 2). Also the subjects gave more details in the second story compared to the first regardless of the truthfulness of the account, however this effect was stronger in order 2 (false story told first, true story told second) than in order 1 (true story told first, false story told second). The fact that there were no clear findings on this item is not surprising as recent research has found that there are sometimes more details in fabricated than true accounts, contrary to what previous research has shown (Porter, 1995; Porter, 1998, Kohnken, 1995). Also, Lamb et al. (1997a) note that “Undeutsch did not claim that credible accounts necessarily contained more details than implausible accounts, and even recognized lack of confidence as an index of credibility rather than its absence” (p.177). Therefore, the literature is inconsistent as to whether or not more details are to be expected in a true versus false statement, and the present research further suggests that the number of details cannot serve as a simple yardstick for assessing credibility.

Previous research has demonstrated that admitted lack of memory consistently discriminates between true and false accounts, however, with the present data there was no difference between true and false stories in admitted lack of memory. In fact, only 11 participants admitted lack of memory in either the true or the false accounts. Two admitted lack of memory for both the true and the false accounts, and there were 7 admissions in the true stories and 6 admissions in the false stories.

As expected, the Cognitive Interview generated additional details that were not reported in the initial free recall version of the statement. Previous research on the Cognitive Interview
has found that it generates between 35% (Kohnken et al., 1995) and 63% (Fisher et al., 1989) new details. In the present study, 21% more details were generated by the Cognitive Interview for the true statements, and 20% more details were generated by the false statements. Encouragement of free recall in version 1 probably increased the number of pre-Cognitive Interview details relative to other studies.

There was a significant interaction effect of order x truth x version for new details: word count ratios as shown in Graph B. For every group there was a higher new details: word count ratio in the first than the second story, regardless of the order in which the true and false stories were told. There was a significant decrease in the new details: word count ratios from the first story to the second. This did not occur in version 3 (backwards recall). There were approximately 20% more new details in the true story than the false story when new details were analyzed as ratios using word count as a denominator. This may appear to contradict the previously stated finding that there were 20% more new details for both the true and the false stories. However, when the lengths of the accounts are factored in, it can be seen that participants who are telling a true story use fewer words to generate the equivalent amount of new details than when they are telling a false story.

There was also a significant order x truth interaction for new details: word count ratios when versions 2 (three combined CI components) and 3 (backwards) are combined. Graph C shows that there is a reduction in new details: word count ratios for the second telling of the story regardless of order, which is consistent with the findings for new details: word count ratios when versions 2 and 3 are not combined.

There was a significant order x truth x version interaction involving word count. Graph D shows that there is a slight increase in the number of words from the first to the second story),
which corresponds to the increase in number details between the first and second story (regardless of whether the true story was told first or second). This trend involving the number of words in the statement and the number of details in the statement does not correspond to the decreasing number of new details: word count ratios, which instead decreased in the second story, regardless of whether the true story was told first or second. It appears that there is no clear pattern of discrimination between true and false statements using word count, which is consistent with findings by Porter et al. (1995). However, other research using word count as a credibility assessment clue has found that there are significant differences between true and false statements (Lamb et al., 1997a; Harrison, Halek, Raney, & Fritz, 1978; Leippe, Manion, & Romanczyk, 1992-all found in Porter et al., 1995; Stiff & Miller, 1985). As with the number of details, it appears that the value of word count in assessing credibility varies with context. More research is needed to determine the context in which these variables assist and those in which they do not assist credibility assessment.

Hypothesis 2 suggested that the various features of the Cognitive Interview, i.e. exhaustive recall, contextual reinstatement, recounting the event from another perspective, and recounting the event in another order (e.g. backwards), may be more difficult for someone who invents a story than someone who is able to draw on an actual memory. It was expected that these people may have more inconsistencies within their stories, be less flexible, and have more filled pauses and more repetitions than when they were telling the truthful account. This hypothesis was not supported.

There were no significant main or interaction effects for filled pauses. Porter et al. (1995) cite a previous study by Comber & Canter, 1983 where filled pauses were analyzed and did not discriminate between true and false accounts. However, the authors note that this was probably
due to small sample sizes. Kohnken et al. (1995) found that there were significantly more
‘clichés’ in fabricated statements than in truthful statements, and defined clichés as “utterances,
expressions or words that were not necessary for the description and that introduced delays into
the report (e.g. ‘more or less’, ‘somewhat’, ‘isn’t it’, or ‘well’). Filled pauses in the present study
included “um”, “ahh”, “you know”, “whatever”, “like”, laughing or coughing, or other
utterances that could be used to ‘buy time’ to think about the story.

There were no significant differences in number of repetitions between the true and false
accounts. Previous research is inconsistent, as it has found that deceptive accounts may have an
increase in repetitions because lying is cognitively more difficult than telling the truth (Kohnken,
1985, as found in Porter et al., 1995), but Kohnken et al. (1995) later found that individuals who
were telling the truth produced more repetitions than individuals who were lying.

As there was a significant difference in the number of new details as a function of total
number of words used between the true and the false accounts, and because it is possible that
there are more subtle differences between the true and false accounts that have not been analyzed
for here, the CI may be useful as a credibility assessment tool. However, further research is
needed before making this determination. Currently, the Cognitive Interview is being used to
help investigators judge the veracity of statements by way of impressions. That is, the
investigator is interested in whether or not the suspect makes more mistakes in his/her story or
gets confused about the story when Cognitive Interview techniques are used. The present study
showed that only 19 people indicated any difficulty with the Cognitive Interview versions of
their stories and 7 of these were telling the truth.

The question of whether or not the CI can be used as a credibility assessment tool has
only been directly addressed in one other study to date. Hernandez-Fernaud et al. (1997), using a
between-subjects design, compared true and false accounts using the Cognitive Interview and a standard interview used by the police force in Spain (STI) and found that there were more details in accounts generated using the CI than using a standard interviewing protocol. The present research compared true and false accounts using only the CI in a within-subjects design, and found that there was no difference in the number of details between the true and the false accounts. Hernandez-Fernaud et al. (1997) speculated that an individual who is fabricating a statement would have difficulty recounting his/her story using the cognitive techniques involved in the CI, and as mentioned above the present research did not support this latter theory.

It might be argued that a lack of significant differences between the true and false accounts for many of the clues analyzed here may be due to the possibility that the subjects were not really telling false stories in the false story condition (that is, they were not lying) and true stories in the true story condition (that is, they were not telling the truth). This was addressed with the manipulation check included on the debriefing form and there was no evidence to suggest that the true stories were not true and that the false stories were not false. Perhaps this type of research should be repeated in a context where ground truth is known. The experimenter could stage an event and therefore know what happened and what did not happen and therefore the truthfulness of the accounts could be verified.

There is a need for more research on the use of the Cognitive Interview as a credibility assessment tool, preferably with more forensically relevant populations. Also, it is possible that this data set would have indicated further differences between true and false statements if it had been coded for CBCA, which is the most consistent statement analysis tool to date (see Raskin & Esplin, 1991) and future research will investigate this hypothesis. However, this exploratory study suggests caution in accepting its utility until further investigation is complete.
5. CONCLUSION

In conclusion, the Cognitive Interview remains a technique that enhances eyewitness recall, but whether or not the CI can be used as a way of discriminating between true and false accounts has yet to be determined. The results presented here are encouraging but equivocal. There were more new details as a function of total word count in the true stories than in the false stories, at least when the true story was told first. There were also some complex interaction effects found in this study, and these involve the order in which the subjects provided their true and false accounts. These interactions do not in any way qualify the main conclusion about the CI. However, these interactions raise the interesting possibility that order may be a subtle but useful variable in credibility assessment that is going to require further research. For example it would be useful to research this question further using a within-subjects design where ground truth is known, for example by staging an event. Also, it would be interesting to have a true-true and a false-false condition (this research used a true-false and false-true condition) to see what impact order by itself, without changing the truthfulness of the account, has on credibility assessment clues.
### Table 1: number of details as a function of story version and story order

<table>
<thead>
<tr>
<th></th>
<th>Order 1</th>
<th>Order 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>s.d.</td>
</tr>
<tr>
<td>True</td>
<td>41.82</td>
<td>21.25</td>
</tr>
<tr>
<td></td>
<td>55.84</td>
<td>25.99</td>
</tr>
<tr>
<td>False</td>
<td>48.24</td>
<td>27.31</td>
</tr>
<tr>
<td></td>
<td>43.52</td>
<td>20.96</td>
</tr>
</tbody>
</table>

*P* < .01
## New Details: Word Count Ratios

<table>
<thead>
<tr>
<th></th>
<th>Version 2</th>
<th></th>
<th>Version 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Order 1</td>
<td>Order 2</td>
<td>Order 1</td>
<td>Order 2</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>s.d.</td>
<td>Mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>True</td>
<td>0.054*</td>
<td>0.030*</td>
<td>0.034</td>
<td>0.023</td>
</tr>
<tr>
<td>False</td>
<td>0.038*</td>
<td>0.023*</td>
<td>0.048</td>
<td>0.025</td>
</tr>
<tr>
<td>Marg.</td>
<td>0.046</td>
<td>0.041</td>
<td>0.016</td>
<td>0.011</td>
</tr>
</tbody>
</table>

* p < .0001

**Table 2:** Ratios of the number of new details: word count as a function of story version and story order.
<table>
<thead>
<tr>
<th></th>
<th>True</th>
<th>False</th>
<th>Marginal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>s.d.</td>
<td>Mean</td>
</tr>
<tr>
<td>Order 1</td>
<td>0.027*</td>
<td>0.016</td>
<td>0.018*</td>
</tr>
<tr>
<td>Order 2</td>
<td>0.016*</td>
<td>0.011</td>
<td>0.022</td>
</tr>
<tr>
<td>Marginals</td>
<td>0.022</td>
<td></td>
<td>0.020</td>
</tr>
</tbody>
</table>

* p<.001

Table 3: ratios of number of new details overall: word count for version 2 + version 3 as a function of story order
# Word Count

<table>
<thead>
<tr>
<th></th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>404.9</td>
<td>204.5</td>
<td>445.9</td>
<td>362.3</td>
<td>234.3</td>
<td>127.4</td>
</tr>
<tr>
<td></td>
<td>559.8</td>
<td>310.3</td>
<td>475.1</td>
<td>276.2</td>
<td>227.2</td>
<td>126.9</td>
</tr>
<tr>
<td></td>
<td>475.5</td>
<td>272.3</td>
<td>437.5</td>
<td>303.4</td>
<td>251.9</td>
<td>159.8</td>
</tr>
<tr>
<td>False</td>
<td>452.9</td>
<td>230.7</td>
<td>453.4</td>
<td>238.0</td>
<td>205.3</td>
<td>111.7</td>
</tr>
<tr>
<td></td>
<td>243.1</td>
<td>506.4</td>
<td>464.3</td>
<td>216.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

P < .02

**Table 4:** total number of words as a function of story version and story order
Reference to difficulty of task

<table>
<thead>
<tr>
<th></th>
<th>Version 2</th>
<th>Version 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>False</td>
<td>4</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 5: total number of times participants referred to the difficulty of recounting the story using the various cognitive mnemonics of the Cognitive Interview as a function of truthfulness of the account.
Figure 1: Number of details as a function of story version and story order.
Figure 2: Ratios of new details to word count as a function of story version and story order.
Figure 3: Ratio of new details overall to word count of V2+V3 as a function of order.
Figure 4: Word count as a function of story version and story order.
REFERENCES


Frye v. United States (1923) 293 F. 1013


