DETECTING DECEIT VIA ANALYSIS OF VERBAL AND NONVERBAL BEHAVIOR

Aldert Vrij, Katherine Edward, Kim P. Roberts, and Ray Bull

ABSTRACT: We examined the hypotheses that (1) a systematic analysis of nonverbal behavior could be useful in the detection of deceit and (2) that lie detection would be most accurate if both verbal and nonverbal indicators of deception are taken into account. Seventy-three nursing students participated in a study about "telling lies" and either told the truth or lied about a film they had just seen. The interviews were videotaped and audiotaped, and the nonverbal behavior (NVB) and speech content of the liars and truth tellers were analyzed, the latter with the Criteria-Based Content Analysis technique (CBCA) and the Reality Monitoring technique (RM). Results revealed several nonverbal and verbal indicators of deception. On the basis of nonverbal behavior alone, 78% of the lies and truths could be correctly classified. An even higher percentage could be correctly classified when all three detection techniques (i.e., NVB, CBCA, RM) were taken into account.

KEY WORDS: detecting deceit; nonverbal behavior; Criteria-Based Content Analysis; Reality Monitoring.

There are, in principle, three ways to catch liars: (1) by observing how they behave (the movements they make, whether or not they smile or show gaze aversion, their pitch of voice, their speech rate, whether or not they stutter, and so on), (2) by listening to what they say (analyzing the speech content), and (3) by measuring their physiological responses. In order to measure physiological responses, several polygraph test procedures have been developed such as the Control Question Test (Raskin, 1979, 1982, 1986; Reid, 1947) and the Guilty Knowledge Test (Lykken, 1960, 1998). Deception detection techniques based on what a person says include Con-

Aldert Vrij and Katherine Edward, Psychology Department, University of Portsmouth, United Kingdom; Kim P. Roberts, National Institute of Child Health and Human Development, Bethesda, Maryland; Ray Bull, Psychology Department, University of Portsmouth, United Kingdom.

This study was sponsored by a grant from the Leverhulme Trust given to the first author.

Address correspondence to Aldert Vrij, Psychology Department, University of Portsmouth, King Henry Building, King Henry 1 Street, Portsmouth PO1 2DY, United Kingdom; e-mail: aldert.vrij@port.ac.uk.

tent-Based Criteria Analysis (CBCA) (Köhnken, 1990, 1996; Steller & Köhnken, 1989) and Reality Monitoring (RM) (Porter & Yuille, 1996; Sporer, 1997). David Raskin and Gunter Köhnken—leading experts in lie detection via physiological responses and via what is said, respectively—both believe that detecting deception via nonverbal behavioral cues is a precarious exercise on which people cannot rely (Köhnken, 1997, personal communication; Raskin, 1996, personal communication).

Research, so far seems to support this pessimistic view. When detecting deceit via nonverbal cues, accuracy rates (percentage of correct answers) usually vary between 45 and 60 percent, when a 50% accuracy rate would be obtained by tossing a coin (DePaulo, Stone, & Lassiter, 1985a; Kraut, 1980; Vrij, 2000). If accuracy at detecting lies is computed separately from accuracy at detecting truths, it emerges that people are particularly poor at detecting lies. In a recent review including approximately 40 studies, Vrij (2000) found a 67% accuracy rate for detecting truths and a 44% accuracy rate for detecting lies. (The high accuracy rate for truths and the low accuracy rate for lies is the result of a 'truth-bias: People's tendency to judge other's messages as being truthful [Levine, Park, & McCornack, 1999; Vrij, 2000]). Although the average hit rate for detecting truths (67%) is reasonably high, it says little about lie detection skills. Good lie detection implies high accuracy rates for both detecting truths and detecting lies. Research has shown that it is possible to detect *both* lies and truths above the level of chance (on average around 70% or above) when conducting polygraph tests (Ekman, 1992; Vrij, 2000) or when using CBCA or RM (Vrij, 2000; Vrij & Akehurst, 1998).

Although observers seem to perform relatively poorly in detecting deceit while paying attention to nonverbal cues compared to detecting deceit by analyzing with CBCA or RM or by measuring physiological responses, we are reluctant to draw any firm conclusions on the basis of such comparisons, as those comparisons are unfair and inappropriate. Studies of detecting deceit via examining physiological responses or via CBCA and RM always include well-trained experts as lie detectors (because they are the only ones who know how to conduct such examinations), whereas lay persons (e.g., college students) are often used as lie detectors in studies involving nonverbal behavior. Some nonverbal behavior studies, however, used professionals such as police officers and customs officers as lie detectors. Although several studies showed that even those professionals perform around the level of chance (DePaulo & Pfeifer, 1986; Köhnken, 1987; Vrij, 1993). Ekman and his colleagues (Ekman & O'Sullivan, 1991; Ekman, O'Sullivan, & Frank, 1999) found that some groups of professionals perform above the level of chance, such as members of the Secret Service and a group of federal officers with a special interest and experience in deception. The latter group obtained an accuracy rate (truths and lies combined) of 73% (Ekman et al., 1999). However, even these hit rates are within the range of scores reported in the literature for untrained human lie detectors with no special experience (DePaulo, Anderson, & Cooper, 1999).

An explanation why professionals also seem to perform poorly in detecting lies via examining nonverbal behavior is that they do not know where to look and have false beliefs about which behaviors might be clues to deception (Akehurst, Köhnken, Bull, & Vrij, 1996; Vrij & Semin, 1996). Vrij and Semin (1996) found that 75% of professional lie detectors (police officers, customs officers and so on) believe that liars look away, although gaze aversion has not been found to be a reliable indicator of deception (DePaulo et al., 1985; Vrij, 2000; Zuckerman, DePaulo, & Rosenthal, 1981). Research has shown that observers improve their skills in detecting deceit if they receive some information about the relationship between nonverbal behavior and deception. Lie detectors in deTurck's (1991) study obtained an accuracy rate of 70% after they were informed to ignore looking at eye contact but to focus their attention on message duration, response latency, pauses, nonfluencies, adaptors and hand gestures. These accuracy rates are probably still not at their potential level due to the fact that judges do not always use the information with which they are provided. Vrij (1994) informed judges that liars generally display fewer subtle hand and finger movements than truth tellers. He then showed judges videoclips of twenty different people. For each person, two video fragments were presented simultaneously (on two different TV screens located next to each other). In one fragment the person was lying and in the other fragment the person was telling the truth. The judges were asked to indicate for each person in which fragment the person was lying. By consequently using the information provided, 75% of the answers could have been correct. The average accuracy rate, however, was only 60%, suggesting that the judges did not consistently apply the information provided.

It is therefore possible that even higher accuracy rates could be obtained when a more sophisticated nonverbal behavior deception detection method is used, excluding any subjective interpretations. The present experiment examines this issue.

The crucial question is to which behaviors attention should be paid. This question is difficult to answer, as research has shown that deception itself is not related to a unique pattern of specific behaviors (DePaulo et al., 1985; Ekman, 1992; Vrij, 1998, 2000; Zuckerman et al., 1981). In other words, there is nothing like Pinocchio's nose. However, liars might experience emotions while lying. The three most common types of emotion associated with deceit are fear, excitement ('duping delight') and guilt (Ekman, 1989, 1992). Liars might be afraid of getting caught, they might become excited at having the opportunity of fooling someone, or they might feel guilty (Ekman, 1992). In some situations, liars also might find it difficult to lie. They have to think of plausible answers, should not contradict themselves, should tell a lie that is consistent with everything the other person knows, should avoid making slips of the tongue, and have to remember what they have said, so that they can say the same things when someone asks them to repeat their story. Experiencing emotions and cognitive load might result in signs of emotion and cognitive load which then gives the lie away (Ekman, 1992; Vrij, 1998, 2000). Experimental studies concerning how people behave under stress have been mainly conducted by Ekman and his colleagues (Ekman, 1992; Frank & Ekman, 1997). They found that under these circumstances it is possible to detect deceit (they reported hit rates around 80%) by paying attention to signs of emotions which emerge via (micro) facial expressions (Frank & Ekman, 1997) or by observing smiles and pitch of voice (Ekman, O'Sullivan, Friesen, & Scherer, 1991). The strongest evidence for the effects of raising the stakes would be obtained by experimentally manipulating the stakes. In a series of experiments conducted by DePaulo and her colleagues in which the stakes were manipulated, it was found that high stake lies were indeed easier to detect than low stake lies (DePaulo, Kirkendol, Tang, & O'Brien, 1988; DePaulo, Lanier, & Davis, 1983; DePaulo, LeMay, & Epstein, 1991; DePaulo, Stone, & Lassiter, 1985b; Lane & DePaulo, 1999). In one of our own recent studies (Vrij, Harden, Terry, Edward, & Bull, in press) this finding was replicated.

The present experiment deals with cognitive load. All participants watched a videotape of a theft in a hospital. In a subsequent interview specific questions about the film were asked. Some participants were requested to recall what they had seen, whereas others were asked to lie without having much time to prepare their lies. The fact that the participants had to lie almost spontaneously makes this task difficult for the liars and we therefore expected liars, compared to truth tellers, to show more behaviors that indicate cognitive load. In particular, we expected liars to show a longer latency period, more 'ah' and 'non-ah' speech disturbances, a slower speech rate and fewer illustrators and hand/finger movements (Hypothesis 1) as these behaviors are associated with thinking hard (Burgoon, Kelly, Newton, & Keely-Dyreson, 1989; Ekman & Friesen, 1972; Goldman-Eisler, 1968; Köhnken, 1989; Vrij, 1998). See the Method section for a description of these behaviors. In order to find out to what extent truth tellers and liars can be correctly classified on the basis of these behaviors, discriminant analyses were conducted with objective truth status as the classifying variable and these six nonverbal behaviors as independent variables. It was expected that the analysis would reveal an accuracy rate (percentage of correct classifications of truth tellers and liars) above the level of chance (Hypothesis 2).

Differences between liars and truth tellers in what they say are often assessed using Criteria-Based Content Analysis (CBCA) (Ruby & Brigham, 1997; Steller & Köhnken, 1989; Vrij, 2000; Vrij & Akehurst, 1998). CBCA was developed in Germany by Steller and Köhnken (Steller & Köhnken, 1989) in order to evaluate statements from children who are witnesses or alleged victims, most commonly of sexual abuse. Many authors still describe CBCA as a technique solely developed to evaluate statements made by children in sexual offense trials (Honts, 1994; Horowitz, Lamb, Esplin, Boychuk, Krispin, & Reiter-Lavery, 1997). Others, however, advocate the additional use of the technique to evaluate the testimonies of adults who talk about issues other than sexual abuse (Köhnken, Schimossek, Aschermann, & Höfer, 1995; Porter & Yuille, 1996; Ruby & Brigham, 1997; Steller & Köhnken, 1989). In CBCA, trained evaluators examine a statement and judge the presence or absence of each of 19 criteria. Appendix 1 provides a brief description of the CBCA criteria used in this study. Vrij and Akehurst (1998) and Vrij (2000) give more detailed descriptions of the CBCA criteria. The underlying hypothesis of CBCA is that a statement derived from an actual memory of an experience differs in content and quality from a statement based on invention or fantasy, and that only a person who has actually experienced an event is likely to incorporate certain types of content into a statement about it. In other words, the presence of each criterion strengthens the hypothesis that the account is based on genuine personal experience. This hypothesis is originally stated by Undeutsch (1967, 1989) and is therefore know as the *Undeutsch-Hypothesis* (Steller, 1989). Following the Undeutsch-hypothesis, it was expected that liars would obtain lower CBCA scores than truth tellers (Hypothesis 3), and that liars and truth tellers could be correctly classified above the level of chance on the basis of their CBCA scores (Hypothesis 4).

Recently, Reality Monitoring has been used as an alternative method to measure verbal differences between responses believed to be true and false (Alonso-Quecuty, 1992, 1996; Hernandez-Fernaud & Alonso-Quecuty, 1997; Höfer, Akehurst, & Metzger, 1996; Manzanero & Diges, 1996; Roberts, Lamb, Zale, & Randall, 1998; Sporer, 1997). The core of Reality Monitoring is that memories of real experiences are obtained through perceptual processes and are therefore likely to contain *perceptual information* (visual details and details of sound, smell, taste, or physical sensations), *contextual information* (details about where and when the event took place), and *affective information* (details about how someone felt during the event). Accounts of imagined events are derived from an internal source and are therefore likely to contain *cognitive operations,* such as thoughts and reasonings ('I can only remember my thinking of what my friend would like to have for a present') (Johnson, Hashtroudi, & Lindsay, 1993; Johnson & Raye, 1981, 1998). It was therefore expected that truth tellers would obtain a higher Reality Monitoring score than liars (Hypothesis 5) and would include more perceptual, contextual and affective information in their statements than liars (Hypothesis 6). Liars, on the other hand, are likely to include more cognitive operations in their statements than truth tellers (Hypothesis 7). It was also expected that liars and truth tellers could be correctly classified above the level of chance on the basis of their Reality Monitoring scores (Hypothesis 8).

Finally, it was investigated whether a combination of the two verbal techniques and the nonverbal technique would classify liars and truth tellers more accurately than the individual techniques. We expected this to be the case. A combined technique takes more information into account than do individual techniques, and, the more aspects of liars that will be scrutinized, the more likely it is that their lies can be detected (Hypothesis 9).

Method

Participants

A total of 73 nursing students participated, 20 males and 53 females. Their average age was $M = 28.89$ years ($SD = 7.9$ years). Originally, 79 participants took part in the experiment. Three participants, however, gave answers which lasted less than 10 seconds. As it is impossible to perform CBCA assessments on very short statements, these participants were disregarded in the analyses. Another two participants did not lie when requested to do so and one participant lied when asked to tell the truth. These participants were also disregarded in the analyses.

Procedure

Nursing students were recruited at the University of Portsmouth nursing school. They were asked to participate in a study about "telling lies." Each student participated individually and received £5 for their participation. First of all, in order to motivate the nurses to try to perform well in the study they were told that the ability to lie successfully is extremely important to nurses and that good nurses may need to be good liars. Previous research has indicated that this information does increase participants' motivation to perform well (DePaulo, Kirkendol, Tang, & O'Brien, 1988; DePaulo, Lanier, & Davis, 1983). They were then told that they would see a video and that they would be interviewed twice about this video. In one interview they had to recall what they had seen and in the other interview they had to lie. The order in which the truthful and deceptive interviews took place was counterbalanced. Only the first interviews (deceptive for some participants and truthful for others) were analyzed, creating a between-subjects design. We introduced the study to the participants as a within-subjects design because we wanted all participants to lie. We did this for motivation purposes: (1) in this case nobody could think that they were allocated to a 'control condition' and (2) the information about the good liar-good nurses relationship was relevant to all participants. The nurses were then shown a video of 118 seconds in length. This videotaped event featured a colour presentation of the theft of a bag from a patient by a visitor. In the video, a woman enters a hospital and walks to the first floor. While walking down the corridor, she notices a patient lying in bed with a handbag next to her. The visitor enters her room, looks at the patient's name plate and pretends that she knows the patient. She then takes the bag and starts to walk out of the room. The patient notices the theft and asks the visitor to return the handbag. A nurse comes in and asks what is going on. The patient tells the nurse that she does not know the visitor and that the visitor is trying to steal her bag. The visitor tells the nurse that she is the patient's neighbour and that the patient is confused. The nurse then leaves the room. The video finishes with the visitor smiling as she opens the patient's purse and notices money in it.

After watching the video, participants in the *truthful condition* (N = 34) were asked the following three general and open-ended questions: What did the nurse do? What did the patient do? and What did the visitor do? They were asked to answer all questions truthfully. Participants in the *deception condition* (N = 39) were asked to lie while answering the same three questions. In order to make the task not too difficult for the participants they were informed about two of the three questions that would be asked before the interview started and they were given approximately 15 seconds to think about an answer. Which questions were told beforehand was counterbalanced. Analyses showed that this manipulation had no effect on either the verbal or nonverbal behavior displayed by the participants and will therefore be disregarded in this article. All interviews were videotaped and audiotaped and were transcribed verbatim from the audiotapes (the transcripts included the stutters made by the participants). The answers of truth tellers were significantly longer (M = 89 seconds, SD = 46) than the answers of liars (M = 42 seconds, SD = 19), $F(1, 71) = 34.06$, $p < .01$).

Dependent Variables

Two observers coded the behavior of the participants independently, and Pearson's correlations were conducted between the two sets of data from the two coders to detect any differences in judgement. The observers were not informed as to whether the participants were lying or telling the truth and had not seen the stimulus video. They employed a coding system used by us in previous studies (Akehurst & Vrij, 1999; Vrij, 1991, 1995; Vrij, Semin, & Bull, 1996; Vrij, Akehurst, & Morris, 1997). The following ten behaviors were coded (the 'ah' and 'non-ah' speech disturbances were scored on the basis of a typed verbatim text):

—*gaze aversion*: number of seconds for which the participant looked away from the interviewer (2 coders, r = .95, $p < .01$)
—*smiling*: frequency of smiles and laughs (2 coders, r = .90, $p < .01$)
—*illustrators*: frequency of arm and hand movements which were designed to modify and/or supplement what was being said verbally (Ekman & Friesen, 1969) (2 coders, r = .96, $p < .01$)
—*adaptors*: frequency of scratching the head, wrists etc. Rubbing one's hands together were not coded as adaptors but as hand and finger movements (2 coders, r = .94, $p < .01$)
—*frequency of hand and finger movements*: Movements of the hands or fingers without moving the arms (2 coders, r = .92, $p < .01$)
—*frequency of foot and leg movements*: Movements of feet or legs. Simultaneous movements of feet and legs were scored as one movement (2 coders, r = .93, $p < .01$)
—*speech hesitations*: frequency of saying 'ah' or 'mm' between words (2 coders, r = .95, $p < .01$)
—*speech errors*: frequency of word and/or sentence repetition, sentence change, sentence incompletion, and slips of the tongue (2 coders, r = .91, $p < .01$).
—*latency period*: period of time between the question being asked and the answer being given (2 coders, r = .98, $p < .01$)
—*speech rate*: number of spoken words (using the count option in WordPerfect) divided by the length of interview minus latency period (2 coders, r = .98, $p < .01$).

The Pearson's correlations show evidence of a strong consistency between the two coders. Thus, the behavioral scores were based on the average scores of the two coders and are presented in Table 1. The reported duration and frequencies of all categories of nonverbal behavior were corrected for the length of the interviews or for the number of spoken words. Patterns listed from gaze aversion down to foot and leg movements were calculated on a per minute basis. Patterns for ah and non-ah disturbances were calculated per 50 words. Latency period scores represent the average latency period per question. Two independent raters received training in CBCA scoring. First, both raters read all the major published papers about CBCA. Second, they were trained in CBCA scoring by a British CBCA expert. Third, both the trainee raters and the expert rater evaluated several example transcripts (from a different study). Fourth, the three raters compared their results and feedback was given by the expert rater. Following common procedure (Craig, 1995; Hershkowitz, Lamb, Sternberg, & Esplin, 1997; Lamb, Sternberg, Esplin, Hershkowitz, Orbach, & Hovav, 1997; Lamb, Sternberg, Esplin, Hershkowitz, & Orbach, 1997; Landry & Brigham, 1992; Ruby & Brigham, 1998; Vrij, Kneller, & Mann, 2000; Winkel & Vrij, 1995; Zaparniuk, Yuille, & Taylor, 1995) the two observers in the present study scored for each of the three answers the presence or absence of each of the CBCA criteria[1] used in this study, with exception of criterion 3 (quantity of details): '1' was assigned when the criterion was present and '0' when the criterion was absent. Per criterion a total score for the whole interview was calculated by adding the three scores for the three individual answers and then dividing this total score by three. In order to score criterion 3, the raters counted per interview the number of details mentioned. Ratings took place by using the transcripts, and the raters were blind to the experimental conditions. Correlations between the two coders for each of the criteria were satisfactory. They were lower than the correlations with regard to nonverbal behaviors, but higher than found by some others in CBCA research (Anson, Golding, & Gully, 1993): logical structure: $r = .55$, $p < .01$; unstructured production: $r = .65$, $p < .01$; quantity of details: $r = .90$, $p < .01$; contextual embedding: $r = .85$, $p < .01$; description of interactions: $r = .90$, $p < .01$; reproduction of speech: $r = .97$, $p < .01$; unusual details: $r = .77$, $p < .01$; superfluous details: $r = .69$, $p < .01$; accounts of mental state: $r = .58$, $p < .01$; attribution of perpetrator's mental state: $r = .71$, $p < .01$; spontaneous corrections: $r = .54$, $p < .01$; admitting lack of memory: $r = .89$, $p < .01$; raising doubts about one's own memory: $r = .70$, $p < .01$; pardoning the perpetrator: $r = 1.00$, $p < .01$. The scores for each of the criteria were therefore based on the average scores of the two coders.

The next step was to calculate a total CBCA score, which is common in CBCA research (Craig, 1995; Esplin, Boychuk, & Raskin, 1988, cited in Raskin & Esplin, 1991; Hershkowitz et al., 1997; Lamb et al., 1997a, b; Vrij et al., 2000; Winkel & Vrij, 1995). In order to calculate the CBCA score, the scores for the 14 criteria were dichotomized. With regard to number of details, a median split (N = 13.50) was used. Those 50% of the participants (N = 37) with a score higher than 13.50 obtained '1' on this criterion, the other 50% obtained a score of '0.' Dichotomizations for the other criteria occurred on the basis of presence or absence of a criterion in the whole interview. A score of '0' was assigned when the criterion was absent, and a score of '1' was assigned when the criterion was present. The total CBCA score was the total score of the 14 criteria and could range from 0 to 14.

Two independent raters received training in RM-scoring. A British RM-expert provided the judges with a detailed description of how the criteria should be scored, including some case examples. On the basis of this information the judges felt capable of scoring the transcripts without any further instructions. This is in agreement with Sporer (1997) who also found that it is much easier to teach (and to learn) Reality Monitoring scoring than CBCA scoring. With regard to the present study, the two raters scored per interview the frequency of occurrence of visual details (which includes actions) ("The visitor[a] came in[b] and kissed[c] the patient[d] are four visual details (a, b, c, and d), sound details ("She said that is my bag" is one sound detail), time details ("When the nurse came in, the patient . . ." is one time detail), details about location ("The visitor walked through the corridor" is one location detail), and cognitive operations ('the patient didn't believe she knew the visitor' is one cognitive operation). Affective information is similar to CBCA criterion 12 (accounts of subjective mental state) and therefore was not scored again, the CBCA score for this criterion being used in the Reality Monitoring scores. Ratings took place using the transcripts and the raters were blind to the experimental conditions. Inter-coder reliability scores (Pearsons's correlations) were calculated for all the individual criteria (visual details: $r=.96$, $p < .01$; sound details: $r=.77$, $p < .01$; details about location: $r=.72$, $p < .01$; time details: $r=.85$, $p < .01$; cognitive operations: $r=.75$, $p < .01$). The correlations showed consistency amongst the two coders and scores for each of the criteria were therefore based on the average scores of the two coders. Table 2 provides the results for the individual Reality Monitoring criteria. In order to create the Reality Monitoring scale each variable was dichotomized. A median split (N = 10.50) was used for visual details. Those 50% of the participants with a score higher than 10.50 obtained '1' on this criterion, the other 50% obtained a score of '0.'[2] Dichotomizations for the other criteria occurred on the basis of absence or presence of each of the criteria in the interview. A score of '0' was assigned when the criterion was absent, and a score of '1' when the criterion was present. Cognitive operations were not included in the total Reality Monitoring score as the presence of this criterion does not indicate truth telling (as is the case with the other criteria). The Reality Monitoring scale therefore contained five criteria (visual details, sound details, details about locations, details about time and affective information) and the total score could range from 0 to 5.

Results

In order to test Hypothesis 1 (liars display a longer latency period, more ah and non-ah speech disturbances, a slower speech rate and fewer illustrators and hand/finger movements than truth tellers), Hypothesis 3 (truth tellers will obtain a higher CBCA score than liars) and Hypothesis 5 (truth tellers will obtain a higher RM score than liars) a MANOVA was conducted with Deception (yes or no) as factor and the nonverbal behaviors, total CBCA score and total RM score as dependent variables. The MANOVA revealed a significant effect, $F(12, 60) = 5.11$, $p < .01$. Table 1 provides the univariate outcomes. As can be seen in Table 1, several significant differences emerged between liars and truth tellers. Compared to truth tellers, liars made fewer illustrators and hand and finger movements, had more ah-speech disturbances, and waited longer before giving an answer. These findings support Hypothesis 1. Furthermore, truth tellers obtained a higher CBCA score[3] and a higher RM score than liars. Therefore, Hypotheses 3 and 5 are also supported.

In order to test Hypothesis 6 (truth tellers will include more perceptual, contextual and affective information in their statements than liars) and Hypothesis 7 (liars are likely to include more cognitive operations in their statements than truth tellers) a MANOVA was conducted with Deception (yes or no) as factor and the individual (not dichotomized) Reality Monitoring criteria as dependent variables. The MANOVA showed a significant effect, $F(5, 67) = 9.61$, $p < .01$. Table 2 gives the results for the individual Reality Monitoring criteria. As can be seen in Table 2, truth tellers included more perceptual details (vision and sound), more information about locations and more information about time in their accounts than liars. This supports Hypothesis 6. In contrast to what was predicted in Hypothesis 7, liars mentioned fewer cognitive operations than truth tellers.

In order to test Hypotheses 2, 4, 8 and 9, four discriminant analyses

TABLE 1

Nonverbal Behavior as a Function of Deception

	Condition				
	Truth		Lie		
Behavior	m	(sd)	m	(sd)	F(1, 71)
gaze aversion	4.66	(6.5)	6.33	(8.2)	.91
smiles	.66	(1.0)	1.66	(3.1)	3.24
illustrators	6.74	(8.1)	1.64	(5.0)	10.86**
adaptors	1.97	(3.8)	.86	(3.0)	1.93
hand/finger movements	15.73	(13.7)	9.17	(13.0)	4.42*
foot/leg movements	11.62	(9.7)	13.78	(16.9)	.43
ah speech disturbances	2.73	(2.7)	4.64	(3.4)	7.00**
non-ah speech disturbances	.98	(1.8)	1.62	(2.3)	1.73
latency period	2.24	(1.4)	3.65	(4.3)	5.73*
speech rate	130.23	(49.4)	142.11	(64.3)	.76
CBCA	5.32	(2.0)	3.31	(1.5)	23.44**
RM	3.20	(1.3)	2.00	(1.1)	14.58**

** $p < .01$; * $p < .05$.

were conducted determining the accuracy of the detection techniques in classifying liars and truth tellers. In these analyses, the objective truth status was the classifying variable and the six nonverbal behaviors mentioned in Hypothesis 1, total CBCA score and total RM score were the independent variables. The results are given in Table 3. First of all, it can be seen that the analysis with nonverbal behaviors as variables yielded a highly significant discriminant function, $\chi2(4, n=73) = 23.57$, $p < .01$. Four variables contributed to this function: Illustrators (Wilks' lambda = .87), ah speech disturbances (Wilks' lambda = .78), hand and finger movements (Wilks' lambda = .73), and latency period (Wilks' lambda = .71). In total, 70.6% of the truth tellers and 84.6% of the liars were correctly classified resulting in a total accuracy score of 78.08%. This supports Hypothesis 2.[4] Also the discriminant analyses with the total CBCA scores and RM scores as variables resulted in highly significant discriminant functions (see Table 3) and correct classifications of the majority of participants (72.60% with CBCA and 67.12% with RM respectively). This supports Hypotheses 4 and 8.

TABLE 2

Reality Monitoring Criteria as a Function of Deception

	Condition				
	Truth		Lie		
Criteria	m	(sd)	m	(sd)	F(1, 71)
Perceptual information: vision	13.13	(6.3)	9.45	(3.7)	9.39**
Perceptual information: sound	3.31	(2.2)	1.26	(1.4)	22.88**
Spatial information	2.03	(2.1)	1.30	(1.2)	3.53*
Temporal information	1.35	(1.4)	.28	(.7)	18.57**
Cognitive operations	3.32	(2.1)	1.71	(1.7)	12.88**

** $p < .01$; * $p < .05$.

The fourth discriminant analysis revealed that, as was predicted in Hypothesis 9, the combination of the two verbal techniques with the nonverbal technique resulted in the highest accuracy scores, in particular a higher accuracy rate for detecting truths was obtained. In that case, 76.5% of the truth tellers and 84.6% of the liars were correctly classified, resulting in a total accuracy score of 80.82%. The discriminant function was highly significant, $\chi 2$ (6, $n=73$) = 38.79, $p < .01$. Six variables contributed to this function: CBCA score (Wilks' lambda = .74), latency period (Wilks' lambda = .67), hand and finger movements (Wilks' lambda = .63), ah-speech disturbances (Wilks' lambda = .60), illustrators (Wilks' lambda = .58) and speech rate (Wilks' lambda = .57).

Discussion

Previous research has created a pessimistic view about the possibility of detecting lies by analyzing nonverbal behavior. We argued that it might be possible to detect lies when the appropriate behaviors are taken into account and subjective interpretations are disregarded. We defined appropriate behaviors as signs of emotion or signs of cognitive load. As mentioned in the introduction, DePaulo's work revealed that high stake lies are easier to detect than low stake lies, and Ekman's work showed that up to 80% of truths and lies can be detected in high stake situations while paying attention to behavioral signs of emotion. Our findings revealed similar high per-

TABLE 3

Discriminant Analyses with Nonverbal Behavior, Criteria-Based Content Analysis and Reality Monitoring

	Hit rates						
Detection technique	Truth	Lie	Total	Eigenvalue	Lambda	df	χ^2
Nonverbal behavior	70.6%	84.6%	78.08%	.41	.71	4	23.57**
CBCA	64.7%	79.5%	72.60%	.35	.74	1	20.91**
RM	70.6%	64.1%	67.12%	.25	.80	1	13.17**
CBCA + RM + nonverbal behavior	76.5%	84.6%	80.82%	.77	.57	6	38.79**

** $p < .01$; * $p < .05$.

centages of accurately detecting truths and lies in situations which require hard thinking while taking signs of cognitive load into account.[5] The findings suggest that, if properly applied, analyzing nonverbal behavior might be an accurate tool to detect deceit.

The liars in this study were facing a difficult task. They had to make up a story (had to tell a so-called 'bold-faced lie' [McCornack, 1997]) and had to do this almost spontaneously. It is therefore perhaps not surprising that they showed signs of cognitive load. To what extent are these bold-faced lies realistic? McCornack (1997) argues that bold faced lies comprise only a small portion of the deceptive messages and that most deceptive messages in daily life involve subtle and complex packaging of both false and truthful elements. We have no reason to dispute this view, but would like to emphasize that bold-faced lies do occur in daily life settings. They also take place in police interviews. For example, in our analysis of police interviews with a convicted murderer (Vrij & Mann, in press) we came across a bold-faced lie, which is related to how he met the victim. Substantial evidence (several independent eye witnesses and physical evidence) has shown that he went to *location A* and that *he made contact with the victim.* He strongly denied this and told the police instead that he met the victim at *location B* (which was totally different from location A) and that *the victim approached him.* He described in detail how the victim contacted him, a story which was entirely fabricated. Another of our studies (Vrij & Mann, in press, b) involved two more apparently bold-faced lies. In one case, a man who was found guilty of the murder of his wife claimed that people forced themselves into his house, killed his wife, beat him until he was unconscious and tied him up. He had several injuries which, he claimed, were the result of the attack. The man, however, had injured himself in an attempt to make his story more plausible. In another case, a woman who was found guilty of killing her boyfriend told the police that they were the victims of road rage and that a stranger chased their car and eventually killed her boyfriend. All these seem to be examples of bold-faced lies. One might argue that even bold-faced lies are often not total fabrications as people, when fabricating a story, could simply describe a situation they had experienced before. It is unlikely that the liars in the present study could do this, as they were forced to fabricate about a specific situation, namely the activities of a patient, nurse, and visitor in a hospital. We acknowledge that some bold-faced lies are probably not total fabrications, but we believe that some are. For example, had the woman ever experienced a road rage event in her life before? Similarly, was the man beaten until unconscious and tied up in his life before? We do not know the answers to these questions but their stories might well have been total fabrications.

We instructed the liars in our study to tell total fabrications not because we are particularly interested in this type of lie, but because we wanted to create a situation in which the lie requires mental effort. We acknowledge that in real life telling lies often does not require more mental effort than telling the truth. McCornack (1997, p. 102) convincingly argued that deception possesses fundamental cognitive efficiency advantages over truth telling within certain contexts. For example, by receiving a present which you do not like from an acquaintance it is often easier to lie and to say that you like the present than to express your true opinion. We believe that in real life some lies do require mental effort, as our own study with the convicted murderer revealed (Vrij & Mann, in press, a). When the police interviewed him the first time, he was asked 'What did you do on that particular day?' He described his activities in detail and in chronological order but his behavior changed as soon as he started to describe his activities during the afternoon. The police later discovered that he lied in that part of the interview. While lying, the man showed more gaze aversion, had longer pauses, spoke slower and made more speech errors than when he was telling the truth. This behavioral pattern is typical for somebody who has to think hard. Perhaps it was surprising that the man gave the impression that he had to think hard. As he told the police, he knew he was a suspect in this case and expected to be interviewed. There is also evidence that he had prepared himself for the police interviews. A possible explanation is that he was not very bright, thus not fully taking advantage of the preparation time that was available to him. As Ekman and Frank (1993) have pointed out, preparation probably does not benefit liars who are not so clever. If intelligence really affects preparations, then criminals or guilty suspects might be in a disadvantageous position during police interviews as their IQ is often rather low (Gudjonsson, 1992).

It also might explain the preliminary findings of our current analyses of interviews with suspects in police interviews (Mann, Vrij, & Bull, 1998). They reveal that when suspects lie they show a decrease in illustrators, an increase in pauses and an increase in latency period. In other words, suspects seem to show signs of cognitive load. In our view, systematic and detailed analyses of nonverbal behavior displayed by suspects and looking for signs of cognitive load are therefore useful to detect deceit.

Obviously, signs of cognitive load per se do not necessarily indicate deception, as truth tellers might have to think hard as well. Detecting deceit by paying attention to nonverbal behavior in real life settings is a two stage process (Vrij, 2000). First, signs of emotion (guilt, fear, excitement or any other emotion) or cognitive load need to be detected and, second, explanations for these signs should be given, where deceit is only one possible explanation. This process may reveal lies. As was mentioned above, the convicted murderer changed his behavior as soon as he started describing his activities during the afternoon (Vrij & Mann, in press, a). We wondered why and hypothesized that he was lying at that particular moment. Our intuition turned out to be correct.

In theory it is possible that in the present experiment the behavior of an extraordinary sample of people was examined whose lies are easy to detect. There is evidence that this was not the case. In a recent study (Vrij & Baxter, 1999), we randomly selected fragments of 10 interviews out of the 73 interviews analyzed in the present study and showed these interviews to 50 college students. We asked the students to indicate for each person in the interview whether the person was lying or not. The students achieved an accuracy rate of 56% for truths and 50% for lies. These outcomes are typically found in this type of detection of deception experiment.

The results of the present experiment further revealed that accuracy rates above the level of chance were obtained with both CBCA assessments and Reality Monitoring assessments. It is difficult to make a comparison of both methods at this stage. CBCA might well have been in a disadvantageous position in this comparison, as it was used in a context other than the one for which it was designed. As stated earlier, CBCA was originally designed for assessing statements of children in sexual abuse cases. In this study it has been used to assess adults' statements. Some researchers have advocated the use of the CBCA technique to also evaluate the testimonies of adults who talk about issues other than sexual abuse. The reasonably high accuracy rate obtained in the present study (72.6%) support their view. Reality Monitoring might have been in a disadvantageous position as well. One important aspect of Reality Monitoring is looking for perceptual information, such as cues of sound, vision, smell, taste or touch. As participants did not actually take part in an event—they were only watching a videotape—they could not smell, taste or touch anything (and therefore did not mention any of these cues). Finally, the comparison of both verbal methods was not entirely fair because accounts of mental state were counted only under CBCA. These scores were then added in the Reality Monitoring lie detection scale as well. CBCA was hereby given a potential advantage.

A logical step is to combine both verbal methods. An interesting addition to the CBCA list of criteria would be the Reality Monitoring criterion 'perceptual information' (criterion 2). For example, pornographic films may increase children's knowledge about sexual acts. As a result, an unexperienced child may give a detailed account about a non-experienced sexual encounter after watching a pornographic film. However, in such a recall details about smell and taste will be missing, as genuine experiences are required for such details. Details about smell and taste in statements about sexual abuse may therefore be a strong indication that the statements are based upon real experiences (unless smell and taste were mentioned by people in the pornographic film).

There was no support for Hypothesis 7 which suggested that liars would include more cognitive operations in their accounts than truth tellers. Previous deception research with Reality Monitoring also could not support this hypothesis (Alonso-Quecuty, 1992, 1996; Hofer et al., 1996; Roberts et al., 1998; Sporer, 1997). One explanation is that people use cognitive operations in order to facilitate and enhance later memory for experienced events (Roediger, 1996). For example, a person who drove fast in Germany might try to remember this in two different ways. First, the person could remember having actually looked at the speedometer to find out how fast he or she was driving. Alternatively, they could remember this by logical reasoning, for example, by thinking that he or she must have driven fast because they used the motorway. The latter alternative, in which a cognitive operation is included, is an easier way of remembering having driven fast than the first alternative. When the person is asked a couple of years later whether he or she drove fast through Germany it is therefore more likely that the person will remember this by thinking that he or she drove on the motorway than by remembering having checked the speedometer. As a result, the person's memory about this experienced event will contain a cognitive operation. Due to the lack of support for cognitive operations found in the present and previous studies, we suggest this variable should not be included in a Reality Monitoring lie detection scale. As was mentioned in the method section, cognitive operations was not included in the Reality Monitoring scale in the present study either. Including this variable in the Reality Monitoring scale and rerunning the discriminant analysis led to a significant discriminant function ($\chi 2(1, n = 73) = 8.53$, $p < .01$, eigenvalue = .13, Wilks' Lambda = .89) but to lower hit rates (truth: 70.6%, lie: 53.8%, total: 61.64%) than the hit rates obtained without cognitive operations (see Table 3).

Although previous studies suggest that verbal cues (CBCA and RM) are more powerful discriminators between truths and lies than nonverbal cues (see introduction), such studies typically do not compare verbal cues and nonverbal cues directly, making it impossible to determine the relative power of both sets of discriminators. Such a direct comparison was made in the present experiment. Interestingly, the present findings did not show superiority of verbal cues above nonverbal cues. However, we acknowledge that more research needs to be done. Such studies should incorporate different types of lie.

Instead of comparing verbal and nonverbal detection methods, in our view a more fruitful approach would be to investigate to what extent a combination of verbal and nonverbal detection methods lead to higher accuracy rates than the two types of method independently. The present study showed that the highest accuracy rates were obtained by combining the verbal and nonverbal techniques. The discriminant analysis that took all three techniques into account could correctly classify 80.82% of the liars and truth tellers and both verbal (CBCA) and nonverbal cues contributed to the significant discriminant function. Some Reality Monitoring researchers (Porter & Yuille, 1996; Roberts et al., 1998) already use a combined instrument by including speech disturbances in their Reality Monitoring scale. Our findings support this idea. In the discriminant analysis in which the three techniques were included, both speech disturbances and verbal cues (CBCA score) contributed to the discriminant function. The discriminant function, however, revealed that in addition to speech disturbances, behaviors such as illustrators, hand and finger movements, and latency period made an important contribution to detecting deceit as well. We therefore recommend to also take behaviors other than speech disturbances into account when attempting to detect deceit.

Concerning the methodology of the study, one issue merits attention. All participants watched the same 2 minute video of a theft in a hospital. It might be that the truths told by the participants bore more similarity to one another than the lies they told. On the basis of this, it might be that a coder could probably tell after a few trials of coding which narrative they were coding were probably truthful and which were lies. This 'knowledge' might have affected their codings. Although this sounds reasonable, we do not think that this actually happened, as the stories of truth tellers did not show too much similarity. Some truth tellers used a 'global approach' and just mentioned in a few sentences what, in their view, were the main events in the video. However, different truth tellers who applied this approach phrased the events differently and mentioned different events. Other truth tellers used a more detailed approach and discussed the video in more detail. Even in this situation the stories of different truth tellers varied, as different truth tellers mentioned different details.

With regard to the ecological validity of the present study, probably the most obvious criticism is that we asked our participants to describe an event they had watched on a video rather than describing a live event in which they had actually participated. There were two reasons why we chose a videotaped event, First, we wanted to create a highly controlled and standardised situation. We believe that it is essential to test innovative ideas (such as comparing different detection of deception techniques) in highly controlled situations first, as the exact impact can only be determined in such situations. Second, recent research has shown that watching an event on a video or actually taking part in such an event results in similar CBCA scores (Akehurst, Köhnken, & Höfer, 1995). This suggests that a method utilizing a videotaped event has a positive effect on the standardization of the study without compromising its ecological validity too much. Despite this, we do acknowledge that deception research should also include studies with higher ecological validity than that of the present study.

Appendix 1:

A Brief Description of the CBCA Criteria Used in This Study

1. Logical structure. Logical structure is present if the statement essentially makes sense, that is, if the statement is coherent and logical and the different segments fit together, that is, for example different segments are not inconsistent or discrepant. *2. Unstructured production.* Unstructured production is present if the information is scattered throughout the statement instead of mentioned in a structured, coherent and chronological order. The incoherent and unorganized manner of presentation is, for instance, caused by digressions or spontaneous shifts of focus. *3. Quantity of details.* This criterion requires that the statement must be rich in detail, that is, specific descriptions of place, time, persons, objects and events should be present. *4. Contextual embedding.* Contextual embedding is present if the events are placed in time and location, and when the actions are connected with other daily activities and/or customs. *5. Descriptions of interactions.* This criterion is fulfilled if the statement contains information about interactions involving at least the accused and witness, and if this information consists of three parts, i.e. an action of actor A leads to a reaction of actor B which leads to a reaction of actor A again. *6. Reproduction of speech.* Reproduction of speech is present if speech, or parts of the conversation, is reported in its original form and if the different speakers are recognizable in the reproduced dialogues. This criterion is not satisfied by a report about the content of a dialogue; it is only satisfied when there is a virtual replication of the utterances of at least one person. *8. Unusual details.* Unusual details refer to details of persons, objects, or events which are unusual and/or unique but meaningful in the context. *9. Superfluous details.* Superfluous details are present if the witness describes details in connection with the allegations which are not essential for the accusation, such as a child who says that the adult tried to get rid of the cat which entered the bedroom because he (the adult) is allergic to cats. *12. Accounts of subjective mental state.* This criterion is present when the witness describes feelings or thoughts experienced at the time of the incident, as well as reports of cognitions, such as thinking about how to escape while the event was in progress. *13. Attribution of perpetrator's mental state.* This criterion is present if the witness describes her or his perceptions of the perpetrator's feelings, thoughts or motives during the incident. *14. Spontaneous corrections.* This criterion is fulfilled if corrections are spontaneously offered or information is spontaneously added to material previously provided in the statement (spontaneous means without any interference by the interviewer). *15. Admitting lack of memory.* This criterion is present if a witness admits lack of memory by either saying "I don't know" or "I don't remember" or by giving a more extensive answer. *16. Raising doubts about one's own testimony.* This criterion is present if the witness expresses concern that some part of the statement seems incorrect or unbelievable. *18. Pardoning the perpetrator.* Pardoning the perpetrator is present if the witness tends to favour the alleged perpetrator in terms of making excuses for the alleged perpetrator or failing to blame the alleged perpetrator.

Notes

1. Given the fact that in this study statements of *adults* were used, we thought that several criteria would be inappropriate and were therefore ignored. These criteria were: accurately reported details misunderstood (criterion 10), related external associations (criterion 11) and details characteristic of the offense (criterion 19). Unexpected complications (criterion 7) and self deprecations (criterion 17) were initially scored but were never present. They were therefore disregarded, leaving a total of 14 CBCA criteria to be assessed.
2. The number of details in the Reality Monitoring scoring differed from the number of details in the CBCA scoring because different definitions are used in both coding systems. For example, "The young nurse . . ." results in two details in CBCA scoring and in one detail in Reality Monitoring scoring.
3. A MANOVA was conducted examining differences between liars and truth tellers with regard to the 14 individual CBCA criteria used in this study. The MANOVA, which was performed on the original, not dichotomized data, showed a significant main effect, $F(14, 58) = 4.74$, $p < .01$. Univariate tests revealed that compared to liars, truth tellers included more details ($M = 21.25$ ($SD = 8.5$) vs $M = 11.30$ ($SD = 4.1$), $F(1, 71) = 42.59$, $p < .01$), more contextual embeddings ($M = .18$ ($SD = .2$) vs $M = .09$ ($SD = .1$), $F(1, 71) = 3.92$, $p < .05$), more reproductions of conversations ($M = .20$ ($SD = .2$) vs $M = .09$, ($SD = .1$), $F(1, 71) = 4.75$, $p < .05$), more unusual details ($M = .07$ ($SD = .1$) vs $M = .00$ ($SD = .0$), $F(1, 71) = 9.02$, $p < .01$, more accounts of other's mental state ($M = .49$ ($SD = .3$) vs $M = .23$ ($SD = .3$), $F(1, 71) = 15.83$,

$p < .01$, and more spontaneous corrections ($M = .23$ ($SD = .2$) vs $M = .11$ ($SD = .2$), $F(1, 71) = 5.74$, $p < .05$).

4. It is important to note that it is the combination of nonverbal behaviors that is powerful, not any one individual behavior. Illustrators obtained a high individual hit rate (69.86%, eigenvalue .15, Wilks' lambda = .87, $\chi 2(1, n = 73) = 10.03$, $p < .01$). However, a distinction between truths and lies resulted in a high hit rate for lies (89.7%) but a particularly low hit rate for truths (47.1%).
5. We assume that our findings are caused by cognitive load. However, as we did not experimentally manipulate cognitive load in this study, we cannot say this for certain.

References

Akehurst, L., Köhnken, G., Vrij, A., & Bull, R. (1996). Lay persons' and police officers' beliefs regarding deceptive behavior. *Applied Cognitive Psychology, 10,* 461–473.

Akehurst, L., Köhnken, G., & Höfer, E. (1995). *The analysis and application of Statement Validity Assessment.* Paper presented at the Fifth European Conference on Psychology and Law, Budapest, Hungary.

Akehurst, L., & Vrij, A. (1999). Creating suspects in police interviews. *Journal of Applied Social Psychology, 29,* 192–210.

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

Alonso-Quecuty, M. L. (1996). Detecting fact from fallacy in child and adult witness accounts. In G. Davies, S. Lloyd-Bostock, M. McMurran, & C. Wilson (Eds.), *Psychology, law, and criminal justice: International developments in research and practice* (pp. 74–80). Berlin: de Gruyter.

Anson, D. A., Golding, S. L., & Gully, K. J. (1993). Child sexual abuse allegations: Reliability of Criteria-Based Content Analysis. *Law and Human Behavior, 17,* 331–341.

Burgoon, J. K., Kelly, D. L., Newton, D. A., & Keely-Dyreson, M. P. (1989). The nature of arousal and nonverbal indices. *Human Communication Research, 16,* 217–255.

Craig, R. A. (1995). *Effects of interviewer behavior on children's statements of sexual abuse.* Unpublished manuscript.

DePaulo, B. M., Anderson, D. E., & Cooper, H. (1999, October). *Explicit and implicit deception detection.* Paper presented at the annual conference of the Society of Experimental Social Psychologists, St. Louis, Missouri.

DePaulo, B. M., Kirkendol, S. E., Tang, J., & O'Brien, T. P. (1988). The motivational impairment effect in the communication of deception: Replications and extensions. *Journal of Nonverbal Behavior, 12,* 177–202.

DePaulo, B. M., Lanier, K., & Davis, T. (1983). Detecting deceit of the motivated liar. *Journal of Personality and Social Psychology, 45,* 1096–1103.

DePaulo, B. M., LeMay, C. S., & Epstein, J. A. (1991). Effects of importance of success and expectations for success on effectiveness at deceiving. *Personality and Social Psychology Bulletin, 17,* 14–24.

DePaulo, B. M., & Pfeifer, R. L. (1986). On-the-job experience and skill at detecting deception. *Journal of Applied Social Psychology, 16,* 249–267.

DePaulo, B. M., Stone, J. L., & Lassiter, G. D. (1985a). Deceiving and detecting deceit. In B. R. Schenkler (Ed.), *The self and social life* (pp. 323–370). New York: McGraw-Hill.

DePaulo, B. M., Stone, J. I., & Lassiter, G. D. (1985b). Telling ingratiating lies: Effects of target sex and target attractiveness on verbal and nonverbal deceptive success. *Journal of Personality and Social Psychology, 48,* 1191–2103.

deTurck, M. A. (1991). Training observers to detect spontaneous deception: Effects of gender. *Communication Reports, 4,* 81–89.

Ekman, P. (1989). Why lies fail and what behaviors betray a lie. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 71–82). Dordrecht, the Netherlands: Kluwer.

Ekman, P. (1992). *Telling lies: Clues to deceit in the marketplace, politics and marriage.* New York: W. W. Norton.

Ekman, P., & Frank, M. G. (1993). Lies that fail. In M. Lewis & C. Saarni (Eds.), *Lying and deception in everyday life* (pp. 184–201). New York: Guildford Press.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, Origins, usage, and coding. *Semiotica, 1,* 49–97.

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

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

Ekman, P., O'Sullivan, M., & Frank, M. G. (1999). A few can catch a liar. *Psychological Science, 10,* 263–266.

Ekman, P., O'Sullivan, M., Friesen, W. V., & Scherer, K. (1991). Face, voice, and body in detecting deceit. *Journal of Nonverbal Behavior, 15,* 125–135.

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

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech.* New York: Doubleday.

Gudjonsson, G. H. (1992). *The psychology of interrogations, confessions and testimony.* Chichester, England: Wiley.

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

Hershkowitz, I., Lamb, M. E., Sternberg, K. J., & Esplin, P. W. (1997). The relationships among interviewer utterance type, CBCA scores and the richness of children's responses. *Legal and Criminological Psychology, 2,* 169–176.

Höfer, E., & Akehurst, L., & Metzger, G. (1996, August). *Reality monitoring: A chance for further development of CBCA?* Paper presented at the Annual meeting of the European Association on Psychology and Law, Siena, Italy.

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

Horowitz, S. W., Lamb, M. E., Esplin, P. W., & Boychuk, T. D., Krispin, O., & Reiter-Lavery, L. (1997). Reliability of criteria-based content analysis of child witness statements. *Legal and Criminological Psychology, 2,* 11–21.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3–28.

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

Johnson, M. K., & Raye, C. L. (1998). False memories and confabulation. *Trends in Cognitive Sciences, 2,* 137–145.

Köhnken, G. (1987). Training police officers to detect deceptive eyewitness statements. Does it work? *Social Behavior, 2,* 1–17.

Köhnken, G. (1989). Behavioral correlates of statement credibility: Theories, paradigms and results. In H. Wegener, F. Lösel, & J. Haisch (Eds.), *Criminal behavior and the justice system: psychological perspectives* (pp. 271–289). New York: Springer-Verlag.

Köhnken, G. (1990). *Glaubwürdigkeit: Untersuchungen zu einem psychologischen konstrukt.* München, Germany: Psychologie Verlags Union.

Köhnken, G. (1996). Social psychology and the law. In G. R. Semin, & K. Fiedler (Eds.), *Applied Social Psychology* (pp. 257–282). London: Sage Publications.

Köhnken, G., Schimossek, E., Aschermann, E., & Höfer, E. (1995). The cognitive interview and the assessment of the credibility of adult's statements. *Journal of Applied Psychology, 80,* 671–684.

Kraut, R. E. (1980). Humans as lie detectors: Some second thoughts. *Journal of Communication, 30,* 209–216.

Lamb, M. E., Sternberg, K. J., Esplin, P. W. Hershkowitz, I., Orbach, Y., & Hovav, M. (1997a). Criterion-based content analysis: A field validation study. *Child Abuse and Neglect, 21,* 255–264.

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

Landry, K., & 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,* 663–675.

Lane, J. D., & DePaulo, B. M. (1999). Completing Coyne's cycle: Dysphorics' ability to detect deception. *Journal of Research in Personality, 33,* 311–329.

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

Lykken, D. T. (1960). The validity of the guilty knowledge technique: The effects of faking. *Journal of Applied Psychology, 44,* 258–262.

Lykken, D. T. (1998). *A tremor in the blood: Uses and abuses of the lie detector.* New York: Plenum Trade.

Mann, S., Vrij, A., & Bull, R. (1998, September). *Telling and detecting true lies.* Paper presented at the eighth annual meeting of the European Association on Psychology and Law, Cracow, Poland.

Manzanero, A. L., & Diges, M. (1996). Effects of preparation on internal and external memories. In G. Davies, S. Lloyd-Bostock, M. McMurran, & C. Wilson (Eds.), *Psychology, law, and criminal justice: International developments in research and practice* (pp. 56–63). Berlin: de Gruyter.

McCornack, S. A. (1997). The generation of deceptive messages: Laying the groundwork for a viable theory of interpersonal deception. In J. O. Greene (Ed.), *Message production: Advances in communication theory.* Mahway, NJ: Lawrence Erlbaum Associates.

Porter, S., & Yuille, J. C. (1996). The language of deceit: An investigation of the verbal clues to deception in the interrogation context. *Law and Human Behavior, 20,* 443–459.

Raskin, D. C. (1979). Orienting and defensive reflexes in the detection of deception. In H. D. Kimmel, E. H. Van Olst, & J. F. Orlebeke (Ed.), *The orienting reflex in humans* (pp. 587–605). Hillsdale, NJ: Erlbaum.

Raskin, D. C. (1982). The scientific basis of polygraph techniques and their uses in the judicial process. In A. Trankell (Ed.), *Reconstructing the past* (pp. 317–371). Stockholm: Norsted & Soners.

Raskin, D. C. (1986). The polygraph in 1986: Scientific, professional, and legal issues surrounding acceptance of polygraph evidence. *Utah Law Review, 29,* 29–74.

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

Reid, J. E. (1947). A revised questioning technique in lie detection tests. *Journal of Criminal Law, Criminology, and Police Science, 37,* 542–547.

Roberts, K. P., Lamb, M. E., Zale, J. L., & Randall, D. W. (1998). *Qualitative differences in children's accounts of confirmed and unconfirmed incidents of sexual abuse.* Paper presented at the biennial meeting of the American Psychology-Law Society, Redondo Beach, LA.

Roediger, H. L. (1996). Memory illusions. *Journal of Memory and Language, 35,* 76–100.

Ruby, C. L., & Brigham, J. C. (1997). The usefulness of the criteria-based content analysis technique in distinguishing between truthful and fabricated allegations. *Psychology, Public Policy, and Law, 3,* 705–737.

Ruby, C. L., & Brigham, J. C. (1998). Can Criteria-Based Content Analysis distinguish between true and false statements of African-American speakers? *Law and Human Behavior, 22,* 369–388.

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.

Steller, M. (1989). Recent developments in statement analysis. In J. C. Yuille (Ed.), *Credibility Assessment* (pp. 135–154). Deventer, the Netherlands: Kluwer.

Steller, M., & Köhnken, G. (1989). Criteria-Based Content Analysis. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217– 245). New York: Springer-Verlag.

Undeutsch, U. (1967). Beurteilung der Glaubhaftigkeit von Aussagen. In U. Undeutsch (Ed.), *Handbuch der Psychologie Vol. 11: Forensische Psychologie* (pp. 26–181). Göttingen, Germany: Hogrefe.

Undeutsch, U. (1989). The development of statement reality analysis. In J. C. Yuille (Ed.), *Credibility Assessment* (pp. 101–121). Dordrecht, The Netherlands: Kluwer.

Vrij, A. (1991). *Misverstanden tussen politie en allochtonen: Sociaal-psychologische aspecten van verdacht zijn.* Amsterdam: VU Uitgeverij.

Vrij, A. (1993). Credibility judgments of detectives: The impact of nonverbal behavior, social skills and physical characteristics on impression formation. *Journal of Social Psychology, 133,* 601–611.

Vrij, A. (1994). The impact of information and setting on detection of deception by police detectives. *Journal of Nonverbal Behavior, 18,* 117–137.

Vrij, A. (1995). Behavioral correlates of deception in a simulated police interview. *Journal of Psychology: Interdisciplinary and Applied, 129,* 15–29.

Vrij, A. (1998). Nonverbal communication and credibility. In A. Memon, A. Vrij, & R. Bull, *Psychology and law: Truthfulness, accuracy and credibility* (pp. 32–59). Maidenhead, Great Britain: McGraw-Hill.

Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and implications for professional practice.* Chichester: John Wiley & Sons.

Vrij, A., & Akehurst, L. (1998). Verbal communication and credibility: Statement Validity Assessment. In A. Memon, A. Vrij, & R. Bull, *Psychology and law: Truthfulness, accuracy and credibility* (pp. 3–31). Maidenhead, Great Britain: McGraw-Hill.

Vrij, A., Akehurst, L., & Morris, P. M. (1997). Individual differences in hand movements during deception. *Journal of Nonverbal Behavior, 21,* 87–102.

Vrij, A., & Baxter, M. (1999). Accuracy and confidence in detecting truths and lies in elaborations and denials: Truth bias, lie bias and individual differences. *Expert Evidence, 7,* 25–36.

Vrij, A., Harden, F., Terry, J., Edward, K., & Bull, R. (in press). The influence of personal characteristics, stakes and lie complexity on the accuracy and confidence to detect deceit. In R. Roesch, R. R. Corrado, & R. J. Dempster (Eds.), *Psychology in the courts: International advances in knowledge.* Amsterdam: Harwood Academic.

Vrij, A., Kneller, W., & Mann, S. (2000). The effect of informing liars about Criteria-Based Content Analysis on their ability to deceive CBCA-raters. *Legal and Criminological Psychology, 5,* 57–70.

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

Vrij, A., & Mann, S. (in press, b). Who killed my relative? Police officers' ability to detect real-life high stake lies. *Psychology, Crime, & Law.*

Vrij, A., & Semin, G. R. (1996). Lie experts' beliefs about nonverbal indicators of deception. *Journal of Nonverbal Behavior, 20,* 1, 65–81.

Vrij, A., Semin, G. R., & Bull, R. (1996). Insight in behavior displayed during deception. *Human Communication Research, 22,* 544–562.

Winkel, F. W., & Vrij, A. (1995). Verklaringen van kinderen in interviews: Een experimenteel onderzoek naar de diagnostische waarde van Criteria Based Content Analysis. *Tijdschrift voor Ontwikkelingspsychologie, 22,* 61–74.

Zaparniuk, J., Yuille, J. C., & Taylor, S. (1995). Assessing the credibility of true and false statements. *International Journal of Law and Psychiatry, 18,* 343–352.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in experimental social psychology, volume 14* (pp. 1–57). New York: Academic Press.