Sidetracked by emotion: Observers’ ability to discriminate genuine and fabricated sexual assault allegations

Kristine A. Peace1*, Stephen Porter2 and Daniel F. Almon3
1Grant MacEwan University, Edmonton, Alberta, Canada
2University of British Columbia, Kelowna, British Columbia, Canada
3Dalhousie University, Halifax, Nova Scotia, Canada

Purpose. Assessing the credibility of reports of sexual victimization – often in the absence of corroboration – presents a significant challenge for legal decision makers. This study examined the accuracy of observers in discriminating genuine and fabricated sexual assault allegations. Further, we examined whether individual differences and cue utilization strategies influenced deception detection accuracy.

Methods. Observers (N = 119) evaluated eight (four truthful and four deceptive) written allegations of sexual assault (counterbalanced), and completed a Credibility Assessment Questionnaire (CAQ) and individual differences measures.

Results. Results indicated that overall accuracy was below chance (M = 45.3%), and a truth bias was evidenced. Examining the Big Five personality traits, we found that openness to experience and neuroticism were positively associated with accuracy, whereas extraversion was negatively related to accuracy. Further, judgement confidence was negatively associated with accuracy.

Conclusions. The present study offers insights into observers’ perceptions of credibility regarding real-life sexual assault allegations. Implications for legal decision making are discussed.

During the past several decades, the detection of deception has come under increasing empirical scrutiny (e.g., see Vrij, 2008). One reason for the enormous scientific interest in deception is the increasingly recognized problems that have arisen in the legal system as a result of failed credibility assessments, such as wrongful convictions (e.g., Porter & ten Brinke, 2010; Vrij, Granhag, & Porter, 2010). In particular, many sexual assault cases pose a major challenge for police and legal decision makers; the offence often is not reported immediately and there is sometimes a complete absence of corroborating evidence, especially when considering increasingly common ‘historical’ sexual assault allegations (e.g., Porter, Peace, Douglas, & Doucette, in press). As a result, legal decision making is discussed.

*Correspondence should be addressed to Dr. Kristine Peace, Department of Psychology, Grant MacEwan University, City Centre Campus, Rm 6–329H, 10700 – 104 Avenue, Edmonton, AB, Canada T5J 4S2 (e-mail: PeaceK@macewan.ca).

DOI:10.1111/j.2044-8333.2011.02031.x
makers often must evaluate the veracity of a claim based on ‘he said, she said’ evidence. While the base rate of false allegations of sexual assault is difficult to establish, some estimate that as many as 20% of them could be false (Greer, 2000; Trocmé & Bala, 2005). As such, research on factors associated with the features of true and false allegations of sexual assault and the accuracy of observers in their discrimination is essential. While the literature on deception detection is extensive (e.g., DePaulo et al., 2003; Vrij, 2008), few studies have focused on ‘emotional’ lies or claims of sexual trauma specifically (Porter & ten Brinke, 2010; cf. Parker & Brown, 2000). The present study examined the ability of lay observers (potential jurors) to discriminate true and false sexual assault reports, and whether individual differences are related to the ability to determine the credibility of such claims.

Credibility assessment and emotion
Studies in the area of deception detection typically indicate that the accuracy of both laypersons and legal decision makers tends to fall around the level of or slightly above chance (see Bond & DePaulo, 2008; Vrij, 2008). Further, a common pattern among evaluators is one of mediocre performance combined with inflated confidence, a formula for disaster in forensic settings. However, the great majority of deception detection studies have examined lies about innocuous, non-emotional events (e.g., someone erasing the blackboard, a film witnessed, mock crime) whereas few have examined lies about powerful emotional events. Previous research has found that subjective ratings of narrative characteristics (i.e., sensory and contextual details) are higher for negative relative to positive events (Barnier, Sharman, McKay, & Sporer, 2005). When Peace and Sinclair (in press) examined deception detection concerning honest or deceptive autobiographical events, they found that the detection of false emotional reports was impaired relative to the neutral reports. Their study showed preliminary evidence of an emotive truth bias, such that emotional stories tend to be inherently ‘believed’ furthering impairing deception ability. In turn, it is possible that the emotional content/intensity of a lie may influence the quality of deception detection judgements. Further, the emotional content of statements (and the emotion aroused in the judge) often is used in judgements of truthfulness (e.g., Semmler & Brewer, 2002). In fact, observers often assess credibility based, in part, on the presence of emotional content (Vrij, 2008). However, Kaufmann, Drevland, Wessel, Overskeid, and Magnussen (2003) reported the ‘emotionality’ of rape allegations as more predictive of credibility determinations than the actual content of the reports. This reflects what Russian novelist Alexandr Solzhenitsyn argued: ‘It is not because the truth is too difficult to see that we make mistakes . . . we make mistakes because the easiest and most comfortable course for us is to seek insight where it accords with our emotions’ (cited from Ericson & Mahoney, 2006; also see Oja, 1988).

One purpose of the present study was to determine whether a truth bias is evidenced when observers are confronted with truthful and fabricated claims of sexual assault. We expected that in evaluating written narratives of powerful and disturbing descriptions of sexual assault, observers would be more likely to believe than disbelieve the accounts.

Individual deception detection ability
In addition to establishing observer accuracy in evaluating sexual assault allegations, we sought to identify possible individual differences among lay observers that might
contribute to differing levels of evaluation proficiency. An ongoing debate concerns the question of whether certain individuals, or ‘wizards’ (e.g., O’Sullivan & Ekman, 2004), are consistently better than the average person at detecting lies. Some researchers have reached the conclusion that there are no individual differences whatsoever in detecting deception (e.g., Aamodt & Custer, 2006). Bond and DePaulo (2008) found that individual differences in deception detection were tiny; standard deviations in judgement ability were less than 1%, ranging no more widely than would be expected by chance. They argued that people only differ in their biases regarding the tendency to believe statements, also known as truth or deception biases (see also Masip, Garrido, & Herrero, 2009). Further, deception detection performance is unreliable over testing sessions (Leach et al., 2009), arguing against the role of stable individual differences. However, in examining the studies included in Bond and DePaulo’s (2008) meta-analysis, the vast majority of lies and truths concerned highly innocuous, non-emotional material. It is possible that individual differences could facilitate or impair evaluators’ ability to discriminate true and false emotional stories specifically.

The potency of the story’s emotional content could interact with the emotional functioning and personality of the judge to influence credibility judgements (Campbell & Porter, 2002). Little research has addressed the Five Factor model’s core personality traits (i.e., openness, conscientiousness, extraversion, agreeableness, and neuroticism; Costa & McCrae, 1992; McCrae & John, 1992) in a deception detection context. In fact, many studies present conflicting results (or limited associations) concerning personality features and detection abilities (Bond & DePaulo, 2008; Porter, Campbell, Stapleton, & Birt, 2002). Such inconsistent empirical findings may result from studies that examined personality in relation to detecting mundane or non-emotional deceptive events (Vrij, 2008). For example, some studies have reported that individual lie detection ‘wizards’ are more introverted (e.g., Campbell & Porter, 2002; O’Sullivan, 2005), whereas others have not confirmed this relationship (e.g., Vrij & Baxter, 1999; Vrij, Harden, Terry, Edward, & Bull, 2001). Yet, it is possible that such personality features, which are associated with particular ways of thinking, feeling, and viewing the world, could influence the manner in which the credibility of others is evaluated. Individuals high in certain personality features also may respond differently to emotionally provocative claims, adding another level of complexity in decision making. For example, individuals high in openness to experience might be more open to the possibility of a horrific sexual assault description being false than someone who views the world in a more concrete black-and-white fashion (and perhaps the latter would show a much higher truth bias in this context). Individuals who are high in conscientiousness or introversion may scrutinize rape allegations more diligently, whereas those high in agreeableness may be more likely to judge sexual assault claims as truthful. High neuroticism - a tendency to persistently experience negative emotional states - could be associated with a superior knowledge concerning the valid and perhaps subtle elements signifying genuine emotional distress. Further, high levels of emotional empathy/contagion may be associated with a greater susceptibility to misjudge emotional reports, especially when confronted with highly evocative claims such as rape allegations. In fact, empathic individuals should demonstrate more of an emotive truth bias as their judgement may be clouded by emotion. Each of these personality features also may be related to differential use of cues in decision making. Collectively, empirical evaluation of deception detection abilities and individual differences has yielded inconsistent results (Aamodt & Custer, 2006; Bond & DePaulo, 2008), and research evaluating such factors in credibility assessments of emotional traumas is warranted.
Cue usage
Research examining cue utilization, in general, has found that participants who rely on a greater variety or number of cues (whether verbal and/or non-verbal) tend to be more accurate at deception detection (e.g., Campbell & Porter, 2002; Porter & ten Brinke, 2010). Further, Peace and Sinclair (in press) found that accuracy for judging negative stimuli improved when participants relied on sensory and emotional cues, reported stress levels, and used more cues overall relative to other emotional conditions. Porter et al. (2002) reported that a greater reliance on content cues was associated with enhanced detection accuracy. Given the emotional nature of sexually traumatic reports, it is unclear how cues associated with truthfulness (i.e., amount of general and emotional detail, admissions of forgetting) and deception (i.e., evidence of cognitive operations) may be utilized.

The present study
The present study had three main objectives: (1) to evaluate the accuracy of credibility judgements of true and false allegations of sexual assault; (2) to investigate the role of personality/individual differences in detection accuracy; and (3) to examine cue utilization strategies and their relation to accuracy in credibility decisions. We predicted that (1) participants would perform at or slightly above the level of chance in detecting deception; (2) participants would demonstrate a truth bias, and perform better at detecting truths relative to lies; (3) individual difference factors may influence perceptions of credibility concerning the sexual assault stories, and perhaps accuracy; specifically, higher levels of openness to experience, neuroticism, conscientiousness could enhance credibility assessment in this context, whereas agreeableness, extraversion, and emotional contagion could impair decision making; and (4) accurate judges would utilize more cues relative to inaccurate judges.

Method
Participants
Undergraduate students (N = 119) participated in this study. In order to comply with ethical standards, all potential participants were warned about the explicit and sexually traumatic nature of stimuli to be evaluated. The sample was comprised of 90 (76%) females and 29 (24%) males, with a mean age of 20.8 years (SD = 5.1). Participation was voluntary, and all participants were provided course credit for their involvement in the study.

Materials
Trauma narratives
Narratives describing incidents of sexual trauma in adulthood were collected as part of two previous studies (see Peace & Porter, in press; Peace, Porter, & ten Brinke, 2008), and the original participants consented to use of their narratives in future studies. Genuine sexual traumas were elicited from women who, via self-referral, sought individual psychological services from a local sexual assault centre. An ad was placed in the clinic seeking participants for a research study on memory for sexual assault, during which victims would be asked to describe everything they could remember about their sexual assault in as much detail as possible. Participants provided police incident reports
of the event, which confirmed that events generally happened as described, and all claims led to convictions. Only those claims with corroborating forensic documentation confirming these events were considered for use in the present study. Further, chances are exceedingly small that someone would fabricate a sexual assault – and necessary corroborating evidence – to seek anonymous treatment for sexual assault victims. Fabricated claims of sexual assault were elicited from a sample of participants who indicated they had never been the victims of sexual assault, but freely admitted other types of traumatic victimization (reducing the likelihood they were lying about these experiences). Participants in this study were asked to report on a ‘fake trauma’ and describe this event convincingly and in as much detail as possible as if it had happened to them (Peace & Porter, in press). Selection criteria for the genuine and fabricated narratives/allegations relating to sexual assault were: the sexual assault had occurred as an adult, consent for use of data in further studies had been provided, and the narratives were of comparable lengths (matched across veracity). Truthful accounts \( (M = 599.25 \text{ words}) \) were not significantly different in number of words from fabricated narratives \( (M = 567.25 \text{ words}) \) of sexual trauma, \( t(3) = .31, p > .05 \). On the advice of a journal reviewer, an independent sample of researchers in the area of emotion and memory working in the authors’ labs, who were blind to veracity, and who had no association with the present research \( (n = 20) \) provided their ratings of how emotionally intense the reports were on a scale from 1 \((\text{low})\) to 7 \((\text{high})\), which were then used to categorize narratives based on high \( (M = 6.38, SD = .07) \) versus lower \( (M = 4.01, SD = .13) \) groupings. In total, eight reports of sexual trauma (four true, four fabricated) were selected. These narratives were presented to the participants in the current study in random order.

**NEO-Five Factor Inventory (NEO-FFI)**
The NEO-Five Factor Inventory (NEO-FFI) (Costa & McCrae, 1992) is a 60-item self-report measure of the ‘Big Five’ personality traits. Each item (i.e., ‘I really enjoy talking to people’) is rated on a scale from 1 \((\text{strongly disagree})\) to 5 \((\text{strongly agree})\), yielding scores for each of the following five indices of personality: neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness. This measure has excellent psychometric properties (McCrae & Costa, 2004) and has demonstrated high reliability and validity across situations (e.g., McCrae et al., 2002).

**The Emotion Contagion Scale (ECS)**
The Emotion Contagion Scale (ECS) (Doherty, 1997) is a 15-item self-report questionnaire that measures the extent to which individuals can identify or empathize with the emotions of others. Participants rated the items (i.e., ‘it irritates me to be around angry people’) on a 5-point scale from 1 \((\text{never})\) to 5 \((\text{always})\). The concept of emotional contagion has been applied in a variety of social and cognitive studies (e.g., Gump & Kulik, 1997), can be used to evaluate empathic responses (e.g., Totterdell, 2000), and has good psychometric properties (e.g., Doherty, Orimoto, Singelis, Hatfield, & Hebb, 1995).

**Credibility Assessment Questionnaire (CAQ)**
The Credibility Assessment Questionnaire (CAQ) (Campbell & Porter, 2002) is a brief 20-item questionnaire used to record judgements and cue utilization with respect to each
trauma narrative. Participants were asked whether the trauma narrative was truthful or fabricated, and their judgement confidence (from 1 [not at all] to 7 [extremely]). In addition, the CAQ asked what cues (e.g., hesitation in writing, uneven flow of words/sentences, amount of details provided) were relied upon in the decision-making process (see Campbell & Porter, 2002; Peace & Sinclair, in press). These cues were drawn from the empirical literature on deception detection (see DePaulo et al., 2003; Porter & ten Brinke, 2010), and participants were asked to indicate whether they used the cue (yes/no) and its perceived usefulness rated on a scale from 1 (not at all helpful) to 7 (extremely helpful). The presence (or absence) of these cues then was rated by 20 independent raters, blind to the veracity of the claims and the purpose of the study, on a scale from 1 (not present) to 7 (highly present).

Procedure

Upon arrival at the experimental session, participants were asked to complete the NEO-FFI and the ECS. Following the personality assessments, each participant was given a booklet containing instructions regarding the decision-making task, and an additional warning about the sensitive nature of the material. This booklet contained eight accounts of sexual victimization, four truthful and four fabricated (counterbalanced). Participants were instructed to read each of the narratives and answer the questions that followed about their credibility judgement and what cues they used to make their decisions (i.e., complete the CAQ). Upon completion of the experiment, participants were debriefed about the nature and hypotheses of the study, and awarded course credit for their involvement.

Results

Credibility assessment and emotion

Overall, the mean level of accuracy at detecting deception for the sample was 45.3% (SD = 16.7%); significantly below the level of chance (Z = 2.66, p < .001). In order to assess differences between correct/accurate credibility judgements of truthful claims relative to deceptive reports, a repeated measures analysis of variance (ANOVA) was conducted with truthful and fabricated accuracy rates as the within-subjects factors. The analysis revealed a significant effect, \( F(1,115) = 8.08, p < .005 \), such that participants were better at correctly judging truthful (\( M = 49.6\% \), \( SD = 25.3\% \)) relative to deceptive (\( M = 41.0\% \), \( SD = 21.2\% \)) accounts.\(^1\) In addition, we examined the relation between accuracy and emotionality; there was a trend for accuracy to be higher for reports rated as lower (\( M = 47\% \); \( SD = 25\% \)) versus higher (\( M = 40.7\% \); \( SD = 25.2\% \)) in emotional intensity, \( t(116) = 1.88, p = .06 \). To determine whether participants displayed a truth bias, we conducted a signal detection (SDT) analysis on the hit/false alarm rates,\(^2\) and calculated estimates of discrimination accuracy (\( A' \)) and response bias (\( B'' \)) (see Meissner & Kassin,

\(^1\) The order of presentation of memory reports (counterbalancing) had no significant effect on total (\( F(7,115) = 1.11, p > .05 \)), truthful (\( F(7,115) = .90, p > .05 \)), or fabricated (\( F(7,118) = 1.44, p > .05 \)) accuracy scores and were not considered further in the analyses.

\(^2\) Hit rates are the proportion of truthful reports that were correctly identified as truthful, and false alarms the proportion of deceptive reports that were incorrectly identified as truthful.
Table 1. Means (and standard deviations) and correlation values for individual difference measures with overall accuracy, as well as truthful and fabricated accuracy

<table>
<thead>
<tr>
<th>Measure</th>
<th>Means (SD)</th>
<th>Overall accuracy</th>
<th>Truthful accuracy</th>
<th>Fabricated accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEO-FFI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuroticism</td>
<td>21.30 (8.25)</td>
<td>.15</td>
<td>.26**</td>
<td>-.07</td>
</tr>
<tr>
<td>Openness</td>
<td>29.72 (6.38)</td>
<td>.18*</td>
<td>.20*</td>
<td>.02</td>
</tr>
<tr>
<td>Extraversion</td>
<td>30.75 (6.44)</td>
<td>-.21*</td>
<td>-.22*</td>
<td>-.08</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>32.34 (6.44)</td>
<td>-.15</td>
<td>-.16</td>
<td>-.06</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>32.24 (6.38)</td>
<td>-.04</td>
<td>-.14</td>
<td>.10</td>
</tr>
<tr>
<td>ECS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score</td>
<td>3.38 (.53)</td>
<td>.01</td>
<td>.02</td>
<td>-.03</td>
</tr>
</tbody>
</table>

Note. *p < .05; **p < .01. 3Approached significance.

The overall level of discrimination accuracy was $A' = .32, SD = .35$ and response bias was $B'' = -.17, SD = .57$, indicating that participants performed poorly at correctly discriminating truth from fiction, and utilized liberal judgement criteria (i.e., participants engaged in the truth bias and were likely to ‘err on the side of caution’ and label reports as truthful more frequently). Bivariate correlations indicated that accuracy was unrelated to perceived ability at deception detection, $r(119) = -.09, p > .05$. However, confidence was negatively related to truth detection accuracy, $r(108) = -.20, p < .05$, and positively related to participants self-rated success on the deception task (i.e., percentage of correct judgements out of eight reports; $r(110) = .28, p < .01$). Confident participants believed that they correctly identified a greater number of judgements, but this was not reflected in their actual performance.

Individual deception detection ability

The relation between mean scores on scales of the NEO-FFI and emotional contagion (ECS) and accuracy scores (total accuracy, accuracy in detecting truthful and deceptive claims) were examined. Levels of conscientiousness and agreeableness were unrelated to all measures of accuracy; however, several domains of personality were related to accuracy (see Table 1 for means and correlations). Specifically, extraversion was negatively correlated with both total accuracy, $r(115) = -.21, p < .05$, and accuracy at detecting truthful claims, $r(115) = -.21, p < .05$, but unrelated to detecting false claims, $r(118) = -.09, p > .05$. Conversely, both neuroticism, $r(115) = .26, p < .01$, and openness, $r(115) = .18, p = .06$ (trend) were positively related to total accuracy scores. Further, higher levels of openness were associated with greater accuracy at correctly identifying truthful claims, $r(115) = .20, p < .05$. A regression analysis was performed in order to examine whether personality features predicted discrimination accuracy ($A'$) and response bias ($B''$). The regression model indicated that openness to experience was the strongest predictor of discrimination accuracy, $\beta = .26, F(1,116) = 8.22, p < .01$, and that neuroticism approached significance for this variable, $\beta = .18, t(115) = 1.95$.

---

3A' represents participants ability to correctly ascertain truth and correctly reject deception (range from −1.0 to +1.0), whereas B' reflects the degree of information required for the participant to decide that a report is truthful (where 0.0 reflects no bias, positive values reflect a conservative bias, and negative values reflect a liberal bias). A correction constant of 0.5 was added to all values to adjust for 0-values in accordance with the formulas used for $A'$ and $B''$ calculations (see Brown & White, 2005).
Table 2. Means (and standard deviations) for presence of cues (objective rating) and perceived usefulness of cues (subjective rating) in determining veracity of the sexually traumatic reports

<table>
<thead>
<tr>
<th>Cue</th>
<th>Real Mean (SD)</th>
<th>False Mean (SD)</th>
<th>Real Mean (SD)</th>
<th>False Mean (SD)</th>
<th>t-Score</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amount of detail</td>
<td>6.12 (.95)</td>
<td>5.74 (1.02)</td>
<td>5.08 (.93)</td>
<td>5.16 (.71)</td>
<td>.93</td>
<td>.354</td>
</tr>
<tr>
<td>Logical consistency</td>
<td>5.61 (1.01)</td>
<td>4.34 (.88)</td>
<td>5.04 (.88)</td>
<td>5.36 (.75)</td>
<td>3.73</td>
<td>.000</td>
</tr>
<tr>
<td>Sensory details</td>
<td>4.22 (.97)</td>
<td>4.48 (.99)</td>
<td>4.83 (1.03)</td>
<td>4.84 (.99)</td>
<td>.10</td>
<td>.920</td>
</tr>
<tr>
<td>Emotional self-detail</td>
<td>5.29 (.83)</td>
<td>5.02 (.90)</td>
<td>4.95 (.94)</td>
<td>4.89 (.98)</td>
<td>.62</td>
<td>.534</td>
</tr>
<tr>
<td>Relevant details</td>
<td>4.65 (1.93)</td>
<td>5.31 (1.09)</td>
<td>4.64 (.94)</td>
<td>4.78 (.94)</td>
<td>1.77</td>
<td>.081</td>
</tr>
<tr>
<td>Reported stress</td>
<td>6.17 (.70)</td>
<td>5.56 (.86)</td>
<td>4.92 (.99)</td>
<td>4.94 (1.03)</td>
<td>.24</td>
<td>.812</td>
</tr>
</tbody>
</table>

*p = .05. Response bias was significantly predicted by levels of extraversion, β = −.22, F(1,116) = 5.84, *p < .05, where higher levels of extraversion were associated with less stringent decision criteria (which negatively influenced accuracy; see above). Analyses of scores on levels of emotional contagion, r(108) = .01, *p > .05, and social desirability, r(116) = −.06, *p > .05, did not yield any significant results.

Cue usage

Overall, participants used a mean total of 84.7 cues (SD = 21.1) across all memory reports, and an average of 10.3 (SD = 2.4) cues per sexual assault evaluation. The mean number of cues used did not differ across truthful (M = 10.29, SD = 2.58) or fabricated (M = 10.31, SD = 2.47) sexual traumas, t(118) = −12, *p > .05. Participants indicated that they used the following cues most often in their decision making: amount of detail, logical consistency, sensory details, self-related emotional details, relevant details, and reported levels of stress (using an arbitrary cut off of 80% endorsement). Independent ratings (n = 20) of cues in the sexual trauma narratives indicated that each of the cues listed above were present in both types of narratives (see Table 2), and ratings of usefulness were correlated with the presence of actual cues in the reports (rs = .56–.84; all *ps < .01). From the perspective of the observers, the only apparent difference between true and false reports was the presence of logical consistency, t(114) = 3.73, *p < .001, which was rated to be less present in fabricated claims (see Table 2). Further, cue use did not vary across reports that were low (M = 10.5, SD = 2.7) and high (M = 10.6, SD = 2.7) in emotional intensity, t(79) = −.49, *p > .05. A series of one-way ANOVAs indicated no significant differences (ps > .05) in all cue use variables (i.e., total, truthful, fabricated) for judges who were more accurate (n = 27 > 50%) versus less accurate (n = 51 ≤ 50%). Pearson correlations revealed no associations with personality variables, and only one significant finding with respect to total cue use on each individual cue: consideration of uneven flow of words and sentences was associated with better accuracy when detecting honest accounts, r(114) = .22, *p < .05 (see Table 2 for means and standard deviations).

Discussion

Deception detection (e.g., Vrij, 2008; Vrij et al., 2010) and memories for sexual trauma (e.g., Alexander et al., 2005; Koss, Figueredo, Bell, Tharan, & Tromp, 1999; Peace et al., 2008) each have received much empirical attention. However, no studies have directly assessed observers’ capacity to detect true and false allegations of sexual victimization.
The present study was designed to examine detection accuracy, truth bias, individual differences, and cues used in decision making in relation to a sample of sexual trauma narratives.

**Credibility assessment and emotion**

Our results support previous research where detection accuracy is generally around the level of chance, and confirmed our prediction regarding a truth bias when judging claims of sexual victimization. In fact, observers performed worse (45.3%) than if they had randomly guessed, showed poor discrimination accuracy and a liberal response bias. That is, they were more likely to judge sexual assault narratives as truthful relative to deceptive. Previous studies that have utilized emotional stimuli also have found detection rates below chance levels (e.g., Campbell & Porter, 2002; Peace & Sinclair, in press). Further, participants were less accurate when judging false allegations (41%) relative to truthful accounts (49.6%). When confronted with claims of sexual assault (all of which appeared plausible), it is reasonable to assume that observers may have felt hesitant to label a report as ‘fake’ when the subject matter is of such an emotional nature. This was reflected in our finding that observers were less accurate when judging reports considered higher (40.7%) versus lower (47%) in emotional intensity. Although this finding approached significance, we think that judgements of emotional intensity may have been constrained by the emotional nature of sexual assault claims overall (i.e., none were rated as truly ‘low’ or a 1 or 2 on a 7-point Likert scale). It appears that allegations concerning traumatic sexual assaults are associated with impaired decision making, and result in a more pronounced truth bias than when non-emotional stimuli are used. Further research comparing judgement accuracy of emotional and non-emotional stimuli is necessary to elucidate the presence of an emotive truth bias.

Interestingly, the current study revealed a small negative correlation between confidence and accuracy at detecting truthful claims of victimization. While previous research has found confidence to be unrelated to accuracy in credibility assessment (e.g., DePaulo, Charlton, Cooper, Lindsay, & Muhlenbruck, 1997; Köhnken, 1987; Porter et al., 2002), Vrij (2008) reported that high confidence in detection abilities can be detrimental; those with high confidence may make decisions without considering information in detail, rely on heuristics and inaccurate cues (i.e., gut instinct, demeanor), and not be motivated to detect deception. This is supported by our finding that participants who felt more confident in their decision abilities (overall) also believed they would be more successful in their veracity determinations across the eight narratives presented, whereas their actual performance was negatively related to their confidence (see Sporer, Penrod, Read, & Cutler, 1995, for similar conclusions in eyewitness studies). As such, confidence in decision making is important to consider given that confidence is associated with perceptions of truthfulness (e.g., Whitley & Greenberg, 1986), despite the lack of relations found in previous studies deception detection studies.

**Individual deception detection abilities**

Individual difference factors such as personality and emotional contagion had not been examined in relation to deception detection involving sexual assault narratives. While previous research has not consistently demonstrated relationships between broad personality dimensions and deception detection abilities (see Bond & DePaulo, 2008), our results indicated several small but significant associations. There was a negative
correlation between extraversion and overall (as well as truth) accuracy, indicating that participants scoring low in extraversion (i.e., introverts) demonstrated enhanced capabilities in the detection task. This outcome is in line with previous observations that introversion is related to more detailed processing of stimuli, greater tolerance of repetition, lower distractibility on cognitive tasks, better verbal and non-verbal abilities, and use of more stringent decision-making criteria (e.g., Aron & Aron, 1997; Koelega, 1992; McCrae, 1987). This advantage appears to be especially pronounced in complex tasks (e.g., Brebner & Cooper, 1978), where introverted individuals may possess more varied cognitive processing abilities that positively influence deception detection accuracy (e.g., Campbell & Porter, 2002; O'Sullivan, 2005). Additional research on introversion (and perhaps critical thinking skills) is necessary to further clarify these associations. Another finding here was a positive association between accuracy and both openness to new experiences and neuroticism. People high in openness to new experience tend to enjoy effortful thinking, creating new solutions to complex problems, adapting their skills to suit new situations/problems, and engaging in divergent thinking (e.g., Colquitt, Hollenback, Ilgen, LePine, & Sheppard, 2002; Flynn, 2005; McCrae, 1996; Sadowski & Cogburn, 1997). Collectively, it appears that individuals high in openness have many important features necessary for veracity determinations and enhanced levels of accuracy. Neuroticism, on the other hand, was only related to accuracy for correctly identifying truthful claims, and unrelated to accuracy for detecting lies or overall accuracy. This trait is strongly associated with negative emotionality (e.g., Izard, Libero, Putnam, & Haynes, 1993), such that when individuals high in neuroticism are in a negative mood, they are better at recognizing negative emotions and classifying affective information (e.g., Carver, Sutton, & Scheier, 2000; Rusting, 1999; Tamir & Robinson, 2004). In our study, it is conceivable that reading reports of sexual victimization induced a negative mood state, and that participants high in neuroticism were able to more efficiently evaluate and recognize plausible negative emotions and content in the narratives, thus improving their accuracy for truthful claims that contained such details.

In general, research has found that fabricated traumas are less emotionally detailed relative to truthful traumas (e.g., Peace & Porter, in press; Porter, Peace, & Emmett, 2007), influencing the availability of this negative emotional content. As such, this helps to explain why neuroticism only was positively associated with truth accuracy (and not overall or lie accuracy).

**Cue usage**

Our analysis of overall cue utilization in the present study did not support our hypothesis; overall accuracy was unrelated to the number or type of cues used. It appeared that heavily utilized cues (e.g., amount of detail, logical consistency, relevant details) were present to a relatively high degree in both truthful and fabricated claims, limiting any associations between cue use and accuracy. Reliance on such cues may further explain the relatively poor hit rates exhibited in our study; participants were relying on cues that did not serve as good discriminators between truths and lies. That being said, participants did indicate that they relied on logical consistency (i.e., does the story make sense, do the details fit together) more in fabricated relative to truthful reports. In fact, fabricated narratives were associated with lower levels of logical consistency, making this a better cue to utilize by our sample. Related to this, reliance on the cue of uneven flow of words/sentences was associated with better accuracy in correctly discriminating truthful accounts. This finding fits with previous studies that have reported
fabricated events contain a rote or ‘story-like’ quality (e.g., Lees-Haley, 1984, Porter, Peace et al., 2007), whereas truthful events are associated with greater unstructured production (DePaulo et al., 2003; Vrij, 2008). As such, it is possible that participants were picking up on these subtle differences between reports, and recognized that truthful accounts from memory may be less structured in their presentation. That being said, the directionality of the cues were not revealed to participants (i.e., whether it was associated with truth or deception), and they may have interpreted the absence of this particular cue as being indicative of truth (e.g., Campbell & Porter, 2002). We also examined whether more accurate judges (overall rate over 50%) utilized different cues relative to inaccurate judges (rates less than 50%). Contrary to our prediction, no difference in cue utilization across judge accuracy was revealed. Previous research has found that accurate judges may rely on a greater number of and more varied cues relative to inaccurate judges (e.g., Campbell & Porter, 2002; Ekman & O’Sullivan, 1991), which also may relate to the constructs of both introversion and openness to experience. Although no personality-cue use associations were found in our study, future research evaluating accuracy and cue use as a function of personality and novelty of the deception detection stimuli may be useful in testing these relationships. Participants may have had problems with interpretation of the cues, as evidenced by the most commonly used cues being applied to both truthful and fabricated claims similarly. In addition, participants may have had difficulties understanding completion of the questionnaire (i.e., cues used/not used, how the cue was used, and the degree of usefulness). In addition, self-reports about cues used may be constrained by our inability to accurately determine what influences our judgements (see Wilson & Nisbett, 1978). As such, our ability to detect true differences in overall cue utilization may have been limited.

**Conclusions**

Some researchers (e.g., Bond & DePaulo, 2008) have concluded that differences in detecting deception result more from the characteristics and behaviours of the deceiver more than the observer. Some argue that individual differences are negligible (Aamodt & Custer, 2006; Bond & DePaulo, 2008), except that people differ in their biases regarding the veracity of statements, also known as a truth- or deception-biases (Masip et al., 2009). Our findings suggest that personality factors have at least a modest influence on deception detection ability and confirm that laypersons assume more often than not that sexual assault allegations are truthful. While in actual cases, this may be indeed be the case in terms of base rate, our results highlight that false allegations may be mistakenly identified as truthful in ‘he said/she said cases’ (Porter, Campbell, Birt, & Woodworth, 2003). Our next step in this research is to examine the ability of police investigators and judicial decision makers to discriminate true and false sexual assault allegations, and to examine whether such professionals exhibit similar or different biases for the credibility of such allegations. The present study offers insights into deception detection abilities when using real life and emotionally provocative stimuli. Participants performed worse than chance level overall, and demonstrated a truth bias in their veracity determinations. These findings have implications in forensic settings (e.g., consideration of emotional intensity of allegations and personality features of the investigator that may enhance or impede accuracy) when investigators are confronted with allegations of sexual assault, and are faced with the challenge of determining veracity.
References


Received 20 December 2010; revised version received 30 March 2011