Pinpointing Reality: How Well Can People Judge True and Mistaken Emotional Childhood Memories?

MARY ANN CAMPBELL and STEPHEN PORTER
Dalhousie University

Abstract
Although legal decision-makers often must judge the credibility of distant emotional memories, little research has examined this competency. The present study evaluated how well people are able to identify completely mistaken emotional memories from childhood. Further, possible individual differences, including personality/interpersonal traits and cue utilization strategies, contributing to this ability were examined. Participants (N = 157 undergraduates) viewed videos of true and mistaken childhood memory reports and judged whether the target had actually experienced the described event. Results indicated that 60% of judges accurately identified mistaken memories (greater than chance) and 53% accurately identified true memories (performing at chance). Compared to inaccurate judges, accurate judges were interpersonally less unassuming-ingenuous and more arrogant-calculating and aloof-introverted. Accurate judges also reported using more overt all cues to formulate their judgments than inaccurate judges. Brief exposure to information about empirically based cues to mistaken memories prior to veracity judgments impaired rather than facilitated judgment ability. Implications for credibility assessment in applied settings and future research are discussed.

Résumé
Même si les preneurs de décisions juridiques doivent souvent juger de la véracité de souvenirs affectifs lointains, il n’existe que peu de recherche à cet égard. La présente étude évalue la mesure dans laquelle les personnes peuvent déterminer des souvenirs affectifs d’enfance entièrement erronés. De plus, des différences individuelles possibles, notamment des traits de personnalité dans les relations interpersonnelles et des stratégies d’utilisation des indices contribuant à cette faculté, ont fait l’objet de l’examen. Les participants (N = 137 étudiants du premier cycle) ont visionné des vidéos relatant des souvenirs d’enfance vrais et erronés et ont dû déterminer si les sujets avaient réellement vécu les événements rapporités. Les résultats indiquent que 60 pour cent des juges ont identifié correctement des souvenirs erronés (possibilité plus grande que) et 53 pour cent ont identifié correctement des souvenirs vrais (selon toute probabilité). Comparativement aux juges qui n’ont pas vu juste, les juges qui ont vu juste sont dans leurs relations interpersonnelles moins sans présomptions et ingénieux, plus arrogants et calculateurs, distants et introvertis. Les juges qui ont vu juste, ont aussi signalé s’être servis davantage des indices d’ensemble pour formuler leur jugement que les juges qui n’ont pas vu juste. Une brève présentation de l’information relative aux indices fondés sur l’empirisme des souvenirs erronés avant les jugements de véracité a davantage compromis que facilité la capacité de juger. Les conséquences sur l’évaluation de la crédibilité dans des situations pratiques et la recherche future sont décrites.

It isn’t so astonishing, the number of things that I can remember, as the number of things I can remember that aren’t so. Mark Twain

The potential fallibility of autobiographical memory is well documented (e.g., Garry & Polaschek, 2000; Loftus, 1998; Payne, Neuschatz, Lampinen, & Lynn, 1997; Schacter, 1999). In fact, both anecdotal and empirical evidence indicate that some individuals can come to construct entire memories for events that never truly occurred (e.g., Kihlstrom, 1997; Loftus, 1998). Unlike intentional deception, during which an individual purposely attempts to deceive others, individuals who experience mistaken memories truly believe that the event or experience they have recounted has occurred. In experimental settings, such complete mistaken memories have been induced in about 20-40% of adult participants for such childhood events as getting lost, embarrassing situations, and negative emotional experiences (e.g., Heaps & Nash, 2001; Hyman, Husband, & Billings, 1995; Hyman & Pentland, 1996; Loftus & Pickrell, 1995; Porter, Birt.
Yuille, & Lehman, 2000; Porter, Yuille, & Lehman, 1999). The use of suggestive interviewing techniques has been widely implicated in the creation of mistaken memories in therapeutic (Lynn, Lock, Myers, & Payne, 1997; Poole, Lindsay, Memon, & Bull, 1995) and forensic (e.g., Quas, Schaaf, & Goodman, 2000) contexts. Recent data suggest that mistaken memories are a common aspect of human cognition. When Porter and Birt (2001) asked 306 participants whether they knew for certain that they had recalled a completely false personal experience at least once, 20% said they were aware of such a mistaken memory.

With the growing recognition of the pervasiveness of mistaken memories, the need to discriminate inaccurate from accurate memories in applied settings has been highlighted (e.g., Lampinen, Neuschatz, & Payne, 1998). Currently, many argue that it is difficult, if not impossible, to accurately make such discriminations in the absence of corroborating information (American Psychological Association, 1996; Blackshaw, Chandarana, Garneau, Merskey, & Moscarello, 1995; Bruck, Ceci, & Hembrooke, 1998; Gov, 1999). Although accurate and mistaken memories may be similar in some respects (Heaps & Nash, 2001; Porter et al., 1999), there are theoretical reasons to predict qualitative differences (Lampinen et al., 1998; Payne et al., 1997). For example, recent neuroimaging studies suggest possible differences in the brain mechanisms involved in recalling true and mistaken events (e.g., Fabiani, Stadler, & Wessels, 2000). Laboratory research has provided evidence of content differences in the memories themselves (for a review see Pezdek & Taylor, 2000). Distorted memory reports tend to contain more verbal hedges and cognitive operations, but fewer perceptual properties than accurate memory reports (Schooler, Gerhard, & Loftus, 1986). Complete mistaken memories tend to contain fewer details and a lower level of clarity than false memories (Hyman & Pentland, 1996; Loftus & Pickrell, 1995; Pezdek, Finger, & Hodge, 1997; Pezdek & Hodge, 1999). Using a detailed content analysis, Porter et al. (1999) compared mistaken childhood memories for negative events (e.g., animal attack, accidents) to both accurate and intentionally deceptive childhood memory reports. Mistaken memories were found to be less coherent, detailed, and vivid compared to true and deceptive reports. There was also a strong trend for mistaken memory reports to contain fewer repeated details than true event descriptions. Collectively, these findings offer compelling evidence that recently created mistaken memories may be distinguishable from true and deceptive accounts of distant events.

Despite recent research demonstrating such qualitative differences, very few studies have evaluated whether individuals are able to detect distorted memories without the benefit of detailed content analysis. Schooler, Clark, and Loftus (1988) had participants judge the credibility of valid and distorted memories of an auto accident and found that they performed at a level only slightly higher than chance in their detection judgments. Recently, Qin and Goodman (2000) investigated the ability of adults to detect accurate and mistaken childhood memories by having judges view videotaped accounts of memory reports. Some of these memory reports were based on true childhood events, while others were mistakenly believed by the memory reporter to have happened in their childhood following experimenter suggestion and repeated interviewing. The authors found that some participants were able to accurately identify memory veracity, in that 64% of the valid memory reports and 59% of the mistaken memory reports were correctly identified. However, Qin and Goodman concluded that their participants were limited in distinguishing valid and mistaken reports because of the high false-alarm rates. Qin and Goodman also found that participants’ perception of memory plausibility and clarity significantly predicted credibility judgments. Additional research is required to identify possible individual differences in the accurate detection of true and mistaken memory reports. Such information may guide the development of future professional training protocols for credibility assessment.

Research on the accuracy and process of credibility assessment has primarily focused on deception detection rather than honestly mistaken reports. Deception research may provide insight into the individual differences that contribute to accurate credibility judgments in the context of mistaken memory detection. It is clear that some individuals detect deceit better than others (Ekman & O’Sullivan, 1991; Ekman, O’Sullivan, & Frank, 1999; Porter, Woodworth, & Birt, 2000). The source of detection differences may rest, in part, with the cues used to decide credibility. Ekman and O’Sullivan (1991) found that more accurate lie detectors tend to use a flexible combination of verbal and nonverbal cues. In addition, the use of invalid deception cues clearly impairs detection (Porter, Woodworth, et al., 2000; Vrij, 2000). Overall, cue use seems to be an important component of deception detection and might also be relevant in detecting mistaken memories.

There is also reason to believe that the personality and interpersonal characteristics of the credibility assessor may influence his or her ability to accurately make judgments of credibility. One of the most supported conceptual approaches for describing personal-
ity is the Five-Factor Model (Digman, 1990; John, 1990; McCrae & John, 1992). According to this model, there are five basic personality dimensions: neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness. One objective of the present study was to determine whether these personality domains, as measured by the NEO-Five Factor Inventory (Costa & McCrae, 1992), impact on the ability to identify mistaken memory reports. For example, individuals with highly agreeable personality traits may be more trusting of accounts reported by others and, thereby, be less proficient credibility judges. In addition, given the interpersonal nature of credibility assessment, the present study evaluated whether a judge’s interpersonal nature influences his or her ability to detect mistaken memory reports. According to the deception detection literature, socially anxious individuals appear to be at a disadvantage in the detection of deception (DePaulo & Tang, 1994), while others have found that interpersonal traits of social submissiveness/dominance and extraversion among judges do not seem to influence detection ability (Keating & Heltman, 1994; Riggs & Friedman, 1983). In the current study, judges’ interpersonal traits were measured using the Interpersonal Adjectives Scale (Wiggins, 1995). This scale is based on a model that conceptualizes interpersonal behaviour as a continuous circular ordering (or circumplex) of interpersonal styles that represent blends of two basic interpersonal dimensions: Dominance-Submissiveness and Hostility-Friendliness (for a detailed description of the interpersonal circumplex model see Orford, 1994 and Wiggins, 1996). Although this model shares some elements with the Five-Factor Model of personality (Pincus & Wiggins, 1992), it is more specifically focused on interpersonal behaviour.

Porter et al. (1999) showed that intensive coding by blind coders (who were not attempting to judge the honesty of the reports) yielded qualitative differences between accurate and mistaken memories. However, no attempts have been made to improve the credibility judgments of true and mistaken memories with brief training. Perspectives on the benefits of training in the related area of deception detection have been mixed. Some researchers argue that existing training programs have not been very helpful (Bull, 1989; DePaulo, 1994) and may even impair the ability to detect deception (Kohnken, 1987; Vrij, 1994). Kassin and Fong (1999) found that participants who had been trained in a widely used traditional police technique performed significantly worse in identifying deception in videotaped mock interrogations than their untrained counterparts (who performed at chance). Others report improvement with the provision of cue information training and feedback (Fiedler & Walka, 1993; Porter, Woodworth, et al., 2000; Vrij, 2000). To our knowledge, only one study has examined the benefits of training in the context of memory distortion. Schooler et al. (Experiment 4, 1986) randomly assigned participants to cue hint and no hint conditions. Participants in the hint condition were provided specific verbal cues that Schooler et al.’s research had identified as suggestive of memory distortion following postevent misinformation (i.e., more words, cognitive operations, verbal hedges, and less sensory information). The no hint group received no such cue information. Participants were then asked to differentiate written descriptions of distorted and true memory reports about an auto accident. The cue hint group accurately detected a higher percentage of the distorted memory reports (60%) than participants in the no hint group (50%). Although the cue information did not improve the judges’ ability to identify the nondistorted memory accounts, it was above chance (63%). It appears that the simple provision of cue information may facilitate the detection of distorted memory accounts, but this issue requires further research.

The first objective of the present study was to investigate how well participants could judge the veracity of accurate and mistaken emotional childhood memories and to identify individual difference factors related to this ability (e.g., personality and interpersonal traits). A second goal was to investigate whether the simple provision of empirically based cue information would facilitate the identification of accurate and mistaken memories. To address these objectives, participants were randomly assigned to either a control or cue information condition. Both groups viewed a video of a target recounting a truthful and mistaken childhood memory (counterbalanced). Given the inconsistent effects of training in deception detection, it was not clear whether cue information would facilitate or impair the detection of mistaken and accurate memory reports. It was possible that simply educating judges about empirically based differences between true and mistaken memories would enhance detection. In contrast, it was also possible that cue provision would significantly impair judgment accuracy (e.g., by causing participants to overfocus on the cues and distracting attention from the decision-making task itself).

**METHOD**

**Participants**

Undergraduate participants (N = 137) participated for course credit or monetary reimbursement. The mean age was 20.0 years (SD = 3.1; range 17-34 years) and 79% were female. Participants were randomly...
assigned to either a cue information or a control group. The two conditions did not differ (p < .05) in age, gender, or number of years in university. Participants were tested in groups of one to six.

Materials

Measures of individual differences. The 60-item NEO-Five Factor Inventory (NEO-FFI; Costa & McCrae, 1992) was used to measure the dimensions of the Five-Factor model of personality as possible contributors to the individual differences in memory detection ability. The five factors measured by the NEO-FFI include Neuroticism, Openness to Experience, Agreeableness, Conscientiousness, and Extroversion. The NEO-FFI takes approximately 15 minutes to complete and has demonstrated adequate reliability and validity (e.g., Costa & McCrae, 1992; Holden & Fekken, 1994; Rolland, Parker, & Stumpf, 1998). Further, given that the process of judging the credibility of another person’s memory report could be construed as a form of social judgment, the interpersonal characteristics of the judges were measured using the Interpersonal Adjectives Scale (IAS; Wiggins, 1995). Sixty-four adjectives related to interpersonal interactions (e.g., “assertive”) are rated by respondents on the IAS according to how accurately each adjective describes them on a scale of 1 (extremely inaccurate) to 8 (extremely accurate). The IAS is based on a circumplex model of interpersonal behaviour. The circumplex model conceptualizes interpersonal behaviour as the ordering of multiple and continuous interpersonal dimensions on a geometric circle. Points on the circle represent variations or blends of two basic interpersonal dimensions: Dominance and Nurturance. Eight main interpersonal dimensions (or octants) on the circumplex model are captured by the IAS and include Assured-Dominant (i.e., forceful, assertive, and dominant in social interactions), Gregarious-Extraverted (i.e., cheerful, friendly, outgoing, and vivacious in social interactions), Warm-Agreeable (i.e., warm, nurturing, sympathetic, sociable, and caring in social interactions), Unassuming-Innocuous (i.e., deferent, obliging, modest, and straightforward in social interactions), Unassured-Submissive (i.e., timid, fearful, shy, and submissive in social interactions), Aloof-Introverted (i.e., avoid social interactions and reject the friendly overtures of others), Cold-hearted (i.e., emphasize autonomy and freedom from social conventions and are interpersonally cold, unsympathetic, and uncaring), and Arrogant-Calculating (i.e., egotistical, arrogant, cunning, and exploitative in social interactions). Higher scores on each of these octant scales reflect a higher level of the corresponding interpersonal trait. About 15 minutes are required to complete the IAS and it has good validity and reliability (e.g., Gurtman & Pincus, 2000; Kurtz, Lee, & Sherker, 1999; Wiggins, 1995).

Videos of true and mistaken memories. Six videos were selected from those created in Porter et al. (1999) as the judgment stimuli. These videos were randomly selected from a pool of ten video reports in which targets recounted a mistaken memory and also agreed to have their video used in future research. In these videos, trained assistants interviewed young adults about two emotional memories of events that may have occurred between the ages of 4 and 10. One of the memories was based on an event described by the target’s parent as having occurred to their child between the specified age period (true memory), while the other was experimentally induced across three interviews (mistaken memory i.e., the target had no initial memory of the event whose core details were concocted by the experimenter). It should be noted that the “true” memory reports were based only on the general details of the parental report and, thus, not all details were confirmed as valid or undistorted. Mistaken memories were created using guided imagery and repeated encouragement to “recover” the memory between interviews. The events described in the memory reports were one of the following: attacked by an animal, a serious medical procedure, an accident, or getting lost. For the present study, the free narrative phase of the interviews was shown to participants during which the targets recounted their event without interruption from the interviewer.

Memory Judgment Questionnaire. After viewing each videotaped account, participants completed a questionnaire inquiring about credibility judgments and cue utilization. First, participants indicated whether they felt each memory report was based on a true experience/event (Yes/No). They also rated the confidence of their judgment on a scale of 0 (not at all confident) to 6 (highly confident). Participants were then presented with a list of possible cues and asked to indicate whether they had used any of them to make their judgment (Yes/No). The perceived usefulness of each cue in judging credibility was rated on a scale of 0 (not at all useful) to 6 (extremely useful). Listed cues consisted of verbal, content, and nonverbal characteristics relating to the target and his/her account and were drawn from the literature on deception detection, memory distortion, and mistaken memories. These cues were conceptually divided into three major categories of cues. Empirically based cues were derived from the distorted/induced memory literature (Hyman & Pentland, 1996; Loftus & Pickrell, 1995;
### TABLE 1
General Content of Empirically Based Mistaken Memory Cue Information

<table>
<thead>
<tr>
<th>General Content Cues</th>
<th>Cue Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Hedges</td>
<td>The mistaken memory may contain more expressions suggesting uncertainty or doubt (e.g., “I think...”; “It might have been...”; “I believe...”).</td>
</tr>
<tr>
<td>Repeated Details</td>
<td>The memory is less likely to contain details that are repeated than a true memory.</td>
</tr>
<tr>
<td>Inferences/Rationalizations</td>
<td>The memory may contain more inferences and rationalizations as to why certain details are remembered (e.g., “That might have been because...”).</td>
</tr>
<tr>
<td>Coherency/Consistency</td>
<td>The content of the memory may appear to be less coherent, consistent, and logical.</td>
</tr>
<tr>
<td>Amount of Details</td>
<td>The memory content tends to contain fewer details than expected had the event actually occurred.</td>
</tr>
<tr>
<td>Vividness</td>
<td>The memory may appear less vivid or clear than would be expected if based on a true event.</td>
</tr>
<tr>
<td>Sensory/Perceptual Details</td>
<td>The memory report tends not to contain many references related to the sensory experience of the event such as description of colours, textures, or what the person saw and felt.</td>
</tr>
<tr>
<td>Cognitive Operations</td>
<td>The memory report tends to contain more references related to what the person or others were thinking or paying attention to (e.g., “I remember thinking that...”).</td>
</tr>
<tr>
<td>Confidence</td>
<td>The rememberer may not be as confident in the memory as compared to a true memory.</td>
</tr>
</tbody>
</table>

Pezdek, Finger, & Hodge, 1997; Pezdek & Hodge, 1999; Pezdek & Taylor, 2000; Porter et al., 1999; Schooler et al., 1986) and were primarily content oriented (i.e., amount of detail, verbal hedges, use of inferences/rationalizations, consistency, target confidence, perceptual/sensory details, relevancy of information, poor flow of words, and number of words). Nonverbal cues referred to the target’s body movements (i.e., movement of arms/hands/head, eye contact, length of gaze, self-manipulation, shifting position, and smiling/laughing). Finally, verbal cues referred to features of the target’s verbal comments (excluding the empirically based cues) and speech quality (i.e., re-experiencing the mental elements of the event, admitted forgetting, amount of reported stress during the experience, filled pauses, reported from an observer perspective, response length, and rate of speech).

**Procedure**

The study was a 2 x 2 mixed design, with cue group (cue information/no information) as the between-subjects factor and memory type (true/mistaken memory) as the within-subjects factor. Participants were told that the experimenter was interested in how people decide whether a memory is either true or based on a false event that the rememberer thinks happened. Participants were informed that they would view two videotaped interviews of the same individual describing two emotional childhood memories. Their task would be to decide whether each description was of a true or mistaken event. Prior to viewing the interviews, participants completed the IAS and NEO-FFI. In the second stage of the experiment, participants in the cue information group were provided with audio-taped information about empirically based verbal and content cues to mistaken memories (see Table 1). They were then asked to read a two-page written summary of the audio-taped information. The cues included in the training were derived from a review by Pezdek and Taylor (2000) and the empirical differences between valid and mistaken/distorted memories reported by Loftus and Pickrell (1995), Pezdek et al. (1997), Porter et al. (1999), and Schooler et al. (1986). An opportunity to ask questions and request elaboration on the various cue definitions was provided. The provision of cue information took about 10 minutes. The control group received no cue information and proceeded to the third stage of the experiment.

In the third stage of the experiment, each participant viewed a randomly selected videotape from one of the six described above that contained two interviews with the same target person. The target described a memory based on an actual event in one interview (True Memory Report), while he or she described a memory based on an event that he or she mistakenly believed to have happened in the other interview (Mistaken Memory Report). Although participants viewed one true and one mistaken memory report, they were instructed prior to the judgment task that one memory could be true, one could be mistaken, both could be mistaken, or both could be based on true experiences. After viewing each report, participants were administered the Memory Judgment
Questionnaire. The order of true and mistaken memory presentation was completely counterbalanced across the cue and control groups.

RESULTS

Accuracy and Report Type

The ability of the entire sample to detect true and mistaken memory reports was generally poor. Only 41% of participants correctly judged the veracity of both reports, while 29% correctly judged the credibility of only one report and 29% misjudged both reports. In terms of memory type, participants correctly identified the mistaken memories at a level significantly above chance, 60%; $\chi^2 (1, N = 136) = 4.97, p = .03$, whereas they only performed at chance in judging the true memory reports, 53%, $\chi^2 (1, N = 134) = 0.48, p = .49$. Judgment accuracy was unrelated to participant gender, $\chi^2 (2, N = 136) = 1.86, p = .39$, age, $F(2, 133) = 0.61, p = .52$, or self-reported experience with detecting deception, $\chi^2 (2, N = 136) = 4.79, p = .09$. Judges reported a moderate to high mean level of confidence in their credibility judgments, $M = 3.91, SD = 1.02$, and there was no significant relation between accuracy and judgment confidence. Participants who incorrectly identified the true memory as mistaken were as confident in their decisions, $M = 3.82, SD = 1.31$, as those who correctly identified it, $M = 4.04, SD = 1.16$, $F(1, 132) = 1.03, p = .31$. Likewise, judges who misidentified the mistaken memory were as confident in their decisions, $M = 3.79, SD = 1.23$, as those who accurately identified it, $M = 4.02, SD = 1.13$, $F(1, 134) = 1.20, p = .27$.

Effect of Cue Information

Collapsed across memory types, the frequency of judges who accurately detected memory veracity was significantly lower in the cue information group (49%) than in the control group (69%), $\chi^2 (1, N = 96) = 3.39, p$
When the memory report under evaluation was based on a true event, the judges' detection accuracy (correct vs. incorrect) did not significantly depend on whether he or she was a member of the cue information group (47%) or the control group (59%), $\chi^2 (1, N = 134) = 1.95, p = .16$. However, there was a nonsignificant trend for judges in the control group to demonstrate a higher level of accuracy (67%) than judges in the cue information group (52%) when the mistaken memory report was being evaluated, $\chi^2 (1, N = 136) = 3.17, p = .07$. The 67% accuracy rate for the control group in detecting mistaken memories was significantly above chance, $\chi^2 (1, N = 67) = 7.90, p = .006$.

Cue Utilization
Judges reported using a similar mean number of total cues when judging true, $M = 15.23, SD = 4.33$, and mistaken memories, $M = 15.36, SD = 4.40$, $t (135) = -0.42, p = .68$. Given this similarity, the data were collapsed across memory type to yield an overall mean number of total cues used. Participants who correctly detected both the true and mistaken memories reported using a significantly higher mean number of cues, $M = 16.19, SD = 3.60$, than those who failed to detect either memory correctly, $M = 14.20, SD = 4.23$, $F(1, 94) = 6.14, p = .01$.

In order to determine whether cue use significantly differed according to the level of judge accuracy (correctly judged both memory reports/incorrectly judged both reports) and judge gender, a MANOVA was conducted with the number of empirically based, nonverbal and verbal cues used as dependent measures. A significant overall main effect was found for the level of accuracy ($ Pillai's Trace = .12; F(3, 90) = 4.13, p = .009$), but not for judge gender ($ Pillai's Trace = .01$), $F(3, 90) = 0.38, p = .77$. No significant interaction was found between the two independent variables ($ Pillai's Trace = .03$), $F(3, 90) = 1.10, p = .35$. Accurate and inaccurate judges reported using a similar mean number of empirically based cues ($Ms = 7.34$ and $7.47$, $SDs = 1.23$ and $1.15$), respectively (except for Warm-Agreeable, where $Ms = 7.74$ and $7.42$, $SDs = 1.30$ and $1.16$), $F(1, 94) = 4.42, p = .038$.
and 1.45) and verbal cues (Ms = 4.95 and 4.34, SDs = 1.41 and 1.52), F(1, 88) = 1.52, p = .22 and F(1, 88) = 4.97, p = .13, respectively. However, accurate judges reported using significantly more nonverbal cues (M = 4.69, SD = 1.77) than inaccurate judges (M = 3.26, SD = 2.36), F(1, 88) = 9.05, p = .003. Figure 1 displays the mean usefulness ratings for individual nonverbal, verbal, and empirically based cues utilized by accurate and mistaken judges to decide memory report credibility.

Accuracy, Personality, and Interpersonal Traits

Two MANOVAs were conducted to examine the impact of judge personality and interpersonal traits on the ability to assess memory credibility. For the first MANOVA, NEO-FFI factor scores did not significantly differ between accurate and inaccurate judges (Pillai’s Trace = .06), F(5, 90) = 1.87, p = .32. The second MANOVA revealed a main effect for accuracy on the multivariate combination of IAS subscale t-scores (Pillai’s Trace = .16), F(8, 87) = 2.03, p = .05. Univariate ANOVAs indicated that accurate and inaccurate judges significantly differed on three IAS scales: Arrogant-Calculating, F(1, 94) = 6.32, p = .01, Alloof-Introverted, F(1, 94) = 6.47, p = .01, and Unassuming-Ingenuous, F(1, 94) = 5.70, p = .02. As shown in Figure 4, accurate judges rated themselves as significantly higher on the Arrogant-Calculating scale (M = 39.5, SD = 11.3) than inaccurate judges (M = 33.9, SD = 10.1). Accurate judges also rated themselves as significantly higher on the Alloof-Introverted scale (M = 41.7, SD = 9.3) than erroneous judges (M = 37.07, SD = 8.1). Finally, compared to inaccurate judges (M = 63.8, SD = 12.2), accurate judges scored significantly lower on the Unassuming-Ingenuous scale (M = 57.7, SD = 12.5). Although participants scoring higher on the Alloof-Introverted scale were more accurate, overall judgment confidence was negatively correlated with this scale, r(135) = -0.23, p = 0.007. Scores on the Arrogant-Calculating, and Unassuming-Ingenuous scales were not associated with judgment confidence, rs (135) = -.10 and .09, respectively, ps > .05.

Discussion

In many criminal cases, judges and juries must evaluate the validity of reports for distant emotional experiences, often in the absence of corroborating evidence. While some legal commentators have expressed the view that jurors are able to do this based on common sense (see Porter, Birt, Yuille, & Hervé, 2001), others claim that it is very difficult or even impossible to distinguish true from mistaken/distorted memory reports (e.g., Blackshaw et al., 1996; Gow, 1999). The present study was one of the first to investigate whether people can actually identify mistaken memories for emotional childhood events in adults. Results indicated that a small majority of participants were able to spot mistaken memories. About two thirds of the control group accurately detected the mistaken memory and slightly fewer correctly identified the true memory. This level of accuracy is higher than that typically reported for deception detection, which tends to be around the level of chance (see Vrij, 2000). However, given the high false positive rates reported here, there is certainly room for improvement in judgment accuracy.

Studies of content/phenomenological features (e.g., Heaps & Nash, 2001; Porter et al., 1999) and neuroimaging findings (Fabiani et al., 2000) have shown that there are important differences between true and mistaken memories. The accuracy of identifying mistaken memories found in our study suggests that some individuals are sensitive to these differences in the absence of training. Although lengthy, in-depth training in memory content analysis may improve the detection of mistaken memories (see Porter et al., 1999), such credibility assessment methods may not be practical for the courtroom or during police interviews. Our results also indicate that brief exposure to cue information may impair rather than facilitate credibility decision-making. It is unlikely that this negative effect was due to the absence of the training cues in the true and mistaken memory reports because many of the training cues were present in a larger sample of the same videos used in the judgment task. It is possible that the brief exposure to cue information led judges to overfocus on the training cues and miss other useful pieces of information. The interfering effects of cue information are similar to those found in Kassin and Fong (1999) but differ from the findings of Schoolder et al. (Experiment 4; 1986), who found a benefit of cue information on the detection of distorted memories. This contrast in findings may be the result of methodological differences, such as the use of complete mistaken childhood memories, the use of text-based versus audio-visual presentation of memory reports, and invalid training material.

Research on the effect of cognitive load on complex judgment tasks may provide some insight into the negative influence of brief cue training on detection accuracy. In general, cognitive load theory (Sweller, 1993, 1994; Sweller, van Merrienboer, & Paas, 1998) suggests that when a task requires a significant amount of cognitive resources to process information (i.e., has a high cognitive load), it can negatively affect learning by “overloading” the limited capacity of working memory. In support of this theory in the context of social judgments, researchers have found that
judgment accuracy can be negatively influenced by a judgment task that has a high cognitive load (Gilbert, Krull, & Pelham, Experiment 2, 1988; Patterson & Stockbridge, 1998). When participants experience a high cognitive load during judgment tasks, they are more likely to use decision-making strategies that simplify the task (e.g., rely on stereotypes or heuristics; Gordon & Anderson, 1995; van Knippenberg, Dijkstra, & Dijksterhuis, & Vermeulen, 1999), but which also limit the amount of information they process. In the present study, judges in the cue information condition were likely to have experienced a higher cognitive load than controls in the judgment task because of the cue information they were asked to remember, attend to, and apply during their decision-making. Given the brief exposure to cue information, the lack of experience using this new information, and the resulting high cognitive load, the training cues may have become distracters rather than facilitators of credibility decision-making. To compensate for the higher cognitive load, judges may have attempted to simplify the judgment task by narrowing their attention and processing information relevant only to the content cues on which they were trained. As a result, these judges were not able to attend to or process other potentially useful pieces of information, unlike the control judges who experienced a lower cognitive load. In sum, it may be useful to consider the impact of cognitive load when developing instructional programs for credibility assessment training. For example, more prolonged training may be necessary to significantly increase a judge’s familiarity with the empirically based cues to mistaken memories and to reduce the effects of cognitive load during judgment tasks. Further, novice trainees may benefit from the use of practice work examples during training, as the use of such learning techniques have been shown to minimize cognitive load (Tuovinen & Sweller, 1999).

In support of the negative influence of narrowed cue focus, accurate judges demonstrated a more flexible pattern of cue utilization than did inaccurate judges. The use of empirically based cues to mistaken memories was similar for both inaccurate and accurate judges. Both tended to use the empirically based cues most frequently and weighted them as the most useful relative to nonverbal and verbal cues. This finding suggests that many individuals appear to be cognizant of the memory content features (e.g., amount of details, consistency, verbal hedges) that would suggest a mistaken memory without being explicitly provided this information. In addition to the use of empirically based cues, accurate judges also reported using more nonverbal cues. This suggests that a more flexible and variable pattern of cue utilization occurred among accurate judges. A similar benefit of combining verbal and nonverbal cues has been found in the detection of deception (Ekman & O’Sullivan, 1991; Ekman et al., 1999; Porter, Woodworth, et al., 2000; Vrij, 2000). There is a caveat regarding the measurement of cue use in the present study. The number of cues endorsed by judges from the cue list was substantially higher than that typically reported among credibility studies using open-ended inquiries of cue use. On the other hand, a list of cues may have helped judges become more cognizant of the cues they considered in making their credibility decisions. On the other hand, a cue list may have led judges to inadvertently inflate the number of cues they had actually used. Future research is required to determine the most efficient and informative means of measuring cue utilization in credibility judgment tasks.

Despite the finding that accurate memory judges used more nonverbal cues, the nature of valid nonverbal cues to mistaken memories has yet to be examined empirically. In the present study, judges rated nonverbal cues as minimally useful. This suggests that judges may use them only as secondary pieces of information relative to information obtained from their evaluation of the report’s content. We are currently in the process of coding and analyzing the memory reports created by Porter et al. (1999) to determine whether nonverbal differences actually exist between mistaken and true memory reports. Preliminary examination of this data suggests that the nonverbal characteristics of targets describing true and mistaken memories are very similar. It is possible that judges reported using nonverbal cues because they were looking for perceived nonverbal cues of deceit (e.g., eye contact; Vrij, 2000), despite being told that individuals with mistaken memories are not being deceptive. Research should be directed towards the determination of whether the use of actual and perceived deception cues hinders or facilitates memory credibility assessments in the case of mistaken memories. It is important to note that the detection of mistaken memory reports may be a qualitatively different task from the credibility evaluation of potentially deceptive memory reports. Although both of these types of “memories” are not based on actual events, the motivation and methods of memory creation are very different. In the case of deceptive reports, individuals knowingly create memory details in an effort to deceive the listener. In the case of mistaken memories, individuals are not being deceptive and truly believe that the events they have related to the listener are indeed based in fact. Porter et al. (1999) have also noted differences in the content and quality of mistaken and deceptive memory reports. Thus, credibility assessment techniques may be better served
by developing methods of evaluation that are specific to the nature of the memory.

Some individual difference factors were found to be associated with judgment accuracy. Although no significant associations were evident between accuracy and the NEO-FFI personality domains, accurate and mistaken judges differed on three IAS octant scales. Accurate judges were interpersonally more Arrogant-Calculating (e.g., egotistical, cunning, exploitative) and Aloof-Introverted (e.g., unsociable), but less Unassuming-Ingenuous (e.g., obliging, nonargumentative, deferential) than judges with poor detection accuracy. In contrast, less proficient judges were interpersonally more trusting, agreeable, and sociable. Although judges with higher levels of Arrogant-Calculating, Unassuming-Ingenuous, or Aloof-Introverted traits were more accurate, only the more introverted judges were less confident in their judgments. This tendency has also been reported in the eyewitness memory literature (Trouvé & Libkuman, 1992; Ward & Loftus, 1985). These findings suggest that confidence in one’s judgment is not necessarily a reliable marker on which to base judgment accuracy. The mechanism by which interpersonal traits influence detection ability is unclear, but it is possible that such traits affect the information-processing and decision-making strategies used by judges. For example, interpersonally trusting and agreeable individuals may use a less critical strategy to evaluate credibility because they are more willing to trust the target’s account as true or are reluctant to challenge the account. Future research is required to confirm the role of interpersonal characteristics in credibility decision-making.

It is unclear as to why interpersonal differences in judgment accuracy occurred in the absence of significant personality differences (NEO-FFI) given the ties of both the NEO-FFI and IAS to the Five-Factor Model of personality. This discrepancy may be attributable to technical and theoretical differences between the two measures. Specifically, the NEO-FFI was designed to provide only a global estimate of personality traits in accordance with five major personality factors (i.e., Five-Factor Model; Costa & McCrae, 1992). In contrast, the IAS was designed to tap the somewhat related, but separate, construct of interpersonal behaviour and was based on a detailed model of interpersonal styles that represents blends of two central interpersonal dimensions: Dominance and Nurturance (i.e., Circumplex Model; Wiggins, 1995). Hence, the NEO-FFI may not have been as sensitive to the interpersonal characteristics measured by the IAS. Given that the model underlying the IAS shares some elements with the Five-Factor Model, the use of the more comprehensive NEO-Personality Inventory (NEO-PI-R; Costa & McCrae, 1992) may demonstrate similar findings to that of the IAS. It has also been argued that personality influences an individual’s interpersonal behaviour (Pincus & Wiggins, 1992). In keeping with this perspective, the above findings suggest that while an individual’s normal personality functioning may not directly influence his or her ability to accurately judge credibility, personality traits may have indirectly influenced detection accuracy by impacting on the judge’s interpersonal behaviour. The influence of interpersonal traits is not surprising given the social and interpersonal nature of the credibility judgment context.

Only a few participants (15%) reported some degree of prior experience with rendering credibility judgments. Given the nature of the present sample (university students), the range of professional experiences with credibility assessment was fairly narrow. A common example of prior experience related to incidents in which a store cashier assessed the credibility of a customer’s account (e.g., customer claims he or she was short-changed or attempted to return items without a receipt). Although more judges with prior detection experience accurately determined target credibility (81%) compared to those without such experience (54%), this difference was not significant and suggests that prior informal experience with credibility judgments does not substantially influence detection ability. This null effect may be due to the very small number of participants with prior experience, but it is consistent with the deception detection literature in that experience with professional lie detection does not typically enhance the ability to detect deceit (e.g., DePaulo & Fiefer, 1986; Porter, Woodworth, et al., 2000; Vrij, 2000). Future research is required to determine whether individuals who have more formal experience with credibility assessment (e.g., police officers) demonstrate an advantage in identifying mistaken memory accounts. Likewise, the influence of other potential individual differences factors related to the judge’s thinking styles (abstract vs. concrete), problem-solving ability, and critical thinking skills should be examined to more directly identify the information-processing strategies used by accurate credibility assessors. Such knowledge can guide the development of credibility assessment training programs.

The present study represents an original attempt at the accurate identification of true versus mistaken memory reports and had several notable strengths. Perhaps most importantly, we used some of the first videotaped reports for complete mistaken emotional memories in adults. Simple nonemotional memory stimuli are not likely to share the complexity and rich-
ness of emotional memory reports, which are more commonplace in forensic contexts. Although the mistaken memories used in the present study were created under controlled settings, the interview context was similar to interrogation and therapeutic interviews. Nonetheless, the use of emotional-laden memory reports in controlled laboratory environments has its limitations. In applied settings, credibility evaluators have the opportunity to spend time with and learn about the target before making informed judgments. In the future, researchers could use actual interviews with true and invalidated claims of memories for historical events to add to the ecological validity of the data. Research using emotional memory reports will more closely represent the type and nature of information that professional credibility assessors will be exposed to in applied settings, but field research will be an essential extension. Another advantage of the present study was the use of a large sample that permitted the investigation of individual differences in the sensitivity to mistaken memories, but it is also important to determine the level of sensitivity among professionals responsible for credibility assessment in legal and therapeutic settings (e.g., judges, police officers, therapists).

In summary, we found that the identification of mistaken memories is difficult but that certain individuals are better at this task than others. Brief exposure to discriminatory cues impairs judgment accuracy. Future studies should extend our findings by examining the ability of “judges” in applied settings and by varying the nature of the events being falsely recalled. For example, would bizarre, implausible, or long-held memories for false events be more easily identified as mistaken? The interaction of judge and target characteristics also requires investigation given the interpersonal nature of credibility judgments. Future work in this area will add to our understanding of the fallibility of memories and how to spot the mistaken ones.

The first author is supported by a Social Sciences and Humanities Research Council of Canada (SSHRC) Doctoral Fellowship, an Isaac Walton Killam Memorial Scholarship, the American Psychology and Law Society (AP-LS), and the American Academy of Forensic Psychology. The second author is supported by a grant from SSHRC. We extend our thanks to the volunteers who helped conduct this study: A.D. Ingram, Amilyn Kearney, Amanda Risser, Cindy Knowles, Lindsey Hawker, Jette Mitgaard, Naomi Doucette, Alicia Murphy, Alicia MacDonald, Rob MacArthur, Yarrow Joyce, and Talia Profit. Thanks to Angela R. Birt for her comments on an earlier draft of this paper. We also thank the reviewers for their comments on an earlier version of this manuscript.

Correspondence should be addressed to the second author at the Psychology Department, Dalhousie University, Halifax, Nova Scotia, Canada, B3H 4J1. Tel.: (902) 494-6934; fax: (902) 494-6585; E-mail: sbporter@is.dal.ca.

References


Received March 21, 2001
Revised August 29, 2001
Accepted March 22, 2002