

**DETECTING SELF-EXPERIENCED
ACCOUNTS USING
THE ABERDEEN REPORT JUDGMENT
SCALES**

**A thesis
submitted in partial fulfilment
of the requirements for the degree
of
Master of Science in Psychology
at the
University of Canterbury
by
Leah R. Havighurst**

University of Canterbury

2002

ACKNOWLEDGEMENTS

There are a number of people I would like to thank for their assistance in the production of this thesis.

Firstly, I would like to thank my supervisor, Dr Lucy Johnston, who provided much needed guidance, patience and time from the moment she became my supervisor.

I would also like to thank my co supervisor, Dr Sigi Sporer, for giving me the opportunity to investigate the topic of truth detection.

Special thanks to my family and friends all over the world, for their patience, knowledge and proof reading skills. Without their continuous positive comments and input, I would not have gotten through the complications, setbacks and interesting twists that made up my thesis year.

Dr Steve Hudson provided enthusiasm and much needed support that I have carried with me from beginning to end. Thank you Steve.

Finally I would like to thank those who participated in the present research for their time, effort and rating skills.

CONTENTS

Chapter		Page
	List of Tables	vii
	List of Figures	viii
	Abstract	1
1.	Introduction	2
	Verbal Verses Nonverbal Cues in Deception	
	Detection	3
	The Effects of Training on Detecting Deception	6
	The Use of Feedback in the Detection of Deception	10
	The Development of Verbal Cues in Detecting	
	Deceit	13
	Criteria-Based Content Analysis and Reality	
	Monitoring	14
	Development of the Aberdeen Report Judgment	
	Scales (ARJS)	24
	The Effects of Preparation Time on Detecting	
	Deception	28
	The Effects of Valence on Deception Detection	31
	Additional Variables in Deception Detection	33
	The Present Research	34

Chapter	Page
2.	
Experiment 1a.	39
Method	39
Participants	39
Materials	40
Procedure	42
Results and Discussion	43
Criteria Ratings	43
Credibility ratings	46
Detection Index	47
The Relationship Between Credibility and Criteria Ratings	48
Control Participants' Reasons for Credibility Ratings	49
Presence of Criteria in Self-Experienced and Transformed Accounts	50
3.	
Experiment 1b.	52
Method	53
Participants	53
Materials	53
Procedure	55
Results and Discussion	56

Chapter	Page
Post-Experimental Questionnaire	56
Criteria Ratings	56
Credibility Ratings	59
Detection Index	62
The Relationship Between Credibility and Criteria Ratings	63
Control Participants' Reasons for Credibility Ratings	65
Presence of Criteria in Self-Experienced and Invented Accounts	66
4.	
General Discussion	70
Detecting Self-Experienced Accounts Using the ARJS Scales	70
The Effects of Preparation Time on Detection	78
The Effects of Feedback on Detection	82
The Effects of Valence on Detection	85
Future Research	88
Conclusion	91
References	95
Appendices	
Appendix A	101

Chapter	Page
Appendix B	107
Appendix C	109
Appendix D	110
Appendix E	111

List of Tables

Table	Page
1. Criteria Included in Criteria-Based Content Analysis.	16
2. Descriptions of the Aberdeen Report Judgment Scales Criteria.	44
3. Mean Criteria and Credibility Ratings For Each Account Version (Experiment 1a)	45
4. Mean Credibility Ratings For Each Account Version According to Group (Experiment 1a)	46
5. Correlations Between ARJS Criteria and Credibility Ratings Across All Accounts (Experiment 1a)	48
6. Mean Criteria and Credibility Ratings For Each Account Version (Experiment 1b)	57
7. Mean Credibility Ratings For Each Account Version According to Group (Experiment 1b)	59
8. Correlations Between ARJS Criteria and Credibility Ratings Across All Accounts	64

List of Figures

Figure	Page
1. Mean Ratings on the Corrections/Memory Failure Criterion for Self-Experienced Short Preparation, Self-Experienced Long Preparation, Invented Short Preparation and Invented Long Preparation Accounts	58
2. Mean Credibility Ratings on the Corrections/Memory Failure Criterion for Self-Experienced Short Preparation, Self-Experienced Long Preparation, Invented Short Preparation and Invented Long Preparation Accounts	60

ABSTRACT

The detection of truthful statements has been investigated over the years using physiological, nonverbal and verbal cues. Researchers have been trying to establish what method of detection is most reliable, with contradictory evidence being found both for and against each of these methods of truth detection. The Aberdeen Report Judgment Scales (ARJS) enable the verbal content of a statement to be assessed using specific criteria thought to be indicative of a self-experienced account. Two experiments were carried out to investigate the ability of the ARJS criteria. Experiment 1a consisted of 48 accounts of either a pleasant or an unpleasant evening dinner. Half of the accounts were true while the other half were false. Those using the criteria, along with outcome feedback, were expected to rate self-experienced accounts as more credible than transformed accounts in comparison to those participants who did not receive guidance. No significant differences in credibility ratings were found between conditions or in relation to truth status. Experiment 1b consisted of 60 accounts of an overnight military officers training exercise. Again, half the accounts were self-experienced while the remaining accounts were invented. Within these two conditions, half of the account tellers had been given a short preparation time while the other half had been given a long preparation time before presenting their accounts. In addition, the effects of feedback on credibility ratings were investigated. A significant difference was found for credibility ratings of self-experienced and invented accounts in the short preparation condition only. No differences were found between groups. These findings, along with the implications they have on use of the ARJS scales are discussed and modifications to the present research are suggested.

CHAPTER 1

Introduction

Over the years, researchers have been trying to find a reliable method of detecting deception. The implications of a method such as this are far reaching, from a courtroom testimony or police interrogation of a suspect, to knowing whether a spouse is being faithful, or a politician honest. Perhaps because of the impact a procedure such as lie detection has on society, numerous methods have been established, including the analysis of physiological responses, behavioural indicators and verbal cues.

Research has found that people rarely produce an average accuracy rate above 60% when trying to detect deception (Ekman & O'Sullivan, 1991). By chance, the average person is expected to achieve a 50% accuracy rate, yet some groups produce an even lower result (Ekman & O'Sullivan, 1991). In fact, researchers have found that the most consistent result of the deception detection area is that people tend to be very bad at detecting lies (Vrij, Edward & Bull, 2001a). Generally, the problem seems to be that people believe almost everything they hear, the majority of the time. When they do attempt to detect a lie, it is thought that they use stereotypical ideas to aid their quest for the truth (DePaulo, 1994; Granhag & Stromwall, 2001). For example, it is believed by some that the eyes hold the key to being able to judge whether someone is lying or telling the truth. Research by Vrij (1998), however, has found this belief to be untrue. Unfortunately, it is often the case that these cues are unreliable or irrelevant when it comes to lie detection (Vrij, 1998).

Those in the field of detecting deception have developed their own methodologies for tasks such as questioning suspects, all of them trying to best distinguish those who are hiding the truth from those who are being honest (Horvath, Jayne, & Buckley, 1994). Of course these detection practises are dependent on there being observable differences in behaviour between the liar and truth teller. Whether these differences exist, and if so, what they are, has driven research in this area for many years. The three most prominent pathways for lie detection in the area of psychology are nonverbal cues (e.g., avoiding eye contact, changes in facial expression), physiological cues (such as galvanic skin response), and verbal cues (such as the inclusion of unusual details in a statement) (Sporer, 1997). The latter two methods together are often referred to as behavioural cues.

The following is a discussion of past research within the area of deception detection. The use of verbal and nonverbal cues in training individuals to better detect true and/or false statements of events (accounts) will be compared and contrasted. In particular, the analysis of a statement's content, in order to identify true accounts, will be examined and discussed in relation to its use in practical areas. In addition, variables believed to affect the ability of detection methods to identify true accounts will be reviewed.

Verbal versus Nonverbal Cues in Deception Detection

An area that has received much interest is that of behavioural cues for detecting deceit. Some research shows that in using both verbal and nonverbal measurements of honest and deceptive behaviour, one is better able to differentiate between the liar and the truth teller than someone using only one mode of lie detection (see Ekman & O'Sullivan, 1991, for a review). Ekman and O'Sullivan (1991) investigated the use of verbal and nonverbal cues in deception detection by evaluating the abilities of 509 people, ranging from the U.S. Secret Service

through to university students. They concluded that the use of both verbal and nonverbal cues were far more reliable than using verbal cues alone. They also found that using nonverbal cues alone produced similar accuracy rates to when the nonverbal and verbal cues were combined. Other researchers have found differing results when examining verbal and/or nonverbal cues in lie detection, with nonverbal cues failing to produce as positive a result as that of Ekman and O'Sullivan (Landry & Brigham, 1992; Steller, 1989; Zuckerman, Koestner, & Colella, 1985).

In Miller and Stiff's (1992) review of the deceptive communication literature, they suggest that nonverbal behaviour may not be as accurate or as useful a cue as is generally thought when investigating deceptive intent. Arousal results from the knowledge that one is about to lie. This, in turn, is what is measured by physiological detectors such as polygraph tests and nonverbal cues such as fidgeting and avoidance of eye contact (Miller & Stiff, 1992). These cues to arousal can also be linked to arousal in non-deceptive circumstances (Miller & Stiff, 1992). An example of this, given by Miller and Stiff (1992), is of an individual who is aroused due to naturally being a socially anxious person who must interact with a stranger, therefore, causing behaviour that could be wrongly construed as deceptive. Further research, in an attempt to avoid this caveat, could examine the individual's behaviour as they communicate in an honest manner (i.e., asking them questions that they will answer honestly, such as demographic information), in order to compare this with the parts of the statement thought to be deceptive. This process is often used in preliminary questioning when implementing polygraph tests or during police interviews (Bull, 1989).

Along with the caveat pointed out by Miller and Stiff, a number of studies have been carried out that fail to replicate the findings presented in Ekman and O'Sullivan's research

(Bauchner, Kaplan, & Miller, 1980; Miller & Burgoon, 1982; Stiff & Miller, 1986). In one experiment, Stiff and Miller (1986) asked undergraduate students to rely on both verbal and nonverbal cues in detecting deception. Previous research had found that the nonverbal cues were related to subjective veracity judgments but not actual deceptive behaviour, while verbal content cues had been related to both perceived and actual message veracity. Stiff and Miller's (1986) findings confirmed this, with none of the nonverbal cues used as indicators of deception related to actual deception, while verbal content was the sole cue associated with message veracity. This research showed that people were using nonverbal cues that were not related to the whether the statement was actually true or false. In addition, the examination of an account's content proved to be the most reliable aspect when detecting deceit.

Not only do those studies mentioned above show the disadvantages of using solely nonverbal cues, but more specifically, some suggest that an advantage may arise from restricting the amount of nonverbal information made available to the perceiver (Bauchner et al., 1980; Stiff & Miller, 1986). This is justified by pointing out that these unhelpful nonverbal cues are in fact detrimental distracters rather than helpful aids in detecting deceit (Miller & Stiff, 1992). Therefore, in order to best detect deception, nonverbal aspects of an account should be eliminated.

In a meta-analysis on the use of nonverbal cues in deception detection, Zuckerman, DePaulo, and Rosenthal (1981) stated that even the most valid nonverbal cues have provided weak correlations with the truth status of an account. Zuckerman et al. (1981) concluded that the use of words, for example in the form of transcripts, enabled deception to be far better detected, from any source of information, compared to any other predictor that did not include words (i.e., facial expressions or body movements). For this reason, it can be concluded that

the use of verbal cues in discriminating between true and false statements is far more reliable and may even result in incorrect cues being used less often. Miller and Stiff (1992) suggested, from their analysis of deception detection literature, that “perhaps nonverbal cues are of limited utility because detectors rely on the wrong ones to make their judgments” (p. 234). In order for the appropriate nonverbal and verbal cues to be used by those detecting deception, training is seen as being a vital research tool.

The Effects of Training on Detecting Deception

While the debate continues as to which method of deception detection is most beneficial, a large quantity of research has been produced that examines the effects of both verbal and nonverbal methods of detecting deceit. A number of studies produced and reviewed by Rosenthal, Hall, DiMatteo, Rogers, and Archer (1979) have indicated that both practice and/or training improved people’s accuracy in detecting deception.

An example of the effects of training is Fiedler and Walka’s (1993) study that had adults judge the veracity of 40 accounts of minor delinquent acts, with the trained group being given information on seven nonverbal cues to be used when making their credibility judgments. These seven cues were taken from a meta-analysis by Zuckerman and Driver (1984) that suggested a number of effective cues for discriminating true from false accounts. Fiedler and Walka (1993) explained that those not given training in the appropriate cues to use for deception detection would resort to the use of stereotypical or heuristic knowledge that was not related to the accurate assessment of authentic cues.

Fiedler and Walka (1993) found that the performance of those not receiving training in the use on nonverbal cues was low in comparison to those who received the training. They

concluded that human lie detectors naturally lack the ability to diagnose, and use, accurate cues for detecting deceit, and hence, when training is received, discrimination between self-experienced and invented accounts is improved.

Even the use of more simple training techniques has been found to produce favourable results. DePaulo, Lassiter, and Stone (1982) used a very simple manipulation, suggesting to their experimental participants that they focus their attention on the sender's tone of voice. Findings showed that this small amount of nonverbal training resulted in participants significantly increasing their ability to detect deceit (DePaulo et al., 1982). Whether this improvement was due to the ability of the cue to indicate deceit, or whether it is merely that the consequence of focusing on a cue is an increase in information processing leading to more accurate detection abilities, is unclear.

Research by Ekman and Friesen (1974) and by Druckman, Rozelle, and Baxter (1982) also investigated this improvement in detection abilities through the use of training. Ekman and Friesen (1974) found that untrained observers could not identify when nurses in the study were lying about a pleasant emotion when viewing their faces, while the trained observers were able to make accurate veracity judgments. This implied that the two groups were using different cues to detect deceit, with only one group adopting accurate deception cues (Ekman & Friesen, 1974). Druckman et al. (1982) gave their participants either an orientation lecture on nonverbal communication, a technical briefing on previous experimental results, or "inference training". Inference training consisted of participants receiving a briefing on past experimental results, information on what nonverbal behaviours have discriminated between groups in past studies, and a procedure for inferring one's intentions from this nonverbal behaviour. All groups were then asked to detect deception from videotapes. Druckman et al.

(1982) found that only the “inference training” group produced significant gains in their lie detection abilities, suggesting that specific types of training are more effective in the deception detection task.

DeTurck, Harszrak, Bodhorn and Texter (1990) and deTurck and Miller’s (1990) studies produced results to suggest that their training of participants was highly effective, despite the limited training and practice time. The trained social perceivers had enhanced detection skills, producing a higher accuracy level for the self-experienced and invented accounts compared to those who received no training (deTurck et al., 1990). The authors noted the raters’ ability to accurately judge the veracity of the communicators, despite a different set of communicators being used in the training session. This suggests that the cues used in this study were not person-specific, and, therefore, could be used with various communicators (deTurck et al., 1990).

Not all studies have found this generalisability across communicators. Zuckerman et al. (1984) concluded from their findings that accuracy in judging statements was only enhanced when the individual was making a credibility rating of the sender they were trained to detect. Differences in procedures may help explain these contradictory results, however. Zuckerman et al.’s (1984) training consisted of feedback as to the account’s veracity, either before or after the account was judged. DeTurck and colleges (1990) were more specific, training their participants to use six nonverbal cues to detect deception. This more focused procedure may have resulted in the participants showing an increase in accuracy in comparison to the control group, while Zuckerman et al.’s (1984) study did not provide accurate enough training in order for it to be beneficial.

Other studies exploring the effect of training on veracity judgments have not always been favourable. In Kassin and Fong's (1999) study, 40 raters were either trained in the use of verbal (e.g., use of first-person pronouns and descriptive verbs) or nonverbal (e.g., posture and eye contact) deception cues, or were given no training. A similar study by Koehnken (1987) used police officers to judge the credibility of eyewitness statements, where experimental groups were trained to use verbal (e.g., amount of details and spontaneous corrections) and nonverbal cues (e.g., head, leg and foot movements). Analyses in both studies found that training raters to use verbal and nonverbal cues resulted in no improvement in accurate veracity judgments. In fact, the training in Kassin and Fong's (1999) study resulted in impaired performance in comparison to the no-training group. Kassin and Fong (1999) suggested that this negative outcome may have been due to the use of cues that had not previously been shown to diagnose veracity, therefore, were not reliably able to discriminate between liars and truth tellers. As no further research has examined the cues used by Kassin and Fong (1999) this justification cannot be confirmed.

Vrij (1994) also provided a possible explanation as to why some researchers were not finding a positive effect for training. He suggested that raters in these studies were not using the information provided in the appropriate way, with individuals either not putting enough weight on the ability of the cues, or failing to use the information altogether due to the belief that the training did not work, instead thinking that their own theories of lie detection were far better. Vrij (1994) found in the manipulation check of one of his studies that many raters did not agree with the information used to later assess statements. In addition to these explanations, Bull (1989) gave an alternative explanation for Koehnken's lack of training effects. He suggested that the training group may have increased their accuracy rates had they been given outcome feedback (informed as to whether the account was true or false) as they

progressed through the detection trials. This, however, does not explain the positive findings from studies that did not use feedback.

The Use of Feedback in the Detection of Deception

One factor that studies examining verbal cues have not addressed is the effect of feedback on guided participants. In contrast, the possible advantages of feedback have been investigated by numerous nonverbal cue researchers over the years. Zuckerman et al. (1985) looked at the difference in accuracy between a control group and two experimental groups who received feedback after each of their credibility ratings, telling them if the statements were self-experienced or invented. Results showed that the experimental groups (speech only, speech plus face and face only), despite receiving feedback only (i.e., no training), provided significantly more accurate credibility ratings than the control group (Zuckerman et al., 1985). This indicates that feedback alone may be a powerful enough tool to provide more accurate findings than someone who is given no guidance whatsoever.

An interesting outcome to Zuckerman et al.'s (1985) study was that the speech only and speech plus face modes of information presentation showed a gradual increase in accuracy over the course of the experiment. Zuckerman et al. (1985) indicated that this lack of increase in accuracy for the face only mode of presentation was in line with the suggestion that the face did not provide a good source of cues about deception. More importantly, however, this research indicates that in order for feedback to be fully taken advantage of, it needs to be applied to multiple statements in order for the user to develop either their own cues in line with the feedback, or become accustomed to the cues given to them with the help of the feedback.

The effects of feedback alone were also documented in Porter, Woodworth, and Birt's (2000) study, that examined the accuracy rates of both federal parole officers and undergraduate university students. All officers received training and feedback, while students either received training and feedback or feedback only. Not only did the officers and those students receiving training and feedback, achieve a higher accuracy rate than the control group, but so did the student feedback only group (Porter et al., 2000). These findings again indicate that feedback is a powerful enough tool that it can be used to improve detection abilities without the help of training. However, this strong support for feedback alone has not always been found to be the case.

Vrij (1994) examined outcome feedback in his study using police detectives. He found that those detectives given both training and feedback achieved a higher accuracy rate than both the control and training only groups (Vrij, 1994). This obviously contrasts with Porter et al.'s (2000) study, suggesting that when paired with appropriate training procedures, such as content analysis, feedback can be a valuable tool in the detection of deception. Vrij did not include a feedback only group, however, so a direct comparison to Porter et al.'s (2000) study is not possible.

Other studies, although not dismissing the use of feedback through negative experimental results, have found that its influence seems to be no more beneficial than when training is used alone. Fiedler and Walka (1993) found that although their training and training plus feedback groups performed above chance in judging credibility, they did not differ from each other. This may have been due to the experiment's design. Feedback was given on the first 16 trials, whereas performance measures were taken from the following 24 trials where

feedback was not being used (Fiedler & Walka, 1993). Therefore, perhaps feedback needs to be continuously supplied throughout the judgment process in order to provide positive results.

Zuckerman et al. (1984) found even less encouraging results than Fiedler and Walka (1993). Participants received outcome feedback on four of their credibility ratings. This did not result in an improvement in accuracy when compared to the control group (Zuckerman et al., 1984), indicating that the use of feedback was not providing participants with the ability to better detect deception. Although unsuccessful in their examination of feedback as a helpful aid in the detection process, reasons for this can be found.

The failure of studies such as Zuckerman et al.'s (1984) to find an improvement with the use of feedback has led some researchers to suggest why this may be happening. Zuckerman et al. (1984) explain that only giving their participants four trials in the feedback condition meant that perhaps too little of this post message feedback was provided in order to see a positive effect. DePaulo and Pfeifer (1986), and Fiedler and Walka (1993), have suggested ways in which to use feedback so that it can help increase accuracy levels in the detection of deception. DePaulo and Pfeifer (1986) state that by providing feedback consistently over many trials, and in an unambiguous manner, participants may find that feedback helps increase their accuracy rates. Fiedler and Walka (1993) add that because feedback reveals little of how one can accurately judge an account, it needs to be matched with an appropriate deception cue system to enable its full potential to be realised. If people are using incorrect cues to deception then feedback will merely be informing them that these cues are incorrect, rather than notifying them of the appropriate cues to be used.

The Development of Verbal Cues in Detecting Deceit

Those studies already discussed have mainly focused on the use of nonverbal cues in detecting deceit. One of the areas where expansion in the research is becoming more evident is in the examination of the content of an account, which Sporer (1997) suggests has, until recently, not been studied thoroughly. The origin of content analysis in this domain can be dated back to Germany during the 1950s with the work of Undeutsch (Landry & Brigham, 1992; Parker & Brown, 2000). The 'Undeutsch hypothesis' suggests that those statements arising from memory of an actual experience have a different content and quality to those statements that are based on fantasy and are, therefore, invented (Craig et al., 1999; Parker & Brown, 2000; Steller, 1989; Undeutsch, 1989).

This view of statement analysis created by Undeutsch was elaborated on by other researchers at the time, such as Arntzen and Szewczyk in Germany, and Trankell in Sweden (Steller, 1989; Undeutsch, 1989; Vrij, 2000). Steller and Koehnken (1989) later integrated the work of all these authors to produce Statement Validity Assessment (SVA). The development of SVA saw a shift away from the former focus on the character of the witness, towards a focus on the characteristics of the actual statement itself (Undeutsch, 1989).

This comprehensive assessment technique was originally used as a way of gathering information and analysing the validity of child witness testimony (Craig, Scheibe, Raskin, Kircher, & Dodd, 1999; Landry & Brigham, 1992; Parker & Brown, 2000). The SVA gave users a guide to the probability that an individual's account was based on fact (i.e., self-experienced) or fiction (i.e., invented) (Parker & Brown, 2000). Using SVA involves the overall assessment of a statement, with the help of all available information, including the

results from a technique called Criteria-Based Content Analysis (CBCA) (Steller, 1989; Vrij, 2000).

Criteria-Based Content Analysis and Reality Monitoring

Use of the CBCA technique has increased over the last 15 years as researchers examine the verbal differences between those who are telling the truth and those who are lying (Vrij et al., 2001a). Research suggests that only a person who actually experiences a situation is likely to incorporate certain elements into their recount of an event (Landry & Brigham, 1992; Vrij, Edward, & Bull, 2001b). The basic assumption of the CBCA criteria is that the elements that are present in self-experienced accounts can be isolated and, hence, used to identify true accounts (Steller & Koehnken, 1989).

Originally used solely for assessing cases of child sexual abuse (Steller, 1989), more recently researchers have suggested that its use for assessing testimonies of adults discussing issues other than sexual abuse is also vital (Sporer, 1997; Steller, 1989; Steller & Koehnken, 1989; Vrij et al., 2001b). Steller and Koehnken (1989) stressed that within a forensic setting the CBCA's usefulness is restricted to accounts of a reasonable length, in order for full use of the criteria to be effective. Therefore, often the simple denial of a suspect's involvement in an event is insufficient for content analysis. This is not to say that this detection device can only be used within forensic settings, however. Instead, it suggests that this tool can be expanded into any area involving the judging of a statement's credibility, as long as the statement is of a length for the CBCA to prove useful. This expansion in the use of CBCA has been shown in numerous studies (Landry & Brigham, 1992; Sporer, 1997; Steller, 1989; Vrij et al., 2001b).

CBCA allows those in a forensic setting to take a statement (either in oral or written form) and systematically analyse its characteristics with the use of a set of defined criteria (Landry & Brigham, 1992; Steller, 1989). The 19 criteria that make up the CBCA are organised into five main categories, which are shown in Table 1 (Steller, 1989; Steller & Koehnken, 1989).

The degree to which each of these criteria is evident within the statement is an indication of the account's truth status. A strong presence of the criteria indicates a self-experienced statement. The ratings of each criterion are combined to give a total credibility score. A set cut-off point, however, has not been developed for establishing an account as either self-experienced or invented, as the final evaluation of credibility is a qualitative one (Steller, 1989). In addition, no guidelines have been set as to how many of these criteria must be present in order to consider an account to be true or false (Steller, 1989).

While the presence of CBCA criteria improves the chances of the statement being self-experienced, absence of these criteria does not necessarily imply that the statement is invented (Steller, 1989; Vrij, 2000). It is important for researchers to be aware that the CBCA cannot act as a "lie detector", as it is not able to detect symptoms of lying (Vrij, 2000). Therefore, an account cannot definitely be classified as invented when analysed by the CBCA scales. More correctly, a low rating on the CBCA scales implies that an account is invented, but this does not mean that this is definitely the case. If the communicator is very young, has poor verbal abilities, or is too distressed to give a detailed account, this may result in low ratings on the CBCA, despite being a true statement (Vrij, 2000). Regardless of this caveat, researchers continue to use this tool, perhaps due to the specific, well-defined criteria that enables easier analysis of a statement.

Table 1. Table of Criteria Included in Criteria-Based Content Analysis

GENERAL CHARACTERISTICS

1. Logical Structure
2. Unstructured Production
3. Quantity of Details

SPECIFIC CONTENTS

4. Contextual Embedding
5. Descriptions of Interactions
6. Reproduction of Conversation
7. Unexpected Complications During the Incident

PECULIARITIES OF THE CONTENT

8. Unusual Details
9. Superfluous Details
10. Accurately Reported Details Misunderstood
11. Related External Associations
12. Accounts of Subjective Mental State
13. Attribution of Perpetrators Mental State

MOTIVATION-RELATED CONTENTS

14. Spontaneous Corrections
15. Admitting Lack of Memory
16. Raising Doubts about One's Own Testimony
17. Self-Deprecation
18. Pardoning the Perpetrator

OFFENCE-SPECIFIC ELEMENTS

19. Details Characteristic of the Offence
-

This statement analysis includes examination of the communicator's original recount of an event, along with any answers to questions that may have followed. The first category of criteria relates to the complete account, which requires the examination of the statement as a whole (Steller, 1989). A Logical Structure is present when all the details in an account can be fitted together describing the same course of events (Steller, 1989). Self-experienced accounts are more likely to contain a logical structure (Steller, 1989). The Unstructured Production criterion suggests that an invented account is more likely to be presented in a very structured, chronological manner, while self-experienced accounts are seen to be more

disorganized and unstructured (Steller, 1989; Steller & Koehnken, 1989). The third criterion, Quantity of Details, states that self-experienced accounts will contain more details about location, people, objects and actions (Steller, 1989; Steller & Koehnken, 1989).

Within the second category, Contextual Embedding refers to whether the account teller includes external facts that anchor the event in time and space, such as the description of habits, relationships between people, or everyday occurrences (Steller, 1989; Steller & Koehnken, 1989). The Descriptions of Interactions criterion suggests that a self-experienced account will produce a chain of actions and reactions between the account teller and a second person (Steller & Koehnken, 1989). The Inclusion of Reproduction of a Conversation and Unexpected Complications During the Incident are both also suggestive of a self-experienced account (Steller, 1989; Steller & Koehnken, 1989).

Accounts with Unusual Details have been related to self-experienced testimonies (Steller & Koehnken, 1989). Because unusual details do not occur often, they are expected to only appear in an account if they have been experienced (Steller & Koehnken, 1989). This is also true for the Superfluous Details criterion. It is not expected that an inventor of an account would think to add irrelevant details that play no part in verifying their account (Steller & Koehnken, 1989). The Accurately Reported Details Misunderstood criterion is more confined to statements given by a child, for which it was originally created. This criterion looks at whether the account teller is describing details that are beyond their comprehension and, therefore, not understood (Steller & Koehnken, 1989). The Related External Associations criterion refers to the inclusion of other events not directly related to the alleged event, however, an external association is present (Steller & Koehnken, 1989). Criterion 12, Accounts of Subjective Mental State, involves the description of feelings and/or cognitions

that were experienced, by the account teller, during the event, whereas, the Attribution of Perpetrator's Mental State criterion refers to the state of mind and motives given to the perpetrator by the communicator (Steller & Koehnken, 1989).

The fourth category within the CBCA looks at the motivation-related contents of an account. The first criterion within this category identifies whether the account teller makes Spontaneous Corrections during the interview, or if they give newer, clearer recollections of the event in question (Steller & Koehnken, 1989). If this is the case, the account teller is believed to be describing a self-experienced event. This is because those inventing an account are assumed to be wanting to provide a statement that does not put their version of an event into question by changing details or correcting themselves (Steller & Koehnken, 1989). This is also the case with the Admitting Lack of Memory criterion. A liar is expected to want to provide a full account rather than seeming vague and not providing specific details (Steller & Koehnken, 1989). Related to this is the criterion examining whether the account teller Raises Doubts as to their Own Testimony. A liar is not thought to want to raise doubts as to the credibility of their account (Steller & Koehnken, 1989). Finally, Self-Deprecation and Pardoning the Perpetrator are also thought to be statements that would be seen as disadvantageous to the invented account teller (Steller & Koehnken, 1989). It is assumed that a liar would not want to make themselves seem unreliable by pointing out self-incriminating details, or providing excuses on behalf of the accused.

The fifth category includes the criterion that addresses Details Characteristic of the Offence, which tends to be relevant only in specific circumstances such as sexual abuse. This suggests that those involved in an event such as sexual abuse will make statements that contradict commonly held beliefs. For example, "sexual abuse of children is believed to be committed

by strangers who use violence”. Self-experienced accounts tend to “embody special qualities characteristic of the offence” (Steller & Koehnken, 1989, p.230). For example, in cases of incestuous relationships, the abuse will tend to happen over a long period of time, with fairly minor sexual acts occurring in the beginning, followed by an escalation of behaviour and a change in the victim’s attitude towards the offender (Steller, 1989). These factors tend to contradict the general belief held about the events within an incestuous relationship (Steller, 1989).

Each criterion is rated on a four-point scale, ranging from “0” (absent), through to “3” (strongly present) (Steller, 1989). By judging the degree to which each of the criteria is present within an account, one can decide whether the statement is likely to be self-experienced or not (Steller & Koehnken, 1989). Interestingly enough, however, without the use of the CBCA criteria, researchers have found that observers are using these cues in the opposite way (Akehurst, Koehnken, Vrij, & Bull, 1996). Participants believed that liars were in fact more likely to give their account in an unstructured way, include more details, make spontaneous corrections, and admit to lack of memory (Akehurst et al., 1996), which may explain why laypersons are so poor at detecting deception without the help of detection tools.

Research using the CBCA criteria has produced supportive results for helping people to successfully use the stated cues to identify true statements. Craig et al. (1999) used police interviews of 48 children involved in sexual abuse claims. Four raters evaluated the transcripts taken from these interviews with the help of the CBCA criteria, resulting in a significant difference being found between the later confirmed and highly doubtful statements (Craig et al., 1999). No control group was used in this study, however, in order to provide a comparison group.

Landry and Brigham's (1992) study produced similar findings. Participants were trained in the use of the CBCA criteria and asked, along with a control group, to judge the credibility of statements made by other adults. The trained group correctly classified 75% of the self-experienced accounts, while the control group were only able to reach a 59% correct classification rate. However, neither group was able to correctly detect the invented accounts, with only a 35% accuracy rate for both groups (Landry & Brigham, 1992).

Further supportive evidence for the use of CBCA criteria was found in Steller, Wellershaus and Wolf's (1988, as cited in Landry & Brigham, 1992) study using 40 accounts produced by young children. Raters either were given a 90-minute tutorial on CBCA, or received no help in making their decisions of credibility. It was found that those given the CBCA tutorial were able to correctly classify 88% of the self-experienced accounts and 71% of the invented accounts, while the control raters achieved a 77% and 53% success rate respectively (Steller et al., 1988, as cited in Landry & Brigham, 1992).

Another supportive study for the CBCA scales was Yuille's (1988b, as cited in Steller, 1989) research, in which elementary school children were asked to tell both self-experienced and invented stories of events that had, or could have, happened. Two raters, using the CBCA criteria, then analysed the transcripts of each of these stories and decided whether they believed the accounts to be true or false. Despite false accounts often containing true elements, the CBCA proved to be very accurate in detecting true accounts (90.9% correct classification rate), and fairly accurate at detecting invented accounts (74.4% correct classification rate) (Yuille, 1988b, as cited in Steller, 1989). It should be noted, however, that there were no control raters included in Yuille's study, therefore, providing no comparison

group. The result of this is that there is no way of knowing whether it was the CBCA criteria that enabled people to correctly classify the accounts, or a separate variable that was not measured.

In comparing studies by Yuille (1988b, as cited in Steller, 1989), Steller et al. (1988, as cited in Landry & Brigham, 1992) and Landry and Brigham (1992), it becomes obvious that self-experienced accounts are more often accurately classified than invented accounts. Using a variety of detection methods, both verbal and nonverbal, researchers have almost consistently found that raters are successfully classifying self-experienced accounts at a higher rate than the deceptive statements (Koehnken, 1987; Levine, Sun Park, & McCornack, 1999). Steller (1989) stresses that the CBCA criteria is a truth verifying tool rather than a lie detection method, which should be kept in mind as this criteria is increasingly used in practical areas. If this method of truth discrimination is promoted as a 'lie detector', there may be a tendency for it to be used as just that, resulting in the misuse of this tool, and perhaps inaccurate classifications of accounts being made.

Despite the recurrent findings of self-experienced accounts being correctly classified more often than invented accounts, one study has been found that produced conflicting results. Parker and Brown's (2000) experiment analysed 43 statements from alleged adult rape interviews. The CBCA criteria were found to enable participants to correctly identify true and false statements, as was determined by additional confirming forensic evidence, guilty pleas, or withdrawal of complaints (Parker & Brown, 2000). However, 87.5% of the self-experienced accounts were correctly judged as being true, while 91.7% of false accounts were accurately judged as being invented (Parker & Brown, 2000).

Although Parker and Brown's (2000) outcome supports the use of CBCA criteria in helping to identify true accounts, what is interesting is that the criteria enabled a higher percentage of the false accounts to be accurately classified than the true accounts. The authors made no attempt to discuss this finding in relation to the original intended use of the CBCA scales, or the conflicting results found in previous successful use of the CBCA method.

What seems to have been overlooked by this large area of research into the CBCA scale's ability to identify self-experienced accounts, is whether the criteria actually targets factors that are present in true accounts and absent in false accounts. Production of the CBCA scales did not involve the analysis of true and false accounts in order to establish how they differed. Instead, experts in the deception detection field selected variables that they believed were vital in the ability to establish an account's veracity, based on the numerous interviews and assessments of credibility they had performed in child sexual abuse cases (Vrij, 2000). The result of this process of forming the CBCA scales is that they were not systematically tested in order to establish whether these criteria were in fact present more often in true accounts than invented accounts. Because of this, studies have produced varying results when investigating the criteria's presence in self-experienced and invented accounts.

Despite the CBCA criteria being developed for child sexual abuse cases, only research by Craig et al. (1999) could be found to investigate the accuracy of the CBCA scales when used with very young children. Other studies have been produced, however, that examine the presence of the CBCA criteria in true and invented statements that do not relate to sexual abuse of children. These studies have analysed the content of true and false statements in order to ascertain how they differed on the CBCA criteria. The number of criteria able to discriminate between true and invented accounts was found to vary greatly. Lamers-

Winkelman (1995, as cited in Vrij, 2000) found only one criteria, Mental State of the Perpetrator, was present more often in self-experienced accounts than invented accounts. Esplin, Boychuk and Raskin (1988, as cited in Vrij, 2000), on the other hand, found that differences between confirmed and doubtful statements emerged for 16 of the 19 criteria. However, other researchers claim that this outcome was due to the doubtful group's cases being made by children who were younger than those in the confirmed group. Younger children may lack the ability to verbalise an event, leading to their statements containing less of the criteria and, therefore, being judged as unconvincing, despite the account possibly being true (Vrij, 2000). Irrespective of these contradictory results, the CBCA scales have continued to be used, despite their ability to identify self-experienced accounts being unverified. This is no doubt due to the successful use of the criteria in both laboratory and real life detection settings.

Much like the CBCA concept, the Reality Monitoring (RM) approach was designed to differentiate between the external and internal generation of memories. The central idea of RM is that memories created from actual experiences are acquired through perceptual information (details of sound, sight, smell, or taste), contextual information (spatial and temporal) and affective information (how one felt during the event) (Johnson & Raye, 1981; Vrij et al., 2001b). Memories created by internal processes, such as one's imagination, are less likely to contain these variables, instead including more subjective idiosyncratic information (Johnson & Raye, 1981; Vrij et al., 2001b).

The scope of RM has been increased over the last 10 years, incorporating it into the deception detection literature. Because of the RM approach's ability to differentiate between perceived and imagined reality, this tool has been used to try and discriminate true and false accounts,

unlike the CBCA scales, which identify solely self-experienced statements (Sporer, 1997). Sporer and Kuepper (1995, as cited in Sporer, 1997), through factor analysis of research by Johnson, Foley, Suengas, and Raye (1988) on RM, developed eight scales to measure RM in the deception context. Using these scales, Sporer and Kuepper (1995, as cited in Sporer, 1997) found that a trained RM rater was able to successfully differentiate between self-experienced and invented accounts based on the RM concepts. The self-experienced accounts were judged as containing more information about time and were rated higher in realism than invented accounts (Sporer & Kuepper, 1995, as cited in Sporer, 1997).

Sporer (1997) also found that with the use of the RM scales, self-experienced and invented transcripts of accounts could successfully be classified on truth status at an above chance level. He went on to suggest that future research should be aimed at incorporating the RM approach with the more forensic content-oriented approach of CBCA in order to develop a tool with improved detection abilities. As these two concepts have common underlying dimensions, such as the measurement of logical consistency and contextual embedding in space and time, integrating them was thought to lead to a new social-cognitive approach of truth detection (Sporer, 1997).

Development of the Aberdeen Report Judgment Scales (ARJS)

The combination of the CBCA criteria with the RM approach is seen to result in both the commonalities, as well as the dimensions represented solely in one or other of the approaches, being collapsed together in order to produce a tool that more effectively helps to identify self-experienced accounts. However, other issues must be taken into account when developing a new method of truth detection.

A variable that must be considered when examining research on appropriate truth detection procedures is the content of the statement being analysed. Zuckerman et al. (1984) asked communicators to describe people that they liked and disliked, therefore, producing accounts where the person described their feelings towards the person. Koehnken (1987) suggested that studies such as this tend not to be practical when applying them to forensic settings. It is seldom the case that a witness or victim must solely report their feelings or emotions rather than factual statements about events (Bull, 1989). Because of this, any new truth detection tool must take into consideration the content of the statement it will be used to assess. A more useful detection method will involve the identification of self-experienced accounts based on factual statements rather than emotions, as this context can be seen as far more beneficial in a forensic situation involving the reporting of an event.

A second consideration for studies in credibility assessment is the motivation of the individual giving the account (account teller). A high motivation level refers to the account teller being highly motivated to lie successfully, for example, a murder suspect. It is very important that one is not found guilty of murder, therefore, they are highly motivated to invent a credible account. In contrast, if there were no negative consequences of a lie being detected, motivation to successfully deceive would be low. Some researchers state that those experiments using high-stake motivational techniques are a far better way of testing a detection approach, as a highly motivated liar will try harder to create a convincing account (DePaulo, Lanier, & Davis, 1983; Vrij & Mann, 2001). Others suggest that using accounts with the low motivation variable could result in smaller differences between self-experienced and invented accounts, making the detection task harder. In turn, this could be more useful if the tools used to detect truthful accounts are found to be able to evaluate the two account types (i.e., self-experienced and invented) differently (Sporer & McCrimmon, 1997). As a

detection tool is developed, the issue of motivation must be considered in order to produce a method that is sensitive to an individual's motivation levels when lying.

With these factors to be considered, on the basis of factor analyses from several previous studies, Sporer (1998) incorporated aspects of the CBCA criteria that did not pertain solely to child sexual abuse statements, with the Reality Monitoring approach. The result was an integrative set of scales, the Aberdeen Report Judgment Scales (ARJS), to be used in the detection of true statements involving factual events. These scales use verbal cues to help identify self-experienced statements, both in written and oral form. The ARJS scales' perceived ability to detect true accounts within a low motivation context, therefore, implying the tool is sensitive to minor differences between true and false statements, enables the use of this method of detection to be used within a wider range of contexts. For this reason, the use of low motivation settings have been suggested when implementing the ARJS scales.

The original 13 scales making up the ARJS were abbreviated, with the abbreviated version being used in a previous study by Sporer and McCrimmon (1997). This version consisted of nine criteria, made up of the following: Logical Structure, quantity and precision of Spatial Details, quantity and precision of Time Details, Sensory Impressions, Emotions and Feelings, description of nonverbal and verbal Interactions, Complications and/or Unusual and/or Superfluous Details, Spontaneous Corrections or Admissions of Memory Failure, and Negative Statements about the Self. As can be seen, these are made up of a combination of both the CBCA and RM approaches.

The ARJS scales were used by Sporer and McCrimmon (1997) to try to identify self-experienced accounts from videotaped statements about either a pleasant or unpleasant dinner

evening. Videotapes were used in order to test the effects of conflicting nonverbal cues on content analysis (Sporer & McCrimmon, 1997). The accounts used were either self-experienced or had been transformed. Transformed accounts refer to those statements where vital aspects of the event are transformed in order to make a pleasant account become unpleasant and vice versa. These accounts are thought to be most comparable to real life lies (Sporer & McCrimmon, 1997). When a person attempts to produce a convincing lie, it is believed that they tend to keep to the truth as much as possible, changing only those aspects of their account that are needed to convince the observer of what they are saying (Sporer & McCrimmon, 1997). Sporer and McCrimmon (1997) found that those who were instructed to use the ARJS criteria ratings to make credibility judgments were able to correctly identify self-experienced accounts (75.8%) significantly more often than the control group (56.7%), who did not use the ARJS scales. This was not found to be the case with transformed accounts.

Despite this encouraging result, of the nine criteria making up the ARJS scales, only ratings for the nonverbal and verbal Interactions criterion produced a significant biserial correlation with the accounts' truth status (Sporer & McCrimmon, 1997). Sporer and McCrimmon (1997) suggested that the lack of significant correlations between the criteria and the accounts was due to the overwhelming amount of information the participants were exposed to through the videotaped accounts. The additional information given in the videotapes (i.e., nonverbal information) was thought to have prevented a person's full cognitive capacity from being focused on the content of the statements (Sporer & McCrimmon, 1997). Because of this outcome, Sporer and McCrimmon (1997) predicted that the relationship between the individual ARJS criteria and the accounts' truth status would be more apparent if the videotapes were replaced by written transcripts.

This suggestion was investigated in an analysis of self-experienced and invented account transcripts of an overnight military exercise, using the original 13 ARJS scales (Sporer, 1998). Sporer (1998) gave no indication as to why there was an inconsistency in the version of ARJS scales used. He found that use of the ARJS scales resulted in participants being significantly better at discriminating self-experienced and invented accounts compared to those participants who did not use the ARJS scales. Whether this difference in outcome was a result of using the non-abbreviated version of the ARJS scales is uncertain, and again, was not discussed by Sporer (1998). In order to make a more direct comparison to the previous study where videotapes were used, research will need to consistently use the abbreviated version of the ARJS scales. Only then can the true effects of nonverbal behaviour in previous ARJS studies be examined.

The Effects of Preparation Time on Detecting Deception

Sporer (1997; 1998) found that the effects of truth detection tools used in his studies were hindered by the presence of preparation time given to the account tellers. In his research using the CBCA criteria, he found that although this method of truth detection could not be used as a basis for identifying self-experienced accounts overall, it was statistically reliable when only a short preparation time was given to communicators.

Similarly, in his 1998 research, Sporer found that the occurrence of participants successfully detecting self-experienced accounts was almost always only when account tellers had been given a short preparation time (i.e., were only given 2-3 minutes). Furthermore, the successful use of his ARJS criteria was compromised by preparation time. Only two of the thirteen criteria used were significantly related to account credibility when communicators

were given the opportunity to prepare their accounts overnight (Sporer, 1998). In comparison, seven of the thirteen criteria produced significant findings in the short preparation condition (Sporer, 1998). No attempt was made by Sporer in either of his studies to examine whether self-experienced accounts did in fact contain more of each of the criteria than invented accounts. If this had been examined then Sporer may have been able to determine whether his results were due to differences in the presence of criteria between true and false accounts, and therefore, the ability of the ARJS criteria to identify true elements of an account.

Other researchers have tried to explain the effect of preparation on deception detection. Vrij et al. (2001a) point out that liars often say things that observers know are implausible, or which sound implausible, because they directly contradict information known to the observer. One of the main reasons for liars giving implausible statements is that they have not been given enough time to prepare themselves or their invented accounts in order to make them sound plausible (Vrij et al., 2001a).

DeTurck and Miller (1990) elaborated on this idea, suggesting that the ability to rehearse an invented message can enhance the liar's chances of successfully deceiving observers. This rehearsal leads to a more relaxed deceiver, in turn producing fewer verbal and nonverbal cues to deception (deTurck & Miller, 1990). Research by Littlepage and Pineault (1982, as cited in Miller & Stiff, 1992) into the effects of preparation found that planned lies were detected less often than spontaneous ones. However, preparation time produced no effect for the self-experienced statements. It was concluded that this was due to planning only being useful for the development of a plausible invented account, as well as giving the deceiver time to mask any nonverbal characteristics of lie telling (Littlepage & Pineault, 1982, as cited in Miller &

Stiff, 1992). This does not necessarily mean that the liar is conscious of the changes they are making to their statement in order to create a more convincing account. Nisbett and Wilson (1977) point out that an individual often is unaware of their cognitive processes, which may include what it is that they must change in order to become more convincing.

The effects of preparation time have even played a part in research that has not intended to measure this variable. Vrij et al. (2001b) found that, in contrast to what was expected, truth tellers spoke slower than liars when giving their statements. Vrij et al. (2001b) suggested that this could have been due to liars being given thinking time before the interviews. Liars would have spent this time formulating a plausible account, therefore, making them well prepared for their interview and as a consequence, not having to slow their speech rate in order to develop the lie as they spoke (Vrij et al., 2001b). Truth tellers, on the other hand, having not been given preparation time, may have been recreating the account while they were talking, explaining why this group spoke at a slower rate than the invented account group.

Very few studies can be found that negate the effects of preparation on statement credibility, however, DePaulo, Lanier and Davis (1983) have done this when examining nonverbal cues to deception detection. In their study, DePaulo et al. (1983) asked communicators to give two accounts of a self-experienced event and two of an invented event. For each of the account types, one statement was prepared and one was not. Results showed that there was no evidence of prepared invented accounts being harder to accurately judge than unprepared invented accounts, with both manipulations producing similar accuracy rates (DePaulo et al., 1983).

Because of this contrast in past research of preparation time, Miller and Stiff (1992) have suggested that future research should continue investigating the effect of this variable on deception detection. It is often the case in forensic settings that an accused may have had time to prepare a credible sounding statement before being questioned. This should be kept in mind when judging the credibility of a suspect's statement. It is because of situations such as these that the effects of preparation need to be thoroughly examined.

The Effects of Valence on Deception Detection

As mentioned earlier, the CBCA scales were originally designed to assess cases of sexual abuse, an event that is associated with extreme negative emotions (Steller, 1989). Because the ARJS scales have been developed using the CBCA criteria, it can be assumed that these criteria will also be useful for identifying self-experienced negative events. This ability is obviously vital for the use of these tools in forensic settings, such as sexual abuse, rape, assault, or other situations involving unpleasant events. However, the use of these detection methods may be beneficial in areas other than those involving a negative valence.

In trying to ascertain whether one's husband/wife has been having an affair, an individual needs to be able to know if their spouse is being honest about where they have been, or who they have been with. This is also the case when trying to evaluate the honesty of a politician. One must assess the statement made by the politician, despite the political promise being a positive one such as lowering crime or reducing tax rates. With the analysis of pleasant statements comes the need to test detection methods already in practise, in order to assess their applicability to a wide range of topics. It is not enough to assume that pleasant accounts can be identified accurately as truthful when using criteria such as the CBCA or ARJS scales, when research has yet to show that this is the case. This research does not appear to have

been carried out using any verbal detection tools, however, nonverbal research has attempted to evaluate the ability of people to lie about an unpleasant event.

Ekman states that by using a tool created by himself and Friesen (1978, as cited in Ekman, Friesen, & O'Sullivan, 1997) called the Facial Action Coding System, involuntary expressions of emotions can be identified despite the liar's efforts to disguise them. Ekman et al. (1997) analysed different types of smiles in an attempt to discriminate between those that were felt and fake, with the belief that those who were lying about their emotions would leak this information through their facial expressions. They found that felt happy expressions occurred more frequently in honest accounts and masking smiles more in invented accounts (Ekman et al., 1997). It should be noted, however, that conditions in this study consisted of either the account teller speaking honestly about pleasant emotions, or the account teller lying about pleasant emotions. Honest accounts of unpleasant emotions and invented accounts of unpleasant emotions were not examined. This could be viewed as a gap in the deception detection literature.

Research later performed by Ekman, O'Sullivan, Friesen & Scherer (1991, as cited in Ekman, 1997) found that the inclusion of voice pitch in an attempt to discriminate true from invented accounts decreased the percentage of classification errors from 16% when using facial expressions only, to 10% when using facial expression with voice pitch. If both facial expressions and voice pitch can be used to determine whether someone is lying about a pleasant or unpleasant event, perhaps the next step is to investigate whether it is easier to detect deceit for pleasant and unpleasant accounts using verbal cues.

Along with the investigation of past nonverbal research into the effects of valence in detecting the veracity of a statement, verbal cues need to be examined for these possible valence effects and a comparison made between the different detection methods. Sporer (1998) examined the effects of valence on judgments of credibility using the ARJS scales. No difference was found between the pleasant and unpleasant accounts, with neither valence leading to a higher accuracy level among self-experienced and transformed accounts. Because this is the only research that can be found which attempts to investigate the effect of this variable in the deception detection context, further examination of this variable is necessary.

Additional Variables in Deception Detection

Along with the investigation of variables such as feedback, preparation time and the valence of an account, additional factors must also be considered when examining the ability of a tool to detect self-experienced statements. For example, as was mentioned earlier, research has shown that people may have a very poor ability to report accurately on their cognitive processes, making them unable to pinpoint what factors have led them to form an impression (Nisbett & Wilson, 1977). Granhag and Stromwall (2000) suggest that this is also the case for judgments made in deception detection circumstances. Their research found that out of the 100 participants given background information that affected their credibility judgments, only one person referred to this information when justifying their judgments (Granhag & Stromwall, 2000). Therefore, if analyses of the cues used by those who have not been trained in a specific detection technique are to be considered, this inability to accurately define one's cognitive processes should be taken into account.

An additional limitation found in people's ability to detect true and/or false accounts is that they tend to hold strong truth biases. With this truth bias affecting credibility ratings, Toris

and DePaulo (1994) suggested that people would be better lie detectors if they were forewarned of the possibility that someone may be trying to deceive them. Their study showed that those who were forewarned became less trusting, therefore, avoiding a truth bias. This ability to decrease the effect of a truth bias should be kept in mind during all investigations of detection methods.

A final variable that must be addressed in the context of truth detection is the possible differences in account length arising from self-experienced and invented accounts. DeTurck and Miller (1985) found that people spent less time talking when they were lying compared to when they were telling the truth. The effects of account length need to be considered by any researcher examining detection methods. If the length of a statement is directly related to its truth status, researchers must be sure that participants are not using this cue in addition to, or instead of, those cues being investigated. Only when the effects of statement length are accounted for can the validity of a detection method be examined.

The Present Research

The present research aims to incorporate a number of issues that have been discussed in this literature review. Investigating the ability to detect true accounts in the first instance seems to be vital, when research such as that mentioned by Akehurst et al. (1996) and DePaulo (1994) suggests that both laypersons and those in forensic fields are using incorrect cues when making their credibility judgments. The main aim of the current research is to investigate whether training participants in the use of verbal cues increases their ability to detect self-experienced accounts. Use of verbal cues will be adopted in this measure of credibility assessment, as opposed to nonverbal cues, due to the considerable amount of literature

confirming the ability of participants trained in the use of verbal cues to accurately detect self-experienced accounts.

Based on research by Sporer and colleagues (1997; 1998), that suggests that the ARJS criteria can be used to identify self-experienced accounts, the present research will employ this method of detection to two different sets of accounts, in two experiments. The current research will investigate the credibility ratings given to these two sets of accounts as a function of account type; self-experienced or transformed (Experiment 1a) and self-experienced or invented (Experiment 1b). Along with this, the present research will investigate whether ratings on the ARJS criteria are related to account credibility ratings. It should be noted, however, that this method of truth detection will be referred to as guidance rather than training. Due to the short time participants are given to familiarise themselves with the criteria, the term training is considered to be misleading, a point which past researchers have failed to consider.

The present study will use accounts taken from Sporer's (1998), and Sporer and McCrimmon's (1997) ARJS studies. By investigating these two very different types of account topics, it is believed that the present study will be able to be generalised to numerous other account topics that may be encountered in future practical use. If the use of this method of truth detection can be generalised to different types of events then this may broaden detection abilities into areas such as the investigation of crimes that are not unpleasant events (e.g., tax evasion or fraud), or the much sought after ability to know if a politician is lying.

Sporer and McCrimmon's (1997) study, examining valence of accounts, implemented the use of videotapes, and found no effects for valence when analysed with the ARJS scales. The

present research, in using transcripts instead of videotapes, will attempt to investigate whether examining the accounts without the interference of nonverbal behaviour will affect the significance of valence in relation to the individual ARJS criteria, as well as the overall effect of the ARJS criteria in helping to detect self-experienced accounts.

The second study completed by Sporer (1998) using the ARJS, replaced investigation of valence with that of preparation. Due to the contradictory evidence produced in past research, along with only a single study measuring the effects of preparation on the ARJS scales, the current research will attempt to replicate Sporer's (1998) study. It is suggested that preparation affects the abilities of those using detection methods to identify true accounts. In addition to this, the present study will investigate the effects of preparation on the relationship between the individual ARJS criteria and the credibility ratings of self-experienced and transformed accounts.

Both of the current study's experiments will investigate the effect of feedback on the detection of true and false accounts, although only Experiment 1b will compare a guidance only group to a guidance plus feedback and control group. Along with this, both experiments in the present research will attempt to use a larger number of accounts with feedback compared to past research. Experiment 1b will provide participants with more accounts than Experiment 1a, in order to see whether the effects of feedback differ between the experiments according to the number of accounts analysed.

The present research will attempt to investigate what cues are being implemented in credibility judgments by those who do not receive guidance. These reasons for credibility judgments will then be compared to those cues described in the ARJS scales. It is thought

that control participants will use differing cues to those included in the ARJS scales, or be unable to provide accurate descriptions of what cues were used. Along with this, account length will be examined, both in terms of whether a difference does exist between the account types (self-experienced and transformed/invented), and if so, whether this may play a part in the rating of credibility.

Evidence indicates that student experimental groups are just as accurate in judging credibility as those working in fields where detecting deception is necessary (DePaulo, 1994). Because of this, the present research will examine the performance of university students. In addition, in order to reduce the possibility of truth bias, participants will be informed of the presence of both self-experienced and transformed/invented accounts within the set of accounts they are judging.

Experiment 1a will consist of accounts describing either a pleasant or an unpleasant evening dinner. Participants will be asked to judge the credibility of these accounts, with the knowledge that some are self-experienced and others have been transformed to make what was originally a pleasant dinner evening an unpleasant evening, and vice versa. Half of the participants will be asked to rate the accounts using the ARJS scales, then base their final credibility ratings on the ARJS ratings. These participants will then be given outcome feedback. It is predicted that those participants given guidance plus feedback, will rate self-experienced accounts as more credible than transformed accounts to a greater extent than control participants. In addition to this, the present research will investigate whether the valence of an account affects the credibility rating given to it.

Experiment 1b compares participants who are given no guidance, guidance, or guidance plus feedback, when rating accounts of overnight military exercises. It is predicted that those receiving guidance and guidance plus feedback will rate self-experienced accounts as more credible than invented accounts to a greater extent than control participants. It is also predicted that those receiving feedback will rate self-experienced accounts higher on credibility than invented accounts in comparison to those not receiving feedback. The present research will also investigate the effects of preparation time on the credibility ratings of accounts. It is predicted that those account tellers who were given longer preparation times (i.e., overnight) for invented accounts would have produced statements that will be rated as more credible than those invented accounts where the account teller was only given 2-3 minutes preparation time.

The overall aim of the current research is to investigate the ability of the ARJS criteria to help participants rate self-experienced accounts as more credible than false accounts, in comparison to those participants not using the criteria. In doing this, the ARJS's ability to be applied to forensic, political and social settings will be assessed. Should this detection method be successful in its ability to identify true elements of an account, its usefulness in each of these areas will be extensive.

CHAPTER 2

Experiment 1a

This experiment investigated the ability of the ARJS criteria to aid in the detection of self-experienced accounts. It was predicted that participants using the ARJS criteria would apply the appropriate verbal cues to enable a higher credibility rating for self-experienced accounts in comparison to transformed accounts of an evening dinner. Based on this reasoning, the following hypothesis was developed.

Hypothesis 1.

Those participants given guidance in the form of the ARJS criteria, along with outcome feedback, will rate self-experienced accounts as more credible than transformed accounts in comparison to those participants who are not given guidance or feedback.

Method

Participants.

A total of 72 University of Canterbury students participated in the experiment. Participants were recruited through an e-mail advertisement, or volunteered from first year Psychology laboratories. Each participant was paid either in the form of a lottery ticket or \$5 in cash.

Experiment 1a consisted of 48 females and 24 males (a ratio of 2:1), with an average age of 23.5 years (SD=9.9 years).

Materials.

Transcripts

Transcripts were taken from a previous experiment by Sporer and McCrimmon (1997). In this experiment, participants were asked to describe either a pleasant or unpleasant evening dinner that they had experienced recently. These events were labeled as self-experienced accounts. All accounts were videotaped. Following this, participants were asked to change as many details as needed about the experienced event in order for the evening dinner to have the opposite valence (pleasant evenings became unpleasant and unpleasant evenings became pleasant). These events were labeled as transformed accounts. These accounts were also videotaped. Therefore, each participant produced two videotaped accounts of an evening dinner, one of which was self-experienced and the other transformed. Participants were all made aware at the time of participation that their accounts would be used in later studies to identify whether others could distinguish the difference between the self-experienced and transformed accounts.

The total of 48 accounts were then transcribed for the present research. As a manipulation check, a 2 (Truth Status: Self-experienced/Transformed) x 2 (Valence: Unpleasant/Pleasant) factorial ANOVA was performed, with the number of words in each account as the dependent variable. This found no significant main or interaction effects. Therefore, accounts did not differ on truth status or valence as a function of word length.

These transcripts were sorted into six sets of eight accounts, making sure that each of the eight accounts within each set were told by different account tellers. Within each set there were two versions of each of the four combinations of accounts (i.e., self-experienced unpleasant, self-experienced pleasant, transformed unpleasant, and transformed pleasant).

The accounts in each set were then placed in a random order. Once this order had been set, half of the accounts for each set were given to participants in a reversed order in an attempt to avoid order effects. Each participant was presented with one set only, as it was believed that one individual completing 48 accounts would be too time consuming resulting in the participant becoming tired, therefore negatively affecting their performance.

Rating Scales

An abbreviated version of the Aberdeen Report Judgment Scales (ARJS; Sporer, 1996) was used to guide participants in their judgments of self-experienced and transformed transcripts. Making up the ARJS were nine criteria shown to be useful in the detection of true verbal accounts (Sporer, 1998; Sporer & McCrimmon, 1997). These criteria were; Logical Structure, quantity and precision of Spatial Details, quantity and precision of Time Details, Sensory Impressions, Emotions and Feelings, nonverbal and verbal Interactions (conversations), Complications and/or Unusual and/or Superfluous Details (referred to as Extra Details), Spontaneous Corrections or Admission of Memory Failure, and Negative Statements about the Self (See Appendix A, page 2-3, for detailed descriptions of each criterion).

Participants were asked to rate each of these nine criteria on a 7-point Likert scale, with '1' indicating a 'very low frequency' of the criterion, and '7' indicating a 'very high frequency'.

Credibility Judgment

Participants were instructed to rate the overall credibility of each transcript. This rating was obtained on a 10-point Likert scale, with '1' indicating a 'freely invented' event, and '10' indicating an 'actual self-experienced' event.

Procedure.

Participants were seated in a research room, with no more than three others simultaneously completing the experiment. In each sitting all participants were in either the control or the guidance plus feedback group. No two participants were ever rating the same set of transcripts in a given experimental setting. The ratio of 2:1 for females to males was kept constant throughout all sets and conditions.

All participants read the instruction sheet, which explained the purpose of the research and gave directions as to how to complete the experiment (see Appendix A and B for each conditions' instruction sheet). Following this, participants were asked to sign a consent form, stating they agreed to participate in the current experiment (Appendix C). Participants in the control group were instructed to read each account, then rate the account on the credibility scale. Following this, control participants were asked to give three reasons for their credibility rating (i.e., why they thought the account was self-experienced or transformed).

Participants in the guidance plus feedback group were informed that in using the ARJS criteria it was thought that one would be better able to identify which accounts were self-experienced. Each criterion was explained, in written form, in detail (see Appendix A). Participants were then told, by the experimenter, that those accounts that possessed a high frequency of each criterion were more likely to be self-experienced. These participants were then invited to read each account and rate each of them using the ARJS criteria on a 7-point rating scale. After completion of these ratings the guidance plus feedback group were instructed, verbally, to rate each account on the credibility rating scale based on their ARJS ratings.

Following the ratings for each account, the participants were asked to open an envelope with the corresponding account number printed on it. Inside each envelope was feedback as to whether the account was self-experienced or transformed (Appendix A, page 4). It was stressed to each participant, by the experimenter, that all ratings must be completed before viewing the feedback form for each account. After receiving the feedback, participants were instructed to place the feedback form back into the envelope and continue on to the next account. When participants had completed their ratings of all the accounts they were thanked for their participation, paid and any questions in relation to the experiment were answered.

Results and Discussion

Criteria Ratings

A mean credibility rating, as well as a mean rating for each criterion, was calculated for each of the forty-eight accounts. Table 2 gives a description of each of the nine ARJS criteria. As a manipulation check, a 2 (Account Type: Self-experienced/ Transformed) x 2 (Valence: Unpleasant/Pleasant) ANOVA was performed for each of the nine ARJS criteria. This was to investigate whether there were any significant differences on any of the individual criteria as a function of account type or valence. Only the guidance plus feedback group participants were included in this analysis as only they rated each account using the criteria. Criteria investigating Spatial Details, Emotions, Interactions, and Negative Statements about Self, produced significant results. See Table 3 for the mean criteria ratings across the four account types.

For the Spatial Details criterion there was a significant main effect for pleasantness ($F(1,44)=4.62, p<.05$). Those accounts that were unpleasant had a significantly higher mean rating on this criterion than those accounts that were pleasant ($M_s=4.07$ vs. 3.60).

Table 2. Descriptions of the Aberdeen Report Judgment Scales Criteria.

Criteria	Description of Criteria
Logical Structure	Description of event is logically structured, with no contradictions.
Spatial Details	Many and precise descriptions of the spatial arrangements of location, people and objects.
Time Details	Many and precise details about time, such as season, date, time of day or duration of event.
Sensory Impressions	Descriptions of sensory impressions about smell, sound, touch, taste, etc.
Emotions	Descriptions of account teller's, or other people's, emotions and feelings.
Interactions	Descriptions of what one person did to another, how they reacted, and/or their conversations.
Complications/Unusual/Superfluous Details	Descriptions of unexpected deviations from normal course of an event, unusual or extraordinary details, or details that one would not normally think of.
Corrections/Admissions of Memory Failure	Account teller spontaneously corrects self or admits that they cannot remember certain details.
Negative Statements about Self	Account teller reports unflattering actions, personal weaknesses, or mistakes made.

For the Emotions criterion there was a significant main effect for pleasantness ($F(1,44)=20.65, p<.01$). Unpleasant accounts were rated more highly for the Emotion criterion than pleasant accounts ($M_s=4.95$ vs. 3.84). Similar results were found for the Interactions criterion, with only a significant main effect for pleasantness found ($F(1,44)=5.49, p<.05$). Unpleasant accounts had a mean rating that was significantly higher than the ratings for pleasant accounts ($M_s=4.03$ vs. 3.35). The criterion focusing on Negative Statements about Self also produced only a significant main effect for pleasantness ($F(1,44)=7.78, p<.01$). Unpleasant accounts were rated more highly for Negative Statements about Self than pleasant accounts ($M_s=2.51$ vs. 1.79). Therefore, unpleasant accounts were judged as having a significantly higher number of spatial details, mentioned emotions more

often, described more interactions, and provided more negative statements about the self than pleasant accounts. There were, however, no significant main effects for account type or any interaction effects.

Table 3. Mean Criteria and Credibility Ratings For Each Account Version.

Criteria	Self-experienced		Transformed	
	Pleasant	Unpleasant	Pleasant	Unpleasant
Logical Structure	4.73	4.98	4.93	4.97
Spatial Details	3.85	4.14	3.34	3.99
Time Details	3.49	3.70	2.99	3.73
Sensory Impressions	3.79	4.04	3.74	3.68
Emotions	3.99	4.92	3.69	4.98
Interactions	3.63	4.29	3.08	3.76
Complications/Unusual/ Superfluous Details	4.05	4.04	3.49	4.31
Corrections/ Memory Failure	3.33	2.73	2.66	3.04
Negative Statements About Self	2.00	2.28	1.58	2.73
Credibility	5.77	6.63	5.95	6.41

This analysis indicated that unpleasant accounts were rated significantly higher on the criteria ratings than pleasant accounts on four of the criteria. Perhaps this was related to the original intended use of the ARJS scale, much like those scales on which it was based. Use of these scales has tended to be designed for situations such as police investigations in which an alleged crime was committed (usually an unpleasant event for those involved).

More importantly, this analysis indicated that no difference could be found, on any of the individual criteria, between self-experienced and transformed accounts. Therefore, expecting

participants to distinguish the difference between the two account types, with the use of the ARJS, was impossible. This analysis, however, did not examine the mean differences across condition, account type, valence and credibility ratings.

Credibility Ratings

In order to examine these differences, a mean credibility rating was created for each participant, for each account version (i.e., self-experienced pleasant, self-experienced unpleasant, transformed pleasant, transformed unpleasant). Using these mean ratings, a 2 (Condition: Control/Guidance plus Feedback) x 2 (Account Type: Self-experienced/Transformed) x 2 (Valence: Pleasant/Unpleasant) ANOVA was performed with repeated measures on the last two factors. See Table 4 for means of the four account types across the two conditions. No significant main effects or interactions were revealed. Credibility ratings did not differ as a function of condition, truth status, or account valence.

Table 4. Mean Credibility Ratings For Each Account Version According to Group.

Group	Self-experienced		Transformed	
	Pleasant	Unpleasant	Pleasant	Unpleasant
Control	6.18	5.75	5.47	6.14
Guidance Plus Feedback	5.92	6.14	6.22	6.18

These findings indicated that there was no difference in credibility ratings between self-experienced and transformed accounts. This, along with the lack of differentiation in ratings of self-experienced and transformed accounts on the ARJS criteria, suggested that the guidance plus feedback group were not able to successfully employ the guidance information

given to them in order to make the credibility judgments. These findings, therefore, did not support Hypothesis 1. Those participants using the ARJS criteria did not rate self-experienced accounts as more credible than transformed accounts in comparison to the control participants. No significant differences were found for valence, despite those effects found on the individual criteria.

Detection Index

In order to investigate whether participants were sensitive to the differences in account type (i.e., self-experienced vs. transformed) a detection index was calculated separately for each participant, for each valence. This index was calculated by subtracting the mean credibility scores for the transformed accounts, for each valence, from the mean credibility scores for the self-experienced accounts for the same valence. Mean scores could range from -9 to +9. The closer the value was to 0, the less differentiation between responses of the participant to the two types of accounts. A high negative score on the detection index indicated that the mean credibility rating for a transformed account was high, while the mean credibility rating for a self-experienced account was low. A high positive score on the detection index indicated the opposite credibility ratings on the transformed and self-experienced accounts.

A 2 (Condition: Control/Guidance plus Feedback) x 2 (Valence: Unpleasant/Pleasant) ANOVA, with repeated measures on the second factor, was then performed. No significant main or interaction effects were found. As all the index values were close to 0, this indicates that individual participants did not differentiate between the self-experienced and transformed accounts, in terms of perceived credibility. Because of the lack of difference between the two account types on the nine criteria, which was already established, this outcome was not surprising.

The Relationship Between Credibility and Criteria Ratings

Correlations were computed between the mean ratings on the individual criteria and credibility ratings across the forty-eight accounts. Only the guidance plus feedback group participants were included in this analysis as only they rated each account using the criteria (see Table 5 for a correlation matrix).

Table 5. Correlations Between ARJS Criteria and Credibility Ratings Across All Accounts. (Highlighted values significant at $p < 0.05$)

Criteria	Correlation
Logical Structure	0.19
Spatial Details	0.55
Time Details	0.48
Sensory Impressions	0.54
Emotions	0.60
Interactions	0.36
Complications/Unusual/ Superfluous Details	0.34
Corrections/ Memory Failure	0.09
Negative Statements About Self	0.35

Correlations showed that those criteria measuring Spatial Details ($r(48) = .547$, $p < .001$), Time Details ($r(48) = .485$, $p < .001$), Sensory Impressions ($r(48) = .536$, $p < .001$), Emotions ($r(48) = .601$, $p < .001$), Interactions ($r(48) = .357$, $p < .01$), Extra Details ($r(48) = .337$, $p < .01$), and Negative Statements about Self ($r(48) = .348$, $p < .01$), all were significantly correlated with credibility ratings. This shows a positive relationship, over all accounts, between the ratings on these criteria and credibility ratings. This means that the more spatial details, time details, sensory impressions, emotions, interactions, extra details, and negative statements about the

self included within an account, the higher the credibility rating the account was given. This result fitted the concept of how the ARJS scale was intended to work, along with the explicit instructions given to participants in the experiment, with participants using their high ratings on the criteria to conclude that the account was more credible.

Control Participants' Reasons for Credibility Ratings

The reasons given by control participants for their credibility ratings were categorized. These categorizations were made by extracting the meanings from each explanation, with similar meanings then grouped and narrowed down into one or two descriptive words. These categories consisted of; details, use of 'um', 'err', 'em', hesitations/pauses, emotions, verbal/nonverbal interactions, boring accounts, account teller laughing, believability, repetition, inconsistencies, knowledge about topic, natural sounding, negative statements about self, account length, memory of event, exaggeration, waffling, clichéd, vague, sounded practiced and logical structure.

The most commonly used reason, given 24.4% of the time, for the control groups' credibility ratings was either the inclusion, or lack of, detail. Often this reason was not elaborated on as to whether the participant meant time details, spatial details, or any other form of detail. A second commonly used reason for one's credibility rating was the general statement of believability, or the account seeming to make sense, making up 21.1% of the reasons given. Again, often this reason was not expanded upon, therefore, making it difficult to interpret this reason more specifically. Other explanations for credibility ratings, however, not as commonly used as the previous reasons, included inconsistencies in the account (8.5%), use of emotions (7.5%), and the use of terms such as "err", "um", and "em" (5.0%).

As can be seen, the most commonly used reason given by control participants also made up two of the ARJS criteria. Other ARJS criteria used by the control group included; Logical Structure (4.7%), Nonverbal and Verbal Interactions (1.0%), Corrections/Memory Failure (2.1%) and Negative Statements about Self (2.0%). If both groups were using similar indicators in deciding the credibility of a statement, this may provide an explanation as to why the two groups did not significantly differ in their credibility ratings of the accounts. The second most common reason given by control participants; believability, was a very broad, unspecific, and even vague rationale for one's rating. If these participants had been asked to be more specific in their explanation then perhaps they would have produced reasons differing from, or more in line with, the ARJS criteria used by the guidance plus feedback group. This explanation, however, is inconsistent with previous research by Nisbett (1977) that suggests that the control group may not have been aware of the cognitive processes being used to decide credibility. These points will be discussed in more detail in a later chapter.

Presence of Criteria in Self-Experienced and Transformed Accounts

An independent coder analysed all accounts, noting the number of times each criterion was present within each account (Logical Structure was coded as present or absent). This coder was blind to the purpose of this procedure, as well as the veracity of each account. In order to investigate whether any differences between self-experienced and transformed accounts existed with regard to the presence of the ARJS criteria, 2 (Account Type: Self-experienced/Transformed) x 2 (Valence: Unpleasant/ Pleasant) factorial ANOVAs were performed separately for each criterion except Logical Structure. For the Logical Structure criterion, no analyses were performed as only two accounts were found to have no logical structure, one of which was self-experienced and one transformed.

Results from the ANOVA^s showed no significant main effects or interaction effects between truth status and valence on any of the criteria except Time Details. For the Time Details criterion a significant interaction effect was found for truth status and valence ($F(1,44)=4.36$, $p<.05$). Planned comparisons found a significant effect of valence for transformed accounts ($F(1,44)=7.06$, $p<.05$). Time details were present more often in transformed unpleasant accounts than transformed pleasant accounts ($M_s=2.75$ vs. 1.25). No significant effect of account type for unpleasant accounts were found. This was also the case for effects of account type for pleasant accounts, and effects of valence for self-experienced accounts.

These results indicate that the self-experienced and transformed accounts did not differ on the criteria in any way. Seeing as these accounts did not differ, the ARJS scales would not have been a useful tool in trying to identify true statements. This would suggest that the ARJS scales could not be used for the purpose they were designed; that is to aid in the detection of true accounts, irrespective of the accounts used. The implications of this finding will be discussed later.

CHAPTER 3

Experiment 1b

This experiment investigated the effect of the ARJS criteria when rating self-experienced and invented accounts. It was believed that in using these criteria, one could better detect true accounts. However, the ability of the ARJS criteria was believed to be affected by original account tellers receiving either long or short preparation times before presenting their account. Long preparation times were thought to result in an increase in credibility ratings. Feedback was also investigated, in order to better examine the beneficial effects of one receiving feedback as they judged the credibility of accounts. With these aims in mind, the following three hypotheses were created.

Hypothesis 1.

Those participants given guidance in the form of the ARJS criteria will rate self-experienced accounts as more credible than invented accounts in comparison to those participants who are given no guidance.

Hypothesis 2.

Those participants receiving feedback will rate self-experienced accounts as more credible than invented accounts in comparison to those participants who do not receive feedback.

Hypothesis 3.

Those invented accounts for which the account teller was given a long preparation time before presenting their account, will be rated as more credible than those invented accounts for which the account teller was given a short preparation time before presenting their account.

Method

Participants.

A total of 90 University of Canterbury students participated in the experiment. Participants were recruited through an e-mail advertisement, or volunteered from first year Psychology laboratories. Each participant was paid either in the form of a lottery ticket or \$5 in cash. Experiment 1b consisted of 60 females and 30 males (a ratio of 2:1), with an average age of 21 years (SD=3.1 years).

Materials

Transcripts

The stimulus materials for Experiment 1b were taken from previous research by Sporer (1998). Participants in the experiment by Sporer were trainees in the Officer Training Corps in Scotland. Half had been on a weekend training activity while the others were yet to experience this activity. Those who had experienced the training exercise were videotaped as they explained the events of the weekend. This was the self-experienced group. Those who had not experienced the training were instructed to lie, talking about the weekend as if they had already completed the training events. These participants were given guidelines as to what to talk about so that they were generally discussing similar events to the self-experienced group. This was to prevent any obvious inconsistencies in knowledge between

the two conditions. This group was also videotaped and was labeled the invented group. Within the self-experienced and invented groups, half of the participants were given 2-3 minutes preparation time before telling their account, while the other half were given instructions the night prior to presenting their accounts. As in Experiment 1a, all participants were informed of how the videotapes would be used in the future.

Following the videotaping of the accounts, each participant's account was transcribed, producing a total of 60 transcripts to be used in the current experiment. As a manipulation check, a 2 (Truth Status: Self-experienced/Invented) x 2 (Preparation Time: Short Preparation/Long Preparation) factorial ANOVA was performed, with number of words in each account as the dependent variable. This found a significant main effect for truth status ($F(1,52)=4.68, p<.05$). There was a significant difference in word length between self-experienced and invented accounts (self-experienced $M=1100.77$, invented $M=863.77$). The effects of this were examined later in the experiment. No other significant main or interaction effects were found.

The transcripts were then sorted into five sets of twelve accounts. In comparison to Experiment 1a, the set size was made larger in order to investigate whether those in the guidance plus feedback group achieved a better result due to being given more practice with the feedback variable. Within each set, there were three versions of each of the four combinations of accounts (i.e. self-experienced long preparation, self-experienced short preparation, invented long preparation, and invented short preparation). The accounts in each set were then placed in a random order. Once this order had been established, half of the accounts for each set were given to participants in a reversed order in an attempt to avoid an order effect. Each participant was presented with one set only, as it was believed that one

individual completing 60 accounts would be too time consuming and result in the participant becoming tired, therefore, negatively affecting their performance.

Rating Scales

The same rating scales that were used in Experiment 1a were also used in Experiment 1b in order to guide participants in their judgments of self-experienced and invented transcripts.

Credibility Judgment

The same credibility judgment scale that was used in Experiment 1a was also used in Experiment 1b.

Post-Experimental Questionnaire

Participants in all conditions were asked to answer an additional question following completion of the experiment (Appendix D). The question asked if the participant had received any training (e.g., military) involving outdoor exercises. This was answered with either a 'yes' or a 'no'.

Procedure

The same procedure as Experiment 1a was used in Experiment 1b, with the exception that as well as the control and guidance plus feedback groups, an additional experimental group was added. The guidance only group received the equivalent instructions to the guidance plus feedback group, however no mention of feedback was given (Appendix E). At completion of the experiment, participants were asked to fill out a post-experiment questionnaire (Appendix D). When participants had completed their ratings of all the accounts along with the post-

experiment questionnaire, they were thanked for their participation, paid and any questions in relation to the experiment were answered.

Results and Discussion

Post-Experimental Questionnaire

A total of twelve participants answered 'yes' to having had past military training (control=1, guidance=5, guidance plus feedback=6). Preliminary analyses, in which these participants were removed, did not change the pattern or significance of the findings and, hence, the following reported results included all participants in each of the conditions.

Criteria Ratings

A mean credibility rating, as well as a mean rating for each of the criteria was calculated for each of the sixty accounts. As a manipulation check, a 2 (Condition: Guidance/Guidance plus Feedback) x 2 (Account Type: Self-experienced/Invented) x 2 (Preparation Time: Short Preparation/Long Preparation) ANOVA was performed individually for all of the nine ARJS criteria. This was to investigate whether there were significant differences for any of the individual criteria as a function of account type or preparation time. Only those participants in the guidance, and guidance plus feedback groups (n=60) were included in this analysis as only they rated the accounts using the ARJS criteria. See Table 6 for mean criteria ratings across the four account types. An ANOVA for the Corrections/Memory Failure criterion produced a significant main effect for preparation ($F(1,56)=4.69, p<.05$). Accounts that contained the long preparation time variable had a significantly higher criteria rating than those accounts with short preparation times ($M_s=3.41$ vs. 2.85). This effect was qualified, however, by a significant interaction effect ($F(1,56)=4.03, p<.05$).

Table 6. Mean Criteria and Credibility Ratings For Each Account Version.

Criteria	Self-experienced		Invented	
	Short Preparation	Long Preparation	Short Preparation	Long Preparation
Logical Structure	4.86	4.72	5.18	4.83
Spatial Details	4.26	4.17	4.37	4.24
Time Details	3.97	3.77	4.25	3.71
Sensory Impressions	3.79	4.09	4.09	3.75
Emotions	3.62	3.96	3.85	3.57
Interactions	3.21	3.49	3.41	3.23
Complications/Unusual/ Superfluous Details	3.72	4.29	3.74	3.78
Corrections/ Memory Failure	3.09	3.13	2.61	3.69
Negative Statements About Self	2.53	2.76	2.58	2.74
Credibility	5.85	5.69	6.35	6.02

Planned comparisons were performed to examine this interaction. These revealed a significant effect of preparation time for invented accounts ($F(1,56)=8.73$, $p<.01$), as can be seen in Figure 1. Participants' ratings on this criterion were higher for invented long preparation accounts than invented short preparation accounts ($M_s=3.69$ vs. 2.61). This indicates that the invented long preparation accounts were thought to contain significantly more corrections/admissions of memory failure than invented short preparation accounts. No significant effect of preparation time was found for the self-experienced accounts. Similarly, no significant effect was found for account type for either the short or long preparation times. No other significant main or interaction effects were found for any of the remaining criteria.

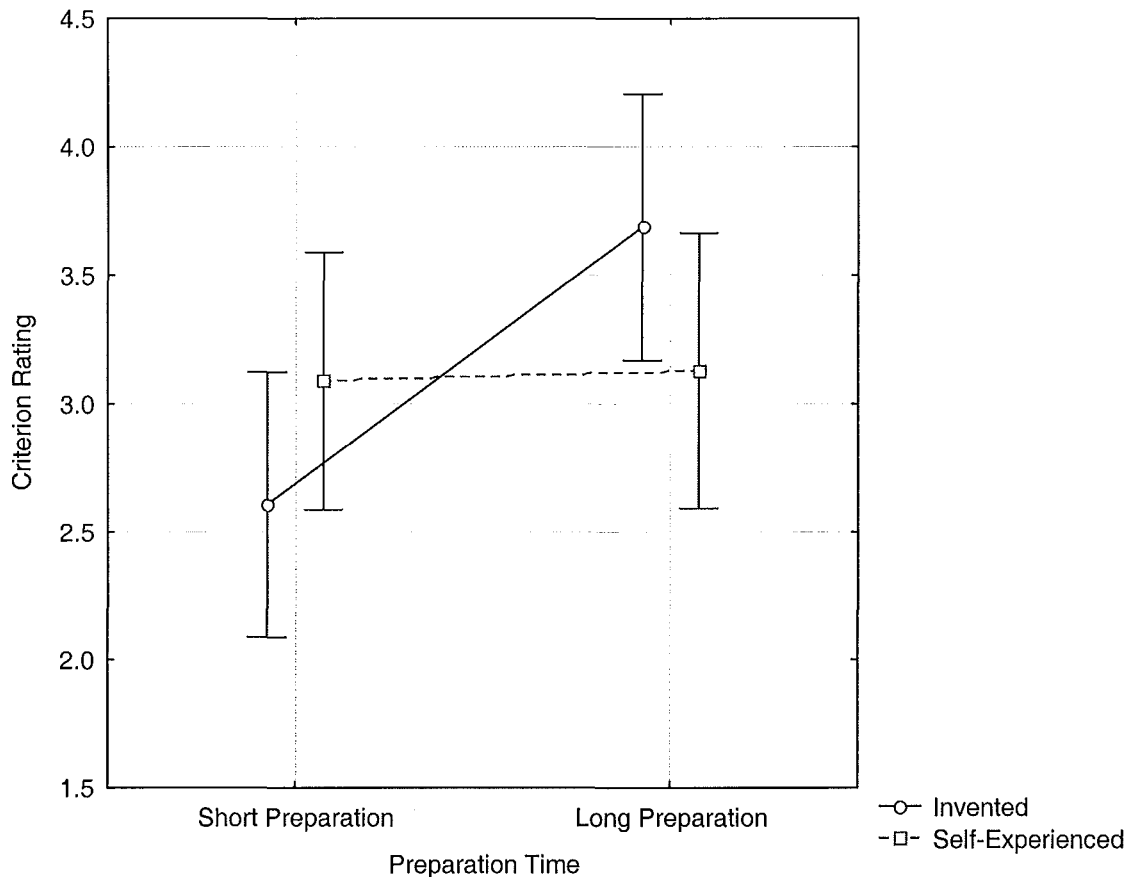


Figure 1. Mean Credibility Ratings on the Corrections/Memory Failure Criterion for Self-Experienced Short Preparation, Self-Experienced Long Preparation, Invented Short Preparation and Invented Long Preparation Accounts.

The significant result found in the planned comparison provided an interesting insight into how the account tellers may have been influenced by preparation time. For a self-experienced account, preparation time may not have had an effect on the account told, hence, this would have explained why there were no significant findings in the self-experienced long preparation versus self-experienced short preparation planned comparison. However, for the invented accounts, having more preparation time seems to have given account tellers the chance to come up with an account that included more corrections or admissions of memory failure. Why this may have been the case, and therefore, why account tellers may not have used the remaining criteria in similar ways, will be discussed in the next chapter.

This analysis indicated that no difference could be found, on any of the individual criteria, between self-experienced and invented accounts or between the control, guidance, and guidance plus feedback groups. As the only significant finding was related to preparation time, this indicated that perhaps the criteria were unable to distinguish a difference between self-experienced and invented accounts. This analysis, however, did not examine the relationship of condition, account type, and valence, with credibility ratings.

Credibility Ratings

A mean credibility rating was created for each participant, for each account version (i.e., self-experienced short preparation, self-experienced long preparation, invented short preparation, invented long preparation). Using these mean ratings, a 3 (Condition: Control/Guidance/Guidance plus Feedback) x 2 (Account Type: Self-experienced/Invented) x 2 (Preparation Time: Short Preparation/ Long Preparation) ANOVA was performed with repeated measures on the last two factors. See Table 7 for means of the four account types across the three conditions.

Table 7. Mean Credibility Ratings For Each Account Version According to Group.

Group	Self-experienced		Invented	
	Short Preparation	Long Preparation	Short Preparation	Long Preparation
Control	5.85	6.04	5.06	6.11
Guidance	6.41	6.40	5.05	6.09
Guidance Plus Feedback	6.65	6.17	5.21	5.86

A significant main effect was found for account type ($F(1,87)=14.37, p<.001$). Self-experienced accounts were rated more highly on the credibility ratings than invented accounts ($M_s=6.26$ vs. 5.56). A significant main effect was produced for preparation time ($F(1,87)=7.84, p<.01$). Those accounts where there was a long preparation time were rated as more credible than those accounts with a short preparation time ($M_s=6.11$ vs. 5.71). These main effects were qualified, however, by a significant interaction effect between account type and preparation time ($F(1,87)=7.04, p<.01$), as can be seen in Figure 2. Planned comparisons were performed to examine this interaction. These revealed a significant effect of account type for short preparation accounts ($F(1,87)=19.02, p<.01$). Self-experienced short preparation accounts had higher mean credibility ratings than invented short preparation accounts ($M_s=6.31$ vs. 5.11).

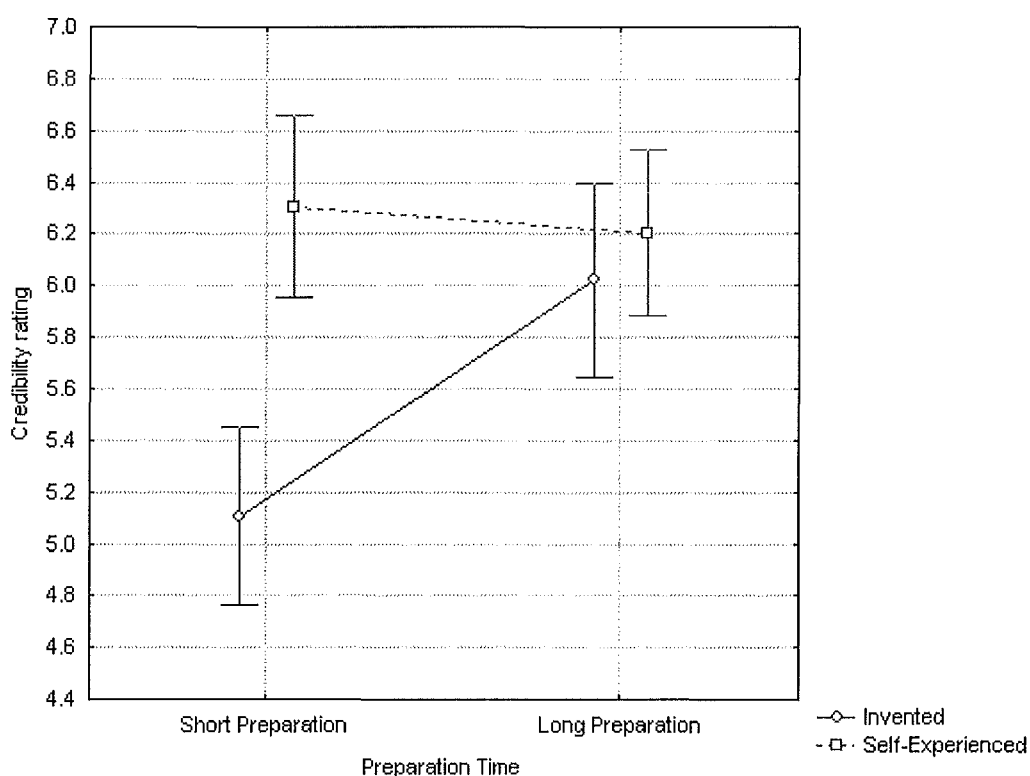


Figure 2. Mean Credibility Ratings for Self-Experienced Short Preparation, Self-Experienced Long Preparation, Invented Short Preparation and Invented Long Preparation Accounts.

A significant effect of preparation time for invented accounts was also found ($F(1,87)=18.89$, $p<.01$). Invented long preparation accounts had higher mean credibility ratings than invented short preparation accounts ($M_s=6.02$ vs. 5.11). No effect of preparation time was found for the self-experienced accounts, along with no effect of account type for long preparation times. There were no other significant main effects.

These findings suggest that despite the criteria being unable to discriminate between self-experienced and invented accounts, participants were still able to judge the true accounts as more credible than the invented accounts, however, this was dependent on the account tellers not having been given a long preparation time. Participants may have ignored the criteria ratings, which resulted in the criteria not being significantly associated with truth status, and instead rated the self-experienced accounts as more credible using a different variable. The presence of an alternative variable, including what the variable may be, will be discussed later. Without the ability to know exactly what other variables are being used by participants, what is evident from these analyses is the great effect preparation time has on the ability to rate credibility for the two account types.

Preparation time made a significant difference to participants' credibility ratings. This could be seen as confirming the previous finding that invented long preparation accounts were rated higher than invented short preparation accounts on the Corrections/Memory Failure criterion. Given preparation time, those account tellers who were inventing an account were able to include factors into their statement that resulted in it being rated as more credible.

The lack of significant findings for the self-experienced long preparation versus self-experienced short preparation accounts indicates that preparation time greatly affects

credibility ratings of invented accounts only. As the self-experienced account tellers were not believed to have used their preparation time in order to develop a more credible sounding account, the credibility ratings for these two account types would not have differed, therefore, suggesting that preparation time was an essential ingredient in making an invented account sound credible.

It should be noted that Hypotheses 1 and 2 were not supported by the results from this analysis. The groups receiving guidance did not rate self-experienced accounts as more credible than invented accounts in comparison to the control group. Along with this, the guidance plus feedback group did not rate self-experienced accounts as more credible than invented accounts in comparison to both the control and guidance only groups. These results will also be discussed later.

In order to examine whether account length was related to credibility ratings, a correlation was computed between these two factors. A significant positive relationship was found ($r(60)=.39, p<.01$). This indicates that those accounts with a longer length were rated as more credible, while the shorter account length was related to lower credibility ratings. This finding is in line with previous research on account length (deTurck & Miller, 1985) and will be discussed later in relation to the detection abilities of the current experiment's participants.

Detection Index

As in Study 1a, a detection index was calculated, for each participant, for each preparation time, by subtracting the mean credibility scores for the invented accounts from those for the self-experienced accounts. A 3 (Condition: Control/Guidance/Guidance plus Feedback) x 2 (Preparation Time: Short Preparation/Long Preparation) ANOVA, with repeated measures on

the second factor, was performed. A main effect for preparation was found ($F(1,87)=7.04$, $p<.01$). The mean scores on the detection index were higher for the short preparation accounts than the long preparation accounts ($M_s=1.20$ vs. 0.18). This shows a greater degree of differentiation between the invented and self-experienced accounts for the short preparation accounts than for the long preparation accounts. No other main or interaction effects were found.

The significant result found in this analysis acted as reconfirmation of the significant difference found previously between self-experienced short preparation and invented short preparation accounts. If the self-experienced accounts were rated as higher than the invented accounts then this would have resulted in a higher detection index score. Therefore, Hypothesis 3 was supported, with those invented transcripts in the long preparation condition rated as being more credible than those invented transcripts in the short preparation condition. Implications of this finding will be discussed in the following chapter.

The Relationship Between Credibility and Criteria Ratings

Correlations were computed between the mean ratings for the individual criteria and credibility ratings across the sixty accounts (See Table 8 for correlations). Only the guidance and guidance plus feedback group participants were included in this analysis as only they rated each account on the criteria. Correlations showed that all criteria were significantly correlated with credibility ratings. Correlations were as follows: Logical Structure ($r(60)=.603$, $p<.001$), Spatial Details ($r(60)=.710$, $p<.001$), Time Details ($r(60)=.716$, $p<.001$), Sensory Impressions ($r(60)=.705$, $p<.001$), Emotions ($r(60)=.624$, $p<.001$), Interactions ($r(60)=.725$, $p<.001$), Extra Detail ($r(60)=.607$, $p<.001$), Corrections/Memory Failure ($r(60)=.289$, $p<.05$), and Negative Statements about Self ($r(60)=.311$, $p<.05$).

Table 8. Correlations Between ARJS Criteria and Credibility Ratings Across All Accounts. (Highlighted values significant at $p < 0.05$)

Criteria	Correlation
Logical Structure	0.60
Spatial Details	0.71
Time Details	0.72
Sensory Impressions	0.71
Emotions	0.62
Interactions	0.73
Complications/Unusual/ Superfluous Details	0.61
Corrections/ Memory Failure	0.29
Negative Statements About Self	0.31

Therefore, this positive relationship indicated that those accounts containing a higher number of each of these criteria were rated as more credible. This showed that participants were using the criteria ratings to make their credibility judgments as they were instructed.

Although participants may be basing their credibility ratings on the criteria ratings, this analysis does not take into consideration the truth status of the accounts. With the criteria ratings alone finding no difference between self-experienced and invented accounts, perhaps participants were using a cue not examined in the ARJS scales, in conjunction with the criteria ratings, in order to make their veracity judgments. This additional cue was able to be used to discriminate between self-experienced and invented accounts, however, it may only have been present when account tellers were given no preparation time. An additional explanation for this outcome was that the control participants were implementing the same

cues making up the ARJS criteria in order to make their judgments, therefore, resulting in both groups rating criteria and credibility to the same extent.

Control Participants' Reasons for Credibility Ratings

The reasons given by control participants for their credibility ratings were categorized. These categorizations were made by extracting the meanings from each explanation, with similar meanings then grouped and narrowed down into one or two descriptive words. Categories included; details, use of 'um', 'err', 'em', knowledge of the topic, logical structure, pauses, emotions, account length, corrections/memory failure, vagueness, extra information, inconsistencies, repetition, account teller laughing, naming person/place, exaggeration, negative statements about self and believability. The most commonly used reason for one's credibility rating was the presence, or lack of, detail, which was used 26.3% of the time. As in Experiment 1a, this reason was seldom elaborated on, therefore, providing no further information as to what sort of detail the participant was referring to. Knowledge of the topic, or lack of, was often used as an explanation for credibility ratings, being given as a reason 12.8% of the time. This may have been due to the nature of the accounts, with its generally uncommon occurrence within participants' past experiences. If the account-teller used terms that were unusual, or unknown to the participant in the current experiment, they may have been more likely to assume that the inclusion of situational knowledge was representative of the accounts truth status.

A third reason often used by control participants for the credibility ratings was an account's inconsistencies (10.6%). Other reasons, although not as commonly used, included the use of pauses (7.7%), logical structure (7.0%), and the account's length (5.6%). The reference to account length that was made by control participants when judging credibility was interesting.

This indicated that participant's were aware of the significant difference in length between the two account types, as was noted earlier. This ability to identify a known factor in the distinction between self-experienced and invented accounts may have contributed to the three groups not significantly differing in their credibility judgments. Despite the control and experimental groups using different criteria to aid their judgments, both groups may have used their own criteria to reach the same conclusion as to the accounts truthfulness.

As can be seen, reasons control participants gave that were included in the ARJS criteria, were the use of details, emotions (3.8%) and extra details (1.1%). The use of details being the most widely used explanation, as in Experiment 1a, could also help to explain why the control group did not significantly differ from the experimental groups in their credibility ratings. However, as mentioned earlier, this contrasts with previous literature on people's knowledge of their cognitive processes. This will be discussed further in the following chapter.

Presence of Criteria in Self-Experienced and Transformed Accounts

An independent coder analysed all the accounts, noting the number of times each criterion was present within each account. This coder was blind to the purpose of this procedure, as well as the veracity of each account. In order to investigate whether any differences between self-experienced and invented accounts existed with regard to the presence of the ARJS criteria, 2 (Account Type: Self-experienced/Invented) x 2 (Preparation Time: Short Preparation/Long Preparation) factorial ANOVAs were performed separately for each criterion except Logical Structure. For the Logical Structure criterion, no analyses were performed as only three accounts were found to have no logical structure, one of which was self-experienced and two invented.

A significant main effect was found for truth on the Time Details criterion ($F(1, 56)=13.25$, $p<.01$). This criterion was present more often in self-experienced accounts than invented accounts ($M_s=8.27$ vs. 4.67). This effect was qualified, however, by a significant interaction effect ($F(1, 56)=7.64$, $p<.01$). Planned comparisons were performed to analyse this interaction. There was a significant effect of account type for short preparation time ($F(1, 56)=20.51$, $p<.01$). Self-experienced short preparation accounts more often included time details than invented short preparation accounts ($M_s=9.13$ vs. 2.80). A significant effect of preparation time was also found for invented accounts ($F(1, 56)=7.13$, $p<.01$). More time details were found in invented long preparation accounts than invented short preparation accounts ($M_s=6.53$ vs. 2.80). There was, however, no effect of preparation for self-experienced accounts, or account type for long preparation accounts. No other main effects were found for the Time Details criterion.

Using the Spatial Details criterion, an ANOVA found a significant interaction effect for truth and preparation ($F(1, 56)=5.43$, $p<.05$). Planned comparisons were performed to examine this interaction. These revealed a significant effect of account type for short preparation accounts ($F(1, 56)=5.95$, $p<.05$). More spatial details were found in self-experienced short preparation accounts than invented short preparation accounts ($M_s=12.60$ vs. 8.80). A significant effect of preparation for self-experienced accounts ($F(1, 56)=5.54$, $p<.05$) was also found. Self-experienced short preparation accounts contained more spatial details than self-experienced long preparation accounts ($M_s=12.60$ vs. 8.93). Effects of account type were not found for long preparation accounts, as well as there being no effect of preparation for invented accounts. No main effects were found for this criterion.

An ANOVA for the Interactions criterion found a significant interaction effect for truth and preparation ($F(1, 56)=4.77, p<.05$). Planned comparisons were performed and found a significant effect of preparation time for invented accounts ($F(1, 56)=4.12, p<.05$). Invented long preparation accounts contained more interactions than invented short preparation accounts ($M_s=2.93$ vs. 1.40). No significant effects of preparation time for self-experienced accounts, account type for short preparation accounts, or account type for long preparation accounts were found. No main effects were found for this criterion.

An ANOVA using the Extra Details criterion produced a main effect for preparation ($F(1, 56)=6.48, p<.05$). Long preparation accounts contained more extra details than short preparation accounts ($M_s=1.17$ vs. 0.60). This effect was qualified, however, by a significant interaction effect ($F(1, 56)=6.48, p<.05$). Planned comparisons were performed, producing a significant effect of preparation time for invented accounts ($F(1, 56)=11.63, p<.01$). Invented long preparation accounts contained more extra details than invented short preparation accounts ($M_s=1.33$ vs. 0.13). A significant effect of account type for short preparation accounts was also found ($F(1, 56)=7.04, p<.05$). More extra details were found in self-experienced short preparation accounts than invented short preparation accounts ($M_s=1.07$ vs. 0.13). Significant effects of preparation time for self-experienced accounts, and of account type for invented accounts were not found. No other main effects were found.

An ANOVA using the Corrections/Memory Failure criterion produced a main effect for truth ($F(1, 56)=6.10, p<.05$). Self-experienced accounts contained more admissions of memory failure than invented accounts ($M_s=1.70$ vs. 0.37). This main effect was qualified by an interaction effect ($F(1, 56)=6.10, p<.05$). Planned comparisons were performed to examine the differences within this interaction. A significant effect of preparation time for self-

experienced accounts was found ($F(1, 56)=8.48, p<.01$). Self-experienced short preparation accounts contained more admissions of memory failure than self-experienced long preparation accounts ($M_s=2.53$ vs. 0.87). A significant effect of account type for short preparation accounts was also found ($F(1, 56)=16.62, p<.01$). Self-experienced short preparation accounts contained more admissions of memory failure than invented short preparation accounts ($M_s=2.53$ vs. 0.20). No significant effect of preparation time for invented accounts, or account type for long preparation time were found. No other main effects were found.

These analyses showed that only two of the criteria (Time Details and Corrections/Admission of Memory Failure) were present more often in self-experienced accounts than invented accounts. However, this difference was only present when the accounts were in the short preparation condition. This indicates that the ARJS scales were not informative enough to discriminate accounts on truth status irrespective of variables such as the amount of preparation time the account teller had before giving their statement. The ARJS scale's inability to be applied to all types of accounts provides large implications in terms of its use. If only accounts including particular factors such as a short preparation time produced differences in the presence of criteria between the true and invented statements, then researchers need to be able to identify what alternative methods would produce a favourable outcome when analysing true and false accounts.

CHAPTER 4

General Discussion

This research found that participants using the Aberdeen Report Judgment Scales did not give higher credibility ratings to self-experienced accounts than transformed/invented accounts. Participants using the ARJS criteria also failed to rate self-experienced accounts higher on the credibility scales than participants who did not use the ARJS criteria. In addition, the use of feedback did not result in higher ratings of credibility for self-experienced accounts than false accounts in comparison to those not receiving feedback. Unpleasant accounts were found to have higher ratings than pleasant accounts on some of the ARJS criteria, however, they were not judged as more credible. Preparation time affected both criteria and credibility ratings for invented accounts, with invented long preparation accounts being rated higher on the ARJS criteria and judged as more credible than invented short preparation accounts. These results will be discussed in relation to the ARJS criteria and their ability to help people detect self-experienced accounts. Further, the effects of preparation, feedback and valence, along with implications for future research into truth detection, will be examined.

Detecting Self-Experienced Accounts Using the ARJS Scales.

Overall, the ability of participants to use the ARJS criteria to rate self-experienced accounts as more credible than false accounts was not found. This is in contrast to findings from previous research into the ARJS (Sporer, 1998; Sporer & McCrimmon, 1997). In order for the concept of the ARJS scales to work, the criteria on which the ARJS scales were based had to have been present more often in self-experienced than invented accounts (Sporer & McCrimmon, 1997). Results showed that those accounts used in Experiment 1a did not differ in terms of

the presence of ARJS criteria as a function of truth status. This explains why participants using the ARJS criteria did not rate self-experienced statements higher on the criteria than invented accounts, and therefore, as more credible than did control participants.

The findings for both experiments in the current research indicated that the use of ARJS scales were insufficient in identifying self-experienced accounts through the presence of the criteria, irrespective of valence or preparation time given to account tellers. This is not because the detection method was unable to find those cues from the ARJS within the accounts. More accurately, those cues included in the ARJS were not ones that were present more often in true accounts and less often in false accounts. Despite identical accounts being used in the previous and present research, participants in previous studies gave true accounts higher ratings on the criteria than invented accounts. There are a number of reasons as to why the results of the present research may have conflicted with previous findings.

In past research where the ARJS scales have been used, analysis of the data focused on participants' accuracy levels of detection. Participants were asked to rate each account on a scale ranging from '1' (transformed/invented) to '10' (self-experienced). Despite this, each individual's rating on the scale was interpreted as a binary decision, with scores of '1' through to '5' interpreted as 'false', while scores of '6' through to '10' were interpreted as 'true'. It is believed that the use of a 10-point scale was an inappropriate measure of a binary decision. Those participants who were unsure as to the veracity of an account were instructed to give a rating of '5' or '6'. Yet, depending on which of these two numbers was chosen, the participant was judged to have made either a true or a false decision. If a dichotomous evaluation of the data had not been made, very little variance between ratings of '5' and '6', in relation to the 10-point scale, would have resulted in a non-significant finding. Therefore,

it is not believed that the variations in ratings given by participants should have been grouped in this way, as it negates the use of a continuous scale and ignores the variance found within both the control and experimental groups.

The present research more appropriately utilised the 10-point credibility scale through the analysis of how high or low an account was rated according to its truth status and in comparison to other experimental groups. The variance in credibility ratings found in the present research, due to scores having a potential range of '1' to '10', made it less likely that a significant difference could be found between true and false accounts. In comparison, previous research had erased any variance contained within the data by turning participants' ratings into a 'true/false' answer, therefore, increasing the chances of a significant rating difference between self-experienced and invented accounts. It is believed that the significant findings found in previous studies of the ARJS criteria were due to this difference in analysis procedures. As a consequence, any variables identified as impacting negatively on the present research are thought to have also affected past studies, despite their conflicting results.

With this difference in analysis, and therefore, outcome of past and present ARJS research addressed, it is possible to focus on explaining the lack of credibility rating difference between the control and experimental groups in the current research. One explanation is that the control group "instinctively" employed the use of cues that were also part of the ARJS criteria. The result of this would have been that all groups were using the same cues and rating accounts similarly on truth status. Examination of the control groups' reasons for veracity judgments in both of the current experiments revealed that cues within the ARJS criteria were identified by this group. Aside from the use of "details" as a cue, however, only a small number of participants identified ARJS criteria as determinants of credibility ratings,

indicating that the majority of the control participants did not knowingly use them. It may be that both the control and experimental participants were using cues found in the ARJS scales and/or additional cues. Irrespective of what cues were used, all groups using the same cues would have resulted in similar credibility ratings.

When examining the cues used by control participants to judge credibility, one must examine the possibility that these participants were unable to articulate exactly what factors were used in making their veracity judgments. If research by Nisbett and Wilson (1977), and by Granhag and Stromwall (2000) is accurate, then control participants could have been incorporating the ARJS cues they identified into their judgments of credibility to either a greater or lesser degree than otherwise found. To the same extent, alternative cues could have been incorporated into credibility judgments without participants being aware of their influence. Further research into what factors shape credibility ratings can begin to explore what cues were actually used by participants, and how these cues affected the credibility ratings of true and false accounts.

The findings from the current study can be used to clarify explanations used in previous research as to why desired results were not obtained. The study by Sporer and McCrimmon (1997) used videotapes of pleasant and unpleasant evening dinners. Only one criterion was found to produce a significant biserial correlation between the mean criterion rating and objective truth status. Sporer and McCrimmon (1997) concluded that interfering nonverbal stimuli in the videotapes had negatively affected participants' ability to accurately rate the criteria. With the use of transcripts for the current study, the suggestion of nonverbal interference can be excluded, as neither the criteria nor credibility ratings differed significantly according to truth status. In other words, rather than the influence of nonverbal

cues resulting in non significant results, the lack of difference between true and false accounts on the criteria were more likely to have produced this finding. The use of videotapes in the previous study may have enabled participants to use nonverbal cues to accurately judge credibility. If this is the case, Sporer and McCrimmon's study provides support for Ekman and O'Sullivan's (1991) findings that nonverbal behavioural cues can be accurately used in the detection of true and false accounts.

Analysis of experimental procedures in previous studies using the ARJS can help explain why differing results were found from the current research. Sporer (1998) used the original thirteen criteria of the ARJS, preventing a direct comparison to Experiment 1b of the current research in which identical accounts were used. It may have been that use of these additional cues within the original version of the ARJS criteria resulted in self-experienced accounts being correctly classified by the experimental group more often than the control participants. Sporer (1998) failed to investigate the presence of the individual criteria in the accounts used, therefore leaving open the suggestion that it was these four criteria removed from the original ARJS that were present more often in true than invented accounts. If this were the case, these would have been the cues used to detect true accounts, explaining why similar results were not found in the present research. As only the cues included in the abbreviated version of the ARJS scales were analysed for their presence in the accounts, the current research was unable to confirm this explanation. Future research implementing the original thirteen ARJS criteria can attempt to investigate this suggestion, with the usefulness of the nine criteria used in the present research being determined. Until then, use of the abbreviated version of the ARJS must be examined.

It has been shown that the true accounts did not contain more of the ARJS criteria than the invented accounts, making detection of true accounts with the use of the criteria impossible. Although the present research has identified this caveat, additional factors included in the current experiments, which affected the abilities of participants to detect true accounts, need to be investigated in an effort to determine whether they played a part in this detection tool not proving to be overly useful.

The ARJS criteria were developed mainly from the Criteria-Based Content Analysis approach. As a way of investigating possible reasons for the ARJS criteria being unable to help participants rate self-experienced accounts as more credible, the CBCA needs to be assessed for its accuracy in detecting true accounts. Research by numerous people has found that the presence of CBCA criteria in true and false accounts varies greatly from study to study (Craig et al., 1999; Esplin et al., 1988, as cited in Vrij, 2000; Lamers-Winkelmann, 1995, as cited in Vrij, 2000). This indicates that the CBCA criteria may not be a reliable instrument for detecting self-experienced accounts. So why has it been continuously used over the years both in research and practical settings? Originally developed by collating ideas and beliefs possessed by expert psychologists in detection (Vrij, 2000), perhaps the supposed accuracy of the CBCA criteria has been assumed by researchers over the years, with supportive research acting as confirmation, while unsupportive research is overlooked, or seen as “the exception to the rule”.

The lack of an experimental development of the CBCA, which would normally systematically measure each of the CBCA criteria, and therefore ARJS criteria, suggests that consideration must be given to the foundations of these methods. Researchers of the CBCA scales have emphasised that it is to be used as a truth detection device only (Steller, 1989; Vrij, 2000).

Therefore, if the criteria are present, the statement is considered more likely to be self-experienced. This form of truth detection, however, can result in the occurrence of a truth bias being overlooked. Participants have a tendency to rate accounts as self-experienced more often (Toris & DePaulo, 1994), resulting in true accounts being accurately classified more frequently than false accounts. As the lower accuracy rates for invented accounts are dismissed due to the device being a truth detector, it is not made obvious that a truth bias exists. The consequence of this is that these truth detection devices are wrongly evaluated as being successful. Although previous research by Toris and DePaulo (1994), along with the present research, attempted to reduce this truth bias by informing the participants that a portion of the accounts were false, perhaps this was not enough to prevent the accounts being rated as self-experienced more often, therefore, producing unwarranted supportive results for scales such as the CBCA and ARJS.

Often interpretations of the ability of scales like the ARJS are exaggerated. This is despite some researchers (Steller, 1989; Vrij, 2000) suggesting that those using truth detection methods should be cautious when dealing with invented statements, instead focusing on judgments of self-experienced statements. Results from studies by Sporer and colleagues (1997; 1998) give the impression that this tool is an accurate method for discriminating true from false accounts. This conclusion is incorrect, as the proposed use of the ARJS criteria merely allows the rater to identify elements thought to be contained within self-experienced accounts. Therefore, no discrimination between self-experienced and invented accounts is taking place.

By detecting self-experienced accounts only, a large gap is left for accounts to be mistakenly judged. False accounts may be given high ratings on the criteria, just as true accounts may be

given low ratings if the detection method is unreliable. Both of these scenarios could result in participants using the criteria ratings to wrongly judge a statement's credibility. This inability of the CBCA and, therefore, the ARJS scales to identify elements present in invented accounts indicates that this tool is not able to classify the truth status of enough accounts to act as a useful detection tool. When discriminating between true and false statements is vital in forensic areas, a detection method that can only detect true accounts is not accurate enough to be considered viable. This is especially the case when detection methods such as these are not guaranteed to identify all true accounts. In order for statements to be accurately identified as self-experienced or invented, researchers need to develop a detection method that is able to recognize elements of both true and false accounts that can then be used to correctly determine a statement's veracity. This may entail the combination of methods that identify elements of either true or false accounts, resulting in a more laborious process of account discrimination, however, a far more accurate detection tool.

Although the Reality Monitoring approach has the ability to discriminate between imagined and experienced memories (Johnson et al., 1988; Johnson & Raye, 1981), little research investigating this approach's capabilities in the deception detection context have been produced. The present research was unable to investigate the ability of the RM scale created by Sporer and Kuepper (1995, as cited in Sporer, 1997) as a detection method, due to its incorporation with the CBCA criteria into the ARJS criteria. Therefore, it is unclear whether the inability of the ARJS criteria to identify true accounts is due to the contribution of the RM or the CBCA approach. Until studies find that the RM scales created by Sporer and Kuepper from the RM approach are a reliable method of detection when used alone, and not adversely affected by variables such as valence or preparation time, the use and possible successes of Reality Monitoring when used within the ARJS should be cautiously interpreted. It should be

highlighted, however, that it may hold the potential to reconcile the issue of truth and lie detection.

The Effects of Preparation Time.

Results from the present research have shown that participants were able to rate self-experienced accounts as more credible than invented accounts, irrespective of whether they were given no guidance, guidance only, or guidance plus feedback. This was only the case, however, when account tellers had not been given a long preparation time before presenting their statements. These findings support those found in previous research investigating the effects of preparation (deTurck & Miller, 1990; Littlepage & Pineault, 1982, as cited in Miller & Stiff, 1992; Sporer, 1997; 1998; Vrij et al., 2001a). The effects of preparation time have significant implications for the detection of true and/or false statements. Whether truth detection tools were being implemented or not, the cues used to make this truth status decision were being affected by preparation time, as was suggested previously by Vrij and colleagues (2001b). This indicates that when given time to prepare their account, liars were in some way changing their statement, resulting in the cues previously used to identify a true account being mistakenly used to judge a false account as self-experienced.

The effect of preparation time, although obviously playing a large part in credibility judgments, has not always been examined thoroughly. Sporer (1998) found only two of the thirteen ARJS criteria produced significant biserial correlations for truth status when participants were given a long preparation time, in comparison to seven of the thirteen criteria producing significant correlations for the short preparation condition. In addition, the rate of correct classification of accounts dropped from 80.6% in the short preparation condition, down to 57.1% in the long preparation condition. The implications of this preparation effect

were not discussed by Sporer (1998) in relation to the reliability of the ARJS scales, however, this issue plays a large part in the potential for this tool's use in practical areas.

A detection tool must have the ability to consistently produce reliable results, despite a change in peripheral factors. The amount of preparation time given to a communicator before presentation of their statement can vary widely in a forensic setting. It is seldom the case that a period of only 2-3 minutes is given to prepare a statement. Therefore, any detection method must be able to account for change in the content of a statement as a function of preparation time. Inflexibility in dealing with this factor can result in a detection tool proving to be unhelpful in the majority of forensic, political and social situations. This is especially the case if, as the amount of preparation time increases (i.e., overnight, one week or one month), so do credibility judgments. Without the ability to account for such a factor, a truth detection method becomes redundant.

An important point that needs to be kept in mind when discussing alternative cues used by participants is that it may not be just control participants using these additional cues.

Although experimental participants were directed to use the ARJS criteria ratings to help make their credibility judgments, it was not stressed that any other cues that the rater thought were important in assessing an account should be ignored. Therefore, cues used by control participants could also have been implemented by experimental participants, resulting in smaller differences between the groups in credibility ratings.

An example of a cue reported to be utilised by control participants was the use of pause fillers, such as "um", "err" and "em". This justification was used solely for accounts being judged as invented, and therefore, were used when making credibility judgments. Although

the presence of this cue was not analysed in true and false accounts, this use of a pause filler is often considered a sign of lying (Koehnken, 1989). Preparation time could have led to participants not using these filler phrases, therefore, resulting in a long preparation time canceling out the ability to use this cue to decipher truth status.

A second alternative variable that may have been used by both control and experimental participants when making their credibility judgments was account length. DeTurck and Miller (1985) found that account length tended to be shorter for invented accounts than true accounts. This variable did not differ according to truth status for Experiment 1a and no significant differences were found for truth status when analyzing credibility judgments. On the other hand, in Experiment 1b there was a significant difference in word length between self-experienced and invented accounts, along with a significant difference in credibility ratings between true and invented statements for short preparation accounts. Therefore, it is possible that participants identified the trend for invented accounts to be of a shorter length and used this as a cue to detecting the truth status of the accounts, hence the self-experienced accounts being rated as more credible when in the short preparation condition.

Those account tellers who provided a self-experienced short preparation account, in comparison to an invented short preparation account, were able to include details that a liar would not have had time to consider when formulating their statement, resulting in a longer account. Those presenting an invented account, yet given a long preparation time, were thought to have used this time to add in information needed for a credible sounding statement, therefore, making the account similar in length to the self-experienced accounts. This explanation, however, does not help account for the lack of significant difference in word

length as a function of preparation time, suggesting that word length alone was not a sufficient enough cue to differentiate true from false accounts.

The identification of word length as a reason for control participants' credibility judgments does not necessarily indicate that these were indeed the only cues used to make veracity judgments. Research by Nisbett and Wilson (1977) and by Granhag and Stromwall (2000) showed that often individuals were unaware of their cognitive processes, making them unable to indicate what cues led them to their veracity judgments. If the control participants were unable to acknowledge what cues they used in making their credibility judgments, alternative variables need to be examined in order to explain the lack of difference in ratings of credibility for true and false accounts between control and experimental groups. If Nisbett and Wilson (1977) and Granhag and Stromwall (2000) were correct, variables such as account length and the use of a pause filler, such as "um", "err" and "em", which have been identified as cues used by the control participants, were not the only factors employed when making veracity judgments.

If participants were using cues other than those stated to judge credibility, the question is what cues were they using? Previous research on the detection of true and false accounts examined possible alternative cues. Kassin and Fong (1999) presented verbal cues such as first-person pronouns, descriptive verbs and unqualified language (e.g., "I would never do anything like that") to their participants. When participants were given a long preparation time to invent an account, it is possible that cues such as these were included, resulting in a higher rating of credibility. The ability of these cues to accurately detect deception is irrelevant, as participants in the present research did not rate true accounts as more credible than false accounts, indicating that the cues used by these participants were ineffective.

The difficulty that participants in both the control and experimental groups had in finding usable cues that enabled them to rate true accounts as more credible, was an interesting discovery. Either participants used the cues contained within the ARJS criteria, or they instinctively identified alternative cues, in an effort to detect true accounts. Despite the broad range of cues being used, what seemed to be the main finding in the current research was that none of these cues enabled participants to identify elements contained solely within true accounts. This indicates that detecting those elements found only in self-experienced statements is a complex task, of which researchers to this day have been unable to unanimously agree upon. In order to train individuals to better identify true accounts, what first must be found are those elements of a true statement that can be accurately perceived. Only then can the true effects of variables such as preparation, valence and feedback, be examined.

The Effects of Feedback on Detection.

The use of ineffective cues in detecting true accounts obscured the benefits of giving participants feedback. Those participants given outcome feedback following each rating of an account did not rate true accounts as more credible than the control or guidance only groups. This does not indicate that feedback was unhelpful, however. First, it must be realised that the feedback groups in both experiments were also groups that used the ARJS criteria when judging the accounts. Because the criteria were found to be unable to help participants detect self-experienced accounts, receiving outcome feedback that was inconsistent with the criteria ratings resulted in the guidance plus feedback participants being given contradictory information.

Guidance plus feedback participants, like the guidance only participants, used the criteria ratings to make their credibility judgments. Yet participants' credibility ratings for self-experienced accounts were no higher than for false accounts. This would have resulted in the outcome feedback not consistently matching criteria ratings. That is, those accounts where the participant gave high criteria ratings were not always found to be self-experienced. This inconsistency in relation to the credibility ratings would have resulted in what seemed like a random truth status outcome. It is because of this that participants receiving feedback would not have been able to use the additional information in order to rate the self-experienced accounts as more credible.

Despite the inability of the present research to investigate the full potential of feedback on the detection of true accounts, other issues can be discussed in relation to this topic. Participants' credibility ratings were assessed, beginning with the first account. However, the first account was rated by the guidance plus feedback group without the benefit of feedback. It was only after rating the initial account, that feedback began and the effects could be observed. The outcome of this was that feedback was affecting only seven of the eight (Experiment 1a) or eleven of the twelve (Experiment 1b) accounts that were rated. As a consequence, the design was not balanced across conditions. This reflects a minor fault in the design of the present research. Omitting the ratings from the first account for each condition would have resolved this problem. In the case of the present research, given the confound between guidance and feedback, further analysis of feedback effects would not have been informative.

This problem of a smaller number of accounts being affected by outcome feedback, in comparison to the number of accounts affected by the criteria in the guidance only group, could also have been resolved by giving participants an opportunity to practice their ratings

using guidance and feedback. This practice time would enable the participant to become accustomed to the use and effects of feedback on credibility ratings. As a consequence, participants may have rated those true accounts assessed at the beginning of each set as more credible due to the effective use of feedback having been established. As deTurck and colleagues (1990) demonstrated, only a limited amount of practice time was needed to produce effective results for their detection training. It is believed that the same positive influence from a small number of trials could be found for feedback.

The use of practice may be necessary for feedback to provide full benefits for those accounts judged during the beginning of the experiment. Zuckerman et al. (1984) suggested that giving participants only four accounts in which to use feedback was not enough to see a positive effect. The initial accounts could have been used by the participants to adapt to, and learn to make full use of, the feedback. Therefore, these accounts would not have received the entire benefit of feedback. The lack of positive effects from using eight or twelve accounts in the present research can be attributed to the ARJS criteria interfering with any positive effects of feedback, as opposed to the use of larger set sizes not allowing participants to adapt to the use of feedback. Therefore, the present research cannot act as confirmation of this suggested revision in experimental procedures. Despite this, it is believed that by combining the comments made by Zuckerman et al. (1984), deTurck et al. (1990) and deTurck and Miller (1990), practice with feedback need only be small, yet the effects robust.

It is thought that a significant difference would be produced between groups if practice with feedback is given to participants before assessing their ability to judge self-experienced accounts as more credible than invented accounts. This effect of practice may have been able to override the contrasting effect of the ARJS criteria on credibility ratings in the present

research, as participants would have had more time and experience with the accounts in order to realise that high ratings on the criteria were not indicative of self-experienced accounts.

Feedback was not found to enhance participants' abilities to detect true accounts in the current research. The majority of past research into the effects of feedback contrast with this finding. It is because of the present study's experimental design, however, that these conflicting results were thought to be produced. Despite this caveat, several factors have been addressed that are believed to be important when assessing the benefits of feedback in future research.

The Effects of Valence on Detection

The effects of feedback in the present research were able to be compared to previous research in the truth detection area. This same comparison cannot be made for the current findings on whether unpleasant accounts were found to be given similar criteria and credibility ratings as pleasant accounts, as no previous research has directly investigated this variable. Despite this, findings from the current research can be discussed in relation to how valence affects detection methods. Unpleasant accounts were found to have higher mean ratings on four of the nine ARJS criteria in comparison to pleasant accounts in Experiment 1a. This may be an effect of the original scales from which the ARJS was created. The CBCA criteria were developed as a means of detecting true accounts of child sexual abuse statements (Steller, 1989). Therefore, this method of detection was designed to pick up those cues within self-experienced accounts specifically for an unpleasant event.

The focus on unpleasant events when developing the CBCA criteria could explain why unpleasant events rated higher on the ARJS scales. Negative events, such as sexual abuse, contain more variables found within the ARJS criteria than pleasant events. Although the

ARJS criteria were unable to detect true accounts, they did result in a differentiation between the two valence types (pleasant and unpleasant). Interestingly, when analysing the presence of criteria in true and false accounts, Time Details was the only criterion present more often in unpleasant than pleasant accounts. This was not, however, one of the criteria rated by participants as being higher in unpleasant accounts. This indicates that for some reason participants were unable to identify the one criterion able to discriminate between pleasant and unpleasant accounts, instead employing the use of criteria found to be unable to make this discrimination. Perhaps participants' own beliefs on appropriate cues were impacting on their ratings, suggesting that an equal weighting was not given to all criteria.

Despite unpleasant accounts producing higher ratings on four of the ARJS criteria, no differences were found for valence on credibility ratings. It may be that high ratings on the four significant criteria were outweighed by ratings on the remaining criteria. This finding again indicates that participants were not giving equal weighting to each criterion when making their credibility judgments.

A further explanation for this inconsistency between criteria and credibility ratings is that although the four significant criteria were rated higher in unpleasant than pleasant accounts when they did occur, they were either not present in all unpleasant accounts or only present in small numbers. As a result, these criteria were either not influential when judging credibility, or were not considered to be as influential as those criteria present in higher numbers in the accounts. By analysing the mean number of times each of the four significant criteria were present in an account, the Emotions, Interactions and Negative Statements about the Self criteria were all found to have appeared infrequently (means ranged from 0.22 to 2.27) in comparison to the remaining criteria (means ranged from 0.88 to 10.15). This indicates that

the above explanation could be used to determine why higher ratings for unpleasant accounts on certain criteria did not translate into higher credibility ratings as a function of valence. In addition, this explanation supports the suggestion that unequal weighting was given to each criterion when making credibility judgments.

From the current research, the effects of valence on detection methods were not able to provide conclusions as to the use of the ARJS criteria in varying contexts. In other words, it is not possible to state whether use of the ARJS criteria is just as effective in detecting true pleasant accounts as it is at detecting true unpleasant accounts. This is due to the ARJS criteria providing participants with inappropriate information on how to judge the credibility of an account. As use of the ARJS criteria resulted in a participants being unable to identify true accounts, the effects of valence could not be examined. As a consequence, the current research attempted to move beyond work by Ekman and colleagues (1997), who examined solely the detection of felt and fake positive emotions, yet still left the need for valence effects to be further investigated.

Examination of a valence effect can have many practical implications. Because of detection methods such as the CBCA and ARJS scales being developed initially with the focus on detecting true unpleasant accounts (Steller, 1989), the ability of these tools to expand into the detection of self-experienced pleasant accounts is still merely suggestive. If detection methods are found to be just as effective for pleasant accounts, use of detection tools can be expanded into numerous areas. These include the assessment of a politician's campaign promises, through to detecting the veracity of a spouse's account of a pleasant evening with friends, when suspecting them of adultery.

Future Research

The accounts used in both the previous and present research consisted of numerous communicators, resulting in participants never rating more than one account told by the same individual. DeTurck and colleagues (1990) found that using different communicators did not result in training being any less effective. Zuckerman and his colleagues (1984), on the other hand, concluded that the effects of training might be limited to judging statements of a communicator that one has already been trained to detect. Future research, using criteria known to be able to differentiate true from false accounts, should attempt to employ accounts told by the same communicator in order to give the rater the opportunity to assess the verbal behaviour of an individual while they are telling the truth. This could then act as a baseline, from which to compare any statements where credibility is being assessed.

Although this repetitive analysis of one communicator appears to reduce the utility of the detection method, this process of assessment is more in line with forensic interviewing procedures (Vrij, 2000). It is often the case that in an attempt to detect when an offender is lying, the suspect's nonverbal and verbal behaviour is compared to previous statements made by the individual that are known to be true (Vrij, 2000). This enables those assessing the account teller's statement veracity to become accustomed to the natural behaviour of the suspect and then use this as a comparison for assessing parts of the statement where credibility is questioned.

Research using the ARJS scales has used accounts that were developed in low motivation contexts. The motivation level of the account tellers may have played a part in the inability of the ARJS criteria to enable participants to identify true accounts. Sporer and McCrimmon (1997) believed that the low motivation context in which the accounts were created meant that

there was less of a difference between the true and false statements. As a result, a sensitive tool would be needed to detect self-experienced accounts. The ARJS criteria were found to be insensitive to the elements contained solely within true accounts. This suggests that this detection tool cannot be used efficiently in a low motivation context. This issue of low motivation reducing the efficacy of the ARJS method was not raised by Sporer and colleagues (1997; 1998), despite the use of high motivation settings being recommended by other researchers (DePaulo, Lanier, & Davis, 1983; Vrij & Mann, 2001). Future research could investigate the ability of alternative detection methods within low motivation settings, along with the effects of high motivation accounts on the ability of the ARJS criteria.

Despite Sporer and McCrimmon (1997) suggesting that use of transformed accounts as opposed to invented accounts was more realistic, this variable may have played a part in the lack of significant results found in Experiment 1a of the present research. Had the ARJS cues been present more often in true than false accounts, Experiment 1a would have resulted in the criteria identifying the self-experienced elements of the transformed account, while ignoring those sections of the account that differentiated it from the true statements. As a result, transformed accounts would have produced no differences as a function of truth status in criteria and credibility ratings, or in the presence of criteria within the accounts. In comparison, the invented accounts in Experiment 1b would not have contained elements of a self-experienced account, resulting in the content of these accounts being different from transformed accounts. Therefore, significant differences would have been identified as a function of truth status. It is because of these findings that truth detection methods may prove to be less helpful in real life settings where communicators are providing transformed rather than invented accounts. Future research into this limitation in the use of ARJS criteria needs

to assess whether the inability to be applied to transformed accounts will result in this detection tool becoming redundant.

The large impact preparation time had on the detection of self-experienced accounts must be examined fully. The fact that a long preparation time resulted in individuals rating self-experienced and false accounts equally on credibility, indicates that people are able to incorporate truthful elements into their account in a successful manner. Future research needs to identify what these elements are, and whether there is any way detection methods can still successfully identify self-experienced accounts despite these truthful elements being included into a false statement.

Within the ARJS criteria, only the Corrections/Memory Failure criterion produced a significant difference in criteria ratings between invented long and invented short preparation accounts. This indicates that when given a longer preparation time, account tellers included more spontaneous corrections or admitted more often to memory failure. This was not found for any of the remaining criteria, suggesting that alternative cues were being included into invented statements in order to make them sound more credible. It is impossible to know from the current research what these alternative cues could have been. Future research into the effect of preparation on statement credibility needs to investigate what elements of an invented account are being changed in order to create a more credible sounding account.

The ability of outcome feedback to help rate self-experienced accounts as more credible was impaired by being used in conjunction with a tool that was unable to detect true accounts through ratings on its criteria. In order to investigate the true effects of feedback, without the risk of the detection method contaminating the outcome, future research needs to have a

group of participants where feedback alone is presented. This would enable researchers to investigate the actual impact of outcome feedback on ratings of credibility, without the risk of the detection method negatively influencing feedback information. Along with this revision in experimental procedures, time series analyses can be used to examine whether the successful employment of feedback increases as the participant continues to rate the accounts. In doing this, the need for practice can be examined, in terms of the amount of accounts to be rated before feedback shows a positive effect.

An important area in the detection literature, that has yet to be fully investigated, is what cues control groups are actually using to make credibility judgments. If participants are unable to consciously access this information (Granhag & Stromwall, 2000), then research needs to examine ways of identifying what factors are influencing credibility judgments. In knowing the actual cues instinctively employed by participants when assessing a statement for credibility, a comparison can be made between control and experimental groups. Researchers can then investigate whether similar or different variables are being considered by the two groups, and whether the cues used are appropriate for differentiating between true and false accounts.

Conclusion

The practical implications of this research for the forensic field of deception detection are important. Ratings of credibility were greatly influenced by preparation time, a factor that is very relevant in scenarios of interviewing victims, witnesses or suspects. Often an account teller has the opportunity to prepare their statement before being assessed on credibility. It is seldom the case that a communicator presents an account of an event as soon as it has occurred.

In the context of sexual abuse, rape, or assault, all those involved are given preparation time before reporting the event in question. Investigations need to be carefully conducted, a process that takes time. Alleged victims may spend time contemplating whether to report a crime, while alleged offenders may use this time to prepare for the possibility of being interviewed by police. In almost all cases within a forensic setting, preparation time plays a role. Whether preparation time is used to develop a more credible account or not, the ability of a detection method is potentially compromised by this factor, resulting in its usefulness being weakened.

As a consequence of the present research, it has been concluded that use of the ARJS criteria for detecting self-experienced accounts should be done so with great caution. The detection method's inability to result in true statements being rated as more credible indicates that the cues identified within the ARJS scales are not accurately focusing on elements contained more often within a self-experienced account. Along with this caveat, the inability of the ARJS criteria to discriminate true from false accounts means that the true purpose for which it would be used in a forensic setting (to detect deception), is overlooked. Therefore, future detection methods would be far more beneficial if used to identify elements of both true and false statements.

It is because of the ARJS criteria's unsatisfactory results that future investigation of this detection method, along with those from which the ARJS was developed, should be questioned. Rather than further examining factors that influence a tool that is unable to discriminate true from false accounts, future research may be better spent developing a new detection method that can reliably identify those elements present within true accounts and

absent within false, or vice versa. The ability of a detection method to accurately discriminate true from false accounts would result in the confidence in credibility assessments increasing. In areas where the repercussions of an inaccurate veracity judgment are large, this confidence in a detection tool is important. In the case of child sexual abuse investigations, incorrectly judging a true account as false, or vice versa, has an enormous impact on both the child and alleged perpetrator. The negative effects on those involved in credibility assessment should be enough to encourage researchers to thoroughly investigate detection tools already in use and not accept those methods found to be inaccurate or unreliable.

When accounts assessed for credibility have the potential to vary as widely as those encountered in forensic, social and political fields, detection methods need to be created that are unaffected by these differences in context and content. The usefulness of detection tools in non-forensic fields is great, despite research having neglected to examine the ability of current detection methods in these alternative areas. If those methods currently being used in deception detection are unable to be expanded into other fields, research needs to attempt to create new detection tools that can. Only then can detection methods be reliably used in a variety of fields.

In summary, although use of the ARJS criteria were not found to enable participants' ratings of self-experienced accounts to be more credible than false accounts in comparison to those not using the detection tool, the effects of important variables such as preparation time were identified. Both in laboratory and practical settings, knowledge of the effects of factors such as preparation are vital to the appropriate, valid assessment of both self-experienced and invented accounts of an event. The present research has paved the way for future investigation into the validity of current detection procedures. Along with this, it has

provided evidence for the effects of variables, such as preparation, that had previously been disputed, as well as the introduction of variables such as valence, that previously had not been examined within the detection context. Any method that is found to be able to successfully incorporate these factors into its detection abilities will act as a vital tool in forensic, political and social settings. For this reason, future detection research needs to concentrate on the development of a reliable, valid method of discriminating the liars from the truth tellers.

References

- Akehurst, L., Koehnken, G., Vrij, A., & Bull, R. (1996). Lay persons' and police officers' beliefs regarding deceptive behaviour. Applied Cognitive Psychology, 10, 461-473.
- Bauchner, J. E., Kaplan, E. P., & Miller, G. R. (1980). Detecting deception: The relationship of available information to judgmental accuracy in initial encounters. Human Communication Research, 6, 251-264.
- Bull, R. (1989). Can training enhance the detection of deception? In J. C. Yuille (Ed.), Credibility Assessment (pp. 83-99). Netherlands: Kluwer Academic Publishers.
- Craig, R. A., Scheibe, R., Raskin, D. C., Kircher, J. C., & Dodd, D. H. (1999). Interviewer questions and content analysis of children's statements of sexual abuse. Applied Developmental Science, 3(2), 77-85.
- DePaulo, B. M. (1994). Spotting lies: Can humans learn to do better? Current Directions in Psychological Science, 3(3), 83-86.
- DePaulo, B. M., & Pfeifer, R. L. (1986). On-the-job experience and skill at detecting deception. Journal of Applied Psychology, 16(3), 249-267.
- DePaulo, B. M., Lanier, K., & Davis, T. (1983). Detecting the deceit of the motivated liar. Journal of Personality and Social Psychology, 45(5), 1096-1103.
- DePaulo, B. M., Lassiter, G. D., & Stone, J. I. (1982). Attentional determinants of success at detecting deception and truth. Personality and Social Psychology Bulletin, 8(2), 273-279.
- deTurck, M. A., & Miller, G. R. (1985). Deception and arousal: Isolating the behavioral correlates of deception. Human Communication Research, 12, 181-201.
- deTurck, M. A., & Miller, G. R. (1990). Training observers to detect deception. Effects of self-monitoring and rehearsal. Human Communication Research, 16(4), 603-620.

deTurck, M. A., Harszlak, J. J., Bodhorn, D. J., & Texter, L. A. (1990). The effect of training social perceivers to detect deception from behavioral cues. Communication Quarterly, 38(2), 189-199.

Druckman, D., Rozelle, R., & Baxter, J. (1982). Nonverbal Communication: Survey, Theory and Research. Beverley Hills: Sage.

Ekman, P. (1997). Smiles when lying (Afterword). In P. Ekman & E. Rosenberg (Ed.), What the Face Reveals. Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS). (pp. 215-216). New York: Oxford University Press.

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

Ekman, P., & O'Sullivan, M. (1991). Who can catch a liar? American Psychologist, 46(9), 913-920.

Ekman, P., Friesen, W. V., & O'Sullivan, M. (1997). Smiles when lying. In P. Ekman & E. Rosenberg (Ed.), 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.

Fiedler, K., & Walka, I. (1993). Training lie detectors to use nonverbal cues instead of global heuristics. Human Communication Research, 20(2), 199-223.

Granhag, P. A., & Stromwall, L. A. (2000). effects of preconceptions on deception detection and new answers to why lie-catchers often fail. Psychology, Crime & Law, 6, 197-218.

Granhag, P. A., & Stromwall, L. A. (2001). Deception detection based on repeated interrogations. Legal and Criminological Psychology, 6, 85-101.

Horvath, F., Jayne, B., & Buckley, J. (1994). Differentiation of truthful and deceptive criminal suspects in behavior analysis interviews. Journal of Forensic Sciences, *39*(3), 793-807.

Johnston, M. K., Foley, M. A., Suengas, A. G., & Raye, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. Journal of Experimental Psychology: General, *117*(4), 371-376.

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

Kassin, S. M., & Fong, C. T. (1999). "I'm innocent!": Effects of training on judgments of truth and deception in the interrogation room. Law and Human Behavior, *23*(5), 499-516.

Koehnken, G. (1987). Training police officers to detect deceptive eyewitness statements: Does it work? Social Behaviour, *2*, 1-17.

Kraut, R. E., & Poe, D. (1980). Behavioral roots of person perception: The deception judgments of customs inspectors and laymen. Journal of Personality and Social Psychology, *39*(5), 784-798.

Landry, K. L., & Brigham, J. C. (1992). The effect of training in Criteria-Based Content Analysis on the ability to detect deception in adults. Law and Human Behavior, *16*(6), 663-676.

Levine, T. R., Sun Park, H., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the "Veracity Effect". Communication Monographs, *66*, 125-144.

Miller, G. R., & Burgoon, J. K. (1982). Factors affecting assessments of witness credibility. In N. L. Kerr & R. M. Bray (Ed.), The Psychology of the Courtroom (pp. 169-194). New York: Academic Press.

Miller, G. R., & Stiff, J. B. (1992). Applied issues in studying deceptive communication. In R. S. Feldman (Ed.), Applications of Nonverbal Behavioral Theories and Research (pp. 217-237). Hillsdale: Lawrence Erlbaum Associates.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. Psychological Review, 84(3), 231-259.

Parker, A. D., & 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.

Porter, S., Woodworth, M., Birt, A. R. (2000). Truth, lies, and videotape: An investigation of the ability of federal parole officers to detect deception. Law and Human Behavior, 24(6), 643-658.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). Sensitivity to nonverbal communication: The PONS test. Baltimore, MD: Johns Hopkins 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. (1998). Detecting deception with the Aberdeen Report Judgment Scales (ARJS): Theoretical development, reliability and validity. Paper presented at the Biennial Meeting of the American Psychology-Law Society, Redondo Beach, California.

Sporer, S. L., McCrimmon, S. (1997). A pleasant or not so pleasant dinner evening? Guiding people to detect what really happened. Paper presented at the Meeting of the Social Psychology Division of the German Psychological Association, Konstanz, Germany.

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

Steller, M., & Koehnken, G. (1989). Criteria-based statement analysis. In D. C. Raskin (Ed.), Psychological Methods in Criminal Investigation and Evidence (pp. 217-245). New York: Springer.

Stiff, J., & Miller, G. (1986). "Come to think of it....": Interrogative probes, deceptive communication, and deception detection. Human Communication Research, *12*, 339-357.

Toris, C., & DePaulo, B. M. (1984). Effects of actual deception and suspiciousness of deception on interpersonal perceptions. Journal of Personality and Social Psychology, *47*, 1063-1073.

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

Vrij, A. (1994). The impact of information and setting on detection of deception by police detectives. Journal of Nonverbal Behavior, *18*(2), 117-136.

Vrij, A. (1998). Nonverbal communication and credibility. In A. Memon, A. Vrij, & R. Bull (Ed.), Psychology and Law. Truthfulness, Accuracy, and Credibility (pp. 32-58). Cambridge: McGraw-Hill.

Vrij, A. (2000). Detecting lies and deceit. The psychology of lying and the implications for professional practice. West Sussex, England: John Wiley and Sons Ltd.

Vrij, A., & Mann, S. (2001). Telling and detecting lies in a high-stake situation: The case of a convicted murderer. Applied Cognitive Psychology, *15*(2), 187-203.

Vrij, A., Edward, K., & Bull, R. (2001a). People's insight into their own behaviour and speech content while lying. British Journal of Psychology, *92*(2), 373-389.

Vrij, A., Edward, K., & Bull, R. (2001b). Stereotypical verbal and nonverbal responses while deceiving others. Personality and Social Psychology Bulletin, *27*(7), 899-909.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology (Vol. 14) (pp. 1-59). San Diego, CA: Academic Press.

Zuckerman, M., & Driver, R. E. (1984). Telling Lies: Verbal and nonverbal correlates of deception. In A. Siegman & S. Feldstein (Eds.), Nonverbal Communication: An Integrated Perspective (pp. 129-148). Hillsdale, NJ: Lawrence Erlbaum.

Zuckerman, M., Koestner, R., & Alton, A. O. (1984). Learning to detect deception. Journal of Personality and Social Psychology, *46*(3), 519-528.

Zuckerman, M., Koestner, R., & Colella, M. J. (1985). Learning to detect deception from three communication channels. Journal of Nonverbal Behavior, *9*(3), 188-194.

Appendix A

For a number of years, psychologists have been trying to study what makes a communicator appear trustworthy, and whether or not we are able to tell if a person is telling the truth or being intentionally deceptive. Thus, for this experiment, we want to find out if people are capable of detecting whether or not a person that was recorded on videotape and later transcribed, is giving an invented or self-experienced account.

In the next several minutes, we would like you to read transcripts of eight different people giving an account of an evening dinner. This person may or may not have actually experienced that event - that is, they will be either telling the truth or being intentionally deceptive.

After reading each account, we would like you to judge the credibility of the account on a 10-point scale according to your personal estimation of whether or not the account was self-experienced or invented. On this scale, a '1' indicates that you are fully convinced that the account has been invented, and a '10' indicates that you are fully convinced that the account has been self-experienced. If you are completely unsure, then take your best guess (i.e., a '5' or a '6'). The scale looks like this:

Freely Invented 1 2 3 4 5 | 6 7 8 9 10 Self-Experienced

CREDIBILITY CRITERIA

Now, it is your task to judge whether or not the person had actually experienced this event or not. To help you in this, we will give you first some guidelines (so-called "credibility criteria") expert psychologists have developed that should be useful for your judgment.

In particular, these experts have proposed that a self-experienced account should differ in quality and/or quantity from an invented account in the following ways. Any of these qualities may also be present in invented accounts to some extent, and they may only be present in self-experienced accounts to a certain degree.

Before you make your final judgment about credibility, you are asked to evaluate each account to the extent you believe these qualities to be present.

In general, a credible account should be judged as more credible if it can be characterized by some of the following qualities:

(1) Logical structure.

- The description of the event is **logically consistent**. It does not contain any obvious contradictions to other elements of the account.
- This does not mean, however, that the account has to be told in a coherent and organized manner. It may well be that it appears chaotic but--when you think about it--reveals an **underlying structure** (the actual sequence of events) that can be discovered or **reconstructed**.

(2) Quantity and precision of spatial details.

- The account contains **many and precise** descriptions of the **spatial arrangements** of the location, setting, objects and people.

(3) **Quantity and precision of time details.**

- The account contains **many and precise** details about **time**.
- Time information can be of the season, the date (year, day) and time of day (hour) as well as the duration (when it started, when it ended).

(4) **Sensory impressions.**

- The account contains **many and precise descriptions** of sensory impressions about **sound, smell, touch, taste etc..**

(5) **Emotions and feelings.**

- The account contains **many and precise** descriptions of the **storyteller's or other people's emotions and feelings** at the time of the event.
- Descriptions can be of the intensity, different types and changes in emotions appropriate to the circumstances.

(6) **Nonverbal and verbal interactions (conversations).**

- The account contains **many and precise** descriptions of **nonverbal interactions**, i.e. observable behaviors of what one person does to another and how this person reacts.
- The account contains **many and precise** descriptions of **conversations** (in direct or indirect speech).

(7) **Complications and/or unusual and/or superfluous details.**

- The account contains descriptions of **complications**, i.e. unexpected deviations from a linear course of events (e.g., a surprising interruption or repeated attempts).
- The account contains descriptions of **unusual or extraordinary** details.
- The account contains descriptions of **superfluous details** of which **one would not have thought of normally**.
- While any of these details may detract from a concise and fluent account they nonetheless fit in within the given context and illustrate knowledge about the specific situation or event. One would not normally have thought of any of these details from a general knowledge of this type of situation.
- Nonetheless, complications, unusual or superfluous details must be reasonably **plausible**. If you try to imagine yourself to have experienced the event described it appears reasonable to you that it may actually have happened the way it did, and that these details are not completely far-fetched or bizarre. Therefore, the description of the event should **not** contain any completely **implausible elements**.

(8) **Spontaneous corrections or admission of memory failure.**

- The storyteller **spontaneously corrects errors** or inaccuracies in his/her report.
- The storyteller **admits a lack of memory** about some details.

(9) **Negative statements about the self.**

- In the account the storyteller makes **negative statements** about him- or herself (**lack of social desirability**); i.e., the storyteller admits unflattering actions due to personal

weaknesses (e.g., bad intentions or maliciousness) or mentions mistakes and errors (e.g. due to incompetence).

To apply this knowledge, please, rate each account on the scales provided, using a 7-point-scale, where 1 indicates a very low frequency, precision or intensity of a certain criterion, and 7 indicates a very high frequency, precision or intensity.

Try to use the whole range of the scale but be aware that some criteria may generally receive higher ratings while others are likely to receive lower ratings.

For example:

Time details: 1 2 3 4 5 6 7 []

On the response sheet itself, you only write your response in the right-hand box!

Finally, you should judge the credibility of each account on a 10-point scale according to your personal estimation of whether you do or do not ultimately believe the account was self-experienced or freely invented.

On this scale, a '1' indicates that you are fully convinced that the account has been invented, and a '10' indicates that you are fully convinced that the account has been self-experienced. If you are completely unsure, then take your best guess (i.e., a '5' or a '6').

The scale looks like this (please, circle one number!):

freely invented 1 2 3 4 5 | 6 7 8 9 10 self-experienced

FEEDBACK

You may then open the attached envelope, which will contain feedback that tells you whether the account you just rated was self-experienced or invented.

After reading the correct answer for each transcript, would you please place the feedback form back into the envelope provided without discussing the information with the experimenter. Following this, you will be presented with the next transcript, where you are asked to repeat the judgment scales.

[Now, read the first account carefully and make your judgment.

Afterwards, you will read the next account, etc. **Do not go back to any previous accounts or change your answers!**

Do not talk to anybody else as their task may differ from yours. Besides, we are interested in your personal judgment. Thank you!]

Appendix B

For a number of years, psychologists have been trying to study what makes a communicator appear trustworthy, and whether or not we are able to tell if a person is telling the truth or being intentionally deceptive. Thus, for this experiment, we want to find out if people are capable of detecting whether or not the account of a person which was recorded on videotape and later transcribed is giving an invented or self-experienced account.

In the next several minutes, we would like you to read transcripts of eight different people giving an account of either a pleasant or an unpleasant dinner evening. This person may or may not have actually experienced that evening - that is, they will be either telling the truth or being intentionally deceptive. After reading each account, we would like you to judge the credibility of each account on a 10-point scale according to your personal estimation of whether you do or do not ultimately believe the account was self-experienced or freely invented.

On this scale, a '1' indicates that you are fully convinced that the account has been invented, and a '10' indicates that you are fully convinced that the account has been self-experienced. If you are completely unsure, then take your best guess (i.e., a '5' or a '6'). The scale looks like this:

freely invented 1 2 3 4 5 | 6 7 8 9 10 self-experienced

We also want to find out what information people use when they judge the credibility of another person's account.

Therefore, briefly note three reasons that have guided you in your decision, each on a separate line. Order the three reasons according to their importance by assigning a number:

1 = most important;

2 = second most important;

3 = third most important.

Before you hand in the answer sheet, make sure that you have written all the account numbers, and your subject number and age on the top line of every answer sheet.

Please, first write down the account number you are judging! Next, give the rating for the account, the 3 reasons and order them.

In my opinion, the account was (please, circle one number!):

Account-No.

freely invented 1 2 3 4 5 | 6 7 8 9 10 self-experienced

Reason:

[Now, read the first account carefully and make your judgment. Afterwards, you will read the next account, etc. **Do not go back to any previous accounts or change your answers!**

Do not talk to anybody else as their task may differ from yours. Besides, we are interested in your personal judgment. Thank you!]

Appendix C

The Use of Verbal Cues to Detect Deception

I agree to participate in the study named above, on the understanding that at any time I wish to withdraw from the study I may, without prejudice, do so. I further understand that if I withdraw, I have the right to have any data collected from me returned.

SIGNATURE _____ DATE _____.

Appendix D

POST EXPERIMENTAL QUESTIONNAIRE

Before we finish, please could you write down whether you have had the following experience:

Have you had any training (e.g., military) involving outdoor exercises?

No Yes

THANK YOU VERY MUCH FOR YOUR PARTICIPATION! I WOULD APPRECIATE IF YOU WOULD NOT TELL OTHER PEOPLE ABOUT DETAILS OF THIS STUDY (EXCEPT, OF COURSE, THAT YOU FOUND IT INTERESTING...)

Appendix E

For a number of years, psychologists have been trying to study what makes a communicator appear trustworthy, and whether or not we are able to tell if a person is telling the truth or being intentionally deceptive. Thus, for this experiment, we want to find out if people are capable of detecting whether or not the account of a person that was recorded on videotape and later transcribed is giving an invented or self-experienced account.

In the next several minutes, we would like you to read transcripts of twelve different people giving an account of an overnight military exercise. This person may or may not have actually experienced that event - that is, they will be either telling the truth or being intentionally deceptive.

After reading each account, we would like you to judge the credibility of the account on a 10-point scale according to your personal estimation of whether or not the account was self-experienced or invented. On this scale, a '1' indicates that you are fully convinced that the account has been invented, and a '10' indicates that you are fully convinced that the account has been self-experienced. If you are completely unsure, then take your best guess (i.e., a '5' or a '6'). The scale looks like this:

freely invented 1 2 3 4 5 | 6 7 8 9 10 self-experienced

CREDIBILITY CRITERIA

Now, it is your task to judge whether or not the person had actually experienced this event or not. To help you in this, we will give you first some guidelines (so-called "credibility criteria") expert psychologists have developed that should be useful for your judgment.

In particular, these experts have proposed that a self-experienced account should differ in quality and/or quantity from an invented account in the following ways. Any of these qualities may also be present in invented accounts to some extent, and they may only be present in self-experienced accounts to a certain degree.

Before you make your final judgment about credibility, you are asked to evaluate each account to the extent you believe these qualities to be present.

In general, a credible account should be judged as more credible if it can be characterized by some of the following qualities:

(1) Logical structure.

- The description of the event is **logically consistent**. It does not contain any obvious contradictions to other elements of the account.
- This does not mean, however, that the account has to be told in a coherent and organized manner. It may well be that it appears chaotic but--when you think about it--reveals an **underlying structure** (the actual sequence of events) that can be discovered or **reconstructed**.

(2) Quantity and precision of spatial details.

- The account contains **many and precise** descriptions of the **spatial arrangements** of the location, setting, objects and people.

(3) **Quantity and precision of time details.**

- The account contains **many and precise** details about **time**.
- Time information can be of the season, the date (year, day) and time of day (hour) as well as the duration (when it started, when it ended).

(4) **Sensory impressions.**

- The account contains **many and precise descriptions** of sensory impressions about **sound, smell, touch, taste etc..**

(5) **Emotions and feelings.**

- The account contains **many and precise** descriptions of the **storyteller's or other people's emotions and feelings** at the time of the event.
- Descriptions can be of the intensity, different types and changes in emotions appropriate to the circumstances.

(6) **Nonverbal and verbal interactions (conversations).**

- The account contains **many and precise** descriptions of **nonverbal interactions**, i.e. observable behaviors of what one person does to another and how this person reacts.
- The account contains **many and precise** descriptions of **conversations** (in direct or indirect speech).

(7) **Complications and/or unusual and/or superfluous details.**

- The account contains descriptions of **complications**, i.e. unexpected deviations from a linear course of events (e.g., a surprising interruption or repeated attempts).
- The account contains descriptions of **unusual or extraordinary** details.
- The account contains descriptions of **superfluous details** of which **one would not have thought of normally**.
- While any of these details may detract from a concise and fluent account they nonetheless fit in within the given context and illustrate knowledge about the specific situation or event. One would not normally have thought of any of these details from a general knowledge of this type of situation.
- Nonetheless, complications, unusual or superfluous details must be reasonably **plausible**. If you try to imagine yourself to have experienced the event described it appears reasonable to you that it may actually have happened the way it did, and that these details are not completely far-fetched or bizarre. Therefore, the description of the event should **not** contain any completely **implausible elements**.

(8) **Spontaneous corrections or admission of memory failure.**

- The storyteller **spontaneously corrects errors** or inaccuracies in his/her report.
- The storyteller **admits a lack of memory** about some details.

(9) **Negative statements about the self.**

- In the account the storyteller makes **negative statements** about him- or herself (**lack of social desirability**); i.e., the storyteller admits unflattering actions due to personal

Now, read the first account carefully and make your judgment.

Afterwards, you will read the next account, etc. **Do not go back to any previous accounts or change your answers!**

Do not talk to anybody else as their task may differ from yours. Besides, we are interested in your personal judgment. Thank you!