The Influence of Judge, Target, and Stimulus Characteristics on the Accuracy of Detecting Deceit

STEPHEN PORTER, MARY ANN CAMPBELL, JENNIFER STAPLETON
Dalhousie University
ANGELA R. BIRT, University of British Columbia

Abstract
Although detecting deception is generally difficult, it appears that some people are better at this task than others. However, the factors contributing to varying accuracy rates in deception detection are not well understood. The purpose of this study was to identify potential factors related to the ability to detect deceit, including characteristics of the credibility assessor (e.g., personality, handedness, cue use), characteristics of the target (attractiveness, gender), and the modality of the report (audiovisual versus audio-only). Participants (N = 310) judged the honesty of eight reports of emotionally laden events (four true reports and four motivated lies) presented in audio-visual or audio-only formats. Results indicated that detection accuracy was significantly higher when the judge was left-handed, the target was unattractive, and the target and judge were of opposite genders. Judges most frequently relied on cues related to the content of the report to decide credibility, but the use of such cues was associated with lower detection accuracy. In contrast, although facial cues were relied upon less often, they were positively associated with accuracy. Implications for the process of credibility assessment are discussed.

Résumé
Bien qu’il soit le plus souvent difficile de détecter la tromperie, il apparaît que certaines personnes sont meilleures à le faire que d’autres. Cependant, les facteurs contribuant à la variance du taux de précision dans la détection de la tromperie ne sont pas bien compris. La présente étude visait à identifier des facteurs possibles liés à la capacité de détecter la tromperie, y compris les caractéristiques de l’évaluateur de la crédibilité (p. ex., la personnalité, la manuelité, l’utilisation des indices), les caractéristiques de la cible (l’attractivité, le sexe) et la modalité du rapport (audiovisuel par opposition à audio uniquement). Les participants (N = 310) ont jugé de l’honnêteté de huit rapports d’événements remplis d’émotivité (quatre rapports vrais et quatre rapports de mensonges motivés) présentés en format audiovisuel et audio uniquement. Les résultats démontrent que la précision dans la détection était significativement plus élevée lorsque le juge était gauche, que la cible était peu attrayante et que la cible et le juge étaient de sexe opposé. Les juges se faisaient le plus souvent à des indices liés au contenu du rapport pour décider de la crédibilité, mais l’utilisation de tels indices était associée à une précision de détection plus faible. En revanche, même si on se fiait moins souvent aux indices faciaux, ils étaient associés positivement à la précision. Des suppositions quant au processus d’évaluation de la crédibilité font l’objet d’une évaluation.

Deception is an important aspect of everyday human social interactions (e.g., DePaulo, Kashey, Kirkendol, Wyer, & Epstein, 1996). Although the consequences of deception are often trivial (as with “white lies”), they can be more serious in certain contexts, such as a police interrogation or in the courtroom (e.g., Porter & Yuille, 1996; Porter, Yuille, & Birt, 2001). Most people, including law enforcement professionals, experience considerable difficulty in accurately detecting deceit. In laboratory studies, most individuals perform around chance in detecting deception (see Vrij, 2000). In a classic study by Ekman and O’Sullivan (1991), 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 the Secret Service performed significantly above what would be expected from guessing alone (64% correct). Similarly, Porter, Woodworth, and Birt (2000) found that Canadian federal parole officers performed significantly below chance at detecting deception prior to participating in a deception detection training workshop.

Despite the general impairment in identifying deceit, some individuals are considerably better at the task than others. For example, Porter et al. (2000) found that some parole officers were much better at detecting deception than others. Likewise, Ekman and O’Sullivan (1991) reported that nearly a third of the Secret Service agents in their sample achieved an accu-
racy rate of at least 80%. More recently, Ekman, O'Sullivan, and Frank (1999) showed a number of professional groups videotapes of speakers relating either a true or a false opinion. Two groups of federal law enforcement officers and a group of clinical psychologists were able to detect deceit at levels above chance (73%, 68%, and 67%, respectively). Collectively, the findings from these studies suggest that individual difference factors influence deception detection. Some of these individual differences may relate to characteristics of the “judge” (i.e., the person evaluating credibility) as well as to characteristics of the target (i.e., the person being judged) and characteristics of the to-be-judged statement.

**Characteristics of the “Judge”**

Given that different judges evaluating the same individuals can vary in their accuracy of detecting deceptive reports, it seems likely that individual differences contribute to accurate deception detection (e.g., Ekman et al., 1999). In theory, this variability may be associated with judge-related factors such as information processing strategies and personality characteristics. In support of this idea, past research suggests that the nature of the information used by judges to evaluate credibility can influence the accuracy of detection. Specifically, it appears that a major impediment to accurate detection is the reliance on cues that are poor indices of deceit (e.g., Porter et al., 2000; Vrij, 1993). For example, there is a widespread, but erroneous, belief that signs of nervousness are reliable cues to deception (see Memon, Vrij, & Bull, 1998; Vrij, 2000). Akehurst, Kohnken, Vrij, and Bull (1996) surveyed police and laypersons about their beliefs regarding cues to deceptive behaviour. Both populations held erroneous assumptions about cues that are suggestive of deception, including such behaviours as clearing the throat, faltering speech, and biting the lips. Consistent with these misconceptions, Ruback and Hopper (1986) found that parole officers relied on such cues as politeness and fidgeting, which are actually unrelated to deception. Ekman and O'Sullivan (1991) found that more accurate judges reported utilizing multiple cues and applied them flexibly to assess credibility, rather than rely on any one cue. Thus, it is clear that mistaken notions about deceptive behaviour can impair the ability to detect deception (e.g., Akehurst et al., 1996; DePaulo, Stone, & Lassiter, 1985a; Ekman, 1988, 1989). The problems with lie detection among experienced law enforcement officers is not surprising given that traditional police training techniques tend to rely on these misconceptions about deception (e.g., Bull, 1989; Kassin & Fong, 1999). It was expected that judges who tend to be more accurate would demonstrate different patterns of cue utilization than judges who tend to make less accurate decisions. Consistent with Ekman and O'Sullivan's (1991) findings, accurate judges were expected to rely on a more flexible combination of cues than inaccurate judges.

A second judge-related characteristic that may be connected to accurate deception detection is handedness, which has implications regarding variations in information processing styles and strategies. In a recent *Nature* article, Etoff, Ekman, Magee, and Frank (2000) compared the deception detection ability of patients with left hemisphere damage (Wernicke’s aphasia), patients with right hemispheric damage, and normal healthy controls. They found that the aphasics were the only group to correctly detect deception above chance (61%). Further, their accuracy levels rose to 7% when they were only exposed to facial expression cues, as opposed to when presented with vocal pitch cues (30%) or a combination of vocal and facial cues (60%). The authors concluded that the aphasic individuals relied more heavily on nonverbal cues (e.g., facial expressions) because their impaired language comprehension limited the ability to utilize verbal cues in making their decisions. From these findings it can be speculated that hemispheric lateralization of functions appears to impact on deception detection ability. A goal of the present study was to extend this research by examining deception detection as a function of handedness. In the present study, handedness was used as a more practical estimate of hemispheric lateralization of functions in nonbrain-damaged individuals. Researchers have found differences in the information processing tendencies/pREFERENCES OF RIGHT- AND LEFT-HANDERS DUE TO HEMISPHERIC LATERALIZATION OF FUNCTIONS (E.G., LEVY, 1969; LEZAK, 1995; MILLER, 1971; SPRINGER & DEUTSCH, 1988; STRACHAN, 2000). For most people, the left hemisphere is dominant for speech/language functions (although 50-67% of left-handers have right hemisphere dominance for language), and the right hemisphere is dominant for processing complex, difficult-to-verbalize stimuli. Further, the left hemisphere is organized for “linear” processing of sequentially presenting stimuli, such as mathematical propositions, and the programming of rapid motor sequences. In contrast, the right hemisphere is superior for “configurational” processing required by material that cannot be described adequately in words or strings of symbols, such as the appearance of a face, or three-dimensional spatial relationships. Although the relationship is certainly not perfect, left-handers sometimes show an information processing style that maps more closely on to functions primarily associated with the right hemisphere.
hemisphere (and vice versa for right-handers). Thus, left- and right-handers show differences across a variety of behavioural and cognitive domains (see Coren, 1992; Levy, 1976; McLaughlin, Dean, & Stanley, 1983; Porac & Coren, 1981; Strachan, 2000). We predicted that left-handed judges would show an advantage in lie detection, particularly in conditions where nonverbal information is available or emphasized due to a possible sensitivity to or bias for processing nonverbal information, such as facial expressions.

A third judge-related characteristic that has received little attention is the influence of the judge's personality characteristics on deception detection. One of the few studies investigating this issue indicates that socially anxious individuals appear to be at a disadvantage in the detection of deception (DePaulo & Tang, 1994). However, others have found that traits of social submissiveness/dominance and extraversion among judges do not seem to influence detection ability (Keating & Helman, 1994; Riggio & Friedman, 1983). More research is required to clarify the role of personality in deception detection. One of the most supported conceptual approaches for describing human personality is the Five-Factor Model (e.g., Costa & McCrae, 1992; John, 1990). This model posits that there are five basic personality factors or domains: neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness. The present study determined whether such personality domains are related to deception detection ability. For example, highly agreeable traits may interfere with one's ability to maintain skepticism during the credibility assessment task.

Vrij (2000) has suggested that overconfidence in one's ability to detect deceit can impair the accuracy of such judgments. According to Vrij, when individuals are overconfident in their ability to detect deception, they may rely on inappropriate heuristics regarding deceptive behaviour (e.g., liars look away). As a result, they may miss important cues to deception/truthfulness and rush to a decision about credibility. However, other research suggests that the ability to detect deceit is unrelated to the individual's level of confidence in their judgment (e.g., Brandt, Miller, & Hocking, 1980; DePaulo, Charlton, Cooper, Lindsay, & Muhlenbruck, 1997). Thus, consistent with the bulk of previous research, confidence was not expected to significantly influence detection accuracy.

Characteristics of the Target
Features of the target of the deception judgment also may influence detection accuracy. In fact, previous research has established that motivated deceivers appear to show more discernible cues indicative of deception than unmotivated deceivers (e.g., DePaulo, Kirkendol, Tang, & O'Brien, 1988). They also 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). In order to decrease the possibility of a confound due to potential differences in motivation, as well as to increase ecological validity, all of the targets in the present study were motivated (with a monetary reward) to appear as credible as possible when relating their deceptive statements.

Stereotyped beliefs about others based on their attractiveness have been well documented and such beliefs may impact on judgments of credibility. In general, more attractive individuals are considered more poised, sociable, mentally healthy, and more intelligent than less attractive individuals (Brigham, 1980; Diener, Wolsic, & Sujita, 1995). Attractive individuals are also less likely to be found guilty of criminal behaviour and tend to be viewed as more honest (e.g., Efrani, 1974; Zebrowitz, Voinescu, & Collins, 1996). DePaulo, Stone, and Lassiter (1985b) found that, in a face-to-face situation, targets were less successful at deceiving attractive individuals than less attractive individuals. However, research is needed to examine the impact of target attractiveness on the accuracy of credibility assessment. This was a goal of the present study. Given that attractive persons tend to be viewed as more likable and trustworthy (e.g., Patzer, 1983), judges may be more inclined to evaluate an attractive individual's report as being more honest than the report of an unattractive target. Therefore, target attractiveness was expected to create a "truth bias" and interfere with the task of credibility assessment.

The interaction between the genders of the target and the judge also may influence the detection of deceit. Although most studies find that males and females do not differ in their general ability to detect deception (e.g., DePaulo, Epstein, & Wyer, 1993; Hurd & Noller; 1988; Porter et al., 2000), there is some evidence that detection accuracy is influenced by whether the target is the same or opposite gender to the judge. In one of the few studies addressing this issue, deceptive targets describing opinions on controversial issues were more easily detected if they were of the opposite gender to the judge rather than of the same gender (DePaulo et al., 1985b). Forrest and Feldman (2000) also found a gender bias in detection accuracy. In their study, judges rated the perceived honesty of targets as they lied or told the truth about their opinions on various issues. Female judges rated same gender targets as more honest than male targets. Similarly, male judges perceived same gender targets as more honest, but only when they were motivated to
do their best at the deception task. Low motivated male judges viewed female targets as more sincere than male targets. This target/judge gender interaction suggests a bias in the detection of deceit. Judges appear to be more inclined to accept statements made by same gender targets than they are of opposite gender targets. The level of motivation to detect deception seems to influence this gender bias for male judges, but not for female judges. It may be the case that the gender interaction effect is at least partly attributable to stereotypes held about the opposite gender. To further investigate the target/judge gender interaction effect, research is required to determine whether this effect generalizes beyond the credibility evaluation of simple opinion statements to situations in which individuals are being deceptive about personal experiences. In forensic settings, there often is a need for investigators to evaluate the veracity of reports concerning emotional or criminal events, rather than opinions or minor forms of deceit. Presumably, in evaluating such an emotional report, the judge would have more information upon which to make a decision concerning credibility and, as such, the gender bias would be minimized. It was hypothesized that the gender interaction effect would not be replicated in the present study given the use of more complex statements about emotional events as stimuli.

**Stimuli Characteristics**

A final issue relating to deception detection considered in the present study was the nature of the deceptive and truthful statements presented to participants. The modality in which the statement is presented to credibility assessors may impact on their ability to detect deceit. However, the research findings on presentation modality have been somewhat inconsistent. Miller, Bauchner, Hocking, Fontes, Kamiński, and Brandt (1981) asked participants to judge the veracity of videotaped targets giving statements about a stimulus they had just witnessed. The statements were either presented to the participants in an audio-only, video-only, or transcript-only format. The authors found that participants exposed to the audio and transcript formats were more accurate than those presented with only the video. However, DePaulo et al. (1985b) found that participants were better at detecting deceit when the accounts were presented in an audio-only, video-only, or audio-visual modality compared to a transcript format. DeTurk and Goldhaber (1988) also reported that judgments of honesty were more accurate when the statement in question was presented in an audio-visual format than when presented in a text-based modality. More recent work provides further evidence of a detection advantage in the audio-visual modality. Atoum and Al-Simadi (2000) presented participants with statements provided by targets who truthfully described individuals they either liked or disliked. Results indicate that judges exposed to the audio-visual format produced the highest (most accurate) ratings of honesty. The authors suggested that, relative to the restricted amount of information in the other modalities, judges were able to acquire a wider range of perceptual cue information from the audio-visual format that may assist in the task of credibility decision-making. However, Atoum and Al-Simadi did not include deceptive statements in their study. A purpose of the present study was to identify the most effective modality for accurate credibility assessment utilizing both honest and deceptive statements. It was predicted that participants exposed to the audio-visual format during the detection task would be more accurate in deciding credibility than those who were exposed only to the target’s voice.

A final consideration related to the nature of the stimulus is the veracity of the experience described. Porter et al. (2000) found that the accuracy rate for identifying truthful accounts (51%) was lower than that of deceptive accounts (59%). Ekman et al. (1999) found a similar pattern and hypothesized that this difference may be due to the fact that there are more cues to deception than to truthfulness. As such, people may be more likely to notice cues to deception. However, others have found a “truth bias” in detection and report lower accuracy rates for deceptive accounts (see Vrij, 2000). This inconsistency may be the result of the heterogeneity in the deceptive/truthful stimuli used across studies investigating deception. For example, it may be easier to identify a description of a deceptive event than to detect a deceptive opinion. In the present study, the relationship between the veracity of the account and judgment accuracy was examined using honest and deceptive depictions of emotionally charged personal experiences. This is an important extension given that in many applied settings, such as an interrogation, the incident at issue is usually a personal emotional event (e.g., a crime). It is possible that descriptions of truthful and deceptive emotional events are associated with greater differences in both verbal (e.g., amount of detail) and nonverbal (e.g., facial expressions) behavior than other more innocuous events. To summarize, participants in the present study were asked to judge the veracity of four truthful and four fabricated descriptions of emotional personal events. Characteristics of the judge, characteristics of the target, and the nature of the to-be-judged report were hypothesized to relate to the accuracy in detecting deception.
METHOD

Participants

The participants were 310 students (244 females and 66 males) recruited from undergraduate psychology courses. Ages ranged from 18-43 years ($M = 21.1$, $SD = 2.6$), and 10% ($n = 31$) reported that they were left-handed. All participants received course credit for their participation.

Materials

True and deceptive stimuli. Statements of persons describing either a true or a fabricated emotionally charged experiences were used. These statements were generated in a previous study (Porter et al., 1999) in which young adults were videotaped during an interview about emotional personal incidents from childhood (e.g., an animal attack, a serious accident, getting lost). The videos displayed the target’s upper body and face as he or she was seated at a table across from the interviewer (who was off screen). Some of these statements were truthful, as indicated by parental report, while others were fabricated. Before the interview about the deceptive account, the target was offered a monetary incentive ($50) to convince the interviewer (blind to statement veracity) that he or she was telling the truth. Fifteen minutes were provided to prepare their account.

In the present study, participants were shown the first two minutes of the videotaped interviews, during which the target freely recounted a truthful or fabricated event without interruption from the interviewer. Twenty targets (consisting of an equal number of male and female targets) were randomly selected from the original pool of videotapes and were rated for attractiveness by a pilot sample of 49 undergraduate students on a scale from 1 (not at all attractive) to 5 (very attractive). Based on these ratings, the two male ($M = 3.3$ and 3.5, $SD = 1.1$ and 1.0, respectively) and two female ($M = 3.8$ and 4.1, $SD = 0.9$ and 0.8, respectively) targets that received the highest attractiveness ratings were selected to represent the “attractive target” condition. Similarly, the two male ($M = 2.5$ and 2.6, $SD = 1.0$ and 0.9, respectively) and two female ($M = 2.4$ and 2.7, $SD = 0.8$ and 0.9, respectively) targets with the lowest attractiveness ratings were selected to represent the “unattractive target” condition. Thus, a total of eight videotapes were selected for use in the current experiment, half of which were honest and half of which were deceptive.

Deception Analysis Questionnaire. A questionnaire designed specifically for this study, the Deception Analysis Questionnaire (DAQ), was administered to all participants. The DAQ included questions pertaining to participants’ age, gender, and dominant handedness. It also asked participants to rate their perceived general ability at detecting deceit on a scale of 0 (not good at all) to 6 (extremely good). After each target’s account was presented, participants used the DAQ to record their credibility judgment (i.e., honest or deceptive), rate their degree of confidence in that judgment on a scale of 0 (not at all confident) to 6 (highly confident), and describe the cues they used to make their decision. The type of cue reported by judges was coded by a scheme similar to Porter et al. (1999). Although multiple cues could be listed, only the first cue reported by each judge was evaluated because this cue likely reflected the most salient feature used by the judge in making each credibility judgment. Cues pertaining to verbal qualities included such features as pitch of voice, details of story, pauses in speech, and laughing. These verbal cues were categorized according to whether they captured features of what was said (i.e., content) versus elements of how it was said (i.e., speech features). Nonverbal cues included features related to the target’s body and body movements, such as posture, eye contact, and hand movements. This category was subdivided into nonfacial body cues versus facial cues. Finally, the term “vague” cues was used to classify nonspecific cues such as “a gut feeling,” intuition, or general nervousness of the target. Twenty percent of these cues were re-coded by a second rater. The coding scheme demonstrated an acceptable level of inter-rater reliability ($M$ Kappa = .80; range = .70-.90). After the type of cue was identified, the total number of cues used across targets within each cue type category was calculated.

Personality characteristics. The “Big Five” Inventory (BFI; John, Donahue, & Kentle, 1991) was used to measure personality traits conceptualized by the Five-Factor Model of personality (e.g., McCrae & John, 1992). The BFI is a 44-item self-report questionnaire designed to assess an individual’s characteristics with regard to five broad personality domains. These domains include Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. On a scale of 1 (disagree strongly) to 5 (agree strongly), participants rated the degree to which each item statement was descriptive of them. The mean score for items related to each of the “Big Five” domains are calculated to determine the individual’s score per domain. The BFI has been used in numerous studies (e.g., Gross & John, 1995; Johnson & Wolfe, 1995; Watson, Clark, & Harkness, 1994) and has been shown to be reliable and valid in diverse cultural samples (e.g., Benet-Martinez & John, 1998). The BFI has demonstrated good levels of internal consistency.
within its scales (coefficient alpha range from .79 to .88 in samples of college students and college-educated individuals). In addition, the factor structure of the BFI supports the “Big Five” model of personality (Benet-Martinez & John, 1998).

**Procedure**

The basic design of the present study was a mixed-subjects design with statement veracity (truth/deceptive), presentation modality (audio-visual/audio-only), target attractiveness (attractive/unattractive), and target gender (male/female) as the within-subjects variables. Judge handedness and judge gender were between-subjects variables. Counter-balancing was used where appropriate. All participants were shown half of the target reports in an audio-visual format and the other half in an audio-only format (one group of participants received the video format first followed by the audio-only format, whereas the other group received the opposite order). The level of attractiveness, gender of the speaker, and the veracity of the memory accounts were counterbalanced across the two modalities. That is, the audio-only condition consisted of an attractive male telling the truth, an unattractive male telling a lie, an attractive female telling the truth, and an unattractive female telling a lie. The audio-visual condition consisted of a different set of four videos with similar characteristics (i.e., attractive male telling a lie, an unattractive male telling the truth, and attractive female telling a lie, and an unattractive female telling the truth). The same eight targets were presented to the second group of participants but were presented in the opposite modality. The order of presentation of the targets within each modality was randomized and all targets were presented individually. After each presentation, participants were given approximately four minutes to judge the veracity of the account and to specify the cues they used to make their decisions. Upon completion of the detection task, participants completed the BFI, and were debriefed.

**RESULTS**

**Overall Accuracy**

Collapsed across all variables and experimental conditions, the mean accuracy for the entire sample in judging target credibility was 57.6% ($SD = 17.77$). Twenty-six percent of the judges demonstrated an accuracy level of 75% or above, while only two judges were 100% accurate across all memory reports. Male judges were significantly more confident ($M = 4.14, SD = .79$) than females ($M = 3.76, SD = .75$) in their judgments of credibility, $t(296) = 3.65, p < .001$, but judgment confidence was not significantly associated with accuracy for male, $r(64) = .14, p > .05$ or female judges, $r(234) = .09, p > .05$. Figure 1 displays the overall mean accuracy rate for modality, attractiveness, target gender, and statement veracity.

**Judge and Target Characteristics**

**Target and Judge Gender.** Table 1 shows the descriptive statistics for judge-related characteristics and their
TABLE 1
Descriptive Statistics for Individual Difference Factors and Correlations with Detection Accuracy

<table>
<thead>
<tr>
<th>Individual Variables</th>
<th>Percentage</th>
<th>M (SD)</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge Gender (0 = male; 1 = female)</td>
<td></td>
<td></td>
<td>-0.01</td>
</tr>
<tr>
<td>21.3% male</td>
<td>78.7% female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judge Handedness (0 = left; 1 = right)</td>
<td></td>
<td></td>
<td>-0.16**</td>
</tr>
<tr>
<td>9.7% left</td>
<td>90.3% right</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judge Age</td>
<td>21.14 (2.17)</td>
<td>-0.01</td>
<td></td>
</tr>
<tr>
<td>Perceived Detection Ability</td>
<td>3.70 (.99)</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Confidence in Judgment</td>
<td>3.83 (.77)</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>BFI Neuroticism Scale</td>
<td>3.21 (.38)</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>BFI Extroversion Scale</td>
<td>3.42 (.36)</td>
<td>-0.06</td>
<td></td>
</tr>
<tr>
<td>BFI Openness to Experience Scale</td>
<td>3.49 (.48)</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>BFI Conscientiousness Scale</td>
<td>3.56 (.33)</td>
<td>-0.04</td>
<td></td>
</tr>
<tr>
<td>BFI Agreeableness Scale</td>
<td>3.39 (.32)</td>
<td>-0.07</td>
<td></td>
</tr>
</tbody>
</table>

* p < .05, ** p < .01, *** p < .001.

associated correlations with accuracy. To test the hypothesis concerning the interaction between judge and target gender on accuracy (expressed as % correct), a within-subjects analysis of variance (ANOVA) was conducted. Gender of judge was the between-subjects variable and target gender was the within-subjects variable. This analysis indicated that male judges were more accurate (M = 58.1, SD = 17.5) as female judges (M = 57.5, SD = 17.9) in their evaluations of target credibility, F(1, 302) = .05, p > .05, η² = .00. However, a main effect of target gender was found in that judges were less accurate at evaluating the credibility of female targets than they were of male targets, F(1, 302) = 56.12, p < .001, η² = .16. In addition, the hypothesized target gender x judge gender interaction was found, F(1, 302) = 3.96, p < .05, η² = .01. Male judges were more proficient at evaluating the credibility of female targets (M = 51.1, SD = 26.7) than were female judges (M = 45.6, SD = 27.3). In contrast, female judges were more accurate in determining credibility of male targets (M = 69.4, SD = 23.9) than were male judges (M = 65.0, SD = 20.6).

Target Attractiveness. A within-subjects ANOVA was conducted with target attractiveness as the within-subjects factor (attractive/unattractive) and percent accuracy as the dependent measure. This analysis was based only on the level of accuracy measured during the audio-visual modality, as judges could not view the target in the audio condition. As predicted, a main effect of target attractiveness was found, F(1, 309) = 4.06, p < .05, η² = .01. Judges were significantly more accurate at determining credibility when the target being judged was unattractive (M = 61.6, SD = 29.2) than attractive (M = 56.9, SD = 33.3).

Judge Handedness. A multivariate analysis of variance (MANOVA) was conducted with handedness as the independent variable and overall accuracy, accuracy for deceptive reports, accuracy for truthful reports, accuracy for audio-visual reports, and accuracy for audio-only statements as the dependent measures. This approach was used to maximize the power of the analysis given the small number of left-handers in the sample and to reduce the chance of Type I errors given the unequal ns and the exploratory nature of this variable. The MANOVA was significant (Hotelling’s Trace = .031), F(3, 299) = 3.13, p < .05, η² = .03. Follow-up analyses indicated that left-handers were significantly more accurate overall (M = 66.2%, SD = 15.4) than right-handers (M = 56.7%, SD = 17.8), F(1, 301) = 7.91, p < .01, η² = .026, as shown in Figure 2. Specifically, left-handers were significantly better at judging honest accounts than right-handers, F(1, 301) = 8.67, p < .01, η² = .028, but the two did not differ in the detection of deceptive accounts, F(1, 301) = 2.00, p > .05, η² = .007. Further, left-handers were better at judging credibility in the audio-visual modality than right-handers, F(1, 301) = 4.86, p < .05, η² = .016. However, handedness did not significantly influence the accuracy of judgments made in the audio condition, F(1, 301) = 3.86, p > .05, η² = .026.

Judge Personality. To examine the relation between personality characteristics and detection accuracy, correlations between BFI scale scores and accuracy rates were examined. As shown in Table 1, the accuracy of credibility judgments was not significantly correlated with any of the five personality domains measured by the BFI (ps > .05) or with baseline ratings of judges’ perceived ability to detect deception, r(309) = .05, p > .05.

Cue Utilization. As shown in Table 2, the mean number of cues reported by judges across targets was relative-
Individual Differences and Deception Detection

targets as the dependent measures. The MANOVA was significant (Hotelling’s Trace = .04), $F(5, 303) = 2.28, p < .05, 
\eta^2 = .04$. Follow-up analyses indicate that left-handers reported using significantly fewer content cues 
($M = 1.40, SD = .84$) than right-handers ($M = 1.74, SD = .89$), $F(1, 307) = 3.93, p < .05, \eta^2 = .01$. However, left- 
and right-handers did not significantly differ in the number of speech, facial, body, or vague cues types 
used, $F(1, 307) = 0.97, 2.64, 0.75$, and $1.86, ps > .05$, respectively.

Stimulus Characteristics

Within-subjects ANOVAs were used to test the hypothesis that judgment accuracy was affected by the veracity of the statement under investigation and the modality in which it is presented. These analyses indicated that judges were significantly more accurate when the target was reporting a deceptive statement ($M = 61.0, SD = 22.9$) rather than a truthful statement ($M = 54.3, SD = 22.8$), $F(1, 303) = 16.59, p < .001, \eta^2 = .05$. In addition, although the mean accuracy for the audio-visual condition ($M = 59.4, SD = 24.2$) was higher than that of the audio condition ($M = 55.9, SD = 23.5$), this difference was not significant, $F(1, 303) = 3.57, p > .05, \eta^2 = .01$.

A number of variables have been identified above as yielding a higher level of accuracy including left-handed judges in the audio-visual modality, judges of the opposite gender to the target, and unattractive targets. Although detailed analyses were not possible due to the small number of left-handed participants, descriptive analyses indicate that of the 22 left-handed female judges, 86% correctly assessed the credibility of unattractive male targets in the audio-visual modality. Similarly, 62% of the 8 left-handed male judges correctly determined the credibility of unattractive female targets when exposed to the audio-visual format. However, even when the targets were of the same gender as the judge, 62% and 75% of female and male left-handed judges (respectively) accurately assessed credibility under these conditions.

DISCUSSION

The accurate detection of deceit is an important skill for professionals working in applied settings, ranging 
from forensic to occupational contexts. Although most people have difficulty in detecting deception, there is 
considerable variation in how well individuals do at this task. We hypothesized that individual differences in 
detection ability would be related to characteristics of the judge and target, as well as aspects of the statement 
der under investigation. To extend the ecological validity of deception detection findings, we used 
statements about emotionally charged personal events...
and motivated deception. Overall, the present findings provided substantial evidence that the process of credibility evaluation can be influenced by judge, target, and statement characteristics.

It was hypothesized that the pattern of information processing strategies (i.e., cue utilization) during judgment formulation contributes to the ability to detect deceit. This hypothesis was supported. Specifically, the reliance on a greater number of facial cues was associated with a higher level of accuracy (in the audio-visual modality), regardless of whether the statement was truthful or deceptive. This finding is consistent with research by Ekman and his colleagues (e.g., Ekman, 1989, 1992; Ekman, Friesen, & O'Sullivan, 1988; Frank & Ekman, 1997) that supports the validity of facial cues (e.g., false smiles and microfacial expressions) in lie detection. Cues related to nonfacial body characteristics, features of speech, or vague intuition, however, were not related to detection ability. In addition, the present study indicated that the reliance on a greater number of cues related to the content of speech was associated with impaired credibility judgments. This seems to contradict research findings that have identified content differences between truthful and deceptive statements (e.g., Porter et al., 1999; for a review see Vrij, 2000). Given such evidence, attention to content cues would be expected to facilitate detection accuracy rather than impair it. However, it is possible that the inflexible reliance or an over-reliance on content cues (the most frequently reported cue) may have prevented judges from attending to other potentially useful indicators of credibility (e.g., facial cues). This possibility seems plausible given that the use of a flexible cue strategy and the consideration of a combination of verbal and nonverbal cues may help increase the accuracy of deception detection (e.g., Ekman & O'Sullivan, 1991).

Therefore, although the associations between certain cue types and detection accuracy were modest in the present study, they imply that attention to facial cues can facilitate detection while an over-reliance on content cues (to the detriment of other possible indicators) can impair this ability. The evaluation of deception cues and their utilization by accurate credibility assessors are important areas of research. Information obtained from such research may help professionals identify situations in which they should be suspicious of a suspect's/witness's statement and inspire further avenues of questioning or investigation. This is particularly important given that Porter et al. (2000) recently found that training professional lie detectors on the valid cues to deception, dispelling myths about deceptive cues, and providing judgment feedback can improve the accuracy of credibility judgments.

Another important judge-related characteristic hypothesized, and found, to be associated with deception detection accuracy was handedness. For the first time, left-handers were shown to be substantially better at detecting deception than were right-handers. As expected, the advantage of left-handers was particularly evident when the target's report was presented in an audio-visual format. In this format, judges were exposed to both the verbal and nonverbal behaviours of the target. Although left- and right-handers reported using a similar number of speech, body, facial, and vague cues, left-handers reported relying less on the content of the statements to decide credibility. This finding suggests that left-handers may differ from right-handed individuals in the manner in which they evaluate credibility and may be more flexible in the cues they use. It is unlikely, however, that less reliance on content cues solely accounts for the more proficient detection accuracy of left-handed judges. Other useful information may have been included in the decision-

<p>| Table 2: Mean Number of Cues Used Under Each Cue Type Across Targets |
|---------------------------------|-----------------|----------------|</p>
<table>
<thead>
<tr>
<th>Cue Type</th>
<th>M Number Used (SD)</th>
<th>r with Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cue Use in the Audio-visual Condition:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td># of content cues</td>
<td>1.48 (1.05)*</td>
<td>-.12*</td>
</tr>
<tr>
<td># of speech cues</td>
<td>0.59 (0.74)*</td>
<td>.01</td>
</tr>
<tr>
<td># of nonfacial body cues</td>
<td>0.62 (0.77)*</td>
<td>.00</td>
</tr>
<tr>
<td># of facial cues</td>
<td>0.68 (0.82)*</td>
<td>.22***</td>
</tr>
<tr>
<td># of vague cues</td>
<td>0.55 (0.79)*</td>
<td>-.06</td>
</tr>
<tr>
<td><strong>Cue use in the Audio-only Condition:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td># of content cues</td>
<td>1.93 (1.13)*</td>
<td>-.14*</td>
</tr>
<tr>
<td># of speech cues</td>
<td>1.38 (1.11)*</td>
<td>.08</td>
</tr>
<tr>
<td># of vague cues</td>
<td>0.48 (0.73)*</td>
<td>.12</td>
</tr>
</tbody>
</table>

*Note. Contrastin letters ("a", "b", or "c") within each modality represent a significant mean difference between cue types. Identical letters (e.g., "b", "b") denote a nonsignificant difference. *p < .05; **p < .01; ***p < .001.
making process of left-handers, but were not well assessed in the present study (e.g., attention to specific micro-facial expressions). Left-handers also may not have been fully cognizant of all the cues they used or may have been using a visual-based processing strategy. As such, their decision-making strategies may have been difficult to verbalize. The use of a visual-based process by left-handers is further supported by the absence of a distinct proficiency in detection when exposed to only the verbal qualities of the report.

Prior research has suggested differences in the information processing tendencies and preferences of right- and left-handers due to hemispheric lateralization of functions (e.g., Lezak, 1995; Porac & Coren, 1981; Strachan, 2000). However, it would be premature to conclude that the advantage of left-handers in lie detection results solely from differences in hemispheric lateralization and dominance. In addition, many left-handers do not differ notably from right-handed individuals in cerebral asymmetry (Lezak, 1995). Nonetheless, there is reason to argue that hemispheric influences may play a role in deception detection. As mentioned, Etcoff et al. (2000) found that the patients with left hemispheric damage were the only group to detect deception above chance (61%) compared to normal controls and right hemisphere damaged individuals. This was especially the case when the to-be-judged accounts allowed the examination of facial expressions in the absence of vocal cues. Ekman and his colleagues (e.g., Ekman et al., 1988) have demonstrated that there are reliable micro-facial cues to deception, although they are difficult for untrained observers to perceive. Perhaps left-handed individuals and/or individuals who have a tendency toward processing strategies associated with the right hemisphere have an advantage in this ability. It seems possible that handedness may be associated with particular cognitive abilities or information processing tendencies/styles (e.g., flexibility in thinking or preference for visuospatial rather than verbal stimuli) that are sensitive to subtleties in communication (e.g., facial expressions) and potentially contribute to lie detection. In fact, we have recently replicated the left-hander finding with suggestions of a facial processing superiority among left-handers (Porter, Birt, & Angelopoulos, 2001). Future research is required to examine the decision-making and information processing strategies of credibility assessors and the manner in which such strategies relate to detection accuracy.

Other judge-related factors examined in this study included self-reported proficiency and confidence in detection and personality features. As found previously (see Vrij, 2000), individuals’ beliefs about their ability to detect deception had little bearing on their actual ability to judge veracity. Similarly, although males reported being more confident in their decisions than did females, confidence was not related to the accuracy of those judgments (see also DePaulo et al., 1997). In contrast to our hypothesis, the “Big Five” personality traits did not appear to have any discernible association with one’s ability to detect deceit or honesty in another individual. Therefore, it appears that personality characteristics may not influence the process of credibility assessment. This does not mean, however, that dysfunctional personality traits, such as psychopathy (or other, more specific traits associated with personality) would have no influence on the ability to detect deception. Future research is required to investigate the role of such personality traits on credibility judgments. In addition, as credibility assessment is a dynamic process influenced by many factors, future research should explore the interaction between the personality traits of the credibility assessor with those of the target.

In addition to judge-related characteristics, the present research investigated the relationship between characteristics of the person being judged and detection accuracy. DePaulo et al. (1985b) found that deception by female targets was generally more detectable than deception by males. However, the present study found that females were more successful liars than males. This discrepancy may be related to the nature of the event or topic under evaluation. In the present study, descriptions of true and deceptive emotional events were the stimuli of interest and it is possible that females are able to deceptively portray emotional material more convincingly than males. To better understand this issue, future research should compare judges’ accuracy in assessing credibility across various types of statement content (i.e., neutral/emotional, positive/negative, or opinion/personal experience).

Results also indicated that the attractiveness of the target was related to the accuracy of credibility judgments. When participants judged an attractive target they were less accurate in their decisions than when judging an unattractive target. One possible explanation for this finding is that the target’s attractiveness interfered with the judge’s decision-making by the activation of common stereotypes about attractive individuals (e.g., more amiable and trustworthy; Zebrowitz et al., 1996) and/or by distracting judges from relevant cues to deception. This explanation appears plausible given the findings of Bodenhausen and Lichtenstein (1987). When faced with complex judgments (e.g., guilt judgments), people often use stereotypes to assist their decision-making. This is especially true when minimal factual information is available, as is often the case in deception detection.
paradigms. When stereotypes are evoked, there is a tendency to attend to stereotype-consistent information, while minimizing or ignoring details that are inconsistent with the stereotype (e.g., Synder, Campbell, & Preston, 1982). Therefore, important cues may be missed when judging an attractive target because the act of deception would be inconsistent with the stereotype of attractiveness. In order to minimize the potential bias of target attractiveness, credibility assessors may benefit from training protocols that facilitate their awareness of possible stereotypes and their potential to influence decisions of credibility.

It is important to note that, contrary to our hypothesis, the use of emotionally based personal statements did not override the gender bias previously observed in the deception detection research with opinion-based statements (DePaulo et al., 1985b; Forrest & Feldman, 2000). Specifically, an interaction between the judge and target gender was found. This interaction indicates that judges performed better when evaluating a target of the opposite gender. Although the gender stereotypes of the present sample were not measured, it is possible that, as with attractiveness, such beliefs biased judges to view same gender targets as more honest and trustworthy, leading to a higher rate of errors for same-gender targets. For example, females are typically viewed as more proficient at expressing themselves nonverbally than males (e.g., DePaulo et al., 1993; Jussim, Milburn, & Nelson, 1991). In addition, compared to males, females tend to be better at reading or decoding the intended messages sent by others through their nonverbal behavior, are less suspicious of others, and are more willing to view others as truthful (see Vrij, 2000). As such, female judges may be more accepting of the misleading messages sent by female targets than they would be of male targets. The applied implication of this finding is that the task of credibility assessment may be facilitated by the involvement of at least one evaluator (as an observer) who is of the opposite gender to the individual under investigation. However, additional research is required to investigate the source of the gender bias in deception detection, the circumstances under which it is most likely to occur, and methods of minimizing its bias in the task of credibility assessment.

Besides judge and target characteristics, the truthfulness of the statement and the format in which it was presented to judges also appears to influence detection accuracy. Deceptive statements were more easily detected than were honest reports. As hypothesized by Ekman et al. (1999), the greater sensitivity to deception may be attributable to the fact that there are more signs of deception than there are of honesty. Although others (see Vrij, 2000) have found that truthful statements are more easily detected than deceptive statements, it is possible that when individuals fabricate more complex scenarios, such as of emotional memories from childhood, the signs of deception are more easily observed and/or more difficult to mask. In addition to the veracity of statements under evaluation, past research suggests that there have been inconsistencies regarding the most effective form of modality to deliver target statements. For example, some have favored audio-only and transcript formats over video-only methods (Miller et al., 1981), while others have suggested that an audio-visual format is better than text-based presentation formats (e.g., Atoum & Al-Simadi, 2000). In the present study, there was a trend for accuracy to be higher for reports presented in the audio-visual format. However, this level of accuracy did not differ significantly from that of the audio modality. In other words, judges may have benefited slightly from exposure to the nonverbal behavior of the target as they recounted an emotional event, but their performance was not substantially impaired (relatively) by the absence of these behaviors in the audio condition. Therefore, when evaluating emotional-oriented reports of historical events, an audio-visual format may be as useful as an audio-only evaluation style. It also may be helpful to have two evaluators, one for each modality, who each form independent judgments of credibility before coming to a consensus concerning the veracity of the statement. Nonetheless, more research is required to determine the most effective mode of statement delivery for credibility evaluations and the conditions under which detection accuracy can be maximized.

Although the present study offers important information about deception, some limitations should be noted. Targets were motivated to deceive with a monetary incentive but this incentive would not equate with the pressure to appear credible that many individuals experience in applied settings (i.e., while being interrogated). Moreover, research suggests that as the level of motivation to deceive increases, deception becomes more discernable (e.g., DePaulo et al., 1988; Forrest & Feldman, 2000). Field studies on deception in realistic contexts (e.g., during interrogations) with an established “ground truth” are required. That is, real-life forensic situations should be used in which the validity of the information evaluated has been clearly confirmed or disconfirmed by external sources of information. In addition, some may argue that the passive detection paradigm used in the current study limited accuracy because the judges were unable to question or challenge the targets as they gave their statements. However, in challenge to this perspective some researchers have demonstrated that
interaction with the target actually decreases deception detection accuracy (Buller, Comstock, Aune, & Strzyzewski, 1989; Granhag & Stömwall, 2001; Ruback & Hopper, 1986). Future research should attempt to clarify the mechanisms underlying the current findings, especially the role of handedness and information processing, and decision-making strategies in accurate credibility assessment. In summary, the present findings offer an original demonstration of a number of factors associated with the judge, the target, and the presentation of the statement under investigation, that appear to be related to the accuracy of lie detection. The current findings also offer promising avenues for future deception detection research that may lead to improved methods of credibility assessment in applied settings.

The first author was supported by an operating grant from the Social Sciences and Humanities Research Council of Canada (SSHRC) and a research and development grant from Dalhousie University. The second author was supported by a doctoral fellowship from SSHRC and an Isaac Walton Killam Memorial Scholarship. The fourth author was supported by a doctoral fellowship from the University of British Columbia. We extend thanks to Mike Woodworth for his contributions during the preparation and implementation of this research study. We also thank Tanya Costain, Naomi Doucette, Rob MacArthur, and Christine Fryer for their assistance in coding data, as well as Talia Profit, Jette Midtgaard, Cathy Stewart, Yarrow Joyce, and Elaine Yandoo for their assistance with scoring questionnaires.

Correspondence should be sent to the first author at the Department of Psychology, Dalhousie University, Halifax, Nova Scotia, Canada, B3H 4J1; E-mail: sbporter@is.dal.ca.

References
Communication Quarterly, 36, 276-289.


University of California, Institute of Personality and Social Research.


Psychology, Crime and Law, 1, 319-331.
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.R. (2001). The discrimina-
tion of deceitful, mistaken, and real memories for
emotional events. In R. Roesch, R.R. Corrado, & R.
Dempster (Eds.), Psychology in the courts: International
of real, implanted, and fabricated memories for emo-
tional childhood events: Implication for the false
memory debate. Law and Human Behavior, 23, 517-537.
Riggio, R.E., & Friedman, H.S. (1983). Individual differ-
ences and cues to deception. Journal of Personality and
Social Psychology, 45, 899-915.
parole interviewers: The effect of case and interview
factors. Law and Human Behavior, 10, 203-214.
hypotheses about human nature: Assessing the accu-
Strachan, A.L. (2000). In the brain of the beholder: The
neuropsychological basis of aesthetic preferences. The
Harvard Brain, 7, 1-10.
impact of nonverbal behavior, social skills, and physi-
cal characteristics in impression formation. Journal of
Social Psychology, 133, 601-610.
Vrij, A. (2000). Detecting lies and deceit: The psychology of
lying and the implications for professional practice.
Chichester, UK: Wiley.
Structures of personality and their relevance to psy-
chopathology. Journal of Abnormal Psychology, 103, 18-
31.
“Wide-eyed” and “crooked-face”: Determinants of per-
cieved and real honesty across the lifespan. Personality

Received October 23, 2000
Revised August 13, 2001
Accepted October 22, 2001