

## **Truth, Lies, and Videotape: An Investigation of the Ability of Federal Parole Officers to Detect Deception**

**Stephen Porter,<sup>1,3</sup> Mike Woodworth,<sup>1</sup> and Angela R. Birt<sup>2</sup>**

---

*The ability of a group of Canadian federal parole officers to detect deception was investigated over the course of 2 days of lie detection training. On the first day of training, 32 officers judged the honesty of 12 (6 true, 6 fabricated) videotaped speakers describing personal experiences, half of which were judged before and half judged after training. On the second day, 5 weeks later, 20 of the original participants judged the honesty of another 12 videotapes (again, 6 pre- and 6 posttraining). To isolate factors relating to detection accuracy, three groups of undergraduate participants made judgments on the same 24 videotapes: (1) a feedback group, which received feedback on accuracy following each judgment, (2) a feedback + cue information group, which was given feedback and information on empirically based cues to deception, and (3) a control group, which did not receive feedback or cue information. Results indicated that at baseline all groups performed at or below chance levels. However, overall, all experimental groups (including the parole officers) became significantly better at detecting deception than the control group. By the final set of judgments, the parole officers were significantly more accurate ( $M = 76.7\%$ ) than their baseline performance ( $M = 40.4\%$ ) as well as significantly more accurate than the control group ( $M = 62.5\%$ ). The results indicate that detecting deceit is difficult, but training and feedback can enhance detection skills.*

---

A man's most valuable trait is a judicious sense of what not to believe.  
Euripides

It is clear that many criminal offenders engage in a high level of manipulation, deception, and malingering (e.g., Hart & Hare, 1997; Porter, 1996; Porter & Yuille, 1995, 1996; Porter, Yuille, & Birt, in press; Rogers, 1997). One context in which deception is expected to occur at a particularly high frequency is the parole interview. The honesty of reports offered by offenders concerning their current offense, degree of remorse, level of rehabilitation, and plans for life in the community

<sup>1</sup>Department of Psychology, Dalhousie University, Halifax, Nova Scotia, Canada.

<sup>2</sup>Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada.

<sup>3</sup>To whom correspondence should be addressed at Forensic Psychology Area, Department of Psychology, Life Sciences Centre, Dalhousie University, Halifax, Nova Scotia, B3H 4J1, Canada; e-mail sbporter@is.dal.ca

must all be assessed to inform and guide parole decision making. The accuracy of judgments of honesty and deception by parole officers can hold important consequences for the offender (e.g., continued detention versus release) and the offender's potential victims in the community. Clearly, parole officers need to keep the number of both false-positive and false-negative errors as low as possible in formulating release decisions. Decisions that could potentially lead to such errors are based in part on the perceived credibility of the offender in the parole interview.

Remarkably, no empirical studies have examined directly the ability of parole officers to detect deceit. There are few data available concerning their detection accuracy, whether level of experience leads to increased accuracy, or whether training can enhance detection skills. Nonetheless, research by Ruback and colleagues has investigated interview factors contributing to parole decision making. Ruback (1981) found that the recommendations of parole officers were based on both their initial recommendation from the review of the offender's file and on their perception of the offender's honesty during the interview. Offenders were rated as being less honest if they smiled less, gestured less, and gave shorter answers during the interview. Ruback and Hopper (1986) coded the nonverbal behaviors of 103 inmates during their parole interviews and obtained information from offender files. They examined the relationship between parole officers' pre- (following a file review) and postinterview judgments of the likelihood of offenders completing parole successfully and their perceptions of honesty during the interview. Results indicated that their honesty ratings were influenced by the severity of the index offense, the amount of information volunteered by the offender, the politeness of the offender, and the level of "fidgeting" during the interview. A comparison of the pre- and postinterview ratings indicated that the interview did not improve, and actually *lowered*, the accuracy of predicting success on parole. After a 32-month follow-up period, it was found that revokees had been given lower preinterview success ratings than nonrevokees. However, following the parole hearing, the prediction ratings for successful and unsuccessful releases did not differ. Thus, the parole officers became less accurate in their predictions following the interview. This pattern of findings suggests that the officers were likely relying on clues (e.g., politeness, fidgeting) that have little relationship with empirically based indices of deception (Memon, Vrij, & Bull, 1998; Vrij, 2000) and/or that insincere offenders may have been successfully prevaricating.

There also is other less direct evidence to suggest that parole officers may experience difficulty in identifying lies. Several studies by Ekman and colleagues have investigated the abilities of diverse professional groups (e.g., clinicians and law enforcement groups) to detect deception. In a classic study, Ekman and O'Sullivan (1991) found that customs officials, police officers, judges, FBI agents, forensic psychiatrists, and other groups were no better than chance at judging the honesty of videotaped speakers. Only one group—the Secret Service—performed significantly, albeit only slightly, above chance (64%). Nonetheless, nearly one third of these agents achieved an accuracy rate of at least 80%, indicative of individual difference factors. Although level of investigative experience played little role in detection success, participants who were most accurate reported relying on multiple clues rather than any one "rule of thumb," suggesting that deception detection may be

a trainable skill. More recently, Ekman, O'Sullivan, and Frank (1999) showed a number of professional groups 10 videotaped speakers relating either a true or a false opinion. Participants were provided 10 sec following each 1-minute segment to make their judgment of honesty. Two groups of federal law enforcement officers and a group of clinical psychologists were able to detect deceit at levels above chance (73%, 66.7%, and 67.5%, respectively). Overall, these groups were more accurate at identifying deceptive than truthful statements. Research is needed to examine whether parole officers' ability to detect deception is similar to other professional groups investigated to date.

Given the relevance of detecting lies to parole decision making, it may be surprising that the training parole officers receive rarely (at least in Canada) addresses the need to detect deception or examines whether credibility judgments can be facilitated. Other professional groups, such as police forces, have long relied upon training to enhance detection skills. However, a significant problem with traditional police training is that it typically has included misleading statements regarding the detection of deceit. Bull (1989) observed that many police training manuals imply that detecting deceit from behavioral clues is a straightforward task and that police training readily enhances detection skills. He concluded that "until a number of publications in refereed journals appear demonstrating that training enhances the detection of deception, it appears that . . . some police training books are deceiving their readers" (p. 83). In support of this contention, Kassin and Fong (1999) found that research participants who had been trained in a widely used traditional police interrogation technique (the "Reid Technique") performed significantly *worse* in identifying deception in videotaped mock interrogations than their untrained counterparts (who performed at chance). Thus, training material with poor validity can be reflected in the performance of those trained. Indeed, it has been established that erroneous preconceived notions about the nature of deceptive behavior can impair the ability to detect deceit (e.g., Akehurst, Kohnken, Vrij, & Bull, 1996; Ekman, 1988, 1989). Akehurst et al. (1996) found that the beliefs of both laypersons and police officers concerning clues to deception (e.g., clearing the throat, biting the lips) differed greatly from empirical findings on deceit. Thus, although traditional training has been of questionable validity, a growing scientific literature on correlates of deception could provide the foundation for improved training approaches.

While it might be expected that level of investigative experience would be associated with a greater proficiency at detecting deceit, this does not appear to be the case. DePaulo and Pfeifer (1986) found that experienced police officers, new police recruits, and university undergraduates did not differ in their detection ability. However, senior officers reported being more *confident* in their decisions. In a meta-analytic review, DePaulo, Charlton, Cooper, Lindsay, and Muhlenbruk (1997) found no relationship between confidence and accuracy in lie detection. One problem in the professional detection of deceit is that interviewers may get little or inconsistent feedback on the accuracy of their deception judgments, or the feedback may come months or years after the judgment. By the time feedback is available, it is usually too late to have any substantial impact on improving judgment accuracy. One study provided evidence that immediate feedback on the accuracy of deception

judgments can facilitate detection ability. Zuckerman, Koestner, and Alton (1984) found that providing information about the truthfulness of a speaker improved subsequent honesty judgments of the particular target. However, there has been little research on the effectiveness of providing professionals with immediate accuracy feedback in detecting deceit.

Given the current literature on deception (see Vrij, 2000, for an excellent overview), there is reason to think that empirically based training could improve deception judgments, particularly when the target has a vested interest in deceiving. It is clear that deceivers, if motivated (e.g., DePaulo, Kirkendol, Tang, & O'Brien, 1988), behave differently from honest speakers both verbally (e.g., Porter & Yuille, 1995, 1996; Porter, Yuille, & Lehman, 1999) and nonverbally (e.g., Ekman, O'Sullivan, Friesen, & Scherer, 1991). One objective of the current research was to investigate whether the accuracy of deception judgments would increase with the provision of information concerning empirically based cues and myths about deception. For example, our previous studies (Porter & Yuille, 1996; Porter et al., 1999) indicate that, relative to honest speakers, deceivers provide more detail (if they have had time to prepare), frequently repeat those details, are less likely to report not remembering something if asked, and tend to exaggerate the overall vividness/quality of their memory. In terms of nonverbal behavior, work by Ekman and colleagues (see Ekman, 1988) indicates that people use fewer "illustrators" (nonverbal gestures normally used to communicate) when being deceptive than when being honest. Prominent myths about deception include the belief that there is a single pattern of nonverbal deceptive behavior (or "rule of thumb") to always to detect deception (Ekman et al., 1999) and that signs of nervousness are reliable cues to deception (e.g., Memon et al., 1998; Vrij, 2000). The current study incorporated these empirically based clues into a novel training approach.

To our knowledge, studies have yet to examine the ability of parole officers to detect deceit. In addition, research has not addressed the questions of whether empirically based training would facilitate detection skills in parole officers and whether any acquired skills would be maintained over time. In 1999, the first author was asked to provide training workshops on deception detection to parole staff of the Correctional Service of Canada. Two days of training, separated by 5 weeks, were carried out and a research component was included. This not only allowed an examination of baseline ability to detect deceit, but also enabled investigation of whether detection skills would increase with training and whether any acquired skills would be maintained over time. To examine the possibility of practice effects and isolate particular aspects of the training, three groups of student participants were tested as well for comparative purposes. These included (1) a feedback group, which received feedback on accuracy following each videotape judgment, (2) a feedback + cue information group, which received feedback and information on empirically based cues to deception, and (3) a control group, which did not receive feedback or the cue information. These groups were tested to compare their performance with the parole officer group and to help determine whether any improvements in deception detection accuracy that might be observed were a result of practice, feedback, and/or the cue information provided.

## METHOD

### Participants

Thirty-two parole officers from the Correctional Service of Canada who volunteered to attend a workshop on deception detection and interviewing participated in the first day of training. Twenty of the original 32 parole officers were available for the second day of training 5 weeks later. Ninety-five undergraduate students also participated in this study, and they received course credit for their participation. As expected, the mean ages of the groups differed significantly,  $F(3, 104) = 56.7$ ,  $p < .001$ . Tukey comparisons indicated that the parole officers were significantly older ( $M = 42.0$  years,  $SD = 9.5$ ) than all three student groups, who did not differ from one another ( $M = 22.1$  years,  $SD = 8.3$ ;  $M = 20.0$  years,  $SD = 4.8$ ; and  $M = 20.5$  years,  $SD = 5.4$ ; for the control, feedback, and feedback + cue information groups, respectively). The students had an average of 1.58 years of university education ( $SD = .97$ ), with a range of 1–5 years. The parole officers had a mean of 11.96 years of job experience ( $SD = 8.29$ ) with a range of .44–33 years. Overall, 67.9% of participants were female (comprising 45.5% of the parole officer sample and 77.7% of the student sample).

### Materials

#### *Videotapes of Truthful and Deceptive Accounts*

The videotaped accounts were generated for previous research (Porter et al., 1999). In this previous study, each of 77 participants related an account of a highly stressful personal experience (e.g., serious medical procedure, serious accident, animal attack, or getting lost). Half of these reports were fictitious and half were true. All original participants had been motivated by a monetary incentive (\$15) to provide narrative accounts that would later appear credible to someone judging their honesty. Further, they had been provided 15 min to prepare their memory accounts for a videotaped interview about their experience. Each interview consisted of two segments: a free narrative phase in which the participant was simply asked to recount what had occurred during the event (with no interruptions from the interviewer) and a specific questions phase.

Twenty-four video clips (12 truthful, 12 fabricated) were randomly selected from the original pool of 77 videos (participants had consented to their tapes being used for further research). The order of presentation of the 12 truthful and 12 fabricated accounts was also selected randomly. For the present research, the first 2 min of each free narrative was shown to the participants on a large-screen television. All participants were informed that there was a 50% chance that each target was relating the truth. All parole officers viewed the tapes together in the workshop. The student participants viewed the videotapes in groups of 4–10. All participants were given 2 min following each presentation to record whether they thought the person was being truthful and to provide a brief rationale for their judgment. To avoid the potential biasing problem of outliers (we previously found that some

speakers are extremely difficult or easy to judge), three video clips, each with an accuracy level of at least one standard deviation above or below the mean judgment accuracy, were omitted. Due to the workshop context of the data collection, it was not possible to counterbalance the presentation order of the videotapes for the parole officers (the population of central interest in this study). Therefore, it was important to present the videos in the same randomized order to the three student groups. Equating the tasks in this way allowed for direct comparison across all groups and precise examination of differences among the experimental conditions. Further, although practice effects are often a confound in designs with a within-subjects component, in this case practice effects were of considerable interest (i.e., to address whether practice alone might facilitate deception judgments).<sup>4</sup>

## The Four Groups

### *Federal Parole Officers*

As mentioned, the data on the parole officers were collected in the context of a 2-day deception detection training workshop. To determine their baseline detection ability, the officers were asked to judge the honesty of 6 (3 true, 3 fabricated) randomly selected videotaped targets (from the pool of 24 videotapes) at the beginning of the workshop. They were informed that they would be asked to judge 12 videos over the course of the first day and that there was a 50% chance that any particular target was relating a truthful experience. They also were asked to record on a handout the clues they used in making each decision. They were provided 2 min to record each judgment and provide a judgment rationale. However, it is important to note that they did not receive any feedback concerning the accuracy of their judgments at baseline (i.e., for the first set of 6 videotape judgments). After the first set of judgments was complete, the officers were surveyed about their opinions on the prevalence of deception by offenders in general and, more specifically, in the parole interview. Following this baseline data collection, the first day of training was conducted.

The training approach utilized was empirically based and comprehensive. The workshop had three major components (for further information on the workshop, contact the first author):

1. *Myth dissolution*: An attempt was made to debunk the prominent myths surrounding deception. For example, the mistaken idea that there is a single “rule of thumb” for detecting lies was discussed in depth with examples.

2. *Information provision*: The officers were given a detailed overview of research on autobiographical memory, history of deception, deception in forensic contexts, verbal and nonverbal cues to deception, fact pattern analysis, hypothesis testing, effective interviewing techniques, and actual cases in which deception was detected. The greatest amount of time was devoted to the empirically based verbal and nonverbal cues to deception. Participants were provided summaries of studies that

<sup>4</sup>It should be noted that for the within-group comparisons that follow, the practice and treatment effects may be confounded by this necessary order of the videotapes. Holding order constant across groups eliminated any confound for between-group comparisons, but not for within-group comparisons.

had focused on various nonverbal and verbal clues to deception. The most and least consistent clues to deception were reviewed with guidelines to their assessment.

3. *Practice judgments, feedback, and knowledge testing:* As mentioned, the officers judged 6 videos in the morning. During the day, actual case materials were presented and discussed. In the afternoon session, they made 6 more judgments and recorded the clues they had utilized for each one. For this set of videos, they received accuracy feedback after they recorded their judgment for each video.

Five weeks later, 20 of the original parole officers were available for a second day of training. Again, at the start of the workshop, 6 more randomly selected videos (3 true, 3 fabricated) were shown and the officers were asked to judge the honesty of each speaker. Feedback on the accuracy of their judgments was provided. The second day was considered a “refresher” course in which the key concepts from the first workshop were reviewed. In addition, much of the day was spent on actual and fictionalized cases in which truthful and deceptive reports were analyzed and discussed. Feedback on these cases was provided and the officers were encouraged to discuss cases and share feedback. Finally, the officers viewed the final set of 6 videos and made their judgments. Again, they received feedback on their accuracy following each judgment.

#### *Control Group*

These randomly assigned participants ( $n = 32$ ) made the same 24 judgments in succession in the absence of any training or feedback concerning accuracy.

#### *Feedback Group*

These randomly assigned participants ( $n = 32$ ) made the same 24 judgments. However, following each judgment, they were informed as to whether their decision had been correct.

#### *Feedback + Cue Information Group*

These 31 randomly assigned participants made the same 24 judgments. Prior to their judgments, they were given a three-page handout providing a brief overview of deception detection highlighting the main points from the above workshop. They were given 15 min to read and study this handout. To ensure that all participants in this condition were exposed to the material, a cassette tape was played during the last 5 min of this period in which a voice reviewed the main points of the handout. After the cue information was studied, the students viewed the videotapes and (as with the feedback group) were informed whether their decision had been correct after each judgment.

## **RESULTS**

### **Preliminary Analyses**

The following variables were unrelated to overall accuracy in detecting deceit: age ( $r = -.03, p > .05$ ), education ( $r = -.037, p > .05$ ), number of years as a

parole officer ( $r = -.08, p > .05$ ), confidence at detecting deceit ( $r = -.06, p > .05$ ), perceived honesty of people in general ( $r = -.19, p > .05$ ), self-reported ability to engage in deception ( $r = -.041, p > .05$ ), and gender ( $t(94) = .002, p > .05$ ).

### *Perceptions of the Prevalence of Deception in Parole Interviews*

The parole officers were surveyed about their beliefs concerning how often offenders use deception. First they were asked to rate how honest they believe offenders to be in general (1 = very dishonest, 2 = often dishonest, 3 = dishonest some of the time, 4 = honest, 5 = very honest). The mean rating was 2.58 ( $SD = .66$ ). They were also asked to estimate how many offenders use deception in the parole interview (1 = very few, 2 = few, 3 = about half, 4 = many, 5 = most or all). The mean rating was 3.73 ( $SD = .72$ ).

### **Baseline Ability to Detect Deceit**

The parole officers' baseline mean accuracy at detecting deception for the initial set of judgments was 40.4% ( $SD = 23.2$ ), with a range of 0–100%. There was no relationship between accuracy and number of years experience as a parole officer,  $r(31) = -.054, p > .05$ , or self-reported (1–5) proficiency at detecting deceit,  $r(31) = -.14, p > .05$ . For the parole officers, there was a trend for males to be more accurate in their judgments at baseline ( $n = 17, M = 47\%, SD = 24\%$ ) than females ( $n = 15, M = 32\%, SD = 19\%$ ),  $t(31) = 1.93, p = .06$ , but by the final set of judgments females tended to outperform males in judgment accuracy.

An analysis of variance (ANOVA) on the first set of baseline judgments indicated that there were significant differences among the groups,  $F(3, 124) = 4.21, p < .01$ . Tukey comparisons ( $p < .05$ ) indicated that the feedback group was significantly more accurate ( $M = 54.8\%, SD = 19.3$ ) than the parole officers ( $M = 40.4\%, SD = 23.2$ ), who had not received any feedback for these first judgments. There also were trends for the feedback group to be more accurate than the control group ( $M = 42.2\%, SD = 17.5, p = .053$ ) and for the feedback + cue information group to be more accurate ( $M = 52.08, SD = 18.33, p = .078$ ) than the parole officers. The baseline accuracy was significantly less than chance (50%) for both the parole officers,  $t(31) = -2.4, p < .05$ , and control group,  $t(31) = -2.5, p < .05$ . The performance of the other two groups did not differ significantly from chance.

### **Effect of Experimental Group Across All Credibility Judgments**

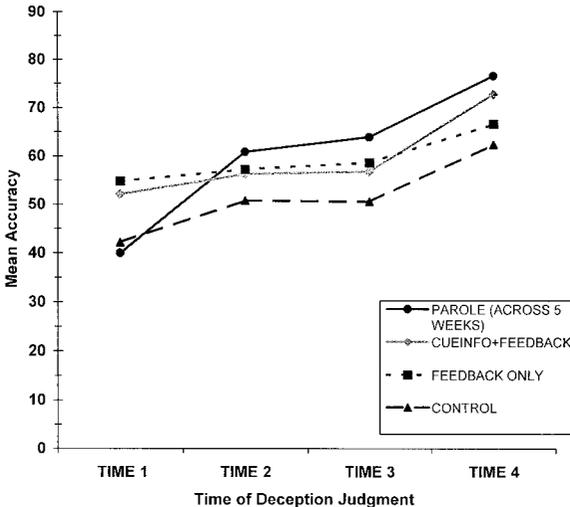
Across all groups and judgments, the overall mean level of accuracy achieved (in terms of percentage correct) was 57.4% ( $SD = 11.4$ ). An ANOVA was conducted to examine whether the four groups differed in their overall ability to detect deception for all of the videotaped targets. This analysis was significant,  $F(3, 111) = 4.4, p < .01$ . Tukey multiple comparisons ( $ps < .05$ ) revealed that all three of the experimental groups—parole officers ( $M = 60.5\%, SD = 9.8$ ), feedback + cue information group ( $M = 59.5\%, SD = 11.9$ ), and feedback only group ( $M = 59.4\%, SD = 9.5$ )—performed significantly better than the control group ( $M = 51.5\%, SD = 11.6$ ) across all sets of videotapes.

### Effect of Time

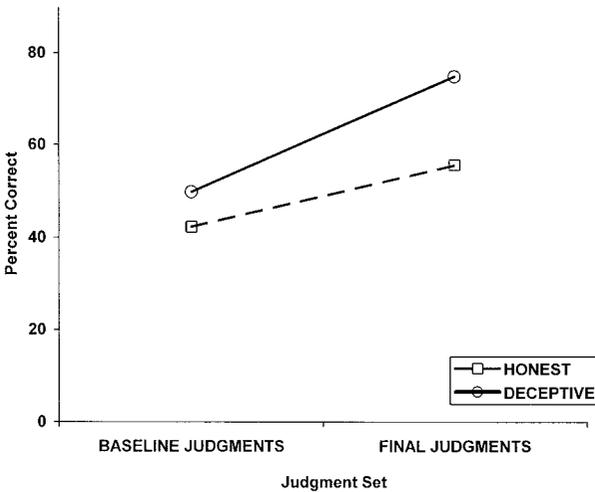
To examine group differences over sets of judgments, a mixed multivariate analysis of variance (MANOVA) was conducted with group (the four experimental conditions) as the between-subjects variable and the deception judgment sets (the four video sets) as the within-subjects variable. Dependent measures included the accuracy rates for the four groups and the accuracy rates for the four sets of judgments. This analysis yielded a significant result,  $F(3, 109) = 23.0, p < .001$ . As depicted in Fig. 1, there was a significant effect of judgment set,  $F(3, 333) = 23.5, p < .001$ . Relative to the first set of judgments ( $M = 48.0\%$ ,  $SD = 20.1$ ), detection accuracy was significantly higher ( $ps < .05$ ) for the second set ( $M = 55.9\%$ ,  $SD = 22.5$ ), the third set ( $M = 56.9\%$ ,  $SD = 21.3$ ), and the final set of judgments ( $M = 69.0\%$ ,  $SD = 18.8$ ). Accuracy for the final set of judgments was also significantly higher than the second and third sets, which did not differ from one another. The interaction between group and time was also significant,  $F(3, 111) = 2.85, p < .05$  (as reported above, there was a significant effect for group). Tukey comparisons ( $p < .05$ ) indicated that for the first set of judgments, the feedback group ( $M = 54.8\%$ ,  $SD = 19.3$ ) outperformed both the control ( $M = 42.2\%$ ,  $SD = 17.4$ ) and parole groups ( $M = 40\%$ ,  $SD = 23.9$ ). For the final set of judgments, the parole officers ( $M = 76.7\%$ ,  $SD = 17.4$ ) outperformed the control group ( $M = 62.5\%$ ,  $SD = 18.0$ ).

### Accuracy of Judgments of Honest Versus Deceptive Targets

Another question of interest concerned whether the accuracy of judgments for honest and dishonest videotaped targets differed (see Fig. 2). Overall, the mean



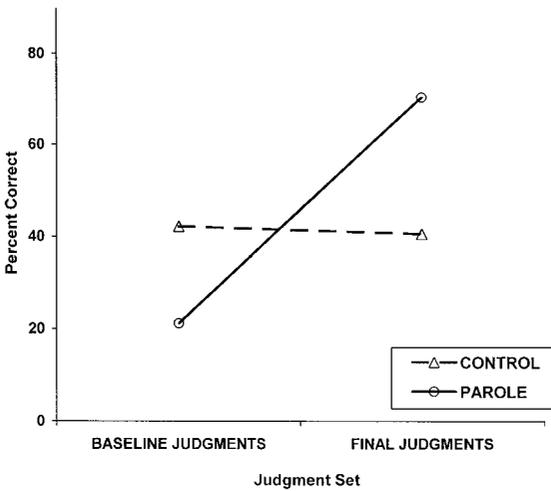
**Fig. 1.** Mean accuracy of deception detection for the experimental groups across four sets of judgments.



**Fig. 2.** Change in accuracy of judgments of all videotaped targets across groups.

level of accuracy in judging the *honest* videos was 51.3% ( $SD = 19.3$ ). An ANOVA was used to examine whether the four groups differed overall in their judgment accuracy of honest targets. This analysis was significant,  $F(3, 111) = 3.14, p < .05$ . Tukey multiple comparisons ( $ps < .05$ ) indicated that the feedback + cue information group ( $M = 56.1\%$ ,  $SD = 18.0$ ) performed significantly better overall than the control group ( $M = 43.0\%$ ,  $SD = 17.01$ ).

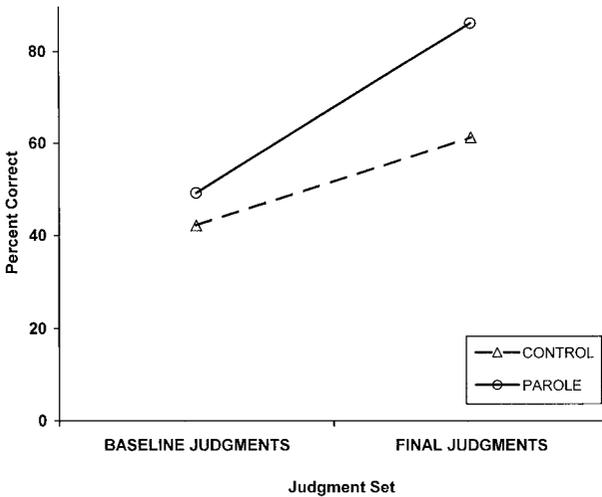
To examine group differences over sets of honest targets, a mixed MANOVA was conducted with group (the four experimental conditions) as the between-subjects variable and the deception judgment sets (the four sets of honest videos) as the within-subjects variable. Dependent measures included the accuracy rates for the four groups and the accuracy rates for the four sets of judgments on the honest videos only. This analysis yielded a significant result,  $F(3, 109) = 8.18, p < .001$ . There was a significant effect of time,  $F(3, 333) = 7.59, p < .001$ . Specifically, detection accuracy was higher ( $ps < .05$ ) for both the second ( $M = 58.3\%$ ,  $SD = 28.2$ ) and final ( $M = 56.1\%$ ,  $SD = 34.5$ ) sets of judgments than for the first set ( $M = 44.3\%$ ,  $SD = 35.5$ ). Accuracy for the final set of judgments was also higher than for the third set ( $M = 46.5\%$ ,  $SD = 32.5$ ). In addition to the main effect of group (above), the interaction between group and time was significant,  $F(9, 333) = 2.36, p < .05$ . Tukey comparisons ( $ps < .05$ ) indicated that for the first (baseline) set of judgments, the feedback ( $M = 59.7\%$ ,  $SD = 35.1$ ) and the feedback + cue information groups ( $M = 46.9\%$ ,  $SD = 33.5$ ) both outperformed the parole group ( $M = 20.0\%$ ,  $SD = 34.03$ ). For the final set of judgments, both the parole officers ( $M = 75.0\%$ ,  $SD = 30.35$ ; see Fig. 3) and the feedback + cue information group ( $M = 65.6\%$ ,  $SD = 32.2$ ) outperformed the control group ( $M = 40.6\%$ ,  $SD = 29.6$ ). The parole officers also performed better than the feedback-only group ( $M = 50.0\%$ ,  $SD = 36.5$ ) in the final judgments of honest targets. Unlike the groups,



**Fig. 3.** Change in judgment accuracy for honest videotaped targets only (percentage correct).

the control group remained constant from the first set of judgments ( $M = 42.2$ ,  $SD = 31.4$ ) to the final set of judgments ( $M = 40.6$ ,  $SD = 29.6$ ) for the honest targets.

Overall, the mean level of accuracy achieved in judging the *deceptive* videos was 59.5% ( $SD = 19.4$ ). An ANOVA was used to examine whether the four groups differed in their overall judgment accuracy of deceptive targets. This analysis was significant,  $F(3, 111) = 8.8$ ,  $p < .001$ . Tukey multiple comparisons ( $ps < .05$ ) indicated that the feedback ( $M = 59.9\%$ ,  $SD = 18.3$ ), the feedback + cue information ( $M = 66.3\%$ ,  $SD = 18.07$ ), and parole officer ( $M = 68.3\%$ ,  $SD = 15.5$ ) groups performed significantly better overall than the control group ( $M = 46.7\%$ ,  $SD = 17.9$ ) (see Fig. 4). To examine group differences over sets of deceptive targets, a mixed MANOVA was conducted with group (the four experimental conditions) as the between-subjects variable and the judgment sets (the four sets of deceptive videos) as the within-subjects variable. Dependent measures included the accuracy rates for the four groups and the accuracy rates at the four sets of judgments on the deceptive videos. This analysis yielded a significant result,  $F(3, 109) = 22.2$ ,  $p < .001$ . There was a significant effect of time,  $F(3, 333) = 16.8$ ,  $p < .001$ . Specifically, detection accuracy was higher ( $ps < .05$ ) for both the third ( $M = 63.8\%$ ,  $SD = 27.2$ ) and the final ( $M = 74.8\%$ ,  $SD = 28.1$ ) sets of judgments than for the baseline set ( $M = 49.8\%$ ,  $SD = 22.3$ ). Accuracy for the third set of judgments was higher than the second set ( $M = 49.6\%$ ,  $SD = 50.2$ ). Accuracy for the final set was also higher than both the second and third sets. Whereas group also had a significant main effect (see above), there was no interaction between group and time,  $F(9, 333) = 1.7$ ,  $p > .05$ . Interestingly, although not statistically significant, the control group improved somewhat in their judgments of the deceptive targets (although not as much as the experimental groups) from the first set of judgments ( $M = 42.2$ ,



**Fig. 4.** Change in judgment accuracy for deceptive targets only (percentage correct).

$SD = 20.5$ ) to the final set of judgments ( $M = 61.4$ ,  $SD = 32.9$ ), whereas they showed very little change for the honest targets.

### Did the Use of Specific Clues Influence Accuracy of Deception Detection?

Participants were asked to describe the type(s) of evidence or clues they used in deciding whether another person is lying or telling the truth. The responses were then coded as either body language cues, speech/verbal cues, or other cues (e.g., “gut feeling”). For the honest videos, there was a significant correlation between the number of times body language was mentioned as a cue and the accuracy of the credibility judgments, but only for the final set of videos,  $r(115) = .21$ ,  $p < .05$ . For the deceptive videos, there was a significant correlation between the number of times body language was mentioned and the accuracy of the credibility judgments, but only for the third set of videos,  $r(115) = .22$ ,  $p < .05$ . Focusing specifically on the parole officers, there was a positive correlation between the number of times speech/verbal clues were cited as cues and the accuracy of judgments for the final set of deceptive videos,  $r(20) = .46$ ,  $p < .05$ . There also were correlations between the number of times speech/verbal clues were reported and the accuracy of judgments for the final set of honest videos,  $r(20) = .76$ ,  $p < .001$ , and across all judgments of honest targets,  $r(20) = .48$ ,  $p < .05$ . Across all target judgments by parole officers, there was a significant negative correlation between the number of times an “other” rule of thumb (neither verbal nor nonverbal) was cited as a rationale for making a judgment and the accuracy of that judgment,  $r(20) = -.52$ ,  $p < .05$ .

## DISCUSSION

This study examined the ability of parole officers to detect deception in persons videotaped while relating either true or fabricated experiences. The first important finding was that the baseline detection ability of the parole officers was significantly below the level of chance. Similar to previous studies, neither level of job experience nor confidence in deception detection ability were associated with accuracy rates. These findings are consistent with those of Ekman and colleagues (e.g., Ekman & O'Sullivan, 1991), indicating that professionals requiring the ability to detect lies are often unable to do so in practice. The fact that these officers performed worse than chance suggests that they were relying on false clues or unreliable "rules of thumb" (Ekman et al., 1999) to guide their credibility judgments. This was further supported by the negative correlation between the number of such cues used and accuracy of judgments by the parole officers. It is also possible that the parole officers held certain expectations or schemas relating to lying and that they engaged in confirmatory hypothesis testing in the deception detection task. For example, perhaps their initial poor performance in judging truthful targets related to the fact that they were "looking for deception" and searching for it using unreliable cues. Indeed, our initial survey indicated that the parole officers believed that the majority of offenders lied during parole hearings, a perception that could lead to a heightened suspiciousness. Future research could examine further the specific beliefs and schemas concerning deception in various professional groups. Despite their poor performance at baseline, the accuracy rates of the parole officers increased from 40.4% (baseline) to 76.7% by the final set of judgments 5 weeks later. This was the highest rate of all groups and significantly higher than the performance of the control group.

Given that the accuracy of the parole officers increased with succeeding judgments, it appeared that training had a positive effect. In order to provide a clearer picture of this process, we isolated the effects of practice, feedback alone, and feedback with cue information (the parole officers received feedback and information for all but the first set of videos) in the student groups. These comparison groups all performed at chance or below chance levels at the baseline measurement. However, overall, both experimental student samples came to perform better than the control group (but did not differ from one another), indicating that the provision of immediate and accurate feedback alone can facilitate judgment accuracy (Zuckerman et al., 1984). It is difficult to ascertain the reason that feedback, in the absence of deception cue information, had a positive effect on judgment accuracy. It is possible that feedback allowed the participants gradually to detect (consciously or unconsciously) valid cues to deception and modify their decision-making accordingly. It is also possible that the provision of feedback increased the motivation of participants to make more careful judgments. That is, the experimenter's feedback may have introduced a social factor (increased pressure on the participant to perform) that increased motivation and facilitated performance. This relationship needs to be explored further. If the performance in the feedback condition increased due to the social consequences of increased motivation to be accurate (and greater concentration on the task at hand), it might be expected that the *accuracy* of the feedback would not be an important factor. In future research, we plan to test this

hypothesis by providing both accurate and inaccurate feedback to deception judges. Whatever the nature of the relationship, in practice it is usually difficult for parole officers or other professionals to obtain accurate and consistent feedback on the validity of information offered in an interview. For example, an offender might lie to an officer in an interview concerning his or her level of rehabilitation and reoffend 3 months later. By this time, the effect of feedback may exert little impact on modifying the judgment process.

Although the parole officers (who had received 2 days of training) were the most accurate by the final set of judgments, their accuracy rate was only modestly higher than the feedback + cue information group, which had been provided with both a summary of the workshop information and feedback regarding accuracy. Given the consequences of judgment errors in applied forensic settings, it might be argued that even a small increase in detection accuracy warrants the intensive training approach. However, it is also important that even a brief training session can improve detection performance. Because intensive training may not always be possible in applied settings, the knowledge that written/taped materials can improve detection ability is of practical utility. These findings are particularly striking when considered in light of recent evidence that traditional police training on deception may *reduce* detection ability to below chance (Kassin & Fong, 1999).

Although various forms of psychological training are common for correctional and investigative professional groups, little research has examined whether any acquired skills are maintained over time. In the present study, it was found that the parole officers had maintained their detection skills after 5 weeks had elapsed. One reason for this skill maintenance may have been that they were encouraged to try to review and implement the workshop materials on the job during the 5-week interim. Importantly, the second “refresher” training session appeared to enhance and consolidate the originally acquired skills. This suggests that a regular training schedule may maintain and facilitate detection skills in legal professionals. As with previous research (e.g., Ekman et al., 1999), lies were more easy to identify than truths for all groups. The parole officers performed very poorly at identifying honest targets at baseline, but improved considerably by the end of the training period. However, whereas the control participants apparently benefited modestly from practice for the deceptive judgments, their accuracy for the honest targets did not change from baseline to the final judgments. This suggests that training may have its greatest effect in facilitating the ability to identify honest reports (and reduce the false-positive rate).

Some limitations of this research should be noted. First, it is possible that some of the present findings were related to a practice effect given that the control group improved somewhat over time (keeping in mind that the control group only improved for deceptive and not honest targets, and that this was not a statistically significant improvement). Because detecting deception is an important part of a parole officer's job, it is plausible that their rate of improvement reflected an amplified practice effect triggered by their increased motivation (relative to students) to succeed on the task over time. Future research could systematically vary motivation level in experimental participants to examine this hypothesis. Second, the videotaped speakers were not offenders and they were not discussing criminal

events. Accordingly, the pressure to lie or tell the truth convincingly by our videotaped targets would not match those of many offenders in the parole interview. In theory, parole officers could have knowledge of a deception pattern specific to offenders that this research did not tap into, although research by Ruback and colleagues suggests otherwise (Ruback, 1981; Ruback & Hopper, 1986). Third, in the parole interview, officers have the opportunity to question and challenge statements by offenders that could influence the accuracy of credibility assessment. In the present study, the participants were passive judges without the ability to view the targets in person or challenge statements.

Nonetheless, this research offers an important and original demonstration of factors contributing to deception detection by both professionals and nonprofessionals. Future studies should build upon this novel approach of examining training effectiveness with correctional staff. Another important remaining task is to identify individual differences contributing to the ability of both professionals and nonprofessionals to detect deception. Following from the current approach, future research on detecting deception could facilitate the ability to catch liars and ultimately improve credibility assessment in forensic interviews.

### ACKNOWLEDGMENTS

The authors would like to express appreciation to Correctional Service of Canada for ongoing support of our research program. Many thanks to Mary Ann Campbell, Prof. Ray Bull, and Dr. Richard Wiener for comments on an earlier draft of this paper. Thanks to Christine Brimson, Rita Gillis, Alexandra Howard, Stephen Lewis, Rob MacArthur, and Jette Midtgaard for their assistance in data collection and management. Thanks to Hugues Hervé for his assistance in conducting the workshops.

The first author was supported by grants from the Social Sciences and Humanities Research Council of Canada (SSHRC) and from the Dalhousie University Faculty of Graduate Studies. The third author was supported by a grant-in-aid from the American Psychology–Law Society (Division 41 of the American Psychological Association).

### REFERENCES

- Akehurst, L., Kohnken, G., Vrij, A., & Bull, R. (1996). Lay persons' and police officers' beliefs regarding deceptive behavior. *Applied Cognitive Psychology, 10*, 461–471.
- Bull, R. (1989). Can training enhance the detection of deception? In J. C. Yuille (Ed.), *Credibility Assessment* (pp. 83–97). Dordrecht, Netherlands: Kluwer.
- DePaulo, B. M., Charlton, K., Cooper, H., Lindsay, J. J., & Muhlenbruk, L. (1997). The accuracy–confidence correlation in the detection of deception. *Personality and Social Psychology Review, 1*, 346–357.
- DePaulo, B. M., Kirkendol, S. E., Tang, J., & O'Brien, T. P. (1988). The motivational impairment effect in the communication of deception: Reproduction and extensions. *Journal of Nonverbal Behavior, 12*, 177–202.
- DePaulo, B. M. & Pfeifer, R. L. (1986). On-the-job experience and skill at detecting deception and truth. *Journal of Applied Social Psychology, 16*, 249–267.

- Ekman, P. (1988). *Telling lies*. New York: Norton.
- Ekman, P. (1989). Why lies fail and what behaviors betray a lie. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 71–82). Dordrecht, Netherlands: Kluwer.
- Ekman, P. & O'Sullivan, M. (1991). Who can catch a liar? *American Psychologist*, *46*, 913–920.
- Ekman, P., O'Sullivan, M., & Frank, M. G. (1999). A few can catch a liar. *Psychological Science*, *10*, 263–266.
- Ekman, P., O'Sullivan, M., Friesen, W. V., & Scherer, K. R. (1991). Face, voice, and body in detecting deceit. *Journal of Non-Verbal Behavior*, *15*, 125–135.
- Hart, S. D. & Hare, R. D. (1997). Psychopathy: Assessment and association with criminal conduct. In D. M. Stoff, J. Brieling, & J. Maser (Eds.), *Handbook of antisocial behavior* (pp. 22–35). New York: Wiley.
- 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.
- Memon, A., Vrij, A., & Bull, R. (1998). *Psychology and law: Truthfulness, accuracy, and credibility*. London: McGraw-Hill.
- Porter, S. (1996). Without conscience or without active conscience?: The etiology of psychopathy revisited. *Aggression and Violent Behavior*, *1*, 179–189.
- Porter, S. & Yuille, J. C. (1995). Credibility assessment of criminal suspects through statement analysis. *Psychology, Crime, and Law*, *1*, 319–331.
- Porter, S. & Yuille, J. C. (1996). The language of deceit: An investigation of the verbal clues to deception in the interrogation context. *Law and Human Behavior*, *20*, 443–458.
- Porter, S., Yuille, J. C., & Birt, A. (in press). The discrimination of deceitful, mistaken, and real memories for emotional events. In R. Roesch, R. R. Corrado, & J. Dempster (Eds.), *Psychology in the courts: International advances in knowledge*. Amsterdam: Harwood.
- Porter, S., Yuille, J. C., & Lehman, D. (1999). The nature of real, implanted, and fabricated memories for emotional childhood events: Implications for the recovered memory debate. *Law and Human Behavior*, *23*, 517–538.
- Rogers, R. (1997). *Clinical assessment of malingering and deception* (2nd ed.). New York: Guilford.
- Ruback, R. B. (1981). Perceived honesty in the parole interview. *Personality and Social Psychology Bulletin*, *7*, 677–681.
- Ruback, R. B. & Hopper, C. H. (1986). Decision making by parole interviewers: The effect of case and interview factors. *Law and Human Behavior*, *10*, 203–214.
- Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and the implications for professional practice*. Chichester, England: Wiley.
- Zuckerman, M., Koestner, R., & Alton, A. O. (1984). Learning to detect deception. *Journal of Personality and Social Psychology*, *46*, 519–528.