# Detecting Deceit Via Analyses of Verbal and Nonverbal Behavior in Children and Adults

Aldert Vrij

Lucy Akehurst

Stavroula Soukara

Ray Bull

University of Portsmouth

Aldert Vrij

University of Portsmouth, Psychology Department

King Henry Building, King Henry 1 Street

Portsmouth PO1 2DY

United Kingdom

Lucy Akehurst

University of Portsmouth, Psychology Department

King Henry Building, King Henry 1 Street

Portsmouth PO1 2DY

United Kingdom

Stavroula Soukara

University of Portsmouth, Psychology Department

King Henry Building, King Henry 1 Street

Portsmouth PO1 2DY

United Kingdom

Ray Bull

University of Portsmouth, Psychology Department

King Henry Building, King Henry 1 Street

Portsmouth PO1 2DY

United Kingdom

Abstract

This experiment examined children's and undergraduates' verbal and nonverbal deceptive behavior, and the extent to which their truths and lies could be correctly classified by paying attention to these responses. A total of 196 participants (aged 5-6, 10-11, 14-15, and undergraduates) participated in an erasing the blackboard event, and told the truth or lied about the event afterwards. Nonverbal and verbal responses were coded, the latter with Criteria-Based Content Analysis and Reality Monitoring. Although children and undergraduates demonstrated different behaviors (for example, the children obtained lower CBCA scores and made more movements), actual cues to deceit were remarkably similar across different age groups (for example, both 5-6-year-olds and undergraduates obtained lower CBCA scores and made fewer movements while lying). A combination of verbal and nonverbal lie detection methods resulted in more correct classifications of liars and truth tellers than the verbal and nonverbal lie detection methods individually, with the combined method obtaining hit rates as high as 88%.

# Detecting Deceit Via Analyses of Verbal and Nonverbal Behavior in Children and Adults

Although it sounds reasonable to suggest that looking at a combination of nonverbal and verbal behaviors ("speech content") will lead to more accurate classifications of liars and truth tellers than investigating nonverbal and verbal behaviors separately, researchers rarely investigate both types simultaneously. Vrij, Edward, Roberts, and Bull (2000) examined (adult) participants' nonverbal and verbal behavior (the latter investigated with Criteria-Based Content Analysis and Reality Monitoring) and obtained the most accurate classifications of liars and truth tellers when both the nonverbal and verbal behaviors were taken into account. Similarly, in the present study, both nonverbal and verbal responses of liars and truth tellers were investigated. However, the present study differs from Vrij et al.'s (2000) study in several ways. Perhaps the most important difference was that in the present study apart from adults (undergraduates) children also participated, and this seems the first study in which truthful and deceptive verbal and nonverbal behaviors of both children and adults were investigated and compared.

Research examining children's nonverbal behavior relating to deceit is rare, and less than a handful of studies have been published to date (see Vrij, 2002, for a review). Child deception research has mainly focused on questions such as whether children lie, why they lie and when they lie (Frank, 1992). At the same time, examining children's behavior while lying is important for both theoretical and practical reasons, the latter due at least in part to the fact that police officers, social workers, judges and juries are sometimes confronted with the question of whether a child is lying or telling the truth.

## Theoretical Reasons to Explain Deceptive Nonverbal Behavior

Three perspectives (the emotional, cognitive complexity and attempted behavioral control perspectives) seem particularly important for predicting and explaining nonverbal deceptive behavior (DePaulo, Stone, & Lassiter, 1985; Vrij, 2000; Zuckerman, DePaulo, & Rosenthal, 1981). Research has revealed that there is no typical deceptive behavior (DePaulo,

Lindsay, Malone, Muhlenbruck, Charlton, & Cooper, in press; Vrij, 2000). Some behaviors, however, are more likely than others to occur during deception, depending on emotions experienced by the liar (Ekman, 1992), cognitive load experienced by the liar (Burgoon, Buller, & Guerrero, 1995; Cody, Marston, & Foster, 1984; Ekman, 1997, Ekman & Friesen, 1972), and the amount of effort liars exert in controlling their behavior (Buller & Burgoon, 1996; Burgoon & Buller, 1994; Burgoon, Buller, White, Afifi, & Buslig, 1999; DePaulo & Kirkendol, 1989; Hocking & Leathers, 1980; Krauss, 1981).

All three processes may occur simultaneously. That is, liars could be nervous, having to think hard, and trying to control themselves all at the same time. Which of these processes is most prevalent should depend on the type of lie. That is, liars will be more nervous when the stakes (negative consequences of getting caught and positive consequences of succeeding) are high, hence, nervous behaviors are more likely to occur in high-stakes lies (Ekman, 1992; Frank & Ekman, 1997). Liars have to think harder when the lie is complicated, therefore indicators of cognitive load are more likely to occur in complicated lies than in easy lies (McCornack, 1997). Liars who are motivated not to get caught may try harder to make an honest impression than those who are less motivated, therefore, attempts to control behavior may especially occur in motivated liars (Burgoon & Floyd, 2000; DePaulo & Kirkendol, 1989).

The best liars are probably those who manage to suppress signs of nervousness and cognitive load and who manage to exhibit, even under difficult circumstances, behavior that looks natural. In order to display natural behavior, three factors are important (Vrij, 2000). First, liars should realize that observers watch their behavior to detect deceit. Second, liars should know which behaviors make an honest impression on others, and, third, liars should be able to control their behavior. The first two issues imply that the effective liar should be able to "take the role of the other", an ability which is largely lacking in young children (Flavell, Botkin, Fry, Wright, & Jarvis, 1968). This might suggest that more cues of nervousness and more cues of cognitive load can be expected in younger children when lying

than in older children and adults, because younger children will not try so hard to suppress these cues.

Children's muscular control ability increases with age as well (Ekman, Roper, & Hager, 1980; Feldman & Phillipot, 1993). For example, Ekman et al. (1980) studied five-, nine- and thirteen-year-olds and found that older children have a greater ability to deliberately produce the component actions involved in facial expression. These findings might also suggest that with increased age, fewer cues of emotions or cognitive load are likely to occur.

DePaulo and Jordan (1982), however, have argued that younger children may experience less emotion when lying. For example, because of their young age they might be less affected by feelings of guilt, and might overlook the consequences of getting caught more often and so experience less fear of getting caught. Also, with increasing age, children show more spontaneous facial emotional expressions, which they need to suppress in order to conceal deceit (Morency & Krauss, 1982). Therefore, older children's and adults' role-taking skills, and increased muscular control that improves their skills in deception may well be counteracted in part by an increase in emotions and an increase in emotional expression while lying.

In the present experiment, participants (5-6-year-olds, 10-11-year-olds, 14-15-year-olds and undergraduates) lied or told the truth about playing a game of Connect 4 (a popular game in the UK for all ages) and wiping a blackboard. Truth tellers were asked to recall truthfully their activities whereas liars were requested to fabricate a story. Fabricating such a lie probably requires some thinking, and liars, compared to truth tellers, are therefore more likely to show behaviors that indicate cognitive load. In order to induce emotions, participants were promised a reward when they could convince the interviewer that they were telling the truth and were told that they could expect a punishment if they failed to convince the interviewer. This might result in signs of nervousness, such as gaze aversion and fidgeting, perhaps most likely in the youngest participants, although this is in no way certain

given our comments above.

Previous research has indicated that several behaviors are associated with cognitive load (Burgoon, Kelly, Newton, & Keeley-Dyreson, 1989; Ekman & Friesen, 1972; Goldman-Eisler, 1968; Köhnken, 1989; Vrij, 2000), and on the basis of these findings we hypothesized that:

H1: Liars (both adults and children) will show fewer movements, slower speech rate, more pauses, more speech disturbances, and a longer latency time than truth tellers.[1]

Apart from a cognitive load based Deception effect, an Age Group effect might occur. Children's nonverbal presentation style differs from the style typically exhibited by adults. For example, children show more limb movements than adults (Eaton, McKeen, & Campbell, 2001). It is also possible that children will find the interview task more difficult than adults, and therefore some age differences in behaviors which indicate cognitive load might emerge. It was therefore hypothesized that:

H2: Children will show more limb movements, a longer latency time, slower speech rate, more pauses, and more speech disturbances than adults.

## Verbal Behaviors Relating to Deception

Children below the age of five do not yet engage in complicated verbal deception (Newton, Reddy, & Bull, 2000). Their lies typically take the form of one-word responses (Bussey, 1992), not providing information (Peskin, 1992), or pointing in the wrong direction (Russell, Mauthner, Sharpe, & Tidswell, 1991; Sodian, 1991). When they get older, however, children's lies become more verbally sophisticated.

Differences between liars and truth tellers in what they say can be assessed with Statement Validity Analysis (SVA), and such assessments form part of experts' reports to criminal courts in several countries, including the United States (Honts, 1994; Ruby & Brigham, 1997), Sweden (Gumpert & Lindblad, 2001) and Germany (Köhnken, 2002). The core of SVA is Criteria-Based Content Analysis (CBCA), a systematic assessment of the credibility of written statements. Steller and Köhnken (1989) compiled a list of 19 criteria

used in the assessment.[2] CBCA is based on the hypothesis, originally advanced by Undeutsch (1967), that a statement derived from memory of an actual experience differs in content and quality from a statement based on invention or fantasy (the "Undeutsch Hypothesis", Steller, 1989). The presence of each criterion strengthens the probability that the account is based on genuine personal experience. In other words, truthful statements will be characterized by more of the elements measured by CBCA than deceptive statements. Köhnken (1989, 1996, 1999) presented a theoretical foundation for the Undeutsch Hypothesis and proposed that both cognitive and motivational factors influence CBCA scores.

With regard to cognitive factors, it is assumed that, compared to those who fabricate a story, those who actually experience an event will be able to produce descriptions about those events which include more CBCA criteria, as some criteria (unstructured production, contextual embedding, reproduction of speech, unusual details, etc.) are believed to be difficult for people to fabricate. For example, regarding unstructured production, truth tellers sometimes tend to give their account in unstructured and incoherent ways, particularly when they talk about emotional events (Boychuk, 1991; Lamb, Sternberg, Esplin, Hershkowitz, Orbach, & Hovav, 1997; Winkel, Vrij, Koppelaar, & van der Steen, 1991). On the other hand, liars tend to tell their stories in a more chronological order than truth tellers (Zaparniuk, Yuille, & Taylor, 1995), as it is often too difficult for them to tell a fabrication in a different order (Köhnken, 1999; Steller, 1989).

Other message features indexed by the CBCA criteria are more likely to occur in truthful statements for motivational reasons. Truthful persons tend not to be as concerned with impression management as deceivers. Compared to truth tellers, deceivers are more keen to try to construct a report which they believe will make a credible impression on others, and will leave out information which, in their view, will damage their image of being a sincere person (Köhnken, 1999). As a result, a truthful person's statement is more likely to contain information that is inconsistent with the stereotypes of truthfulness. The CBCA list includes several so-called "contrary-to stereotype" criteria (Ruby & Brigham, 1998):

"spontaneous corrections," "admitting lack of memory," "raising doubts about one's own testimony," etc. For example, liars who claimed to have seen non-existing persons near the scene of the crime, might be reluctant to include spontaneous corrections in their stories ("He wore black trousers, no sorry, they were green") as they might be afraid that this will make their stories sound less convincing.

CBCA/SVA was developed (in Germany) in order to evaluate the statements from children who are witnesses or alleged victims, most commonly of sexual abuse. Many authors have described CBCA as a technique solely developed to evaluate statements made by children in sexual offense trials (e.g., 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). The latter group of authors have pointed out that the underlying Undeutsch Hypothesis is restricted neither to children, witnesses and victims, nor to sexual abuse. This point of view has received some empirical support to date. Significantly higher CBCA scores for truth tellers than for liars have not only been found in studies with child witnesses (e.g., Akehurst, Köhnken, & Höfer, 2001; Lamb, Sternberg, Esplin, Hershkowitz, & Orbach, 1997; Lamb et al., 1997a; Lamers-Winkelman & Buffing, 1996; Tye, Amoto, Honts, Kevitt, & Peters, 1999; Winkel & Vrij, 1995) but also in studies with adult witnesses (e.g., Akehurst et al., 2001; Höfer, Akehurst, & Metzger, 1996; Köhnken et al., 1995; Landry & Brigham, 1992; Porter & Yuille, 1996; Ruby & Brigham, 1998; Sporer, 1997; Vrij, Edward, & Bull, 2001b).[3]

Therefore, in the present study we predicted that:

H3: CBCA scores would be significantly higher for truth tellers than for liars in both adult statements and child statements.

Cognitive abilities and command of language develop throughout childhood, making it gradually easier to provide detailed accounts of what has been witnessed (Davies, 1994).

Also, children are probably less aware or are less concerned with impression management than adults (Flavell et al., 1968; Vrij, 2002) and may not fully realize yet that brief accounts might look suspicious. It was therefore predicted that:

H4: Older participants would have higher CBCA scores than younger participants.

Recently, Reality Monitoring has been used as an alternative method to examine verbal differences between responses believed to be true and false (Alonso-Quecuty, 1992, 1996; Alonso-Quecuty, Hernandez-Fernaud, & Campos, 1997; Höfer et al., 1996; Manzanero & Diges, 1996; Roberts, Lamb, Zale, & Randall, 1998; Sporer, 1997; Vrij et al., 2000; Vrij et al., 2001b). The core of Reality Monitoring is the claim that memories of experienced events differ in quality from memories of imagined (e.g., fabricated) events. Memories of real experiences are obtained through perceptual processes and are therefore likely to contain, amongst others, *perceptual information*: details of smell, taste or touch, visual details and auditory details (details of sound) and *contextual information*: spatial details (details about where the event took place, and details about how objects and people were situated in relation to each other, e.g., "He stood behind me"), and temporal details (details about time order of the events, e.g., "First he switched on the video-recorder and then the TV", and details about duration of events). Accounts of imagined events are derived from an internal source and are therefore likely to contain *cognitive operations*, such as thoughts and reasonings ("I must have had my coat on, as it was very cold that night") (Johnson, Hashtroudi, & Lindsay, 1993; Johnson & Raye, 1981, 1998). One might argue that "experienced events" reflect truth telling whereas "imagined events" reflect deception. Therefore, differences between truth tellers and liars could be expected regarding Reality Monitoring criteria. It was hypothesized that:

H5: Truth tellers are likely to include more perceptual and contextual information in their statements than liars.

H6: Liars are likely to include more cognitive operations in their statements than truth tellers.

Regarding Hypothesis 6, previous studies often have failed to find the expected

difference between liars and truth tellers regarding cognitive operations (Alonso-Quecuty, 1992, 1996; Höfer et al., 1996; Vrij et al., 2000, 2001b). This might have been caused by the nature of the event. For example, in Vrij et al.'s (2000, 2001b) studies, participants were requested to give truthful or deceptive factual accounts of aspects of a film they had seen. This allows little room for cognitive operations. Truth tellers were asked to recall and liars were asked to fabricate what people in the film were doing, not what they, themselves, were thinking. Cognitive operations, however, are related to people's own thinking rather than recall of factual information about others. In the present study, participants were asked to describe their own activities during a certain period of time which provides more opportunity to include reports of cognitive operations.

For the same reason that we expected age differences in CBCA scores (i.e., cognitive abilities and command of language develop throughout childhood which makes it gradually easier to provide detailed accounts), we expected age differences in Reality Monitoring scores as well:

H7: The older participants would have a higher RM score than the younger participants.

Finally, the study was designed to ascertain whether a combination of the two verbal indices (i.e., CBCA and RM) and the nonverbal cues would classify liars and truth tellers more accurately than any of the techniques separately. A combination of techniques takes more information into account (both nonverbal and verbal) than the separate techniques, and, the more characteristics of lies that are scrutinized, the more likely it is that they will be detected. It was therefore hypothesized that:

H8: A combination of the two verbal techniques (CBCA and RM) and the nonverbal technique would result in superior hit rates for classifying liars and truth tellers than any of the three approaches alone.

## Method

### Participants

A total of 196 participants took part in the study (55% males). Their mean age was

M=14.16 years, SD=6.28. There were four different age groups: 35 5-6-year-olds (16 males and 19 females, all but one were 6 years old), 54 10-11-year-olds (22 males and 32 females, all but one were 11 years old), 55 14-15-year-olds (32 males and 23 females, 7 14-year-olds and 48 15-year-olds), and 52 undergraduate students (38 males and 14 females, M=22.37 years, SD = 5.1).

## Procedure

The experiment took place at a Students' Union (for undergraduates) and at two local schools (for children). Undergraduates were recruited under the guise of participating in an experiment about "telling a convincing story" with the possibility of earning £5 (approximately $7.5). Children were asked by their teacher to go to see "a woman to play a game." Participants participated individually. After entering the experimental room, the female experimenter made a fairly brief, polite conversation (exchanged names, What have you been doing?, Do you know how to play Connect 4?, etc.). From that moment events differed for the participants depending on which of the two conditions (lying or truth telling) they were in. Random assignment was made to these conditions.

Participants in the truthful condition (N = 102) played a game of Connect 4 with the experimenter (all 196 participants in this study knew how to play this game). During the game, a confederate (we used different persons, see below) entered the room, said "Hello", and "Excuse me for interrupting", and also said "Ah! You are playing Connect 4, I'm hopeless at that game, I always lose!", walked to the blackboard and erased some information (complicated math formulas) off the blackboard, and then left the room. The experimenter remained immersed in the game. However, after the game she looked at the blackboard and noticed that the information had been erased. She then asked the participant whether he or she saw who erased the information. All participants in this condition told the experimenter that the person who came in during the game erased the blackboard. Then, the experimenter gave the following instructions: Actually, I know that the information should not have been wiped off the blackboard, as it is needed for a lesson later on. In a minute you will be

interviewed by another woman. Her task is to find out who wiped the information off the blackboard. Now, you know you did see who it was, so your task is to convince her that you did. All you need to do is be truthful about everything that happened while you were in this room. So say that we played a game of Connect 4 and that someone came in to wipe the blackboard. This is very important as, if you are successful in convincing her that you are telling the truth, we will give you (undergraduates - £5; 14-15-year-olds - £2; 10-11-year-olds - £2; 5-6-year-olds - present). At the end of the interview, she will tell you whether she believes you or not. If she does believe that we played Connect 4 and that someone came in to wipe the blackboard, we will give you the money/present when you come out. If she doesn't believe you, you will not get any money/present at all and that you might have to write a statement about everything that happened.[4]

Participants in the deception condition (N = 94) were told by the experimenter that she (the experimenter) had earlier wiped some important information off the blackboard which was supposed to stay there for a lesson later on. The participants were told that they would be interviewed by another woman whose task was to find out who wiped the information off the blackboard. The experimenter then asked the participants to pretend that it wasn't the experimenter who wiped the information off the blackboard. Instead the participants were asked to pretend that they (experimenter and participant) played Connect 4 and that someone else entered the room and erased the information during the game. Identical to the truthful condition, it was stressed that if the participant was successful in convincing the interviewer that it was someone else (other than the experimenter or participant) who erased the information from the blackboard, he or she would receive £5/£2/present, and that he or she would receive nothing at all and might have to write a statement about everything that happened if the interviewer did not believe the participant. Again, participants were informed that the interviewer would tell the participants at the end of the interview whether or not she believed them.[5]

The participants were then given the opportunity to prepare themselves for the

interview.[6] After the participants indicated that they were ready for the interview, they were shown into the interview room. The (female) interviewer was unaware of the experimental condition (truthful or deception) for each participant. After building rapport with each participant, the interviewer gave the following instructions: The reason I am interviewing you is that I have heard that someone wiped some important information off the blackboard in the room you were just in. I understand that you were in there just now with (name of experimenter). I need to know if you saw anything that will help me to find out who it was. Also, remember that my colleagues need to know whether I think you are telling the truth or not.[7] If you do not know the answer to one of my questions please do not be afraid to say you don't know. If you don't understand one of my questions please let me know and I will try to explain what I mean. If I ask you the same question twice do not worry. It is not that you got it wrong the first time. First of all, I would like you to tell me, in as much detail as possible, everything you can remember about what happened when you were in that room just now. Give me as much information as you can, even small details you do not think are very important.

After the free recall, three further questions were asked. (However, a question was not asked if the topic of the question had not been mentioned in the free report): (1) "So you played a game? Can you tell me again, in lots of detail exactly what happened when you played the game?, (2) And you said you were interrupted by someone coming in? Can you tell me again, as much as you can remember about what the person looked like?, (3) And lastly, can you tell me, in lots of detail, exactly what the person who interrupted did?"

After this interview the interviewer told each participant that they had convinced her (the interviewer) that they told the truth. For ethical reasons the interviewer was instructed to tell all children that she believed them, regardless of how convincing their stories were. With regard to the undergraduates, she was instructed not to give money to any who gave very short statements. However, all undergraduates put effort in giving extensive and credible statements (as did all other participants), and each was paid £5. The average length of the

deceptive and truthful interviews were M = 106.77 seconds (SD = 44.1) and M = 116.07 seconds (SD = 47.5) respectively. The difference in length between the truthful and deceptive interviews was not significant, F(1, 194) = 2.01, ns.[8]

## CBCA Scoring

Two raters received training in CBCA scoring. Both raters first read several major published papers about CBCA (Raskin & Esplin, 1991; Steller, 1989; Steller & Köhnken, 1989; Vrij, 2000; Vrij & Akehurst, 1998). They were then trained in CBCA scoring by a British CBCA expert. The expert explained each criterion under investigation in this study (see below) and gave examples of each criterion. Third, both the trainee raters and the expert evaluated one example transcript individually (from a different study). These three raters compared their results and feedback was given by the expert rater. Fourth, the trainees received more transcripts and were asked to rate these transcripts at home. In a follow up meeting, the results were evaluated and, again, feedback was given by the expert. After that meeting the expert felt that the two raters had been adequately trained, and it was decided that they could commence their coding task for the present experiment. Coding was carried out individually by the two trained coders (they coded the statements at home) and involved written transcripts of the interviews. The raters were blind to the hypotheses under investigation, to the staged event, and to the experimental conditions (although they were aware that some scripts would be truthful and some not). Some criteria ("accurately reported details misunderstood," "pardoning the perpetrator" and "details characteristic of the offense" were not scored, as they are specifically related to (sexual) crimes. "Superfluous details," "related external associations" and "self deprecations" were to be scored but, in fact, were never present. They were therefore disregarded, leaving a total of 13 CBCA criteria to be assessed. CBCA scoring took place in two different ways. The coders scored the strength of presence of each criterion in each statement, 1 = absent, 5 = is strongly present (see also Akehurst et al., 2001; Köhnken, 1999; Köhnken et al., 1995). They also scored the frequency of occurrence of each criterion in each statement (although this was not possible for the

criteria "logical structure" and "unstructured production" which are general criteria). Pearson correlations were calculated between the two coders: logical structure, r(scale) = .49; unstructured production, r(scale) = .08; quantity of details, r(scale) = .56, r(frequency) = .90,; contextual embedding, r(scale) = .76, r(frequency) = .93; description of interactions, r(scale) = .55, r(frequency) = .58; reproduction of conversation, r(scale) = .52, r(frequency) = .62; unexpected complications, r(scale) = .30, r(frequency) = .51; unusual details, r(scale) = .05, r(frequency) = .17; subjective mental state, r(scale) = .68, r(frequency) = .61; attribution of other's mental state, r(scale) = .20, r(frequency) = .21; spontaneous corrections, r(scale) = .57, r(frequency) = .71; admitting lack of memory, r(scale) = .66, r(frequency) = .63; raising doubts about one's own memory, r(scale) = .68, r(frequency) = .78.[9]

The correlations indicate that "frequency scoring"' gave higher correlations between raters than "'scale scoring." This is intriguing (but not surprising, see endnote 9) since "scale scoring" is the common practice in CBCA assessments. Moreover, most correlations were lower than we have obtained in the past (Vrij, Edward, & Bull, 2001a, b; Vrij et al., 2000), although higher than some others have found (e.g., Anson, Golding, & Gully, 1993). There are several reasons why some of the correlations may have been low: (1) some statements (particularly of 5-6-year-olds) were rather short which made it difficult to score criteria such as "unstructured production," (2) some criteria ("unusual details" and "attribution of other person's mental state") rarely occurred (in less than 10% of the statements), and (3) the questions the interviewer asked often resulted in participants repeating some information they had already given which probably made the task more difficult for CBCA coders because common practice (Köhnken, 1999; Steller, 1989; Steller & Köhnken, 1989) requires that repetitions should not be counted. However, sometimes new information was embedded in a repetition, which should be scored, but could easily have been overlooked by (one of) the coders.

Nevertheless, the correlations reveal acceptable interrater reliability scores (r's > .49, Anson et al., 1993; Fleiss, 1981) for nine criteria: logical structure, quantity of details,

contextual embedding, description of interactions, reproduction of conversation, accounts of subjective mental state, spontaneous corrections, admitting lack of memory, and raising doubts about one's own memory. Following previous examples (Craig, Scheibe, Raskin, Kircher, & Dodd, 1999; Hershkowitz, Lamb, Sternberg, & Esplin, 1997; Lamb et al., 1997a, b; Parker & Brown, 2000; Tye et al., 1999; Vrij et al., 2000, 2001a, b), we calculated a total CBCA score on the basis of these nine criteria. The total CBCA score was the combined score of the nine criteria on the 5-point scales and therefore could range from 9 to 45. In order to check reliability for the total CBCA score, total scores for both coders were calculated. The correlation between these two CBCA scores was very high (.85). This result is consistent with previous studies where interrater agreement ratings on a total CBCA score are higher than the interrater agreement ratings on individual criteria (see Vrij & Akehurst, 1998, for a discussion of this issue). In the present analyses we used as total CBCA score the average score of the two coders.

RM Scoring

Two other raters received training in Reality Monitoring (RM) scoring. A British RM expert (another person than the CBCA expert) provided the raters with a detailed description of how the criteria should be scored, including some case examples. Then, both the trainee raters and the expert evaluated some example transcripts individually (from a different study). The three raters compared their results and feedback was given by the expert. At this stage the expert and the two raters felt that the raters were 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) RM scoring than CBCA scoring. With regard to the present study, coding was carried out individually by the two trained raters (they coded the statements at home) and involved the written transcripts of the interviews. The raters were blind to the hypotheses under investigation, to the staged event, and to the experimental condition (although they were aware that some scripts would be truthful and some would be not). The two raters scored per interview the frequency of occurrence of visual details (e.g.,

"I walked in to the room" contains three visual details), auditory details (e.g., "She said to sit down" contains one sound detail), temporal details (e.g., "We started playing" is one temporal detail), spatial details (e.g., "And then the pieces fell on to the floor" contains one spatial detail) and cognitive operations (e.g., "Because she was quite clever, she won the game" contains one cognitive operation; so does "I presume that the two people knew each other" and "She was quite tall for a girl"). Intercoder reliability scores (Pearson's correlations) were satisfactory for all the individual criteria (visual details: r=.80; auditory details: r=.92; spatial details: r=.61; temporal details: r=.78; cognitive operations: r=.54). The scores for each of the criteria were therefore based on the average scores of the two raters.

Following previous examples (Vrij et al., 2000, 2001b) a RM total score was calculated. In order to create the RM scale the visual, auditory, spatial and temporal variables were dichotomized (see also Vrij et al., 2000, 2001b).[10] Dichotomizations for auditory, temporal and spatial details were based on the 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. For dichotomization of visual details, a median split was used.[11] Those 5-6-year-olds (N = 18) with a score higher than 18.00 obtained 1 on this criterion, the other 17 5-6-year-olds obtained a score of 0. Those remaining participants (10-11-year-olds, 14-15-year-olds and undergraduates) with a score higher than 37.50 (N = 80) for visual details obtained a 1 on the visual details criterion, whereas the 81 remaining participants received a score of 0. Cognitive operations was not included in the total RM score as this criterion is conceptually different from the other criteria: Its presence does not indicate truth telling (as is the case with the other criteria), but lying. The RM scale therefore contained four criteria (visual details, auditory details, spatial details and temporal details) and the total-score could range from 0 to 4. Throughout this article, the total RM score refers to the RM score without cognitive operations.[12]

## Nonverbal Behavior Scoring

Two other observers coded the nonverbal behavior of the participants independently

by watching the videotaped interviews. The observers were not informed as to when the participants were lying and when they were telling the truth. Neither were they informed about the hypotheses or the staged event. They used a coding system previously employed in a number of studies (Akehurst & Vrij, 1999; Vrij, 1991, 1995; Vrij, Akehurst, & Morris, 1997; Vrij, Semin, & Bull, 1996; Vrij et al., 2000, 2001a, b). The following behaviors were coded[13] (speech hesitations and speech errors were scored on the basis of a typed verbatim text): gaze aversion (number of seconds for which the participant looked away from the interviewer, $r = .84$); illustrators (frequency of arm and hand movements which were designed to modify and supplement what was being said, Ekman & Friesen, 1969, $r = .93$); adaptors (frequency of scratching the head, wrists, etc. Rubbing one's hands together were not coded as adaptors but as hand and finger movements, $r = .73$); hand and finger movements (frequency of movements of the hands or fingers without moving the arms, $r = .81$); foot and leg movements (frequency of movements of feet or legs. Simultaneous movements of feet and legs were scored as one movement, $r = .86$); speech hesitations (frequency of saying 'ah' or 'mm' between words, $r = .99$); speech errors (frequency of word or sentence repetition, sentence change, sentence incompletion, and slips of the tongue, $r = .89$); latency period (period of time between the question being asked and the answer given, $r = .78$); speech rate (number of spoken words divided by the length of interview minus latency period).

The Pearson's correlations indicate considerable agreement between the two coders, and composite scores were taken as the average of the two coders. Following Eaton et al. (2001) different types of limb movements (illustrators, adaptors, hand and finger movements and foot and leg movements) were clustered into one category: limb movements. Two reasons were behind this clustering. First, it reduced the number of dependent variables and, second, we felt that distinguishing between different types of movement was less appropriate for the very young children (see also Eaton et al., 2001). Following Kasl and Mahl (1965) (and again to reduce the number of variables), speech hesitations and speech errors were

clustered in to one category: speech disturbances. If relevant, results for the original behaviors are given in the discussion of the results. 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 for gaze aversion, limb movements, frequency of pauses, latency period and speech rate were calculated on a per minute basis. Patterns for speech disturbances were calculated per 100 words.

## Results

### Verbal and Nonverbal Responses: Age Differences and Differences between Liars and Truth Tellers

In order to examine participants' verbal and nonverbal responses, we first looked at the total RM scores (which does not contain cognitive operations) and total CBCA scores. Two ANOVAs were conducted with Age Group and Veracity as between-subjects factors and the total CBCA and total RM scores as dependent variables. Significant main effects for Age Group and Veracity emerged in both ANOVAs (see Tables 1 and 2), whereas both Age Group X Veracity interaction effects were not significant, $F(3, 188) = .92$, ns (CBCA) and $F(3, 188) = .89$, ns (RM). Tukey HSD post hoc tests were conducted to further investigate the significant findings. Regarding the Age Group factor, the CBCA scores of the youngest age group were significantly lower than the CBCA scores of the remaining three groups, and the CBCA scores of the oldest age group were significantly higher than the CBCA scores of the remaining three groups (see Table 1). The CBCA scores of the two middle groups did not differ significantly from each other and were significantly higher than the CBCA score of the youngest age group and significantly lower than the CBCA score of the oldest age group. This provides support for Hypothesis 4.

With regard to the total RM score, the youngest age group obtained the lowest score and this score was significantly lower than the scores of each of the three remaining groups (see Table 1). The scores for these three groups did not differ from each other. This provides support for Hypothesis 7.

Regarding the Veracity factor (see Table 2), truth tellers obtained a higher total CBCA score and a higher total RM score than liars, which supports Hypotheses 3 and 5.

In order to examine the individual verbal and nonverbal responses, a MANOVA was carried out with Age Group and Veracity as between-subjects factors and the nine CBCA criteria, five RM criteria, and six nonverbal responses as dependent variables. At a multivariate level the analyses revealed a significant main effect for Age Group, Wilks' lambda = .17, $E(60, 505) = 6.72$, $p < .01$, $eta^2 = .44$, a significant main effect for Veracity, Wilks' lambda = .73, $E(20, 169) = 3.10$, $p < .01$, $eta^2 = .27$, and a significant Age Group X Veracity interaction effect, Wilks' lambda = .64, $E(60, 505) = 1.38$, $p < .05$, $eta^2 = .14$. Regarding the Age Group factor, at a univariate level almost all effects were significant. Tukey HSD post hoc tests were conducted to further investigate the significant findings (see also Table 1).

With regard to the verbal responses (CBCA and RM), the lowest scores were always obtained by the youngest age group and, except for visual details, the highest scores were always obtained by the oldest age group. With the exception of "admitting lack of memory," the score of the oldest age group was always significantly higher than the score for the youngest age group. The scores for the two middle age groups did not differ significantly from each other. They were typically between the scores of the two extreme age groups, and in some cases significantly higher than the youngest age group but significantly lower than the oldest age group. These findings reveal an age effect as was predicted in Hypotheses 4 and 7.

Regarding nonverbal responses, similar to the verbal responses, the most extreme scores were obtained by the youngest and oldest participants. The youngest participants showed most gaze aversion, although their score differed significantly only from the 10-11 year-olds. The youngest age group had the longest latency period, significantly longer than any of the other three groups, which is in agreement with Hypothesis 2. They also had the slowest speech, and their speech was significantly slower than the speech of the three other

groups, which also supports Hypothesis 2. The oldest participants made fewer limb movements than the other three groups (which supports Hypothesis 2)[14] and, contrary to what was predicted in Hypothesis 2, exhibited more speech disturbances than the other three groups.[15]

Table 2 shows the findings for the Veracity factor. For five of the nine CBCA criteria significant differences were found between truthful and deceptive accounts. All five differences were in the predicted direction (Hypothesis 3), with higher scores for truth tellers than for liars.

All effects with regard to the Reality Monitoring variables were significant, and all the differences were in the predicted direction. Compared to truth tellers, liars obtained lower scores for visual details, auditory details, spatial details and temporal details, which supports Hypothesis 5, and higher scores for cognitive operations, which supports Hypothesis 6.

Regarding the nonverbal responses, only one significant finding emerged: Liars exhibited fewer limb movements than truth tellers, which is in line with Hypothesis 1.[16]

At a univariate level, three Age Group X Veracity interaction effects were significant, and they were all Reality Monitoring variables; auditory details: $F(3, 188) = 3.13$, $p < .05$, $eta^2 = .05$; spatial details: $F(3, 188) = 3{,}23$, $p < .05$. $eta^2 = .05$; temporal details: $F(3, 188) = 3.10$, $p < .05$, $eta^2 = .05$. For each of the three variables, the pattern was the same. For each age group, fewer auditory, spatial and temporal details were present during deception than during truth telling. For each detail, the largest differences were found for college students, whereas the details were not often present in the statements of either truthful or lying 5-6-year-olds.[17]

Table 3 shows differences between truth tellers and liars for each age group. In order to reduce the number of variables, only the total CBCA score and total RM score are included in Table 3 together with cognitive operations and limb movements, the only two remaining variables which obtained a significant Veracity effect (see Table 2). The pattern of means are very consistent and all are in the predicted direction, however, not all effects were

significant. As can be seen in Table 3, CBCA scores were higher for truth tellers than for liars in each of the four age groups. The differences were significant for the youngest and the oldest age groups and were marginally significant for 10-11-year-olds.

Total RM scores were higher for truth tellers than for liars in each of the four age groups and the differences were significant for the oldest age group and 14-15-year-olds and marginally significant for 10-11-year-olds.

In all four age groups liars reported more cognitive operations than truth tellers, but the difference was only significant for the 14-15-year-olds.

In all four age groups, liars made fewer limb movements than truth tellers, and the differences were significant for the youngest and oldest age groups.[18]

## Classifying Liars and Truth Tellers

In order to determine the usefulness of the detection techniques in classifying truth tellers and liars, stepwise discriminant analyses utilizing the Wilks' Lambda method were conducted. With this technique the variables remaining in the final analysis are those which contribute to maximizing the correct assignment of the cases to the objective truth status. In the analyses, the objective truth status was the classifying variable and the six nonverbal behaviors, plus total CBCA score, total RM score, and cognitive operations were the dependent variables. Analyses were carried out for the whole sample and additionally per age group. The results are given in Table 4.[19]

Hypothesis 8, a combination of the two verbal techniques and the nonverbal technique would result in the most accurate classifications, received strong support. The combined detection technique was the only technique which resulted in significant discriminant functions in all four year groups. Moreover, in each age group the combined technique resulted in the highest total hit rate. These hit rates varied from a modest 65% for 10-11-year-olds to a very high 88% for undergraduates (and 89% hit rate for lie detection and 88% hit rate for truth detection). Also in the discriminant analysis for the whole sample, the combined technique resulted in the highest hit rate.

However, for each age group a different set of variables contributed to the discriminant function. In 5-6-year-olds, four variables (total CBCA score, Wilks' lambda = .82, limb movements, Wilks' lambda = .71, cognitive operations, Wilks' lambda = .63, and speech disturbances, Wilks' lambda = .60), contributed to the significant discriminant function; in 10-11-year-olds three variables (latency time, Wilks' lambda = .90, total CBCA score, Wilks' lambda = .88, and cognitive operations, Wilks' lambda = .84); in 14-15-year-olds five variables (cognitive operations, Wilks' lambda = .77, total RM score, Wilks' lambda = .60, frequency of pauses, Wilks' lambda = .56, speech disturbances, Wilks' lambda = .52, and limb movements, Wilks' lambda = .50), and in undergraduates seven variables (total RM score, Wilks' lambda = .74, frequency of pauses, Wilks' lambda = .64, cognitive operations, Wilks' lambda = .58, latency period, Wilks' lambda = .54, limb movements, Wilks' lambda = .51, total CBCA score, Wilks' lambda = .49, and gaze aversion, Wilks' lambda = .48).

Although there is overlap between the discriminant functions of the four age groups (i.e., cognitive operations appeared in all four functions, total CBCA score and limb movements in three functions and both are missing in the discriminant function for 10-11-year-olds), the discriminant functions do not replicate each other. On the one hand, it might be that the differences in discriminant functions are caused by the age differences in behaviors between the four groups. Perhaps different age groups have different clusters of cues to deceit. On the other hand, we cannot rule out that the apparent differences may simply be an artifact of high multicollinearity.[20]

DePaulo, Anderson, and Cooper (1999) correctly point out that a lack of overlap between discriminant functions is problematic, as it does not provide a set of variables to reliably build upon in lie detection. In order to seek replication of previous findings obtained with undergraduate participants (Vrij et al., 2000), an additional discriminant analysis for undergraduates was conducted. This time we included the variables which were included in Vrij et al.'s (2000) study, which were total CBCA score, total RM score, latency period, speech errors, speech hesitations, speech rate, illustrators and hand/finger movements.[21] Vrij

et al. (2000) reported a 81% total hit rate (77% truth hit rate and 85% lie hit rate) with CBCA score, latency period, hand and finger movements, speech hesitations, illustrators and speech rate contributing to the discriminant function. The results for the present experiment are shown in Table 4. As can be seen, the hit rates were very similar, 79% total hit rate, 77% hit rate for truth detection and 81% hit rate for lie detection. This time four variables contributed to the discriminant function, total RM score (Wilks' lambda = .74), latency period (Wilks' lambda = .65), hand and finger movements (Wilks' lambda = .62), and speech errors (Wilks' lambda = .61). In other words, there was an overlap of two cues (latency period and hand and finger movements), and in both analyses the cues were in the same direction. Again, the differences between the two discriminant functions might have been the result of multicollinearity.

## Discussion

The present experiment examined children's and undergraduates' verbal and nonverbal responses during deception, and to what extent truths and lies could be correctly classified by paying attention to these verbal and nonverbal responses.

### Age Differences in Verbal and Nonverbal Responses

Considerable age differences were found in verbal and nonverbal responses. As predicted in Hypothesis 4, a CBCA age effect was found, with the oldest participants (undergraduates) having the highest CBCA scores and the youngest participants (5-6-year-olds) obtaining the lowest CBCA scores. In more concrete terms, the accounts of undergraduates were often more detailed, and were, for example, more likely to include descriptions of interactions ("She asked me to sit down, so I sat down"), reproductions of conversations ("And the woman said: do you fancy a game of Connect 4?"), and how they felt during the event ("I couldn't stand that I was losing that game"). They also included more spontaneous corrections ("She was probably five foot eleven, no, more likely five foot seven") and raised more doubts about their own memory ("I think she had black shoes on").

RM total scores for 5-6-year-olds were significantly lower than the RM total scores

for the other age groups (in line with Hypothesis 7). In concrete terms, it means that 5-6-year-olds were least likely to include, for example, auditory details ("She just walked in, she didn't knock or anything"), temporal details ("After about four pieces went in, a lady entered the room") and spatial details ("She walked behind us") in their accounts. Undergraduates' scores for cognitive operations were significantly higher than the scores for the other three age groups. That is, they were more likely to include phrases such as "She was very polite," "He seemed very confident," "You could tell that he knew what he was doing," etc. in their statements. Similar to the results of verbal responses, nonverbal behavioral responses of children and adults differed considerably. For example, 5-6-year-olds waited longer before giving an answer and spoke slower than any of the other three age groups (which provide some support for Hypothesis 2) and undergraduates made fewer limb movements (particularly hand and finger movements) than any of the other three age groups, which is in agreement with Hypothesis 2. An unpredicted finding, for which we as yet have no plausible explanation, was that adults included more speech hesitations (mm's, er's, etc.) in their accounts than any other age group.

## Verbal and Nonverbal Differences between Liars and Truth Tellers

Despite these age differences in verbal and nonverbal responses, cues to deceit were remarkably similar across the different age groups. For example, as was predicted in Hypothesis 3, in all four age groups, liars obtained lower CBCA scores than truth tellers, although the differences were not significant for 14-15-year-olds. We can only speculate about the absence of a significant CBCA deception effect for 14-15-years-olds. Perhaps, we did not succeed in motivating this particular age group enough. Perhaps, the prize they could win, (i.e., £2), was not sufficient for them to put much effort into the task. For this group a CBCA score in between the CBCA scores of 10-11-years-olds and undergraduates could be expected. However, their actual CBCA score (see Table 1) was much closer to the scores of 10-11-year-olds, and did not differ significantly from the CBCA score of that group (but was significantly lower than the CBCA score of undergraduates). This suggests that the 14-15-

year-olds underachieved. Alternatively, it might be that the game of Connect 4 was not that attractive for this age group compared to the other age groups. Although the game is played in the UK by young and old (it is, for example, a popular pub game for undergraduates), 14-15-years-olds, might feel themselves "too old" for this game.

In all four age groups, liars obtained lower total RM scores than truth tellers which was predicted in Hypothesis 5. When analyses were conducted per age group it appeared that the differences were significant for the two oldest age groups, and marginally significant for 10-11-year-olds. No significant differences emerged for 5-6-year-olds. In other words, it appears that RM assessments are not useful below a certain age. This reflects Reality Monitoring memory research (see Lindsay, 2002, for an overview). Young children are worse than adults in differentiating between memories of actual and imagined events. In order to explain this, it has been suggested that "children may be better than adults at imagining themselves performing actions, such that their memories of imagined and actual self-performed actions are more similar" (Lindsay, 2002, p. 88). Since RM assessments are based upon differences in people's memories between actual and imagined events, it is reasonable that if their memories are more similar, RM assessments will be less successful in distinguishing truth from deception.[22]

For the first time to our knowledge, a significant difference was found between liars and truth tellers in cognitive operations. As was predicted in Hypothesis 6, liars included more of these operations in their statements than truth tellers. Also, cognitive operations was the only variable that appeared in the discriminant functions for each of the four Age Groups. These findings are noteworthy because CBCA assessments are prone to a truth bias (i.e., CBCA assessments are typically more accurate for detecting truths than for detecting lies; see Vrij, 2000, for a review of CBCA hit rates). One possible reason is that all CBCA criteria are "truth telling" criteria, that is, their presence might indicate that someone is telling the truth. The technique might result in more balanced and higher hit rates when lie telling criteria (criteria which indicate deception) are included. The cognitive operations criterion is an

example of a lie telling criterion. The present data set supports this idea. Additional discriminant analyses were conducted with total CBCA score and cognitive operations as dependent variables. All analyses (for the total sample and for each of the four age groups) resulted in higher hit rates than the analyses in which just the CBCA score was included.[23]

Finally, in agreement with Hypothesis 1, in all four age groups a decrease in movements (particularly in hand and finger movements) occurred during deception, and this finding (regarding hand and finger movements) was significant for 5-6-year-olds and undergraduates and marginally significant for 14-15-year-olds. A decrease in hand and finger movements, as well as a decrease in other movements such as illustrators, has been found in deception research before, and they appeared in recent meta-analyses of deception research (DePaulo et al., in press; Vrij, 2000) as some of the very few nonverbal behaviors which significantly discriminated between liars and truth tellers. A possible explanation for why, in this study, the expected decrease in movements did occur, but other expected findings, such as a slower speech rate, more pauses and a longer latency period did not emerge, is that a decrease in movements could have been the result of two processes (cognitive load and attempted behavioral control), whereas the other effects could just have been the result of cognitive load. Perhaps, liars experienced both processes, which might have made their impact on movements strong enough for behavioral differences between liars and truth tellers to occur.

The popular belief that liars look away and fidget could not be supported. No significant differences for gaze aversion or adaptors were found for any of the four age groups. These findings are also consistent with prior research (DePaulo et al., in press; Vrij, 2000).

## Classifying Liars and Truth Tellers

Concerning classifying liars and truth tellers on the basis of verbal and nonverbal responses, two results are worth mentioning. First, hit rates based on RM assessments were as good as, and especially for older participants, even better than, hit rates based on CBCA

assessments. This makes Reality Monitoring a promising method of lie detection.

Second, the hit rates for a combination of verbal and nonverbal techniques were always higher than the hit rates for the individual verbal and nonverbal techniques, which supports Hypothesis 8. The combined technique resulted in a very high hit rate of 88% for undergraduates. Also, a combination of verbal and nonverbal responses which resulted in a hit rate of 81% in a previous study (Vrij et al., 2000), resulted in a hit rate of 79% in this study.

## Methodological Considerations

Three methodological issues merit attention. First, the way we motivated participants might raise some concerns. Different age groups received different rewards, therefore effects of age were confounded with type of motivation. We varied the reward so as to avoid such a confound. We believe that giving all participants the same reward would have created a confound because the same reward (for example £5) would be perceived as substantially higher by young children than by undergraduates. However, it is difficult to determine what would be comparable rewards for different age groups, and, as discussed above, the award for 14-15-years-olds might have been too low.

Also, we did not introduce different rewards and punishments for liars and truth tellers. We believe that sometimes the rewards and punishments for liars and truth tellers are identical. For example, if a mother believes that her child has finished his or her homework, both the honest and deceptive child might get the same reward (for example, money to go to the cinema), whereas if, on the other hand, the mother does not believe that the child has finished his or her homework, both may get the same punishment (not allowed to go out that night). Sometimes, however, the situation is different. For example, sometimes the liar has more to gain than the truth teller. The guilty suspect who has hidden stolen goods has more to gain by being believed (i.e., that he or she was not involved in the crime) than the innocent suspect who was not involved and therefore has no hidden valuables. In other words, the rewards and punishments for liars and truth tellers differ per situation, and it would be a very

difficult task to incorporate all possible situations in one experiment. With hindsight it would have been better to introduce a manipulation check and to ask participants to what extent they were motivated. We recommend the inclusion of such a motivation check in future research.

Second, asking participants about their motivation might have clarified another issue. Research has shown that people are more motivated to protect a friend than to protect a stranger, if necessary by telling lies (Cole, 2001; Metts, 1989), among other strategies, because they care more about their friends. In the present study, participants in the deceptive situation might have had more positive feelings towards the person they were asked to protect (confederate 1) than the participants in the truthful situation who had to speak about a total stranger (confederate 2). Although the difference in affection towards these two confederates was probably small (the participants in the deceptive situation did not really interact with the confederate they had to protect, they did not play the game of Connect 4 with that person, only the truth tellers did), with hindsight we could have asked participants about their affinity towards the confederates.

Finally, more self-selection might have taken place in the two oldest age groups than in the two youngest age groups. Specifically, because participants volunteered for a study about "telling a convincing story" it might be that in the older age groups mostly participants who thought they would be convincing signed up, whereas in the younger age groups parental permission to participate was the dominant factor to participate. However, self-selection might have occurred within the younger age groups as well. Perhaps only the parents who believed that their children would be good at the task might have given parental consent to take part in the study. It is therefore not certain that self-selection differed between the participating age groups. Perhaps this self-selection worked in our benefit. In daily life, it is mostly people's own free choice to lie and people probably decide to lie only if they think that have a chance to get away with it, and if they think that they will be able to tell a convincing story. Therefore, those who believe that they will not be able to tell a convincing story might have decided against participation in our study, but neither are they likely to lie

in real-life. Obviously, deception studies with participants who almost never lie in daily life are ecologically less valid.

## Future Research

The findings suggest that combining verbal and nonverbal lie detection methods is a worthwhile method of lie detection. Obviously, the question then arises which verbal and nonverbal cues are particularly relevant in such a combined method. We have only just started to investigate this, and conclusions can not be drawn at this stage. However, the findings so far are promising, and it might be that some suggestions can eventually be made. For example, in most discriminant functions discussed in this article, total CBCA scores, total RM scores, cognitive operations and movements were included, making these cues relevant cues to focus upon. Also, some cues, such as gaze aversion, are mostly absent in these discriminant functions, suggesting that lie detectors could ignore such cues without negative consequences.

The next step would be to investigate whether snap judgments of relevant cues will increase lie detection. For example, research has shown that people are particularly poor at detecting lies in friends and partners, due to a truth-bias, the belief that the other is telling the truth (Levine, Park, & McCornack, 1999; McCornack & Parks, 1986; Stiff, Kim, & Ramesh, 1992). Making snap judgments of some relevant verbal and nonverbal cues might negate the tendency for a truth-bias and will therefore increase accuracy in lie detection.

References

Akehurst, L. Köhnken, G., Vrij, A., & Bull, R. (1996). Lay persons' and police officers' beliefs regarding deceptive behaviour. Applied Cognitive Psychology, 10, 461-471.

Akehurst, L., Köhnken, G., & Höfer, E. (2001). Content credibility of accounts derived from live and video presentation. Legal and Criminological Psychology, 6, 65-83.

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: Walter 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: Walter de Gruyter.

Alonso-Quecuty, M. L., Hernandez-Fernaud, E., & Campos, L. (1997). Child witnesses: Lying about something heard. In S. Redondo, V. Garrido, J. Perez, & R. Barbaret (Eds.), Advances is psychology and law (pp. 129-135). Berlin: Walter 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.

Boychuk, T. (1991). Criteria-Based Content Analysis of children's statements about sexual abuse: A field-based validation study. Unpublished doctoral dissertation, Arizona State University.

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. Communication Theory, 6, 203-242.

Burgoon, J. K., & Buller, D. B. (1994). Interpersonal deception: III. Effects of deceit on perceived communication and nonverbal behavior dynamics. Journal of Nonverbal Behavior, 18, 155-184.

Burgoon, J. K., Buller, D. B., & Guerrero, L. K. (1995). Interpersonal deception IX: Effects of social skill and nonverbal communication on deception success and detection accuracy. Journal of Language and Social Psychology, 14, 289-311.

Burgoon, J. K., Buller, D. B., White, C. H., Afifi, W., & Buslig, A. L. S. (1999). The role of conversation involvement in deceptive interpersonal interactions. Personality and Social Psychology Bulletin, 25, 669-685.

Burgoon, J. K., & Floyd, K. (2000). Testing for the motivational impairment effect during deceptive and truthful interaction. Western Journal of Communication, 64, 243-267.

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.

Bussey, K. (1992). Children's lying and truthfulness: Implications for children's testimony. In S. J. Ceci, M. DeSimone Leichtman, & M. Putnick (Eds.), Cognitive and social factors in early deception (p. 89-110). Hillsdale: Erlbaum.

Ceci, S. J., & DeSimone Leichtman, M. (1992). 'I know that you know that I know that you broke the toy': A brief report of recursive awareness among 3-year-olds. In S. J. Ceci, M. DeSimone Leichtman, & M. Putnick (Eds.), Cognitive and social factors in early deception (p. 1-9). Hillsdale, NJ: Erlbaum.

Cody, M. J., Marston, P. J., & Foster, M. (1984). Deception: Paralinguistic and verbal leakage. In R. N. Bostrom, & B. H. Westley (Eds.), Communication Yearbook 8 (pp. 464-490). Beverly Hills: Sage.

Cole, T. (2001). Lying to the one you love: The use of deception in romantic relationships. Journal of Social and Personal Relationships, 18, 107-129.

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, 77-85.

Davies, G. M. (1994). Children's testimony: Research findings and police implications. Psychology, Crime, & Law, 1, 175-180.

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

DePaulo, B. M., & Jordan, A. (1982). Age changes in deceiving and detecting deceit. In R. S. Feldman (Ed.), Development of nonverbal behavior in children (pp. 151-180). New York: Springer-Verlag.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. Journal of Personality and Social Psychology, 70, 979-995.

DePaulo, B. M., & Kirkendol, S. E. (1989). The motivational impairment effect in the communication of deception. In J. C. Yuille (Ed.), Credibility assessment (pp. 51-70). Dordrecht, The Netherlands: Kluwer.

DePaulo, B. M., Lindsay, J. L., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (in press). Cues to deception. Psychological Bulletin.

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

Eaton, W. O., McKeen, N. A., & Campbell, D. W. (2001). The waxing and waning of movement: Implications for psychological development. Developmental Review, 21, 205-223.

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

Ekman, P. (1997). Deception, lying, and demeanor. In D. F. Halpern & A. E. Voiskounsky (Eds.), States of mind: American and post-soviet perspectives on contemporary issues in psychology (pp. 93-105). New York: Oxford University Press.

Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. Psychiatry, 32, 88-106.

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

374.

Ekman, P., Roper, G., & Hager, J. C. (1980). Deliberate facial movement. Child Development, 51, 886-891.

Feldman, R. S., & Phileppot, P. (1993). Children's deception skills and social competence. In G. Goodman & B. Bottoms (Eds.), Child victims, child witnesses (pp. 80-99). New York: Guilford Press.

Flavell, J. H., Botkin, P. T., Fry, C. L., Wright, J. C., & Jarvis, P. T. (19968). The development of role-taking and communication skills in children. New York: Wiley.

Fleiss, J. L. (1981). Statistical methods for rates and proportions. New York: Wiley.

Frank, M. G. (1992). Commentary: On the structure of lies and deception experiments. In S. J. Ceci, M. DeSimone Leichtman, & M. Putnick (Eds.), Cognitive and social factors in early deception (p. 127-146). Hillsdale: Erlbaum.

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.

Gumpert, C. H., & Lindblad, F. (2001). Expert testimony on child sexual abuse: A qualitative study of the Swedish approach to statement analysis. Expert Evidence, 7, 279-314.

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.

Hocking, J. E., & Leathers, D. G. (1980). Nonverbal indicators of deception: A new theoretical perspective. Communication Monographs, 47, 119-131.

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, 3-29.

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.

Kasl, S. V., & Mahl, G. F. (1965). The relationship of disturbances and hesitations in spontaneous speech to anxiety. Journal of Personality and Social Psychology, 1, 425-433.

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. (1996). Social psychology and the law. In G. R. Semin & K. Fiedler (Eds.), Applied social psychology (pp. 257-282). London: Sage.

Köhnken, G. (1999, July). Statement Validity Assessment. Paper presented at the pre-conference program of applied courses "Assessing credibility" organised by the European Association of Psychology and Law, Dublin, Ireland.

Köhnken, G. (2002). A German perspective on children's testimony. In H. L. Westcott, G. M. Davies, & R. H. C. Bull (Eds.), Children's testimony: A handbook of psychological research and forensic practice (pp. 233-244). Chichester, United Kingdom: Wiley.

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.

Krauss, R. M. (1981). Impression formation, impression management, and nonverbal behaviors. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), Social cognition: The Ontario Symposium (Vol 1, pp. 323-341). Hillsdale, NJ: Erlbaum.

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

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

Lamers-Winkelman, F., & Buffing, F. (1996). Children's testimony in the Netherlands: A study of Statement Validity Analysis. In B. L. Bottoms & G. S. Goodman (Eds.), International perspectives on child abuse and children's testimony (pp. 45-62). Thousand Oaks, CA: Sage.

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.

Levine, T. R., Park, H. S., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the "veracity effect". Communication Monographs, 65, 125-144.

Lindsay, D. S. (2002). Children's source monitoring. In H. L. Westcott, G. M. Davies, & R. H. C. Bull (Eds.), Children's testimony (pp. 83-98). Chichester, UK: Wiley.

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, Germany: Walter 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 (pp. 91-126). Mahway, NJ: Erlbaum.

McCornack, S. A., & Parks, M. R. (1986). Deception detection and relational

development: The other side of trust. In M. L. McLaughin (Ed.), Communication yearbook 9 (pp. 377-389). Beverly Hills, CA: Sage.

Metts, S. (1989). An exploratory investigation of deception in close relationships. Journal of Social and Personal Relationships, 6, 159-179.

Morency, N. L., & Krauss, R. M. (1982). Children's nonverbal encoding and decoding of affect. In R. S. Feldman (Ed.), Development of nonverbal behaviour in children (pp. 181-199). New York: Springer-Verlag.

Newton, P., Reddy, V., & Bull, R. (2000). Children's everyday deception and performance on false-belief tasks. British Journal of Developmental Psychology, 18, 297-317.

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.

Peskin, J. (1992). Ruse and representations: On children's ability to conceal information. Developmental Psychology, 28, 84-89.

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

Roberts, K. P., Lamb, M. E., Zale, J. L., & Randall, D. W. (1998, March). 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.

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.

Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The 'windows task' as a measure of strategic deception in preschoolers and autistic subjects. British Journal of Developmental Psychology, 9, 331-349.

Sodian, B. (1991). The development of deception in young children. British Journal of Developmental Psychology, 9, 173-188.

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, NJ: Springer-Verlag.

Stiff, J. B., Kim, H. J., & Ramesh, C. N. (1992). Truth biases and aroused suspicion in relational deception. Communication Research, 19, 326-345.

Tye, M. C., Amato, S. L., Honts, C. R., Kevitt, M. K., & Peters, D. (1999). The willingness of children to lie and the assessment of credibility in an ecologically relevant laboratory setting. Applied Developmental Science, 3, 92-109.

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.

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

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

Vrij, A. (2000). Detecting lies and deceit: The psychology of lying and its implications for professional practice. Chichester, UK: Wiley.

Vrij, A. (2002). Deception in children: A literature review and implications for children's testimony. In H. L. Westcott, G. M. Davies, & R. H. C. Bull (Eds.), Children's testimony (pp. 175-194). Chichester, UK: Wiley.

Vrij, A., & Akehurst, L. (1998). Verbal communication and credibility: Statement Validity Assessment. In A. Memon, A. Vrij & R. Bull (Eds.), Psychology and law: Truthfulness, accuracy and credibility (pp. 3-26). Maidenhead, UK: McGraw-Hill.

Vrij, A., Akehurst, L., & Morris, P. (1997). Individual differences in hand movements during deception. Journal of Nonverbal Behavior, 21, 87-103.

Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2002). Will the truth come out? The effect of deception, age, status, coaching, and social skills on CBCA scores. Law and Human Behavior, 26, 261-283.

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

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

Vrij, A., Edward, K., Roberts, K. P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. Journal of Nonverbal Behavior, 24, 239-263.

Vrij, A., & Semin, G. R. (1996). Lie experts' beliefs about nonverbal indicators of deception. Journal of Nonverbal Behaviour, 20, 65-80.

Vrij, A., Semin, G. R., & Bull, R. (1996). Insight into behaviour 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

(CBCA). Tijdschrift voor Ontwikkelingspsychologie, 22, 61-74.

Winkel, F. W., Vrij, A., Koppelaar, L., & Van der Steen, J. (1991). Reducing secondary victimisation risks and skilled police intervention: Enhancing the quality of police rape victim encounters through training programmes. Journal of Police and Criminal Psychology, 7, 2-11.

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 (1-57). New York, NJ: Academic Press.

Table 1.

Verbal and Nonverbal Responses as a Function of Age Group.

| 10-11 | 14-15 | | undergraduates | | F(3, 188) | | eta$^2$ | | 5-6 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | m | sd | m | sd | m | sd | m | sd | | | |
| CBCA | | | | | | | | | | | |
| CBCA (total score) | 16.10$^{a}$ | 2.5 | 20.44$^{b}$ | 3.0 | 21.28$^{b}$ | 3.0 | 24.14$^{c}$ | 4.2 | 47.56** | .43 | |
| logical structure | 3.90$^{a}$ | .9 | 4.06$^{ab}$ | .8 | 4.32$^{b}$ | .6 | 4.35$^{b}$ | .6 | 4.68** | .07 | |
| quantity of details | 2.97$^{a}$ | .5 | 4.06$^{b}$ | .6 | 4.05$^{b}$ | .6 | 4.40$^{c}$ | .6 | 46.27** | .43 | |
| contextual embedding | 1.90$^{a}$ | .9 | 3.19$^{b}$ | .9 | 3.15$^{b}$ | .8 | 3.50$^{b}$ | .9 | 28.10** | .31 | |
| description of interactions | 1.13$^{a}$ | .5 | 1.45$^{ab}$ | .6 | 1.52$^{b}$ | .7 | 1.94$^{c}$ | .8 | 11.68** | .16 | |
| reproduction of conversation | 1.13$^{a}$ | .3 | 1.66$^{b}$ | .6 | 1.77$^{bc}$ | .8 | 2.02$^{c}$ | .9 | 12.54** | .17 | |
| accounts of own mental state | 1.00$^{a}$ | .0 | 1.10$^{a}$ | .3 | 1.18$^{a}$ | .5 | 1.51$^{b}$ | .9 | 8.06** | .11 | |
| spontaneous corrections | 1.59$^{a}$ | .6 | 1.94$^{ab}$ | .6 | 2.04$^{b}$ | .6 | 2.54$^{c}$ | .9 | 14.84** | .19 | |
| admitting lack of memory | 1.34$^{a}$ | .5 | 1.37$^{a}$ | .5 | 1.38$^{a}$ | .5 | 1.60$^{a}$ | .6 | 2.35 | | |
| raising doubts | 1.14$^{a}$ | .3 | 1.60$^{b}$ | .7 | 1.88$^{b}$ | .8 | 2.28$^{c}$ | .8 | 20.85** | .25 | |
| Reality Monitoring | | | | | | | | | | | |
| Reality Monitoring (total score) | 2.37$^{a}$ | 1.2 | 3.35$^{b}$ | .9 | 3.27$^{b}$ | .8 | 3.12$^{b}$ | .9 | 11.83** | .16 | |
| visual details | 19.79$^{a}$ | 10.9 | 43.77$^{b}$ | 18.5 | 39.63$^{b}$ | 12.7 | 36.75$^{b}$ | 17.9 | 19.44** | .24 | |
| auditory details | .56$^{a}$ | 1.0 | 3.24$^{b}$ | 2.7 | 3.54$^{b}$ | 3.1 | 5.00$^{c}$ | 4.1 | 17.53** | .22 | |
| spatial details | .86$^{a}$ | .9 | 2.09$^{a}$ | 1.7 | 1.84$^{a}$ | 1.2 | 4.13$^{b}$ | 4.6 | 13.51** | .18 | |
| temporal details | .67$^{a}$ | .9 | 3.16$^{b}$ | 2.5 | 2.85$^{b}$ | 1.9 | 5.11$^{c}$ | 3.8 | 22.63** | .27 | |
| cognitive operations | .03$^{a}$ | .2 | .46$^{ab}$ | .7 | .65$^{b}$ | 1.0 | 1.46$^{c}$ | 1.6 | 15.38** | .20 | |
| nonverbal behavior | | | | | | | | | | | |
| gaze aversion | 38.77$^{b}$ | 9.3 | 33.94$^{a}$ | 6.2 | 35.13$^{ab}$ | 7.9 | 35.42$^{ab}$ | 7.2 | 2.99* | .05 | |
| movements | 80.23$^{b}$ | 26.9 | 82.93$^{b}$ | 29.9 | 84.67$^{b}$ | 49.1 | 49.40$^{a}$ | 21.7 | 12.27** | .16 | |
| latency time | 3.86$^{b}$ | 5.9 | 1.65$^{a}$ | 1.7 | 1.99$^{a}$ | 1.6 | 1.71$^{a}$ | 1.8 | 5.67** | .08 | |
| pauses | 1.74$^{a}$ | 1.6 | 1.54$^{a}$ | .9 | 1.43$^{a}$ | 1.1 | 1.89$^{a}$ | 1.0 | 1.58 | | |
| speech disturbances | 5.95$^{a}$ | 5.3 | 6.75$^{a}$ | 3.9 | 5.53$^{a}$ | 3.8 | 11.73$^{b}$ | 5.8 | 18.41** | .23 | |
| speech rate | 77.32$^{a}$ | 33.2 | 121.97$^{b}$ | 34.9 | 135.73$^{bc}$ | 34.0 | 145.41$^{c}$ | 42.4 | 28.28** | .31 | |

* $p < .05$, ** $p < .01$

Table 2.

Verbal and Nonverbal Responses as a Function of Veracity.

| | truth | | | lie | | | | |
|---|---|---|---|---|---|---|---|---|
| | m | sd | | m | sd | | | |
| | CBCA | | | | | | | |
| CBCA (total score) | 20.08 | 3.9 | | 21.62 | 4.4 | | 17.25** | .08 |
| logical structure | 4.08 | .8 | | 4.27 | .6 | | 5.52* | .03 |
| quantity of details | 3.86 | .7 | | 4.04 | .8 | | 8.93** | .05 |
| contextual embedding | 2.86 | 1.0 | | 3.19 | 1.1 | | 11.27** | .06 |
| description of interactions | 1.38 | .5 | | 1.70 | .8 | | 12.04** | .06 |
| reproduction of conversation | 1.48 | .6 | | 1.89 | .8 | | 18.14** | .09 |
| accounts of own mental state | 1.30 | .6 | | 1.14 | .5 | | 2.94 | |
| spontaneous corrections | 1.99 | .7 | | 2.13 | .8 | | 3.34 | |
| admitting lack of memory | 1.29 | .5 | | 1.38 | .6 | | 1.76 | |
| raising doubts | 1.76 | .8 | | 1.80 | .8 | | .77 | |
| Reality Monitoring | | | | | | | | |
| Reality Monitoring (total score) | 2.85 | .9 | | 3.36 | .9 | | 19.63** | .10 |
| visual details | 33.07 | 16.4 | | 39.59 | 18.1 | | 11.96** | .06 |
| auditory details | 2.19 | 2.3 | | 4.34 | 3.9 | | 28.11** | .13 |
| spatial details | 1.80 | 1.7 | | 2.83 | 3.5 | | 8.67** | .04 |
| temporal details | 2.57 | 2.2 | | 3.67 | 3.5 | | 11.29** | .06 |
| cognitive operations | .97 | 1.3 | | .46 | .9 | | 8.13** | .04 |
| nonverbal behavior | | | | | | | | |
| gaze aversion | 34.84 | 8.1 | | 36.17 | 7.3 | | .67 | |
| movements | 66.95 | 30.4 | | 80.58 | 41.5 | | 7.60** | .04 |
| latency time | 2.24 | 3.7 | | 2.08 | 2.3 | | .95 | |
| pauses | 1.60 | 1.1 | | 1.67 | 1.2 | | .37 | |
| speech disturbances | 8.06 | 5.4 | | 7.15 | 5.3 | | 1.56 | |
| speech rate | 121.40 | 45.1 | | 126.54 | 41.4 | | 3.13 | |

* $p < .05$, ** $p < .01$

Table 3.

Verbal and Nonverbal Responses for Each Age Group as a Function of Veracity.

| | 5-6-year-olds | | | | | | | | 10-11-y | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | lie | | truth | | | | lie | | truth | | | | lie | | truth | | | |
| | m | sd | m | sd | F | eta$^2$ | m | sd | m | sd | F | eta$^2$ | m | sd | m | sd | F | eta$^2$ |
| CBCA (total score) | 14.82 | 1.5 | 16.95 | 2.67.44** | | | | | | | | | | | | | | |
| .18 | 19.71 | 2.8 | 21.21 | 3.1 | 3.54$^t$ | .06 | 20.73 | 2.4 | 21.78 | 3.4 | 1.71 | | 22.65 | 4.1 | 25.61 | 3.7 | 7.35** | .13 |
| Reality Monitoring (total) | 2.21 | 1.1 | 2.48 | 1.2 | .41 | | 3.14 | 1.0 | 3.58 | .6 | 3.50$^t$ | .06 | 2.88 | .8 | 3.62 | .6 | 14.32** | .21 |
| cognitive operations | 2.85 | .7 | 3.58 | .5 | 17.59** | .26 | .07 | .3 | .00 | .0 | 1.52 | | .57 | .7 | .35 | .8 | 1.24 | |
| movements | 1.15 | 1.2 | .21 | .4 | 15.46** | .23 | 69.19 | 25.9 | 87.60 | 25.5 | 4.31* | .12 | 82.05 | 26.8 | 83.88 | 33.3 | .05 | |
| | 75.72 | 29.4 | 92.70 | 61.1 | 1.66 | | 1.69 | 1.8 | 1.23 | 1.3 | 1.12 | | 40.71 | 20.0 | 58.08 | 20.0 | 9.78** | .16 |

$^t$ .5 < p < .10,* p < .05,** p < .01

Table 4.
Discriminant Analyses with Nonverbal Behavior, Criteria-Based Content Analysis and Reality Monitoring.

| Detection technique | hit rates: lie | truth | total | Eigenvalue | Lambda | df | X2 |
|---|---|---|---|---|---|---|---|
| **Total sample** | | | | | | | |
| CBCA[1] | 62% | 61% | 61% | .04 | .97 | 1 | 6.66** |
| RM[2] | 67% | 72% | 69% | .17 | .85 | 2 | 30.51** |
| Nonverbal behavior | 61% | 61% | 61% | .05 | .95 | 3 | 8.97* |
| CBCA + RM + nonverbal behavior | 68% | 71% | 69% | .25 | .80 | 6 | 42.63** |
| **5-6-year-olds** | | | | | | | |
| CBCA | 64% | 71% | 69% | .23 | .82 | 1 | 6.61* |
| RM | | | | | | 1 | 1.47 |
| Nonverbal behavior | 79% | 86% | 83% | .27 | .79 | 2 | 7.59* |
| CBCA + RM + nonverbal behavior | 79% | 91% | 86% | .68 | .60 | 4 | 16.00** |
| **10-11-year-olds** | | | | | | | |
| CBCA | | | | | | 1 | 3.39 |
| RM | | | | | | 2 | 5.37 |
| Nonverbal behavior | 43% | 81% | 61% | .12 | .90 | 1 | 5.66* |
| CBCA + RM + nonverbal behavior | 64% | 65% | 65% | .19 | .84 | 3 | 8.86* |
| **14-15-year-olds** | | | | | | | |
| CBCA | | | | | | 1 | 1.67 |
| RM | 69% | 92% | 82% | .68 | .60 | 2 | 26.85** |
| Nonverbal behavior | | | | | | 1 | 1.62 |
| CBCA + RM + nonverbal behavior | 81% | 93% | 87% | .99 | .50 | 5 | 34.76** |
| **undergraduates** | | | | | | | |
| CBCA | 65% | 58% | 62% | .15 | .87 | 1 | 6.79** |
| RM | 73% | 81% | 77% | .51 | .66 | 2 | 20.17** |
| Nonverbal behavior | 77% | 65% | 71% | .25 | .80 | 2 | 10.87* |
| CBCA + RM + nonverbal behavior | 89% | 88% | 88% | 1.09 | .48 | 7 | 34.29** |
| **adult participants, replication Vrij et al. (2000)** | | | | | | | |
| CBCA + RM + nonverbal behavior | 81% | 77% | 79% | .64 | .61 | 4 | 23.83** |

[1] CBCA represents the total CBCA score
[2] RM represents two variables: the total Reality Monitoring score and the cognitive operations variable
* $p < .05$, ** $p < .01$

## Authors' Notes

Aldert Vrij (PhD, 1991, Free University, Amsterdam) is a Professor of Applied Social Psychology, Lucy Akehurst (PhD, 1997, University of Portsmouth) is a Senior Lecturer, Stavroula Soukara (MSc, 1998) is a PhD student and Ray Bull (DSc, 1995, University of Portsmouth) is a Professor of Criminological and Legal Psychology.

This study was sponsored by a grant (R000222820) from the Economic and Social Research Council.

Correspondence concerning this article should be addressed to Aldert Vrij, University of Portsmouth, Psychology Department, King Henry Building, King Henry 1 Street, Portsmouth PO1 2DY, United Kingdom. Electronic mail may be sent via Internet to aldert.vrij@port.ac.uk

1. See the Method section for a definition of the behaviors.

2. See Vrij (2000) for detailed descriptions of the CBCA criteria.

3. In these studies which involve adults as participants who talk about non-sexual events (including the present experiment) several CBCA criteria are typically not examined as they are probably specific for children who talk about sexual events. For example, criterion 10, "accurately reported details misunderstood", is fulfilled if someone speaks of details that are beyond the comprehension of the person, such as a child who describes the adult's sexual behavior but attributes it to a sneeze or pain. This criterion is typically ignored in experimental studies with adults as participants.

4. It might seem suspicious from the participant's perspective that the experimenter would know that an interview was about to ensue in order to ascertain the truth of the blackboard mishap. We had prepared an answer which the experimenter would give in case a participant asked a question about this. However, none of the participants in this study raised this issue. In other words, we have no evidence to suggest that the participants were suspicious.

5. In the procedure just described the liars were 'fabricating witnesses'. We also created a 'fabricating suspect condition'. In this condition the experimenter asked participants to wipe the blackboard clean while she

(experimenter) had to pop out for something. The experimenter then left the room, and all participants in this condition erased the information from the blackboard. Participants were asked to pretend that they (experimenter and participant) played Connect 4 and that someone else entered the room and wiped off the information, and so on. The suspect/witness manipulation was introduced for a purpose which goes beyond the scope of this article. Although this manipulation might be of theoretical interest for the present article as well (e.g., suspects might be more aroused than truth tellers and their responses might subsequently differ), the manipulation did not result in significant effects on the variables presented in the present article. Since we do not know why the effect was not significant (we cannot rule out that the manipulation was unsuccessful) this factor is ignored throughout this article.

6. At this stage another manipulation took place which was, again, beyond the scope of this article. In order to facilitate the participants' preparation for the interview, they were 'lightly' or 'heavily' coached. Participants in the light coaching condition (N = 127) were told that it would be more likely that the interviewer would believe the participant if he/she told in lots of detail what happened when he/she was in the room. In addition to these light coaching instructions, participants in the heavy coaching condition (N = 69) were taught some of the CBCA criteria. This coaching manipulation is irrelevant for the present article, and had only little effect on the findings presented in this article. This factor is therefore ignored throughout the article, but information will be given about the one occasion where this manipulation did have an effect on the findings. (See endnote 19 and see also Vrij, Akehurst, Soukara, & Bull, 2002).

7. Again, the procedure described in the text was the procedure for the witness condition. After building rapport with the participant, the interviewer gave participants in the suspect condition the following instructions: "The reason I am interviewing you is that I have heard that YOU wiped some important information off the blackboard in the room you were just in. I understand that you were in there just now with (name of the experimenter). I need to know whether it was you or not!" For the remaining part, the suspect interviews and instructions were identical to the witness interviews and instructions. The interviewer knew the status (witness/suspect) of each participant when s/he arrived for his/her interview.

8. A study like this might raise ethical concerns. For example, we instructed participants, including children, to lie. We thought that this was acceptable for the following two reasons. First, the lie was not serious. Second, we know that telling lies is a regular event in life both for children (Ceci & DeSimone Leichtman, 1992) and adults (DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996).

9. Differences between correlations in scale scoring and frequency scoring are probably caused by the fact that scale scoring is more subjective than frequency scoring, That is, using Likert scale scoring (the common procedure in CBCA research) involves two subjective processes: (1) indicating whether or not a criterion is present in a transcript, and (2) deciding to what extent a criterion is present (rarely present, strongly present, etc.). Frequency scoring only contains the first step, and is therefore, compared to Likert scale scoring, less subjective. Perhaps not surprisingly, the correlations in frequency scoring were generally higher than the correlations in Likert scale scoring. Analyses on the basis of frequency scoring are not reported in this article, as this type of scoring is not common practice in CBCA assessments. Also, analyses based upon scale scoring (presented in this article) were highly similar to findings which were obtained on the basis of frequency scoring. Therefore, presenting the frequency scoring results would not add any valuable information to this article. In other words, frequency scoring was solely conducted for inter-rater reliability purposes.

10. RM scores cannot be averaged (like CBCA scores) as this is impossible with scores based on frequency scoring: It would give a variable with a high frequency of occurrence more weight than a variable with a low frequency of occurrence.

11. Separate median splits were used for 5-6-year-olds and the remaining participants for the following reason: An ANOVA with Age Group as factor and visual details as dependent variable revealed a significant effect, $F(3, 192) = 17.86$, $p < .01$. Tukey HSD tests revealed that 5-6-year olds included significantly fewer details in their statements ($M = 19.79$, $SD = 10.9$) than any of the other three groups (10-11-year-olds: $M = 43.77$, $SD = 18.5$; 14-15-year-olds: $M = 39.63$, $SD = 12.78$; undergraduates: $M = 36.75$, $SD = 18.0$). These

three groups did not differ significantly from each other. A median split for the whole group would therefore imply that almost all 5-6-year-olds would be allocated to the "low score" group and relatively many other participants to the "high score" group (as the median split would be relatively low). We therefore decided to use two different median splits, one for the 5-6-year olds (M = 18.00) and one for the other participants (M = 37.50).

12. All truthful participants participated in more or less the same staged event. It might therefore be that the truths told by these participants bore certain similarities. These similarities could be picked up by the CBCA, Reality Monitoring and nonverbal raters after a few trials of coding, and this "knowledge" might have affected their codings. We do not think that this actually happened. Although there were similarities in the staged event, there were also differences which were purposefully introduced by us in order to prevent this happening. For example, different people were used to come into the room and wipe the blackboard. Also, the same person wore different clothes at different times. As a result, the descriptions of the "actor" differed considerably even in the truthful condition. (Because variations of the same staged event were introduced, we videotaped all participants while they were in the room with the experimenter, and checked the veracity of their stories afterwards by comparing their statements with what actually had happened. We did not come across any commissions, purposefully distorting the truth, in the truthful reports). Additionally, not all truthful participants gave a complete account, with some participants describing some features and other participants describing totally different features.

13. All behaviors, except gaze aversion, were selected because we expected that liars had to think hard in the present study, and these behaviors are particularly associated with cognitive load (Burgoon et al., 1989; Ekman & Friesen, 1972; Goldman-Eisler, 1968; Vrij, 2000). Gaze aversion was selected because this is probably the most powerful subjective cue to deception, that is, when people are asked how they think liars behave, their most likely answer is that liars look away (Akehurst, Köhnken, Vrij, & Bull, 1996; Vrij & Semin, 1996).

14. Fewer limb movements were particularly caused by fewer hand and finger movements ($F(3, 192) = 9.90$, $p < .01$, $eta^2 = 13\%$: 5-6-year-olds: $M_{ab} = 24.53$, $SD = 26.6$; 10-11-year-olds: $M_c = 38.91$, $SD = 22.1$; 14-15-year-olds: $M_{bc} = 33.01$, $SD = 27.1$, undergraduates: $M_a = 15.90$, $SD = 15.6$) and fewer foot and leg movements ($F(3, 192) = 7.85$, $p < .01$, $eta^2 = 11\%$: 5-6-year-olds: $M_c = 37.35$, $SD = 18.6$; 10-11-year-olds: $M_{ab} = 24.16$, $SD = 21.5$; 14-15-year-olds: $M_{bc} = 29.38$, $SD = 32.5$; undergraduates: $M_a = 13.93$, $SD = 14.6$). (Only mean scores with a different subscript differ significantly, $p < .05$, from each other).

15. More speech disturbances were caused by more speech hesitations ($F(3, 192) = 26.05$, $p <. 01$, $eta_2 = .29$: 5-6-year-olds: $M_a = 3.68$, $SD = 4.0$; 10-11-year-olds: $M_a = 4.28$, $SD = 3.3$; 14-15-year-olds: $M_a = 3.51$, $SD = 3.3$; undergraduates: $M_b = 9.46$, $SD = 5.0$). (Only mean scores with a different subscript differ significantly, $p < .05$, from each other).

16. Although there was a decrease during deception of all types of limb movements (adaptors, illustrators, hand and finger movements and foot and leg movements), only the decrease in hand and finger movements was significant (liars: $M = 24.93$, $SD = 22.7$, truth tellers: $M = 31.95$, $SD = 25.8$, $F(1, 188) = 6.97$, $p < .01$, $eta^2 = .02$).

17. Auditory details, 5-6-year-olds: liars, $M = .46$, $SD = .6$, truth tellers, $M = .62$, $SD = 1.1$; 10-11-year-olds: liars, $M = 2.29$, $SD = 2.2$, truth tellers, $M = 4.27$, $SD = 2.8$; 14-15-year-olds: liars, $M = 2.13$, $SD = 2.3$, truth tellers, $M = 4.79$, $SD = 3.3$; undergraduates: liars, $M = 3.08$, $SD = 2.5$, truth tellers, $M = 6.92$, $SD = 4.5$.

Spatial details, 5-6-year-olds: liars, $M = .68$, $SD = .5$, truth tellers, $M = .98$, $SD = 1.1$; 10-11-year-olds: liars, $M = 1.88$, $SD = 1.8$, truth tellers, $M = 2.33$, $SD = 1.6$; 14-15-year-olds: liars, $M = 1.52$, $SD = .9$, truth tellers, $M = 2.12$, $SD = 1.4$; undergraduates: liars, $M = 2.62$, $SD = 2.3$, truth tellers, $M = 5.63$, $SD = 5.7$.

Temporal details, 5-6-year-olds: liars, $M = .43$, $SD = .5$, truth tellers, $M = .83$, $SD = 1.0$; 10-11-year-olds: liars, $M = 2.91$, $SD = 2.3$, truth tellers, $M = 3.42$, $SD = 2.7$; 14-15-year-olds: liars, $M = 2.37$, $SD = 1.6$, truth tellers, $M = 3.28$, $SD = 2.1$; undergraduates: liars, $M = 3.58$, $SD = 2.3$, truth tellers, $M = 6.63$, $SD =$

4.4.

18. Analyses for each type of limb movement (adaptors, illustrators, hand and finger movements and foot and leg movements) per Age Group revealed that significant findings only emerged for hand and finger movements. Both 5-6-year-olds and college students showed fewer hand and finger movements when lying (5-6-year-olds: M(lie) = 12.42, SD = 11.7, M(truth) = 32.61, SD = 30.7, F(1, 33) = 5.45, p < .05, $eta^2$ = ,14; college students: M(lie) = 10.55, SD = 8.5, M(truth) = 21.24, SD = 19.2, F(1, 50) = 6.77, p < .05, $eta^2$ = .12). The difference in hand and finger movements was marginally significant for 14-15-year-olds, M(lie) = 26.32, SD = 21.8, M(truth) = 39.01, SD = 30.1, F(1, 53) = 3.14, p = .08, $eta^2$ = .06.

19. For undergraduates, hit rates for CBCA scores were different for lightly or heavily coached participants (see endnote 6). When heavily coached participants were excluded from the analysis (69 participants were heavily coached), the total hit rate was 71% with 75% for lie detection and 67% for truth detection.

20. Most, but not all, intercorrelations between the nine variables were low. Thirty out of 36 correlations were r's < .30, but three correlations were r's > .50. These were total CBCA - total RM score, r(196) = .60, total CBCA score - speech rate, r(196) = .70, and total RM score - speech rate, r(196) = .51.

21. In the discriminant analysis of the present study, described before, total CBCA score, total RM score, cognitive operations, gaze aversion, movements, latency time, pauses, speech disturbances and speech rate were included.

22. An alternative explanation for the lack of differences in RM scores between young liars and truth tellers is that both groups lack the ability to provide details. However, CBCA scores between truthful and deceptive statements did differ for young children which seems to contradict this explanation.

23. Total sample: lie hit rate = 65%, truth hit rate = 64%, total hit rate = 64%, eigenvalue = .16, Lambda = .86, $X^2$(2, N = 196) = 29.05, p < .01; 5-6-year-olds: lie hit rate = 71%, truth hit rate = 71%, total hit rate = 71%, eigenvalue = .29, Lambda = .77, $X^2$(2, N = 35) = 8.23, p < .05; 10-11-year-olds: lie hit rate = 68%,

truth hit rate = 58%, total hit rate = 63%, eigenvalue = .14, Lambda = .88, $X^2(2, N = 54) = 6.67$, $p < .05$; 14-15-year-olds: lie hit rate = 65%, truth hit rate = 76%, total hit rate = 71%, eigenvalue = .42, Lambda = .70, $X^2(2, N = 55) = 18.31$, $p < .01$; undergraduates: lie hit rate = 58%, truth hit rate = 73%, total hit rate = 65%, eigenvalue = .23, Lambda = .81, $X^2(2, N = 52) = 10.09$, $p < .01$.