

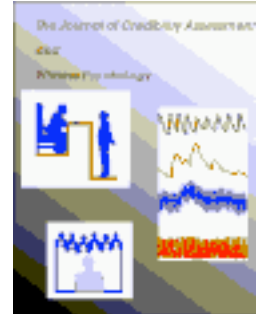
The Journal of Credibility Assessment and Witness Psychology

2006, Vol. 7, No. 1, pp. 1-16

Looking Through the Eyes of an Accurate Lie Detector

Samantha Mann¹, Aldert Vrij¹, and Ray Bull²

¹University of Portsmouth, ²University of Leicester



Abstract

Previous research has indicated that teaching people how to detect deception often results in little success (Bull, 1989, 2004). This experiment sets out to improve lie detection ability in a novel way, by asking participants to “look through the eyes of an accurate lie detector”. In a prior experiment (Mann, Vrij, & Bull, 2004) participants were exposed to clips of suspects in their interviews who either lied or told the truth, and asked to make veracity judgements. They were also asked to select fragments within those clips which contained behaviour that they considered to be relevant to their decision. Those fragments, as selected by the most accurate lie detectors, are the basis of this experiment. Lie detectors in this study were exposed to either the full-length original clips (control group) or fragments of the original clips as selected by previously accurate lie detectors (experimental groups). It was hypothesized that, by eliminating supposedly 'white noise' ambiguous behaviour in the clip, participants should achieve higher accuracy than participants who saw the whole clip. Results indicated that experimental lie detectors could detect truths and lies above the level of chance, but not significantly better than the control group. In addition, confidence scores for correct judgments were consistently higher than confidence scores for incorrect judgments.

Notes: Correspondence concerning this article should be addressed to: Samantha Mann, University of Portsmouth, Psychology Department, King Henry Building, King Henry 1 Street, Portsmouth PO1 2DY, United Kingdom or via email: Samantha.mann@port.ac.uk This research project was sponsored by a PhD studentship grant from the Economic and Social Research Council (R000429734727).

Copyright 2006 Boise State University and the Authors. Permission for non-profit electronic dissemination of this article is granted. Reproduction in hardcopy/print format for educational purposes or by non-profit organizations such as libraries and schools is permitted. For all other uses of this article, prior advance written permission is required. Send inquiries by hardcopy to: Charles R. Honts, Ph. D., Editor, *The Journal of Credibility Assessment and Witness Psychology*, Department of Psychology, Boise State University, 1910 University Drive, Boise, Idaho 83725, USA.

Looking Through the Eyes of an Accurate Lie Detector

Detecting deception on the basis of someone's demeanor is a difficult task. In a typical lie detection study, observers (typically college students) are shown videotaped clips of liars and truth tellers and have to indicate after each clip whether it contained a truth or a lie. In those studies, the observers typically do not know the liars and truth tellers they have to judge and have no other information (factual evidence, statements of third parties, etc.) to rely upon. Guessing in such a task is likely to result in 50% accuracy. Bond and DePaulo (2005) reviewed the results of 186 samples, including 22,282 observers. On average, observers achieved a 54% accuracy rate, correctly classifying 47% of the lies as deceptive and 61% of truths as truthful.

In some lie detection studies, the observers are professional lie catchers, such as police officers, rather than college students. Granhag and Vrij (in press) reviewed ten of those studies and found an overall accuracy rate very similar to those obtained with college students: 55% (with a 55% lie accuracy rate and a 55% truth accuracy rate). However, these studies also suggest that some professional groups are better than others. For example, in their lie detection study, Ekman, O'Sullivan and Frank (1999) found that US federal officers (police officers with a special interest and experience in deception and demeanor) and sheriffs (police officers identified by their department as outstanding interrogators) were considerably better at detecting lies (73% and 67% accuracy respectively) than 'mixed' law-enforcement officers (officers who had not been specifically chosen because of their reputation as interrogators, 51% accuracy).

However, lie detection in the work environment of these professionals differs in some aspects from lie detection in the experimental lie detection studies (Vrij, 2004). For example, in almost all these studies, the truths and lies the professionals were asked to detect were told by college students who did so for the sake of the experiment. The stakes (positive consequences of getting away with the lie and negative consequences of being caught) in those studies are typically low and probably considerably lower than the stakes that, for example, suspects face in their police interviews. Miller and Stiff (1993) suggested that in low stakes laboratory studies the stakes are not high enough for the liar to elicit clear cues to deception. Indeed, in experimental studies where the stakes were manipulated (although they were never really high), it has been found that high-stakes lies were easier to detect than low-stakes lies (Vrij, 2000). Moreover, the average suspect in a police interview probably differs from the average college student. For example, college students are on average more intelligent than suspects in police interviews (Gudjonsson, 2003), and this difference in intelligence might affect the way they tell lies (Ekman & Frank, 1993). Finally, college students in laboratory settings typically lie about different topics (for example, what their attitude is towards the

death penalty) than suspects in police interviews. Because of these differences, examining police officers' ability to distinguish between truths and lies told by students in laboratory settings might not be a valid test of their true ability to detect truths and lies. Mann, Vrij, & Bull (2004) therefore showed 99 police officers, not identified in previous research as belonging to groups which are superior in lie detection, videotapes consisting of 54 truths and lies told by suspects during their police interviews. A total accuracy of 65% was found, with 66% lie accuracy and 63% truth accuracy. Also in this experiment police officers were exposed to the real life material used in Mann et al.'s (2004) study.

Several studies have attempted to investigate the feasibility of teaching the art of detecting deception. However, such efforts have borne little fruit. In a review of the literature, Bull (2004) maintained that the effectiveness of such training was minimal. Some researchers, (DePaulo, Lassiter, & Stone, 1982; deTurck, 1991; deTurck & Miller, 1990; deTurck, Feeley, & Roman, 1997; deTurck, Harsztrak, Bodhorn, & Texter, 1990) have tried to instruct participants to focus on specific cues that often occur during deception, and ignore other cues which they might erroneously believe to occur during deception. Others (Fiedler & Walka, 1993; Vrij, 1994; Zuckerman, Koestner, & Colella, 1985) have hoped to teach participants through providing them with feedback by showing them a clip, asking them to make a judgment, and then telling them if they are accurate or not before showing them the next clip. There are several reasons why the success rates for these methods are disappointing. For example, asking participants to focus on certain cues may fail because there are no hard and fast rules to deceptive behaviour (DePaulo, Lindsay, Malone, Muhlenburck, Charlton, & Cooper, 2003), so the cues which participants are given may not actually occur when some of the communicators lie. Moreover, participants may become confused about the suggested cues, or choose to disregard them completely. Additionally, the stimulus material in those training studies is normally people who have been asked to lie about trivial matters, hence any behavioural differences that do exist may be too subtle for observers to notice (Miller & Stiff, 1993). Furthermore, giving participants feedback about their judgements will not be effective if participants are unable to identify consistent patterns in the communicators' behaviour.

The present experiment is designed to investigate whether a training effect is attainable by asking observers to 'look through the eyes of an accurate lie detector' by asking participants to look only at pieces of behaviour that previously accurate lie detectors had selected as containing important indicators of deception or honesty. This is a novel idea that previous researchers seem not to have examined. We expected this to be beneficial to the observers, as it will focus their attention to 'crucial' parts of the interview, leaving out any 'white noise' behaviour that previously accurate lie detectors had considered to be irrelevant for making truth/lie judgments. We therefore predicted that participants who will see the selected pieces will be more accurate than those who will see original full length clips (Hypothesis 1).

Participants' confidence in the decisions they made was also investigated. Previous research has not found a relationship between confidence and accuracy (see DePaulo, Charlton, Cooper, Lindsay, and Muhlenbruck, 1997, for a meta-analysis), but observers are typically more confident when they watch a truthful statement than when they watch a deceptive statement (regardless of whether they judge that statement as truthful or deceptive). We examined these aspects in the present experiment. It could well be that we will obtain different findings to those in previous studies. In the present study police officers were asked to detect truths and lies told by suspects during their police interviews. Unlike the observers (college students and professionals) in most other studies, the police officers in the present experiment might well find that the task they are asked to do is relevant and familiar to them, and this might result in them being more aware of the accuracy of their own judgements. In other words, a significant relationship between confidence and accuracy may well arise (Hypothesis 2).

Method

Participants

Seventy-two police officers participated, of whom 49 (68%) were male and 23 (32%) were female. Two of the 72 participants were Sergeants, the remainder being Police Constables. Participants were representative of the following divisions: CID (12) (Criminal Investigations Department – these officers are normally plain clothes detectives), Police Trainers (1) and Uniform Patrol (59). Participants' age ranged from 18 to 54, with a mean age of 28.65 years (SD = 7.92, median = 26.50). Length of service ranged from 1 to 27 years (a number (51) of officers were still in 'probation' - i.e. the first two years of the job), with a mean of 4.1 years (SD = 6.5, median score = 1). Participants' self-perceptions of experience in interviewing suspects ranged from the minimum of 1 (totally inexperienced) to the maximum of 5 (highly experienced) with a mean of 1.97 (SD = .95, median score = 2). Permission to approach police officers was granted by the Chief Constable in the first instance, and then by appropriate Superintendents. Participants were recruited on duty from the training college where they were attending courses. They were informed that they would be taking part in a totally anonymous, brief deception detection task.

The Stimulus Materials

Participants in this study were asked to judge the veracity of people in real-life, high-stake situations. Participants saw 8 video clips (4 truthful and 4 deceptive) of 4 suspects, (two male adults, one male juvenile and one female adult) in their police interviews. Specifically suspects saw one truth and one lie of a female adult, two truths and one lie of a male juvenile, one truth and one lie of one male adult, and one lie only of another male adult. The crimes about which the suspects were being interviewed included theft, possession of illegal substances, arson, and

murder. Cases had been chosen where the experimenters knew for sure that the suspects told the truth or lied at various points within their interview. Then, only those particular clips where each word was known to be a truth or a lie were selected. This was known through reliable independent witness statements, and/or forensic evidence (for more detail see Mann, Vrij, & Bull, 2002). In each case, both the lie and the truth clips related to questioning about the case, and hence were emotion-provoking. In some cases they consisted of lie clips taken when a false story was given about the suspect's activities during the time in question, which could then be directly compared to truthful clips where the suspect confessed to the crime having been presented with evidence against him. In other cases where the suspect consistently denied having committed the crime and did not confess, truthful elements of their story were compared with deceptive elements. E.g. One suspect gave truthful details of how she knew the murder victim and when she had seen him, but then lied about a specific visit there to find him threatening to kill himself.

Actual behaviour exhibited by the suspects is included in Table 1 below. The behaviour of participants was analysed in a previous study (for details about behaviours, see Mann et al., 2002). Table 1 includes mean occurrences per minute, or per 100 words as appropriate, of a number of behaviours for those truthful and deceptive clips that were included in this study, to give an indication of the behaviour that participants in the present study were exposed to.

Table 1 demonstrates that as a group, the suspects averted their gaze for just under half the time during truthful clips, but only a third of the time during deceptive clips. Suspects hardly smiled at all throughout all clips, and shook their heads more, made slightly more hand/finger movements, illustrators, and paused in speech more when lying. Suspects blinked slightly more, moved and nodded their heads more, made slightly more illustrators and self-manipulations, shrugs, speech fillers and speech errors when telling the truth. As a subgroup, the behaviour in these clips is a reflection of that found in Mann et al. (2002), with the exception of increased eye contact during deceptive clips.

Procedure

In a previous study (Mann et al., 2004) we showed 99 police officers a number of video clips of suspects who were either lying or telling the truth in their police interviews, and asked them to make a veracity judgement after watching each clip. Additionally participants were asked to select (by using 'the Observer' software) those samples of behaviour exhibited by the suspect (verbal or nonverbal) that they believed motivated their veracity judgement. The 'Observer' allows us to know

Table 1. Mean occurrences per minute of suspect behaviour in truthful and deceptive clips.

Behaviour	Truthful Clips <u>M</u>	Deceptive Clips <u>M</u>
Gaze aversion (in seconds)	29.02	19.61
Blinks	11.00	10.00
Smiles	0.00	0.33
Head nods	2.20	2.00
Head shakes	1.60	6.00
Other head movements	13.60	9.67
Illustrators	4.00	4.33
Self manipulations	2.80	0.00
Shrugs	0.40	0.33
Hand/finger movements	2.00	2.67
Speech fillers	0.60	0.38
Speech errors	0.80	0.67
Pauses in speech (in seconds)	2.58	5.34

within a fraction of a second, exactly which fragments of the video clip participants selected. Ninety-nine police officers each saw one of four tapes containing between 10 to 16 clips. Overall accuracy averaged 64%, significantly above the level of chance. Some participants however were particularly accurate in their veracity judgements, with 22% correctly judging 75% or more clips. The four highest scoring officers attained equal accuracy of 90% (each correctly identified the veracity of suspects in nine out of a possible ten clips), hence we examined the data from these four participants in order to develop the present study. Two of the four were eliminated from this study as each had selected so many impracticably short fragments (e.g. ten per clip with an average length of 0.2 seconds). We used the data from the remaining two highest scoring participants (hereon participants #1 and #2). Participants #1 and #2 had seen the same ten original clips. Both participants correctly identified the veracity of the suspects in nine out of the ten clips, though the clip that each participant did not identify correctly varied. These two clips were excluded from the current study, leaving eight clips, four truths and four lies. The eight clips were shown in the following standard order: lie, lie, truth, truth, truth, lie, lie, truth.

Participants were randomly allocated to one of three conditions, one being control and the other two being experimental. Twenty-two participants (15 male, 7 female) participated in the Control Condition (A), 25 (18 males, 7 females) participated in one of the Experimental Condition groups (B) and 25 in the other Experimental Condition group (C) (16 males, 9 females).

Participants in the control group of the current study were exposed to the eight original full-length video clips that participants #1 and #2 had both

previously identified correctly. These varied in length with a range from 11.5 seconds to 45.2 seconds, with an average of 28.6 seconds.

Participants in experimental groups in the current study were exposed to the selections, or fragments, of the same eight clips previously chosen by participants #1 and #2. One experimental group was exposed to fragments selected by accurate participant #1, the other experimental group was exposed to fragments selected by accurate participant #2. Participants #1 and #2 had selected only one fragment for some clips, with a maximum of three fragments from any one clip. Fragment lengths seen by the experimental groups ranged from 1 second to 25 seconds with a mean of 7.67 seconds. Participants were exposed to all fragments relating to each clip before making a judgement. An advantage of using two experimental conditions rather than just one is that comparing results of two groups to one control group will tell us more about our manipulations than comparing one experimental group. If participants of both experimental groups were more accurate than the control group then this would reveal that their positive scores are more likely a result of our manipulations than luck than if just one experimental group had been used. However, since we cannot predict any differences between the two experimental groups, their results will be compared in preliminary analyses, and the two groups will be combined if no significant differences between the two groups emerge (see Results section for these analyses).

Table 2 below illustrates the length of footage each group saw for each clip. The control group saw the full length clips. Fragment groups (experimental groups) #1 and #2 saw between 1 and a maximum of 3 fragments taken from each of those clips. So, for example for Clip 3, the control group saw the full 37.4 second clip. Fragment group #1 saw a 25 second fragment as selected by Accurate Participant #1 from a previous experiment (as described above), and Fragment

Table 2: Length of clip/fragments exposed to each of the three condition groups.

Clip	T/L	Suspect Info	Length of clips/fragments as seen by participants (in seconds)		
			Control Group	Fragment group #1	Fragment group #2
1	Lie	Suspect 1 - Female Adult	11.50	5.50	8.00
2	Lie	Suspect 2 - Male Juvenile	12.80	3.00	5.00
3	Truth	Suspect 3 - Male Adult	37.40	25.00	14.00 and 6.00 and 4.00
4	Truth	Suspect 1 - Female Adult	20.10	7.00	4.00 and 1.00
5	Truth	Suspect 2 - Male Juvenile	42.50	19.00	4.00 and 12.00 and 13.00
6	Lie	Suspect 4 - Male Adult	24.90	11.00	9.00
7	Lie	Suspect 3 - Male Adult	45.20	1.00 and 15.00	4.00 and 6.00
8	Truth	Suspect 2 - Male Juvenile	34.40	4.00 and 4.00	5.00 and 3.00 and 7.00

group #2 saw a 14 second fragment, a 6 second fragment and a 4 second fragment, all selected by Accurate Participant #2 from a previous experiment.

Participants were asked if they would partake in a brief, anonymous task to test their deception detection skills. Between three and ten participants were tested simultaneously. This variation in group size was only a result of the number of officers that trainers were willing to release from class at that time and did not in any way affect the running of the experiment. The clips were shown on a large screen (approximately 2m x 1m), in a large classroom that would have enabled twenty participants to see the screen clearly sitting far enough apart so as not to see each other's answers. Participants were asked not to talk throughout the experiment, and were given questionnaires whereupon they completed the first section relating to age, gender, rank, division, length of service and perceived experience in interviewing suspects. They were then given one of the following sets of instructions depending on whether they were in the control or an experimental group.

Control Condition Instructions: Participants in the control condition were given the following instructions: "You are about to see a selection of clips of suspects who are either lying or telling the truth. The clips vary considerably in length, and the suspects may appear on several occasions. This is irrelevant. They will either be lying the whole length of the clip or truth-telling for the length of the clip. After watching each clip the video will pause, and I would like you to complete the two questions on your questionnaire relating to that clip. You are asked whether you think the suspect is lying or telling the truth, by circling truth or lie. Question 2 asks how confident you are of your decision on a seven-point scale where 1 = not at all confident and 7 = extremely confident. When everyone has done this then we will move on to the next clip. Please fill in the questions in section 1 first". After each clip was shown the experimenter checked that all participants were filling in their responses with regard to the correct clip number.

Experimental Conditions Instructions: Participants in both experimental conditions were given the following instructions: "You are about to see a selection of clips of suspects who are either lying or telling the truth. The clips vary considerably in length, although they are all rather short. After watching each clip the video will pause, and I would like you to complete the two questions on your questionnaire relating to that clip. You are asked whether you think the suspect is lying or telling the truth, by circling truth or lie. Question 2 asks how confident you are of your decision on a seven-point scale where 1 = not at all confident and 7 = extremely confident. Your questionnaire asks you for the same information with regard to 8 clips. The suspects that you will see may appear on several occasions. This is irrelevant. They will either be lying the whole length of the clip or truth-telling for the length of the clip. On a few occasions you will see two or more rather brief clips that will all relate to one clip number on your questionnaire. I will warn you in advance when this will happen. For example, for most clips you will only see one brief episode of footage, but for some there may be

up to three very small clips all taken from the same interview. In these instances I will ask you to see each of these mini-clips before making a decision. Again, in all of these mini-clips the suspect will either be telling the truth or a lie. One of the intentions of this research is to investigate whether you are able to detect deception based on only a small portion of behaviour. I realise that this is a difficult task in that you have only a small example of behaviour on which to base your decision rather than being able to hear the facts of the case. However some research would indicate that such information may cloud the judgement of a lie-detector rather than enhance it and this is something that we are investigating”.

The latter sentences were added on the basis of the experimenter's prior experience of asking police officers to judge the veracity of suspects based on full-length clips. Previously, the complaint from officers that they cannot be expected to make a judgement based on such a short piece of information, and that a longer observation would surely result in higher accuracy, was all too common. Hence asking participants to make judgements on considerably shorter clips would only serve to demotivate and agitate them. To point out in advance that we are aware that the task is difficult seemed appropriate in order to neutrally motivate the officers and make them feel more positive about the experiment. Possibly it could be argued that the information supplied to participants gives away the intentions of the experiment. However, the experimenter sees no reason why this would affect participants' scores, since the intention of the experiment is to investigate ability and not opinions or attitudes. It was considered that to forgo the forewarning would only impede performance.

The clips were shown via a large screen [approx 72” screen]. After each clip (or each set of fragments relating to each clip) the videotape was stopped for participants to make their judgement and complete the appropriate section of the questionnaire. Hence all participants made eight veracity judgements, one for each total clip, regardless of the number of fragments they saw. At the end of the task participants were asked to complete the last section, which asked them to estimate what lie detection accuracy percentage they believe they attained in the task. They were then thanked and debriefed.

Results

Preliminary analyses

Analyses were carried out to explore differences between the two experimental groups in terms of accuracy and confidence scores. In the first MANOVA the two experimental groups were compared in terms of accuracy regarding the eight clips. The multivariate effect of that analysis was not significant, $F(8, 41) = 1.03$, ns. In the second MANOVA the confidence scores regarding the eight clips were included as dependent variables. That analysis also revealed a non significant effect, $F(8, 41) = .59$, ns. In other words, the results showed similar findings for the two

experimental groups, and they were therefore combined in the subsequent analyses, resulting in an experimental group with $N = 50$ participants.

Descriptive analyses

Figure 1 shows the mean accuracy scores for each of the eight clips for both the control group ($N = 22$) and experimental group ($N = 50$). It also shows which accuracy scores differed significantly from chance (50%). Figure 1 shows that most accuracy scores were significantly above the level of chance. The accuracy score for clip 7 (experimental group), however, was significantly below chance. The total accuracy score of the whole sample was 68.75% ($SD = .14$), where truth accuracy and lie accuracy were 73.26% ($SD = .23$) and 64.24% ($SD = .22$) respectively.

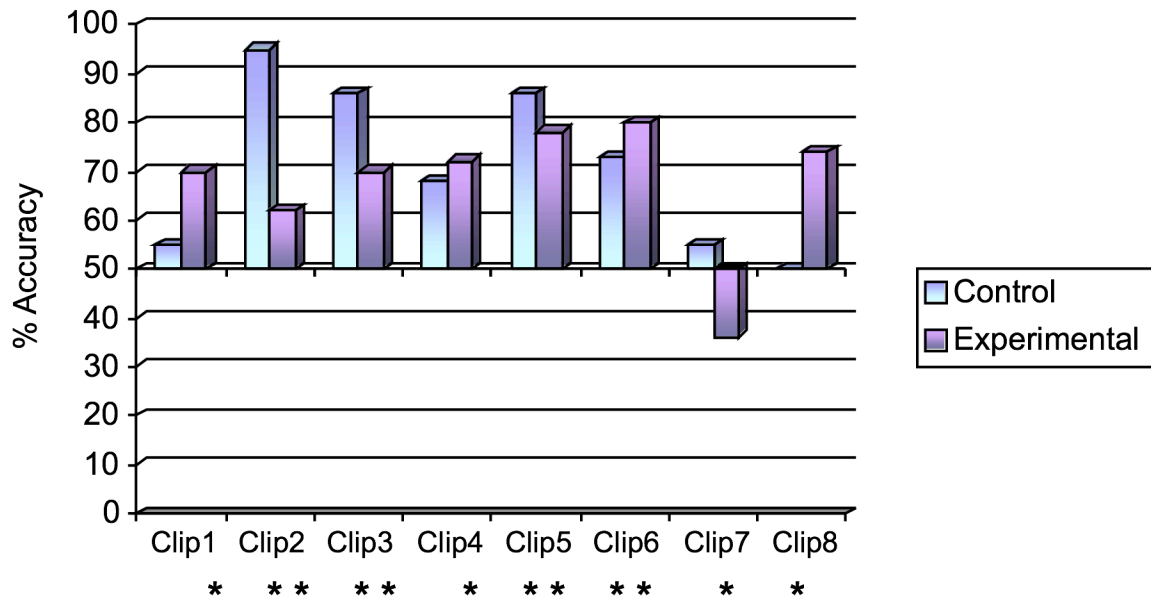


Figure 1: Percentage accuracy for the Control and Experimental groups for each clip.

Note: Stars indicate scores which significantly differed from the 50% level of chance.

Figure 2 shows the mean confidence scores for each of the eight clips for both the control group and the experimental group. It also shows whether the confidence score was significantly above '4' that was the neutral point on the 7-point Likert scale. Figure 2 shows that participants in both the control and experimental groups felt reasonably confident in their judgments. All scores were above 4 and most of them were significantly above 4.

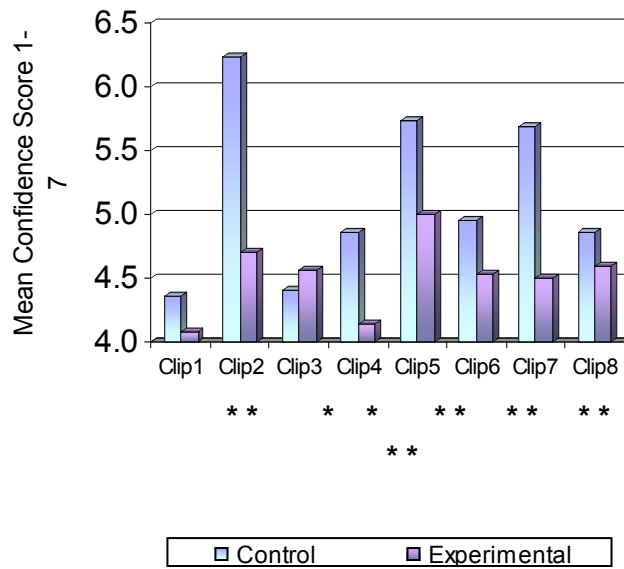


Figure 2: Mean Confidence scores for Control and Experimental groups for each clip.
 Stars indicate scores which significantly differed from the mean likert scale score of 4.

Figure 3 shows for each clip the mean confidence scores for those who made a correct judgement and for those who made an incorrect judgement. Figure 3 demonstrates that confidence scores for correct judgements were consistently higher than confidence scores for incorrect judgements.

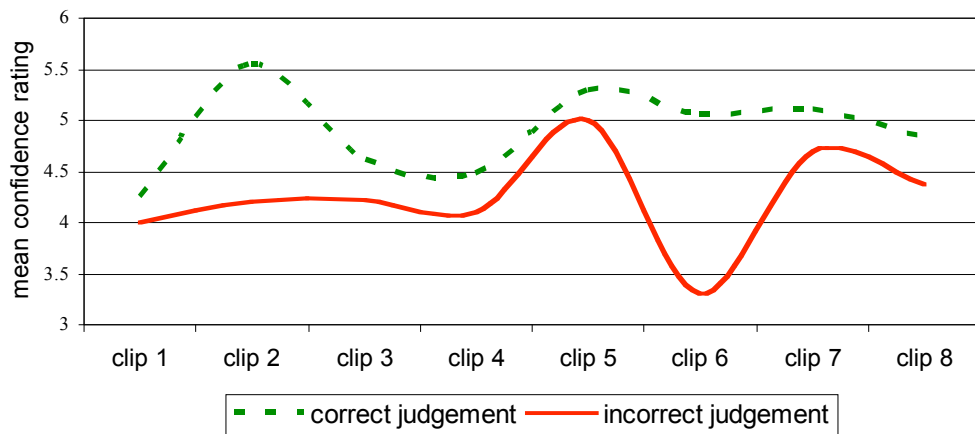


Figure 3: Mean confidence scores for correct and incorrect accuracy judgements.

Hypothesis-testing

In order to test whether the experimental group obtained higher accuracy scores than the control group (Hypothesis 1), an ANOVA was carried out with Veracity (truth vs. lie) and Group (control vs experimental) as factors. Veracity was a within-subjects factor, Group a between-subjects factor, and Accuracy score the dependent variable. Both main effects were non significant, F 's < 2.74 , p 's $> .10$. The Veracity X Group interaction effect was not significant either, $F(1, 70) = .81$, ns. Hypothesis 1 is therefore not supported. The truth and lie accuracy scores for the experimental group were 74% ($SD = .22$) and 62% ($SD = .21$) respectively, those for the control group were 73% ($SD = .27$) and 69% ($SD = .23$) respectively. All these scores were significantly above the 50% accuracy score that could be expected by chance (all t 's > 3.92 , all p 's $< .01$).

Participants were asked at the end of the task to estimate what lie detection accuracy percentage they believed they attained in the task. This percentage, which did not differ between the control group and experimental group, $F(1, 70) = 3.04$, ns, was rather modest, 55.71% ($SD = 17.7$), and this score was significantly lower than the actual total accuracy score (68.75%) obtained in this experiment, $F(1, 71) = 25.55$, $p < .01$, $\eta^2 = .27$.

In order to test for differences in confidence scores a second ANOVA was conducted, again utilizing a 2 (Veracity) X 2 (Group) factorial design. In this analysis confidence score was the dependent variable. The analysis revealed a significant main effect for Group, $F(1, 70) = 5.44$, $p < .05$, $\eta^2 = .07$. Also the Veracity X Group interaction effect was significant, $F(1, 70) = 4.78$, $p < .05$, $\eta^2 = .06$. The veracity main effect was not significant, $F(1, 70) = 1.10$, ns.

The Group main effect revealed that participants in the control group felt more confident ($M = 5.14$, $SD = .7$) about their judgments than participants in the experimental group ($M = 4.51$, $SD = 1.2$). The interaction effect revealed that the control group was significantly more confident than the experimental group in judging lies ($M = 5.32$, $SD = .8$ vs $M = 4.46$, $SD = 1.2$, $F(1, 70) = 8.59$, $p < .01$, $\eta^2 = .11$), whereas no difference emerged between the two groups in judging truths, $F(1, 70) = 1.92$, ns.

In the third ANOVA differences in confidence between incorrect and correct judgements were tested (Hypothesis 2). In order to do this, two new variables were created. 'Confidence scores in incorrect judgments' was the mean confidence score for the incorrect judgments each participant made, whereas 'confidence scores in correct judgements' was the mean confidence score for the correct judgments each participant made. A 2 (Judgment (incorrect vs correct) X 2 (Group (control vs experimental)) ANOVA was conducted and revealed a significant main effect for Judgment, $F(1, 70) = 26.35$, $p < .01$, $\eta^2 = .27$. Neither the Group main effect nor the Judgment X Group interaction effect was significant, both F 's < 3.74 , both p 's $> .056$. The Judgment main effect revealed that participants were more confident

in their correct judgments ($M = 4.87$, $SD = 1.1$) than in their incorrect judgments ($M = 4.29$, $SD = 1.3$). This supports Hypothesis 2.

Discussion

We predicted in Hypothesis 1 that lie detectors in the experimental group who watched statements of truth tellers and liars ‘through the eyes of accurate lie detectors’ would be more able to discriminate between those liars and truth tellers than lie detectors in the control group who were exposed to original full length clips. However, this was not the case. Those who looked through the eyes of accurate lie detectors were able to detect lies and truths significantly above the level of chance (62% and 74% respectively), but so could the participants who saw the original full length clips (69% and 73% respectively), and the differences between participants in the experimental group and control group were not significant. A possible explanation for the absence of a significant effect was the good performance of the control group. Their accuracy scores are amongst the highest ever found in this type of deception research and therefore hard to outperform. The results of the control group were similar to the results of lie detectors in Mann et al.’s (2004) study who were exposed to the same material. This replication of Mann et al.’s (2004) findings strengthens the conclusion that ordinary police officers are to some extent capable of detecting the lies and truths in the stimulus material used in the present experiment. This material contained statements of suspects during their police interviews. The present experiment and Mann et al.’s (2004) experiment do not allow us to point out a reason why observers were reasonably good. It could be one or more of the following reasons. First, the statements were made in high-stakes situations and it could be that these high stakes facilitated lie detection (if this is true, then also laypersons could be good at distinguishing between truths and lies in the present stimulus material). Second, it could be that the stimulus material is atypical and that we selected statements of particularly poor liars (this again would result in laypersons also being good at the task). We find it hard to believe that this explanation is true as we made a random selection of statements. Third, it could be that the observers were good because they faced a lie detection task which resembles more the lie detection activities in their work environment than previous laboratory based lie detection studies do (if this is true, than laypersons would not be so good at distinguishing between these truths and lies). To test these assumptions, further research is needed, for example by exposing observers to both high-stakes and low-stakes lies (first explanation), by exposing observers to a second, independent, sample of suspects’ statements (second explanation) or by showing the stimulus material used in this study also to a group of laypersons (third explanation). This latter suggestion, however, is not possible as, due to the sensitive nature of the stimulus material we used, we don’t have permission to show it to anyone but police officers.

A second explanation for the absence of a significant difference between the control group and experimental group in the present study is that the two accurate lie detectors in Mann et al.'s study (2004) whose selected fragments the participants in the present experiment were exposed to, were accurate just by chance in Mann et al.'s study (2004). If the two selected observers were 90% correct in a new sample of cases, then it would be more convincing that these individuals possess some special lie detecting ability. If however, they did not score as well on another test, indicating that their first good score was down to chance, then this would indicate that they do not have any particular skill and so there's no reason why the clips that they had selected would benefit future lie detectors. These explanations make us think that, although the findings did not support Hypothesis 1, this hypothesis might still hold true and is worth testing in future experiments. In such experiments stimulus materials need to be used that are least likely to result in floor or ceiling effects. In addition, the skills of the accurate lie detectors on which information the experiment will be based need to be more thoroughly tested.

Previous research has indicated that there is no relationship between accuracy in deception detection and confidence (DePaulo, Charlton, Cooper, Lindsay, & Muhlenbruck, 1997). Furthermore, a study by DePaulo and Pfeifer (1986) implied that, though no more accurate, police officers are more confident in making veracity judgements than students. Heightened confidence, and misplaced belief in one's ability to detect deception, could result in complacency, which in the case of professional lie-catchers could be harmful since they may be more inclined to neglect sufficiently scrutinising a person's behaviour (Levine & McCornack, 1992; Lord, Ross, & Lepper, 1979). The relationship between confidence and accuracy in this experiment differs from those found in previous studies. The actual veracity of the suspect made little difference to confidence scores. Participants were as confident of their score regardless of whether they were watching a truthful or a deceptive clip. However, confidence scores for correct judgements were significantly higher than those for incorrect judgements. This would indicate that participants were (perhaps unconsciously) actually aware of when they were making correct veracity judgements. This result is encouraging, and the findings of this study refute the notion that police officers make self-assured but inaccurate lie detectors. Conversely, participants were confident of their decisions when it was appropriate to be so, i.e. when they were correct. Finally, the idea that police officers are overly confident in their lie detection skills was not supported in the present experiment. On the contrary, they significantly outperformed their own expectations.

The suspects in this study differed from communicators in other studies as they had so much at stake if their lies failed, and therefore their behaviour may have betrayed them more than communicators' behaviour in other studies. Although the suspects' behaviour in the present study was investigated, and the findings were similar to those found in laboratory studies in the deceptive behaviour literature, this is not to say that cues did not exist which we did not

measure, or which may not be quantifiable. This could explain why the officers in this study not only had fairly high accuracy scores, but were more likely to rate themselves as confident of their decisions when they actually were accurate. Some clips that participants judged will certainly have been guesswork, but for other clips, participants must have felt certain that they were correct, and the stimulus material in this study may have given participants more to grasp than judging communicators in a sterile and irrelevant situation.

References

- Bond, C. F., & DePaulo, B. M. (2005). *Accuracy of deception judgments*. Manuscript submitted for publication.
- Bull, R. (1989). Can training enhance the detection of deception? In J. C. Yuille (Ed.), *Credibility assessment*. (pp. 83-97). Dordrecht: Kluwer.
- Bull, R. (2004). Training to detect deception from behavioural cues: Attempts and problems. In P. A. Granhag & L. A. Stromwall (Eds.), *The detection of deception in forensic contexts* (pp. 251-268). Cambridge: Cambridge University Press.
- DePaulo, B. M., Lassiter, G. D., & Stone, J. I. (1982). Attentional determinants of success at detecting deception and truth. *Personality and Social Psychology Bulletin*, 8, 273-279.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 129, 74-118.
- deTurck, M. A. (1991). Training observers to detect spontaneous deception: effects of gender. *Communication Reports*, 4, 81-89.
- deTurck, M. A., Feeley, T. H., & Roman L. A. (1997). Vocal and visual cue training in behavioural lie detection. *Communication Research Reports*, 14, 249-259.
- deTurck, M. A., Harsztrak, J. J., Bodhorn, D. J., & Texter, L. A. (1990). The effects of training social perceivers to detect deception from behavioural cues. *Communication Quarterly*, 38, 189-199.
- deTurck, M. A. & Miller, G. R. (1990). Training observers to detect deception: effects of self-monitoring and rehearsal. *Human Communication Research*, 16, 603-620.
- Ekman, P., & Frank, M. G. (1993). Lies that fail. In M. Lewis & C. Saarni (Eds.), *Lying and deception in everyday life* (pp. 184-201). New York, NJ: Guilford Press.
- Ekman, P., O'Sullivan, M., & Frank, M. G. (1999). A few can catch a liar. *Psychological Science*, 10, 263-266.
- Fiedler, K. & Walka, I. (1993). Training lie detectors to use nonverbal cues instead of global heuristics. *Human Communication Research*, 20, 199-223.
- Frank, M. G., & Feeley, T. H. (2003). To catch a liar: Challenges for research in lie detection. *Journal of Applied Communication Research*, 31, 58-75.
- Granhag, P. A., & Vrij, A. (in press). Deception detection. In N. Brewer & K. Williams (Eds.), *Psychology and law: An empirical perspective*. New York, NJ: Guilford Press.
- Gudjonsson, G. H. (2003). *The psychology of interrogations and confessions*. Chichester: Wiley & Sons.
- Mann, S., Vrij, A., & Bull, R. (2002). Suspects, lies and videotape: An analysis of authentic high-stakes liars. *Law and Human Behavior*, 26, 365-376.

- Mann, S., Vrij, A., & Bull, R. (2004). Detecting true lies: Police officers' ability to detect suspects' lies. *Journal of Applied Psychology, 89*, 137-149.
- Miller, G. R., & Stiff, J. B. (1993). *Deceptive communication*. Newbury Park, CA: Sage.
- Vrij, A. (1994). The impact of information and setting on detection of deception by police detectives. *Journal of Nonverbal Behavior, 18*, 117-137.
- Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and its implications for professional practice*. Chichester: John Wiley and Sons.
- Vrij, A. (2004). Invited article: Why professionals fail to catch liars and how they can improve. *Legal and Criminological Psychology, 9*, 159-181.
- Zuckerman, M., Koestner, R. & Colella, M. J. (1985). Learning to detect deception from three communication channels. *Journal of Nonverbal Behavior, 9*, 188-194.

Manuscript Submitted: 19 October 2004

Revision Accepted for Publication: 14 March 2006

Published: 5 May 2006