Law Hum Behav (2007) 31:499–518
DOI 10.1007/s10979-006-9066-4

ORIGINAL ARTICLE

Cues to Deception and Ability to Detect Lies as a Function of Police Interview Styles

Aldert Vrij · Samantha Mann · Susanne Kristen · Ronald P. Fisher

Published online: 9 January 2007

Abstract In Experiment 1, we examined whether three interview styles used by the police, accusatory, information-gathering and behaviour analysis, reveal verbal cues to deceit, measured with the Criteria-Based Content Analysis (CBCA) and Reality Monitoring (RM) methods. A total of 120 mock suspects told the truth or lied about a staged event and were interviewed by a police officer employing one of these three interview styles. The results showed that accusatory interviews, which typically result in suspects making short denials, contained the fewest verbal cues to deceit. Moreover, RM distinguished between truth tellers and liars better than CBCA. Finally, manual RM coding resulted in more verbal cues to deception than automatic coding of the RM criteria utilising the Linguistic Inquiry and Word Count (LIWC) software programme.

In Experiment 2, we examined the effects of the three police interview styles on the ability to detect deception. Sixty-eight police officers watched some of the videotaped interviews of Experiment 1 and made veracity and confidence judgements. Accuracy scores did not differ between the three interview styles; however, watching accusatory interviews resulted in more false accusations (accusing truth tellers of lying) than watching information-gathering interviews. Furthermore, only in accusatory interviews, judgements of mendacity were associated with higher confidence. We discuss the possible danger of conducting accusatory interviews.

Keywords Interviewing to detect deception · Criteria-based content analysis and reality monitoring · Accusatory interviews, information-gathering interviews and the behaviour analysis interview

This Project was sponsored by a grant from the Economic and Social Research Council (RES-000-23-0292).

A. Vrij (✉) · S. Mann · S. Kristen
University of Portsmouth, Psychology Department, King Henry Building, King Henry 1 Street, Portsmouth, PO1 2DY, United Kingdom
e-mail: aldert.vrij@port.ac.uk

R. P. Fisher
Florida International University, Miami, FL, USA

Cues to deception and ability to detect lies as a function of police interview styles

Information-gathering and accusatory interview styles

Police interviewers frequently need to determine whether a suspect is lying (Horvath, Jayne, & Buckley, 1994). In the service of detecting deception, researchers have monitored a variety of verbal and nonverbal behaviours which they believe might discriminate between liars and truth tellers (DePaulo et al., 2003). Only recently, however, have researchers taken a more proactive role in developing interview protocols to magnify the differences between liars and truth tellers. For example, Vrij (2006) examined whether the style of the interview influenced the likelihood of suspects displaying nonverbal cues to deception. Very few studies have examined the impact of interview style on verbal and nonverbal lie detection (Colwell, Hiscock, & Memon, 2002; Hartwig, Granhag, Strömwall, & Vrij, 2005; Hernandez-Fernaud & Alonso-Quecuty, 1997; Horvath, Jayne, & Buckley, 1994; Levine & McCornack, 2001), and in those that have, different interview styles often used by the police were not directly compared. We examine here whether the interview style affects the likelihood that verbal cues to deception will occur in suspects' statements (Experiment 1), and whether using different interview styles affects police officers' ability to detect deceit (Experiment 2).

In their analysis of audiotaped interviews with suspects in England and Wales, Moston and Engelberg (1993) observed that the police commonly use two types of interview formats: Information-gathering and accusatory. In the information-gathering style, interviewers request suspects to give detailed statements about their activities through open questions (e.g., "What did you do between 3pm and 4pm?"; "You just mentioned that you went to the gym last night; Who else was there?"). By comparison, in the accusatory style, interviewers confront suspects with accusations (e.g., "Your reactions make me think that you are hiding something from me"). Not surprisingly, interview style affects suspects' verbal responses, with accusatory interviews often yield short denials (e.g., "I am not hiding anything".) whereas information-gathering interviews promote longer, more elaborate responses.

The more words there are in the suspect's response, the more verbal cues there should be to discriminate between liars and truth tellers, simply because words are the carriers of verbal cues. Two such discriminating verbal cues are contextual embedding (reference to place and time) and conversational reproductions (e.g., reproducing speech in its original format), both of which appear more frequently in truth tellers' statements than in liars' statements (Vrij, 2005b). Obviously, these cues will be more available if interviewees provide long, elaborate responses than if they provide only short denials. We expect, therefore, that information-gathering interviews, which generate longer answers, will contain more verbal cues to deception than will accusatory interviews.

A third interview style: Behaviour analysis interview

Inbau, Reid, Buckley, and Jayne (2001) describe in their influential police manual a third interview style: The behavioral analysis interview (BAI). Similar to an information-gathering interview, a typical BAI interview starts with an open-ended question inviting suspects to describe their activities during a specific period of time. After this initial open-ended question, information-gathering and BAI interviews take different paths. In information-gathering interviews, interviewers continue with follow-up questions that are based on the suspect's initial statement, thereby allowing suspects to elaborate on their initial statement. BAI interviewers, however, continue by asking a predetermined standardised list of 15 questions, starting with the

question: "What is your understanding of the purpose of this interview?", followed by questions such as: "Who would have had the best opportunity to commit the crime if they had wanted to?"

Despite its name, *behaviour* analysis interview, Inbau et al. (2001) predict that liars and truth tellers will differ in both their nonverbal behaviour and in their verbal responses. We focus here on the verbal responses. One key assumption behind the BAI interview is that, compared to liars, truth tellers expect to be exonerated and therefore should be inclined to offer helpful information (Horvath, Jayne, & Buckley, 1994; Inbau et al., 2001). Thus, truth tellers should be more likely to name possible suspects, more likely to divulge who had an opportunity to commit the crime, etc. There is no empirical evidence, however, to support these claims. Vrij (2005a) examined the cooperativeness of truth tellers and liars in a mock crime scenario, and found just the opposite: Truthful suspects were *less* co-operative than deceptive suspects. Similarly, Vrij, Mann, and Fisher (2006) examined the BAI technique and found that truthful mock suspects were *less* helpful than deceptive mock suspects. In summary, the experimental research refutes Inbau et al.'s (2001) predictions about verbal cues to deception. These findings do not rule out the possibility that BAI interviews may reveal verbal cues to deceit *other* than those that Inbau et al. focus on. We explored the availability of such verbal cues in the present experiment.

There is good reason to believe that the BAI technique is used frequently. Reid and Associates claim to have trained more than 300,000 law enforcement professionals worldwide, and the BAI technique is part of their training package (see http://www.reid.com/training_programs/r_interview.html). The BAI technique is also believed to be one of the two most commonly taught questioning methods in the US (Frank Horvath, 2006, personal communication).

Verbal cues to deception: Criteria-Based Content Analysis and Reality Monitoring

Different tools exist to examine verbal indicators of deceit. Probably the most widely known and most widely researched tool is Criteria-Based Content Analysis (CBCA, Steller & Köhnken, 1989), which emerged from German psychologists' practice in interviewing children (Masip, Sporer, Garrido, & Herrero, 2005). CBCA was originally developed to assess the veracity of statements made by children in alleged sexual abuse cases, but researchers have advocated and demonstrated that it can also be used to assess the veracity of statements made by adults who talk about issues other than sexual abuse (Vrij, 2005b).

CBCA experts assess transcribed oral statements and judge the presence of 19 criteria. These criteria include *logical structure* of the statement, *contextual embeddings* (references to time and space), *descriptions of interactions*, *reproduction of speech* (speech in its original form), *accounts of subjective mental state* (feelings experienced), *spontaneous corrections* (corrections made without prompting from the interviewer), and *admitting lack of memory* (expressing concern that some parts of the statement might be incorrect). See Köhnken and Steller (1988), Ruby and Brigham (1997), Steller and Köhnken (1989), Raskin and Esplin (1991) and Vrij (2005b) for detailed descriptions of these criteria. The underlying hypothesis is that these criteria occur more frequently in truthful than in fabricated stories (Steller, 1989).

A second verbal tool to assess veracity is Reality Monitoring (RM). The core of RM is 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 other features, *sensory 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 the time order of events, e.g., "First he switched on the video-recorder and then the TV", and details about the duration of events). By contrast, 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 may argue that "experienced events" reflect truth telling whereas "imagined events" reflect deception. Consequently, compared to liars' statements, truth tellers' statements will include more sensory and contextual information and fewer cognitive operations. See Masip et al. (2005), Sporer (2004) and Vrij (2000) for reviews of RM deception research.

RM has several advantages over CBCA. It has a stronger theoretical foundation (Masip et al., 2005; Sporer, 1997, 2004); it is relatively easy to teach and to learn, and less time consuming to apply (Sporer, 1997; Vrij et al., 2004b); and often leads to higher inter-rater reliability (Sporer, 1997; Strömwall, Bengtsson, Leander, & Granhag, 2004; Vrij, Edward, Roberts, & Bull, 2000; Vrij, Akehurst, Soukara, & Bull, 2004a). Finally, studies where statements were analysed with both the CBCA and RM tools have revealed that either RM discriminates between liars and truth tellers better than does CBCA (Vrij et al., 2004a, b; Granhag, Strömwall, & Landström, 2006; Strömwall et al., 2004) or that both tools have similar ability (Sporer, 1997; Vrij et al., 2000). It is therefore remarkable that, of the two tools, CBCA is currently used in criminal investigations (Köhnken, 2004), whereas, to our knowledge, RM is not.

The first experiment

In the first experiment, participants were requested to lie or tell the truth about a staged event in one of three interview settings: Information-gathering, accusatory or BAI. The oral statements were transcribed and assessed via the CBCA and RM methods. We predicted that the accusatory interviews would elicit shorter responses than the information-gathering and behaviour analysis interviews, and that they would reveal fewer CBCA- and RM- related verbal cues to deceit.

Bond and Lee (2005, p. 326) recently concluded that "the jury is still out" regarding whether RM could be coded by using a computer software programme (Linguistic Inquiry and Word Count, LIWC) rather than manually, which is the typical method of RM coding. Automatic coding would have advantages. Texts could be scrutinised more quickly, and more objectively, because human coders, whose coding necessarily depends on their interpretations of the transcripts, are not necessary. LIWC (Pennebaker, Francis, & Booth, 2001) contains word categories that correspond with RM labels such as "senses," "space," "time" and "cognitive mechanisms." Bond and Lee (2005) obtained mixed success in differentiating between liars and truth tellers by using these RM-type LIWC word categories. In their experiment, truth tellers obtained a significantly higher score for sensory details than liars, but a significantly lower score for spatial details than liars. The latter finding contradicts RM theory. The only other research project (comprising five experiments) that we are aware of where the RM-type LIWC categories were examined did not yield significant effects for these LIWC categories (Newman, Pennebaker, Bery, & Richards, 2003).

Bond and Lee (2005) and Newman et al. (2003) did not carry out manual RM coding on their data, so their studies do not indicate how effective automatic RM coding (with LIWC) is compared to manual RM coding. We therefore examined this in the present experiment. We doubted whether automatic RM scoring would be as effective as manual scoring for a variety of reasons. First, the results of Newman et al. (2003) and Bond and Lee (2005), described above, do not give much reason for optimism. Second, although the LIWC categories may resemble the RM categories, they are not developed on the basis of RM theory. This lack of theoretical foundation may cause error. For example, the LIWC cognitive mechanism category includes words such as "think". Thus, the sentence "I think she had dark hair" would produce a hit in the

LIWC cognitive mechanism category. By comparison, human RM coders do not count this as a cognitive operation. We therefore predict that LIWC coding will be less successful than manual coding in discriminating between liars and truth tellers.

Experiment 1

Method

Participants

The participants were 120 undergraduate students, of whom 58% were male and 42% were female. Their average age was $M = 22.07$ ($SD = 6.46$) years.

Procedure

The experiment took place at a Students' Union in a British university. Undergraduates were recruited under the guise of participating in an experiment about "telling a convincing story" with the possibility of earning £15. The participants signed an informed consent form, and then were randomly allocated to the truth telling condition or the deception condition.

The 60 truth tellers participated in a staged event in which they played a game of Connect 4 with a confederate (who posed as another participant). (Connect 4 is a popular two-player game where players drop counters into a slotted grid to achieve, and simultaneously prevent their opponent from achieving, four of their counters in a row). During the game they were interrupted twice, first by a second confederate who came in to wipe a blackboard and later by a third confederate who entered looking for his or her wallet. Upon finding the wallet, this latter confederate then claimed that a £10 note had gone missing from it. The participant was then told that s/he would be interviewed about the missing money. This event is a modification of Vrij, Akehurst, Bull, and Soukara (2002).

The 60 liars did not participate in this staged event. Instead, they were asked to take the £10 from the wallet, but deny having taken this money in a subsequent interview. They were told to tell the interviewer that they played a game of Connect 4 like the truth tellers had. The liars were then presented with a sheet containing the following information about the staged event that the truth tellers had participated in.

> You enter the room to find another participant, 'Sam'. The two of you are instructed by the experimenter to play Connect 4 for a while, which you do alone together. The other participant sat where the experimenter was just sat and you sat where you are sitting now. You had a general conversation with the other participant as you played (e.g. about your courses/life as a student in Portsmouth/TV/the weather). Then the other participant's mobile phone rang and, clearly an important personal call, they excuse themselves and leave the room, leaving you alone for a minute or so. Then they return and you both continue playing the game. Then someone else entered the room, made a comment about you playing the game, wiped the mathematical formulas that you can see off the board and then left. You continued playing the game for a few moments when someone else entered the room looking for his/her wallet. The person made several comments when they entered and had clearly been looking for it for a while. The wallet which you can see in front of you, is found somewhere around the room (up to you to decide where – it was varied in the scenario). You continue playing the game when the experimenter came back in, with

> the wallet-owner, and informs you and the other participant that some money had gone missing from the wallet and you are both to be interviewed. You now have a few minutes to familiarise yourself with this alibi before going into the next room to give your story to the interviewer. Remember to give as much detail as you can about what happened in order to make your story convincing.

Just before the interview started, both liars and truth tellers were told that if they convinced the interviewer that they did not take the money, they would receive £15 for participating in this study. If they did not convince the interviewer, they would have to write a statement about what actually occurred.

In summary, the liars did not engage in any of the activities the truth tellers were engaged in (playing Connect 4, etc.). Instead, the liars took the money out of the wallet, hid it somewhere on their person, and pretended that they had been engaged in the truth tellers' activities. They therefore lied about the entire scenario, including taking £10 from the wallet. The procedure reflects a situation where a liar is familiar with the event s/he described but lacks the experience of true participation in that event.

All 120 participants were interviewed by the same uniformed, male, British police officer. The interviewer was blind to the participants' condition (truth telling or lying). The interviewer started the interview by saying "£10 has gone missing from a wallet in the room next door and I have to find out whether or not it was you who took it." After several introductory questions, the actual interview commenced. Participants were allocated randomly to the information-gathering, accusatory and behaviour analysis interview conditions.

Participants in the *information-gathering* condition (20 liars and 20 truth tellers) were asked to tell in as much detail as possible what happened when they played Connect 4. Several questions followed, such as: "You just mentioned that someone came into the room who rubbed information off the board. Can you describe that person in detail?". Participants in the *accusation* condition (20 liars and 20 truth tellers) were asked eleven questions adapted from Vrij and Winkel (1991), including: "Are you sure you're telling me the truth?," "You forgot to mention the £10 note that you took from the wallet, didn't you?," "Your reactions make me think that you are hiding something from me," etc. Participants in the *behaviour analysis interview* condition (20 liars and 20 truth tellers) were asked first to report in as much detail as possible what happened when they played Connect 4. After this recall, they were asked the 15 behaviour analysis interview questions. They were directed towards the critical event, the theft of money from a wallet. Examples of the questions include: "Is there anyone other than yourself who you feel certain did not take the money?", "Do you think that someone did actually purposefully take the money?", and "Who would have had the best opportunity to have taken the money if they had wanted to?" See Vrij, Mann and Fisher (2006) for a description of the 15 questions asked.

After the interview the police officer gave each participant a questionnaire, which he or she completed in another room. Participants were asked (i) to what extent they were motivated to appear convincing during the interview, (ii) what they thought the likelihood was of getting the £15, and (iii) what they thought the likelihood was of being made to write a statement. Answers were given on Likert scales ranging from (1) very unlikely to (7) very likely. After each participant completed the questionnaire, the experimenter told him or her that the police officer had been convinced by his or her story. Thus, both the truth tellers and the liars were paid £15.

Manual CBCA and RM coding

The interviews were simultaneously videotaped and audiotaped, and then transcribed. These transcripts were the basis for CBCA and RM coding. The transcripts were scored by a CBCA

expert and a second rater. The second rater received training in CBCA scoring by the expert. First, the rater read several descriptions of the CBCA criteria provided by Raskin and Esplin (1991), Steller (1989) and Vrij (2000). Then, the expert explained and gave examples of each criterion. After that, both trainee and expert worked together to evaluate several example scripts together (from Vrij et al., 2002, study). Finally, the expert and trainee rated a few scripts individually. The trainee and expert compared their results, and feedback was given by the expert rater. Then, the trainee received more transcripts which she rated herself. In a follow-up meeting, the results were evaluated and again the expert provided feedback. After this meeting, both the rater and the expert felt that the trainee had been adequately trained and that the coding could commence. Both expert and second rater coded the scripts individually. Both 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 not). One of the CBCA criteria (number 10) "accurately reported details misunderstood" was not scored as it is specifically relates to young children. Criterion 19 "Details characteristic of the offence" was not used either because it specifically relates to sexual crimes. The CBCA expert scored the frequency of occurrence of each criterion in each statement. Following common procedure, repeated information was not counted twice. After completing this frequency scoring, the expert then scored the presence of each criterion on 5-point Likert scales, (1) = absent, and (5) = strongly present, by using the frequency scores. This was carried out in a mechanical manner. For example, regarding information-gathering and BAI-interviews, when more than 56 details (criterion 3) occurred in a statement, a score of '5' was given, when 41 to 55 details were mentioned a score of '4' was given, etc. The cut-off points were derived following inspection of the frequency distribution and assurance that we achieved a reasonable spread on the 5-point Likert scales. A total CBCA score was calculated by adding the Likert scale scores of these 17 CBCA criteria, and this is the score we used throughout the analyses. We used total CBCA scores because total CBCA scores are typically used in real-life cases (Gumpert & Lindblad, 1999; Köhnken, 2004).

In order to check for inter-rater reliability, the second coder also conducted the frequency scoring ratings, and conducted the Likert scale transformations on her frequency scores on 50% of the transcripts. We then calculated the total CBCA scores for this second coder. These CBCA scores correlated highly with the CBCA scores of the CBCA expert ($r(60) = .89$). The correlations were also satisfactory if broken down per interview style (all $r(20)$'s $> .67$.[1]

Two other raters received training in Reality Monitoring (RM) scoring. An RM expert (who was also the CBCA expert) provided the raters with a detailed description of how the criteria should be scored, including some case examples. Then, both the trainees and the expert evaluated some example transcripts individually (from Vrij et al., 2002, 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

[1]The Pearson correlations between the two coders for the frequency scores were as follows: logical structure, $r = .36$; unstructured production, $r = .50$; quantity of details, $r = .98$; contextual embedding, $r = .96$; description of interactions, $r = .51$; reproduction of conversation, $r = .97$; unexpected complications, $r = .61$; unusual details, $r = .71$; superfluous details, $r = .60$; related external associations, $r = .37$; subjective mental state, $r = .79$; attribution of other's mental state, $r = .85$; spontaneous corrections, $r = .67$; admitting lack of memory, $r = .82$; raising doubts about one's own memory, $r = .44$; self-deprecation, $r = .62$; pardoning the perpetrator, $r = .53$. The correlations indicate fair to very good inter-rater reliability (Fleiss, 1981; Gödert, Gamer, Rill, & Vossel, 2005). The relatively low agreement scores for "logical structure" and "related external associations" are probably due to the low frequency of occurrence of these criteria. In low frequency distributions the correlations tend to underestimate the true inter-rater agreement (Gödert et al., 2005). Spearman correlations between the two coders revealed a similar pattern to the Pearson correlations.

agrees with Sporer (1997) who also found it relatively easy to teach (and to learn) RM scoring. RM scoring is probably less standardised than CBCA scoring and different researchers use somewhat different RM scales (see Masip et al., 2005; Sporer, 2004). We used the scales used by Vrij et al. (2000, 2004a, b).

Two trained raters individually coded the statements from the present study. 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 not). The two raters coded per interview the frequency of occurrence of visual details (e.g., "He walked over to the whiteboard" contains three visual details), auditory details (e.g., "She said to sit down" contains one auditory 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 (observations that indicate cognitive suppositions of sensory information, e.g., "She seemed quite clever" contains one cognitive operation). Again, repeated information was counted only once. One rater transformed her frequency scores into 5-point Likert scale scores ((1) = absent and (5) = strongly present) in a mechanical manner (see above). The average Reality Monitoring score was based on these Likert scores and calculated as follows: visual score + auditory score + spatial score + temporal score − cognitive operations score. The second rater also conducted the frequency scoring and conducted the Likert-scale transformations on 50% of the transcripts. We then calculated a total RM score on the basis of the second coder's ratings and this total RM score correlated highly with the total RM score of the first rater $r(60) = .92$. Correlations were also high if broken down per interview style, all $r(20)$'s $> .82$. RM total score and CBCA total scores were also significantly correlated ($r(120) = .67$, $p < .01$).[2,3]

LIWC coding

The transcripts were prepared for LIWC analyses according to the LIWC manual (Pennebaker et al., 2001). Thus, all interviewers' texts were deleted, the interviewees' texts were searched for spelling errors and were corrected, fillers such as "you know" were changed into "you know," etc.

Results

Manipulation checks

A 2 (Veracity) × 3 (Type of Interview) MANOVA with the three manipulation checks as the dependent variables, did not result in any main or interaction effects (all F's < 1.04). The vast majority of participants (85%) reported that they were motivated to appear convincing during the interview (a score of 5 or higher on the 7-point scale); 28% thought that it was unlikely that they

[2]Intercoder reliability scores (Pearson's correlations) on the frequency scores were good for all the individual criteria (visual details: $r = .98$; auditory details: $r = .98$; spatial details: $r = .89$; temporal details: $r = .95$; cognitive operations: $r = .94$). Spearman correlations between the two coders revealed a similar pattern to the Pearson correlations.

[3]The CBCA and RM inter-rater reliability scores were also calculated per interview condition. The correlations for the information-gathering and behaviour analysis interview conditions were very similar to the correlations reported in the text. Several reliability scores could not be calculated for the accusatory condition because several criteria were never present in that condition. Those that could be calculated were very good for CBCA scores (all r's $> .80$) and good for RM scores (all r's $> .65$).

would be getting the £15 (a score of 3 or lower on the 7-point Likert scale); and 29% thought that it was likely that they would have to write a statement about the event (a score of 5 or higher on the 7-point Likert scale). In summary, the participants were motivated to be convincing and the incentive and threat appeared realistic.

Length of interview

In order to examine differences in length of interview, a 2 (Veracity) × 3 (Type of Interview) ANOVA was carried out with number of words spoken by the interviewee as the dependent variable. The length of truthful ($M = 379.72$, $SD = 303.8$) and deceptive ($M = 341.13$, $SD = 227.5$) statements did not differ significantly from each other, $F(1, 114) = 1.38$, *ns*, $eta^2 = .02$. However, as expected, there was a significant main effect for Type of Interview, $F(2, 114) = 73.58$, $p < .01$, $eta^2 = .56$. Tukey posthoc tests revealed that the accusatory interviews ($M = 79.20$, $SD = 54.7$) were significantly shorter than the information-gathering ($M = 514.23$, $SD = 255.7$) and behaviour analysis interviews ($M = 487.85$, $SD = 169.3$). The latter two types of interview did not differ significantly from each other. The veracity × Type of Interview interaction effect was not significant, $F(2, 114) = .82$, *ns*, $eta^2 = .02$.

Manual CBCA and RM coding

Two 2 (Veracity) × 3 (Type of Interview) ANOVAs were carried out with the CBCA and RM scores as dependent variables. The CBCA analysis resulted in a significant Type of Interview main effect, $F(2, 114) = 43.66$, $p < .01$, $eta^2 = .43$. Tukey post hoc tests revealed that, as predicted, CBCA scores in the accusatory interview ($M = 23.03$, $SD = 2.1$) were significantly lower than those in the information-gathering ($M = 31.10$, $SD = 6.2$) and BAI interviews ($M = 30.82$, $SD = 4.1$), with the latter two conditions not significantly differ from each other. The main Veracity effect, $F(1, 114) = 3.82$, *ns*, $eta^2 = .03$, and The Veracity × Type of Interview interaction, $F(2, 114) = 1.28$, *ns*, $eta^2 = .02$, were not significant.

In order to test our hypothesis, planned comparisons were conducted. CBCA scores differed significantly between liars and truth tellers in the information-gathering condition, as predicted, with the CBCA scores being higher in the truth telling condition.[2] By comparison, CBCA scores did not differ significantly between liars and truth tellers in the accusatory and BAI conditions (see Table 1).

The ANOVA of RM scores revealed main effects for Veracity, $F(1, 114) = 12.67$, $p < .01$, $eta^2 = .10$ and Type of Interview, $F(2, 114) = 16.78$, $p < .01$, $eta^2 = .23$. Conforming to RM theory, RM scores were significantly higher for truth tellers ($M = 8.77$, $SD = 3.9$) than for liars ($M = 6.67$, $SD = 3.5$). Post hoc Tukey tests showed that RM scores were significantly lower in the accusatory interview ($M = 5.30$, $SD = 2.5$) than in the information-gathering ($M = 8.93$, $SD = 3.6$) and BAI ($M = 8.93$, $SD = 4.$) interviews, whereas the scores in the latter two conditions

[2] Univariate tests on the individual CBCA criteria (frequency scores) revealed that liars and truth tellers significantly differed on contextual embeddings, $F(1, 38) = 4.11$, $p < .05$, $eta^2 = .10$; description of interactions, $F(1, 38) = 4.33$, $p < .05$, $eta^2 = .10$; reproduction of conversations, $F(1, 38) = 3.08$, $p < .05$, one-tailed, $eta^2 = .08$; unusual details, $F(1, 38) = 3.10$, $p < .05$, onetailed, $eta^2 = .08$; and admitting lack of memory, $F(1, 38) = 8.11$, $p < .01$, $eta^2 = .18$. For all these criteria, truth tellers obtained higher scores than liars (contextual embeddings: $M = 20.85$ ($SD = 9.5$) vs $M = 15.40$ ($SD = 7.4$); reproduction of conversations: $M = 2.65$ ($SD = 4.3$) vs $M = .90$ ($SD = 1.1$); unusual details: $M = 3.20$ ($SD = 3.2$) vs $M = 1.85$ ($SD = 1.3$); and admitting lack of memory: $M = 4.30$ ($SD = 3.3$) vs $M = 1.90$, $SD = 1.7$)). The exception was description of interactions where, in contrast to CBCA predictions, truth tellers obtained a lower score than liars: $M = .10$ ($SD = .3$) vs $M = .45$ ($SD = .7$).

Table 1. CBCA and RM Scores as a function of veracity and type of interview

	Truth		Lie			
	M	*SD*	*M*	*SD*	*F*(1, 38)	eta^2
CBCA-scores						
Info-gathering	32.75	8.1	29.45	2.9	2.96[a*]	.07
Accusatory	23.15	1.6	22.90	2.5	.14	
Bai	31.40	5.0	30.25	2.9	.80	
RM-scores						
Info-gathering	10.15	3.2	7.70	3.7	5.04 *	.12
Accusatory	5.40	1.7	5.20	3.2	.06	
Bai	10.75	3.8	6.67	3.5	10.26[**]	.21

[a]One-tailed test.

*$p < .05$.

**$p < .01$.

did not differ significantly from each other. The Veracity × Type of Interview interaction effect just failed to reach a significant effect, $F(2, 114) = 2.94$, $p = .057$, $eta^2 = .05$.

Planned comparisons (Table 1) revealed that RM scores differed significantly between liars and truth tellers in the information-gathering and BAI conditions, and, as predicted, the RM scores were higher in the truth telling condition.[5] Accusatory interviews did not reveal significant differences between liars and truth tellers.

In order to examine which part of the BAI interview (the information-gathering part, the 15 questions part, or both), caused the difference in RM scores between truth tellers and liars, RM total scores were calculated for the two parts (information-gathering and 15 questions) separately. Two ANOVAs with Veracity as factor and the RM scores for the separate phases as dependent variables revealed that the RM scores between liars ($M = 7.70$, $SD = 3.0$) and truth tellers ($M = 10.80$, $SD = 3.5$) significantly differed in the information-gathering part of the interview, $F(1, 38) = 8.92$, $p < .01$, $eta^2 = .19$, but not in the 15 questions part of the interview, $F(1, 38) = 2.76$, *ns*, $eta^2 = .07$.

Manual vs automatic RM coding

A 2 (Veracity) × 3 (Type of Interview) MANOVA was carried out with the four LIWC categories as the dependent variables: senses, time, space and cognitive mechanisms. At a multivariate level the analysis revealed a significant main effect for Type of Interview, $F(8, 222) = 46.67$, $p < .01$, $eta^2 = .63$, and a significant Veracity × Type of Interview interaction, $F(8, 222) = 2.11$, $p <$

[5]Univariate tests on the individual RM criteria (frequency scores) in the *information-gathering* condition revealed that liars and truth tellers significantly differed on auditory details, $F(1, 38) = 7.45$, $p < .01$, $eta^2 = .16$, spatial details, $F(1, 38) = 16.62$, $p < .05$, $eta^2 = .30$, and temporal details, $F(1, 38) = 7.73$, $p < .01$, $eta^2 = 16$. For all these variables, truth tellers obtained higher scores than liars (auditory details: $M - 18.40$ ($SD - 12.7$) vs $M = 10.35$ ($SD = 3.7$); spatial details: $M = 6.35$ ($SD = 3.7$) vs $M = 2.75$ ($SD = 1.5$); and temporal details: $M - 11.95$ ($SD - 6.8$) vs $M - 7.00$ ($SD - 4.4$)). Univariate tests on the individual RM criteria (frequency scores) in the behaviour analysis interview condition revealed that liars and truth tellers significantly differed on auditory details, $F(1, 38) = 13.69$, $p < .01$, $eta^2 = .27$, spatial details, $F(1, 38) = 10.18$, $p < .01$, $eta^2 = .21$, temporal details, $F(1, 38) - 4.35$, $p < .05$, $eta^2 - 10$, and cognitive operations, $F(1, 38) - 4.05$, $p < .05$, one-tailed, $eta^2 - .10$. In agreement with RM theory, truth tellers obtained higher scores than liars for auditory details ($M = 8.75$ ($SD - 3.9$) vs $M - 4.85$ ($SD - 2.6$), spatial details ($M - 5.45$ ($SD - 2.8$) vs $M - 2.90$ ($SD - 2.5$), and temporal details ($M = 10.25$ ($SD = 4.9$) vs $M = 7.50$ ($SD = 3.3$)). Also in agreement with RM theory, liars ($M = 2.85$ ($SD = 2.1$) obtained higher scores for cognitive operations than truth tellers ($M = 1.65$ ($SD = 1.7$)).

Table 2 LIWC Categories as a function of interview type

	Info-gathering		Accusatory		Bai	
	M	*SD*	*M*	*SD*	*M*	*SD*
Senses	6.75^{c}	2.3	1.80^{a}	1.7	5.11^{b}	1.5
Space	7.64^{b}	3.1	1.65^{a}	1.8	7.14^{b}	2.6
Time	6.64^{b}	3.1	1.15^{a}	1.7	6.98^{b}	2.2
Cognitive mechanisms	9.86^{b}	2.7	6.06^{a}	4.3	14.66^{c}	3.1

Note. Only mean scores with a different superscript differ significantly ($p < .05$) from each other.

.05, $eta^2 = .07$. The Veracity main effect was not significant, $F(4, 111) = .17$, *ns*, $eta^2 = .01$. Univariate tests showed that Type of Interview effect was significant for all four dependent variables (senses: $F(2, 114) = 79.78$, $p < .01$, $eta^2 = .58$; time: $F(2, 114) = 73.11$, $p < .01$, $eta^2 = .56$; space: $F(2, 114) = 67.29$, $p < .01$, $eta^2 = .54$; cognitive mechanisms, $F(2, 114) = 62.40$, $p < .01$, $eta^2 = .52$). Table 2 reveals that the lowest scores were always obtained in the accusatory condition.

At a univariate level, a Veracity × Type of Interview interaction emerged for senses, $F(2, 114) = 6.92$, $p < .01$, $eta^2 = .11$. Table 3 shows that no differences between liars and truth tellers emerged in the accusatory condition. Significant differences emerged in the two other conditions but the findings were contradictory. Truth tellers included more sensory details in their statements than liars in the BAI interview but fewer sensory details in the information-gathering interview.

Discussion

Manual CBCA and RM coding

Information-gathering interviews elicited more verbal cues to deception than did accusatory interviews, as predicted. In fact, accusatory interviews did not result in any verbal cues to deceit, whereas information-gathering interviews led to significant differences in both CBCA and RM scores.

The behaviour analysis interview also resulted in verbal cues to deception, with truth tellers obtaining significantly higher RM scores than liars. Hover, these verbal cues to deception emerged only in the information-gathering part of the BAI method. In other words, the 15 questions part did not add new information about verbal cues to deception. There is also no evidence that the 15 BAI questions are useful for nonverbal lie detection purposes. Inbau et al. (2001) assume that liars feel less comfortable than truth tellers while answering (some of) the 15 questions and, as a result, guilty suspects are more likely to cross their legs, shift in their chair, and perform grooming behaviours. There is no empirical evidence to support these claims. None of these behaviours have been identified as diagnostic cues to deception in deception research

Table 3 LIWC Category senses as a function of veracity and interview type

	Truth		Lie			
	M	*SD*	*M*	*SD*	$F(1, 38)$	eta^2
Information-gathering	5.95	2.4	7.55	1.9	5.55^{a}	.13
Accusatory	2.19	2.0	1.42	1.3	2.01	
Behaviour analysis interview	5.68	1.3	4.55	1.5	6.43^{a}	.15

$^{a}p < .05$.

(DePaulo et al., 2003). In fact, lie detectors who pay attention to such cues perform significantly worse than lie detectors who do not (Kassin & Fong, 1999; Mann, Vrij, & Bull, 2004). In summary, the merit of using the 15 BAI questions as a verbal and nonverbal lie detection tool has not yet been demonstrated.

Many deception studies revealed that liars include fewer details in their statements than truth tellers (Vrij, 2005b). Interestingly, in this experiment neither the broad CBCA-criterion "quantity of details" (Criterion 3) nor the similar RM category "visual details" yielded significant differences between liars and truth tellers (see endnotes 6 and 7). Specific details, however, did differ significantly between liars and truth tellers, such as "reproductions of conversations", "unusual details" (CBCA, endnote 4) and "auditory," "spatial" and "temporal" details (RM, endnote 5). In the present experiment, liars were informed about the details of the staged event truth tellers were involved in. Perhaps such "coaching" provides liars with the opportunity to include much detail in their stories. Nevertheless, the quality of these details gave their lies away.

Our results suggest that RM was a more successful tool to detect deceit than CBCA: RM significantly differentiated between truths and lies in both the information-gathering and BAI conditions, whereas CBCA significantly differentiated only in the information-gathering condition. These favourable findings for RM are in agreement with recent findings (Granhag et al., 2006; Strömwall et al., 2004; Vrij et al., 2004a,b). This is encouraging, particularly because the RM instrument is easy to teach and straightforward to apply. We believe that there is sufficient evidence for the discriminative power of RM to encourage those with a professional interest in lie detection to make themselves familiar with this verbal veracity detection tool and to use it in their daily work. Obviously, applying RM coding could become even more straightforward if such coding could be carried out automatically via computer software programs without any human interference. Our automatic RM coding analyses revealed that this is not feasible and that human interpretation of transcripts is necessary.

Experiment 2

Accuracy in lie detection and false positives

The findings of the first experiment suggest that information-gathering interviews have more potential to discriminate between liars and truth tellers than the accusatory style, because information-gathering interviews result in more verbal cues to deceit than accusatory interviews. This finding converges with Vrij's (2006) experiment showing that information-gathering interviews also revealed more nonverbal cues to deception than accusatory interviews.

The fact that information-gathering interviews reveal more cues to deception than accusatory interviews does not automatically imply that observers will be able to discriminate better between truths and lies in information-gathering interviews. Observers tend to look at nonverbal cues that do not actually discriminate between truth tellers and liars, such as gaze aversion and fidgeting (Vrij, 2000, 2004). They are also largely unaware which verbal cues are related to deception (Strömwall, Granhag, & Hartwig, 2004; Vrij, Akehurst, & Knight, 2006). Lack of knowledge about cues to deception may be one reason why in experimental lie detection studies accuracy rates (i.e., correct classifications of liars and truth tellers) are typically just above the level of chance (Vrij, 2000), even when the observers are professional lie catchers such as police officers (Vrij & Mann, 2005).

In addition to accuracy, false positive accusations (accusing truth tellers of lying) are important, because of the negative consequences they may have for truth tellers. For example, Kassin, Goldstein and Savitsky (2003) found that innocent suspects who are presumed to be

guilty elicit highly confrontational interrogations, and certain commonly used techniques in such interrogations lead innocent suspects to confess to crimes they did not commit (see also Kassin, 2005). We predict that, for a variety of reasons, more false positives will be made in accusatory interviews. First, truth tellers may be more taken aback (than liars) when they are falsely accused (e.g., accusatory interviews) than when they are invited to tell what happened (i.e., information-gathering), and this emotional reaction shines through in their responses. Second, an accusatory interview creates a negative atmosphere (i.e., accusing others) that could easily lead to negative judgements (i.e., judging someone to be lying). Third, when accusatory questions are asked, observers have more opportunities to listen to responses related to guilt than to listen to responses related to innocence. Fourth, the interviewer in the accusatory interviews may be perceived as having an orientation toward guilt. In fact, Inbau et al. (2001) recommend using an accusatory interview style if the interviewer believes that the suspect is guilty. The lie detectors in our study may be sensitive to the beliefs of the interviewer.

One could argue that the BAI technique falls in between the information-gathering and accusatory techniques in terms of interview style. The BAI interview style differs from the information-gathering interview style in asking a predetermined standardised list of 15 questions. Although these 15 questions clearly put the suspect on the spot, they differ from the kinds of questions asked in accusatory interviews in that the suspect is at no point actually accused of wrongdoing or lying. The resultant number of false positives may reflect this mixture of the two other interview styles and may fall in between the information-gathering and accusatory techniques.

Confidence – veracity judgements correlations

DePaulo, Charlton, Cooper, Lindsay and Muhlenbruck's (1997) meta-analysis regarding confidence measures revealed a relationship between confidence scores and type of veracity judgements. Judges are typically more confident in their decision making when they judge someone as telling the truth than when they judge someone as a liar. We believe that this may be true for information-gathering interviews. In such interviews, suspects are encouraged to talk and to discuss what happened. They are neither challenged nor accused of lying. When basing a decision on this, observers may have lower confidence when they decide that someone is lying. However, the opposite could be true for accusatory interviews. In those interviews, the interviewer is searching for signs of guilt, and, when they believe they have found them, may be more confident about the decision they make. Because BAI interviews include elements of both approaches, it is more difficult to predict the relationship between confidence and judgements in these interviews.

Method

Participants

The participants were 68 British police officers, of whom 62% were male and 38% were female. The largest group (49%) were general uniformed officers; an additional 40% were specialised in CID, and 11% were police trainers (police officers who have specialised in providing training courses such as probationer, or interview training, for other police officers). None of the participants had received training in lie detection (such training does not exist in England and Wales). Their average age was $M = 32.87$ years ($SD = 7.5$). Most of the police officers (85%) were Constables and the remaining 15% were Sergeants. Their average length of service in the police was $M = 6.92$ years ($SD = 8.5$). When asked to indicate on a 5-point Likert

scale how experienced they considered themselves in interviewing ($M = 2.53$, $SD = 1.3$), 34% rated themselves as 'inexperienced' (a score of 1 or 2 on the 5-point Likert scale) whereas 25% declared themselves as 'experienced' (a score of 4 or 5 on the 5-point Likert scale). When asked to indicate on a 5-point Likert scale how motivated they were to perform well on the task, 81% reported themselves as fairly or highly motivated (a score of 4 or 5 on the 5-point Likert scale, $M = 4.00$, $SD = .8$).

Procedure

The study took place at training colleges with police constabularies in the South of England. Between seven and fifteen participants were tested simultaneously. This variation in group size reflected only the number of officers that trainers were willing to release from class at that time. It did not in any way affect the running of the experiment. The videotaped interviews (e.g., 'clips') were shown on a large screen (approximately 2 m $\times$ 1 m), in a large classroom that would have enabled twenty participants to have seen the screen clearly, sitting far enough apart so as not to see each other's answers. Participants were given questionnaires and asked to complete the first section relating to the details discussed in the participants section above. They were then informed that they were about to see a selection of clips of students who were either lying or telling the truth, about a scenario that involved the theft of money from a wallet. The scenario involved their playing a game of Connect 4 with another participant (actually a stooge) whilst various people entered or exited the room. Truth tellers had actually participated in this event, and truthfully had not taken any money; liars were merely informed about the event, and had actually taken the money from the wallet. The experimenter did not tell the participants how many clips they would see, or what percentage were truths or lies, so as to avoid participants calculating how many truths and lies they were probably actually being shown, and hence deliberately trying to achieve a certain number of truth/lie responses for just that reason. Instead they were told that although they would not be told how many clips they would see, there would not be as many clips as were in their questionnaire (there was space in the questionnaire for 16 clips). They were told that after each clip the tape would be stopped, and when everybody had completed all questions on the questionnaire relating to that clip, the next clip would be shown. They were then shown one of the three tapes (26 officers saw Tape 1, 18 saw Tape 2, and 24 saw Tape 3), and each tape consisted of 12 videoclips, two lies and two truths of each of the three interview types. After watching each clip the observers were asked to answer the following two questions: (i) Do you think that the suspect is telling ... (dichotomous answer, the truth/a lie), and (ii) How confident are you of your decision? (7 point Likert scale, ranging from (1) not at all to (7) extremely). The study took about one hour to conduct.

Accuracy was measured by calculating the percentage of correct veracity judgements given by each participant in judging the truthful clips (truth accuracy) and deceptive clips (lie accuracy). We further calculated *hits* (percentage of correct classifications of liars) and *false positives* (percentage of truth tellers falsely accused of lying). The confidence in making the veracity judgement was measured by calculating the average confidence scores allocated to liars and to truth tellers.

Results

Overall accuracy scores, and percentages of hits and false positives did not differ per tape, all F's < 2.50, all p's $> .09$, and so we combined the results of the three different tapes.

Table 4 Judgements scores as a function of interview style

	Info-gathering		Bai		Accusatory	
	M	*SD*	*M*	*SD*	*M*	*SD*
Overall accuracy	.52	(.26)	.51	(.23)	.48	(.25)
Hits	.36	(.38)	.39	(.30)	.40	(.35)
False positives	.32	(.35)	.37	(.34)	.43	(.39)
Confidence in decision	4.36	(1.1)	4.69	(1.8)	4.60	(1.0)

Overall accuracy

The overall accuracy scores ranged from a low of 25% to a high of 83% with an average of $M = .50\%$ ($SD = .13\%$). This percentage did not differ significantly from the level of chance (50%). $t(67) = .31$, *ns*.

A 3 (Interview Style) × 2 (Veracity) ANOVA revealed that the Interview Style main effect was not significant, $F(2, 66) = .41$, *ns*, $eta^2 = .01$. Indeed, overall accuracy scores for the three types of interviews were similar (see Table 4) and none of these accuracy scores differed significantly from chance (all $t(67)$'s<.70, all p's > .49). The Veracity main effect was significant, $F(1, 67) = 45.84$, $p < .01$, $eta^2 = .41$, with truths being more accurately judged ($M = .63$, $SD = .21$) than lies ($M = .38$, $SD = .19$). The Interview Style × Veracity interaction effect was not significant, $F(2, 66) = 1.49$, *ns*, $eta^2 = .04$.

A superior truth accuracy often indicates a truth bias, i.e., a tendency to judge clips as truthful. Indeed, the observers thought that the persons in the clips were telling the truth 62% of the time ($SD = .15$), which is significantly more than the level of chance (50%), $t(67) = 6.77$, $p < .01$.

Hits and false positives

An ANOVA with Interview Style as the only factor and hits as dependent variable did not show a significant result, $F(2, 66) = .18$, *ns*, $eta^2 = .01$, and the percentages of hits were very similar across the three interview conditions (see Table 4). As predicted, the percentage of false positives in the accusatory interviews was significantly higher than the percentage of false positives in the information-gathering interviews, $F(1.67) = 3.31$, $p < .05$, $eta^2 = .05$. The percentage of false positives in the BAI interviews fell between those two scores and did not differ significantly from either of them.

Confidence measures

An ANOVA was conducted utilising a 3 (Interview Style) × 2 (Veracity) factorial design with the confidence in decision making as the dependent variable. The analysis resulted in a main effect for Interview Style, $F(2, 66) = 3.57$, $p < .05$, $eta^2 = .10$. The Veracity main effect, $F(1, 67) = .01$, *ns*, $eta^2 = .00$, and the Interview Style X Veracity interaction effect, $F(2, 66) = 2.62$, *ns*, $eta^2 = .07$, were not significant. Confidence scores for the information-gathering interviews were significantly lower than confidence scores for the BAI (see Table 4), $F(1, 67) = 6.15$, $p < .05$, $eta^2 = .08$, and the accusatory interviews, $F(1, 67) = 3.98$, $p < .05$, $eta^2 = .06$, whereas the confidence scores for the accusatory and BAI interviews did not differ from each other, $F(1, 67) = .46$, *ns*, $eta^2 = .01$.

We then conducted three Pearson correlations (one for each interview style) to examine the relationship between confidence scores and veracity judgements. For the accusatory interviews,

as we predicted, the more lie judgements the participants made, the more confident they were in their decisions, $r(68) = .25$, $p < .05$. The correlations for information-gathering interviews, $r(68) = -.05$, and BAI interviews, $r(68) = .03$, were not significant.

Discussion

The present experiment revealed that style of interviewing did not affect on overall accuracy (ability to distinguish between truths or lies) or on lie detection accuracy (ability to correctly identify liars). In fact, the overall accuracy rates were low and did not differ from the level of chance. This study, like so many previous studies (Vrij, 2000), thus shows the difficulty police officers face when discerning truths from lies by observing the suspect's verbal and nonverbal behaviours.

Our study was the first experiment testing the efficiency of BAI interviewing in discriminating between liars and truth tellers. Inbau et al. (2001) suggested that BAI interviews could be used effectively for verbal and nonverbal lie detection purposes. The results of our experiment did not support this claim. Perhaps Inbau et al. (2001) based their claim on the results of the only other, observational, study where the BAI technique has been tested (Horvath, Jayne, & Buckley, 1994). That study, where fragments of real-life suspect interviews were used, revealed that the BAI technique was successful in detecting liars and truth tellers. However, Horvath et al. study had a fundamental methodological weakness: The ground truth (true, actual, status of guilt or innocence of the suspect) was not known for certain. Horvath, Jayne, and Buckley (1994) themselves acknowledge that the interpretation of their own findings "would be less problematic" (p. 805) if the ground truth could have been established. In other words, the ability of the BAI technique to correctly classify liars and truth tellers has not yet been demonstrated.

The inability of police to discriminate between liars and truth tellers, although in itself undesirable, might have only limited consequences if police were aware of their poor performance, i.e., they had good metacognition. Metacognition is important because it often controls behaviour (Koriat & Goldsmith, 1996; Nelson & Narens, 1990). For example, if police officers believe that they are not confident enough to make a veracity judgement then they may refrain from making such judgements and instead decide to further investigate the case (see also Levine & McCornack, 1992). Perhaps evidence about the involvement of the suspect in the case will arise from such investigations. In the present study, police observers were the least confident in their decision making after watching the information-gathering interviews. Given the poor accuracy in discerning truths from lies obtained in this experiment, we believe that this is a positive result for information-gathering interviewing.

In contrast to the null-findings regarding accuracy, interview style did affect false positives (false accusation of truth tellers). As predicted, accusatory interviews resulted in more false positives than information-gathering interviews. This is worrying, particularly because accusatory interviews are typically conducted when police interviewers commence the interview with the opinion that the suspect is guilty (Inbau et al., 2001; Kassin, 2005; Kassin & Gudjonsson, 2004; Moston et al., 1993). In case the suspect is actually innocent, our findings thus suggest that interviewers are more likely to maintain their incorrect assumption of guilt when they conduct accusatory interviews than when they conduct information-gathering interviews. Our additional finding, that in accusatory interviews, judgements of mendacity were associated with higher confidence, further indicates that it is unlikely that interviewers will change their mind in accusatory interviews once they have decided that someone is guilty. These findings have important implications. If police officers think they "know" that a suspect is lying in an accusatory interview and do not change their opinion about this perception of guilt in such an interview, they may well be inclined to put pressure on suspects in order to elicit a confession. This may result in

coerced confessions of innocent suspects (Kassin, 2005). A crucial aspect of police interviewing is to provide safeguards for innocent suspects. In this regard, our data show that accusatory interviews are more dangerous than information-gathering interviews. The false positives rate of BAI interviewing fell in between those of information-gathering and accusatory interviewing. This finding fits well in our observation that BAI interviewing could be seen as a mixture of the other two types of interviewing.

The observers in this study were prone to a truth bias and were inclined to believe that the suspects were telling the truth. The occurrence of a truth bias is a well-established finding in deception research (Vrij, 2000), however, this is typically the case when the observers are laypersons. Professional lie catchers, like the police officers in our study, are usually less inclined to a truth bias (Vrij & Mann, 2005). Sometimes they are prone to a lie bias (Meissner & Kassin, 2002). There may be two reasons why we did not obtain such a lie bias. First, Meissner and Kassin (2002) suggested that (i) general law enforcement experience, and (ii) being trained in lie detection increases the likelihood of judging someone as a liar (i.e., lie bias). Meissner and Kassin (2002) found a positive correlation between general law enforcement experience and increased bias in judging someone as a liar. When we carried out a correlation between length of service and the tendency to judge someone as a liar on our data, it was not significant, $r(68) = .05$, *ns*. The correlation between experience in interviewing and the tendency to judge someone as a liar was not significant either in our experiment, $r(68) = .11$, *ns*. We thus found no support for the suggested link between experience and being prone to a lie bias. Training programmes in lie detection do not exist in the United Kingdom so we could not test the suggested relationship between being trained in lie detection and being prone to a lie bias.

A second explanation why we did not find a lie bias is that this may be a cultural phenomenon. Perhaps American police officers (Meissner and Kassin's participants) are more inclined to a lie bias than British police officers (our participants). At least in publications about police interviewing, there seems to be a cultural difference. British publications emphasise an 'ethical approach' to police interviewing that has 'open mindedness of the interviewer' as a core aspect (e.g., Williamson, 1993). American manuals, on the other hand, mainly emphasise tactics that could be used to break a suspect's resistance in order to obtain confessions (e.g., Inbau et al., 2001). Those tactics assume guilt of the suspect.

The benefits of an information-gathering interview style

There is an increasing body of literature pointing out the benefits of using an information-gathering style of police interviewing. From previous research we already know the following benefits: First, it encourages suspects to talk, and therefore it may provide the police with more information about the alleged event (Fisher, Brennan, & McCauley, 2002). Second, because it does not involve accusing suspects of any wrongdoing, it may be a safeguard against false confessions (Gudjonsson, 2003). Third, this approach may be seen as more ethical (Williamson, 1993). Fourth, compared to accusatory interviewing, it results in more nonverbal cues to deceit (Vrij, 2006). The present experiments revealed three more advantages: Compared to accusatory interviewing, information-gathering interviews result in (i) more verbal cues to deceit, (ii) less confidence in detecting deceit, and, hence, more awareness of the difficulties in detecting deceit, and (iii) it provides safeguards against false accusations of lying.

A final comment

It is perhaps unfair to suggest that police officers use either an entirely accusatory style or an entirely information-gathering style or an entirely behaviour analysis interview style. In practice

elements of all three styles may well be incorporated in one interview. We distinguished between the three styles in our experiments because we can only draw conclusions about the effects of such styles only by examining them in their purest form. On the basis of this distinction we can now predict that the more information-gathering these interviews are, the more verbal cues to deception are likely to occur and the less likely it is that innocent suspects are accused of lying.

References

Bond, G. D., & Lee, A. Y. (2005). Language of lies in prison: Linguistic classification of prisoners' truthful and deceptive natural language. *Applied Cognitive Psychology, 19*, 313–329.

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

DePaulo, B. M. (1994). Spotting lies: Can humans learn to do better? *Current Directions in Psychological Science, 3*, 83–86.

DePaulo, B. M., Charlton, K., Cooper, H., Lindsay, J. L., & Muhlenbruck, L. (1997). The accuracy – confidence correlation in the detection of deception. *Personality and Social Psychology Review, 1*, 346–357.

DePaulo, B. M., Lindsay, J. L., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*, 74–118.

Fisher, R. P., Brennan, K. H., & McCauley, M. R. (2002). The cognitive interview method to enhance eyewitness recall. In M. L. Eisen, J. A. Quas, & G. S. Goodman (Eds.), *Memory and suggestibility in the forensic interview* (pp. 265–286). Mayway, NJ: Lawrence Erlbaum.

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

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

Granhag, P. A., Strömwall, L. A., & Landström, S. (2006). Children recalling an event repeatedly: Effects on RM and CBCA scores. *Legal and Criminological Psychology, 11*, 81–98.

Gudjonsson, G. H. (2003). *The psychology of interrogations and confessions: A handbook*. Chichester: Wiley.

Gumpert, C. H., & Lindblad, F. (1999). Expert testimony on child sexual abuse: A qualitative study of the Swedish approach to statement analysis. *Expert Evidence, 7*, 279-314.

Hartwig, M., Granhag, P.A., Strömwall, L.A., & Vrij, A. (2005). Detecting deception via strategic disclosure of evidence. *Law and Human Behaviour, 29*, 469–484.

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

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

Inbau, F. E., Reid, J. E., Buckley, J. P., & Jayne, B. C. (2001). *Criminal interrogation and confessions* (4th ed.). Gaithersburg, Maryland: Aspen Publishers.

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.

Kassin, S. M., (2005). On the psychology of confessions: Does innocence put innocents at risk? *American Psychologist, 60*, 215–228.

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

Kassin, S. M., Goldstein, C. J., & Savitsky, K. (2003). Behavioral confirmation in the interrogation room: Compliance, internalisation, and confabulation. *Law and Human Behavior, 27*, 187–203.

Kassin, S. M., & Gudjonsson, G. H. (2004). The psychology of confessions: A review of the literature and issues. *Psychological Science in the Public Interest, 5*, 33–67.

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. (2004). Statement Validity Analysis and the "detection of the truth". In P. A. Granhag & L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 41–63). Cambridge: Cambridge University Press.

Köhnken, G., & Steller, M. (1988). The evaluation of the credibility of child witness statements in German procedural system. In G. Davies & J. Drinkwater (Eds.), *The child witness: Do the courts abuse children?*

(Issues in Criminological and Legal Psychology, no. 13) (pp. 37–45). Leicester, United Kingdom: British Psychological Society.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*, 490–517.

Levine, T. R., & McCornack, S. A. (1992). Linking love and lies: A formal test of the McCornack and Parks model of deception detection. *Journal of Social and Personal Relationships, 9*, 143–154.

Levine, T. R., & McCornack, S. A. (2001). Behavioral adaptation, confidence, and heuristic-based explanations of the probing effect. *Human Communication Research, 27*, 471–502.

Mann, S., Vrij, A., & Bull, R. (2004). Detecting true lies: Police officers' ability to detect deceit. *Journal of Applied Psychology, 89*, 137–149.

Masip, J., Sporer, S. L., Garrido, E., & Herrero, C. (2005). The detection of deception with the Reality Monitoring approach: A review of the empirical evidence. *Psychology, Crime, & Law, 11*, 99–122.

Meissner, C. A., & Kassin, S. M. (2002). "He's guilty!": Investigator bias in judgments of truth and deception. *Law and Human Behavior, 26*, 469–480.

Moston, S. J., & Engelberg, T. (1993). Police questioning techniques in tape recorded interviews with criminal suspects. *Policing and Society, 6*, 61–75.

Moston, S. J., Stephenson, G. M., & Williamson, T. M. (1992). The effects of case characteristics on suspect behaviour during police questioning. *British Journal of Criminology, 32*, 23–39.

Nelson, T. O. & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The Psychology of Learning and Motivation, 26*, 125–141.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. N. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin, 29*, 665–675.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic Inquiry and Word Count (LIWC): LIWC 2001 manual*. Mahwah, NJ: Erlbaum.

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.

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.

Sporer, S. L. (1997). The less traveled road to truth: Verbal cues in deception detection in accounts of fabricated and self-experienced events. *Applied Cognitive Psychology, 11*, 373–397.

Sporer, S. L. (2004). The detection of deception in forensic contexts. In P. A. Granhag & L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 64–102). Cambridge: Cambridge University Press.

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.

Strömwall, L. A., Bengtsson, L., Leander, L., & Granhag, P. A. (2004). Assessing children's statements: The impact of a repeated experience on CBCA and RM ratings. *Applied Cognitive Psychology, 18*, 653–668.

Strömwall, L. A., Granhag, P. A., & Hartwig, M. (2004). Practitioners' beliefs about deception. In P. A. Granhag & L. A. Strömwall (Eds.), *Deception detection in forensic contexts* (pp. 229–250).

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

Vrij, A. (2003). 'We will protect your wife and child, but only if you confess': Police interrogations in England and the Netherlands. In P. J. van Koppen & S. D. Penrod (Eds.), *Adversarial versus inquisitorial justice: Psychological perspectives on criminal justice systems* (pp. 57–79). New York: Plenum.

Vrij, A. (2004). Invited article: Why professionals fail to catch liars and how they can improve. *Legal and Criminological Psychology, 9*, 159–181.

Vrij, A. (2005a). Cooperation of liars and truth tellers. *Applied Cognitive Psychology, 19*, 39–50.

Vrij, A. (2005b). Criteria-Based Content Analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law, 11*, 3–41.

Vrij, A. (2006). Challenging interviewees during interviews: The potential effects on lie detection. *Psychology, Crime, & Law, 12*, 193–206.

Vrij, A., Akehurst, L., & Knight, S. (2006). Police officers', social workers', teachers' and the general public's beliefs about deception in children, adolescents and adults. *Legal and Criminological Psychology, 11*, 297–312.

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., Akehurst, L., Soukara, R., & Bull, R. (2004a). Detecting deceit via analyses of verbal and nonverbal behavior in adults and children. *Human Communication Research, 30*, 8–41.

Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2004b). Let me inform you how to tell a convincing story: CBCA and Reality Monitoring scores as a function of age, coaching and deception. *Canadian Journal of Behavioural Science* (special issue on Forensic Psychology), *36*, 113–126.

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., & Mann, S. (2005). Police use of nonverbal behavior as indicators of deception. In R. E. Riggio & R. S. Feldman (Eds.), *Applications of nonverbal communication* (pp. 63–94). Mahwah, NJ: Lawrence Erlbaum Associates.

Vrij, A., Mann, S., & Fisher, R. (2006). An empirical test of the behaviour analysis interview. *Law and Human Behavior*, *30*, 329–345.

Vrij, A., & Winkel, F. W. (1991). Cultural patterns in Dutch and Surinam nonverbal behavior: An analysis of simulated police citizen encounters. *Journal of Nonverbal Behavior*, *15*, 169–184.

Williamson, T. (1993). From interrogation to investigative interviewing: Strategic trends in police questioning. *Journal of Community and Applied Social Psychology*, *3*, 89–99.