Remembrance of Lies Past: A Comparison of the Features and Consistency of Truthful and Fabricated Trauma Narratives

KRISTINE A. PEACE1* and STEPHEN PORTER2

1Grant MacEwan University, Edmonton, Alberta, Canada
2University of British Columbia—Okanagan, Kelowna, British Columbia, Canada

SUMMARY

The credibility of reports of victimization must be evaluated by police and adjudicators. The present prospective study investigated the features of truthful and fabricated narratives of trauma and their relative consistency over a 6-month period. Participants described both a genuine and fabricated traumatic experience on three occasions over the 6 months. The narrative features were coded at each phase and the stories’ consistency was examined at Times 2 and 3. We found that truthful trauma narratives contained more: details, contextual information, and emotional details, and were rated as more ‘plausible’, relative to fabricated traumas. The details of truthful narratives also were more factually consistent than fabricated narratives over the 6 month interval. This research offers basic information on the nature and consistency of traumatic memories, and potentially applicable information for discriminating true and false allegations. Copyright © 2010 John Wiley & Sons, Ltd.

Fabricated allegations of victimization occur on a regular basis in legal settings, with estimates ranging from 2 to 40% of all allegations (e.g. Anthony & Watkeys, 1991; Kanin, 1994; Sheridan & Blaauw, 2004; Yuille, Tymofievich, & Marxsen, 1995). In the first national study to document the rate of false allegations of abuse investigated in Canada, Trocmé and Bala (2005) examined data from the Canadian Incidence Study of Reported Child Abuse and Neglect. They found that more than a third of maltreatment cases were ‘unsubstantiated’ (i.e. no corroboration was found), and 4% of the cases had been intentionally fabricated. Credibility assessment in such contexts is a process fraught with error and there is no ‘Pinocchio’s nose’ for identifying liars (e.g. DePaulo et al., 2003; Sporer, 2008; Vrij, 2008). While certain professional groups (e.g. see Ekman, O’Sullivan, & Frank, 1999; Mann, Vrij, & Bull, 2004) and individuals, or ‘wizards’ (O’Sullivan & Ekman, 2004), appear to be better than chance at detecting lies (but see Bond & Uysal, 2007), both laypersons and legal professionals generally perform around chance, which can be disastrous in forensic settings (see Porter & ten Brinke, 2009). Yet, the Supreme Court of Canada concluded that common sense is sufficient for jurors to identify a deceptive witness (R. v. Marquard, 1993), as long as facial cues and demeanour are observable (R. v. B. (K.G.), 1993). Further, researchers have found that legal decision-
makers often report beliefs concerning ‘valid cues’ to deception that are completely inaccurate (e.g. Porter & ten Brinke, 2010).

There is little empirical knowledge concerning the features of false allegations of victimization or their consistency over time (e.g. Vrij, 2008). Despite an enormous body of research on behavioural cues to deception (see Porter & ten Brinke, 2010, for an overview), few reliably discriminate truth and lies (e.g. DePaulo et al., 2003). In forensic contexts, the features of deceptive narratives are particularly relevant given that verbal statements typically are provided as evidence. The utility of attending to language (relative to non-verbal and physiological cues) to catch liars has been supported by both empirical and anecdotal evidence (e.g. Sporer, 2008; Vrij, 2008). Criteria-based content analysis (CBCA) is the most widely studied and best-validated statement analysis approach, based on the idea that accounts derived from memory for an experienced event differ qualitatively and quantitatively from fictitious reports (e.g. Köhnken, 2002). Research on CBCA has indicated that in the majority of field and experimental studies reviewed by Vrij (2008), the criteria are present to a greater extent in genuine versus fabricated statements. In particular, studies consistently have found that the quantity of details, unstructured production, contextual embedding and reproduction of conversation are useful in differentiating truths and lies (Vrij, 2008). However, few studies on CBCA have examined its utility with real and false traumas. One notable exception was a study comparing CBCA characteristics of genuine vs. deceptive allegations of rape (Parker & Brown, 2000). Their results indicated that corroborated claims of rape contained more unstructured production, details, descriptions of interactions, reproductions of conversation and descriptions of both victim and perpetrator mental states. In another study, Lamb et al. (1997) evaluated interviews with alleged victims of child sexual abuse, and found that cases rated as being plausible/likely (i.e. substantive corroboration existed) contained more CBCA criteria than those that were unlikely to have occurred (and lacked corroboration). That being said, only some criteria were predictive of veracity (e.g. unstructured production, quantity of details, contextual embedding), whereas others were rated similarly across plausible/impossible accounts.

Another approach to evaluating the credibility of memory narratives derives from the reality monitoring (RM) framework (Johnson & Raye, 1981). Proponents of this perspective argue that the origins of a ‘memory’ are reflected in its narrative characteristics; with internally generated (imagined) events differing from externally generated (perceived) events (e.g. Johnson, Foley, Suengas, & Raye, 1988). In particular, experienced events contain more perceptual, contextual, affective and semantic detail, whereas imagined events require more cognitive processes (i.e. records of organization and thought, reflective construction). Early RM studies involved subjective assessment of one’s own memory experiences (meta-memory level, see Sporer, 2004). For example, Johnson et al. (1988) had participants recall perceived events (e.g. trip to the dentist, social occasion, library visit), as well as a fantasy or dream they had experienced. Results indicated that perceived events differed from imagined events on a variety of measures: higher levels of sensory details, realism, spatial arrangement of environment, and time and place details. In addition, real events included surrounding details of memories preceding and following the event being described (e.g. contextual embedding). Conversely, imagined events were rated as being more intense, consequential (i.e. had more implications at the time), complex, thought about more often, and less realistic. More recently, Kealy, Kuiper, and Klein (2006) also found that real events were associated with greater visual, emotional and contextual
(time and place) details relative to made-up events. Yet, such studies have not focused on emotionally provocative, personally salient or traumatic events.

Further research has combined CBCA and RM approaches in an attempt to assess self- and other-judgments of memory. For example, Sporer (1997; 2004) reported that ‘reality criteria’ can be applied to meta-memory judgments (self-rated) and meta-meta memory judgments (other-rated). The present study utilized a similar approach to assessing the qualities of traumatic memories. The Memory Assessment Procedure (MAP; Porter, Yuille, & Lehman, 1999; Porter & Peace, 2007) is based on a combination of CBCA and RM criteria to provide a global assessment of both qualitative characteristics (assessed by trained coders) and phenomenological characteristics (as rated by the rememberer). Using this measure, Porter et al. (1999) reported that genuine and fabricated experiences differed on both other- and self-rated variables. Specifically, truthful reports tended to contain higher scores on other-rated criteria, whereas self-rated properties of fabricated experiences were exaggerated and considered ‘over the top’ (i.e. participants tried overly hard to make their fake traumas appear genuine and assigned them higher stress, vividness and clarity ratings relative to truthful reports). Overall, these findings suggest that features of reports for truthful and deceptive events are different and vary across self-rated vs. other-rated characteristics. In another study using the MAP, Porter, Peace, and Emmett (2007) demonstrated that fabricated reports were associated with higher levels of self-reported emotionality (e.g. emotional intensity, feeling detached), fewer time and place details, shorter reports and a greater endorsement of post-traumatic stress disorder (PTSD) symptoms. As such, the MAP was reliably used to reveal characteristic differences in both the generation and judgement of truthful and deceptive statements (see Sporer, 2008).

There also has been little research on how true and false reports ‘evolve’ over time (Granhag & Stro¨mwall, 2001, 2002; Granhag, Stro¨mwall, & Landstrom, 2006). Witnesses to a crime often provide multiple statements about the event and are re-interviewed on several occasions prior to giving court testimony. Testimonial inconsistencies (i.e. differences between an initial statement and testimony in trial) appear to be one of the prime perceived indicators of memory veracity (Brewer, Potter, Fisher, Bond, & Luszcz, 1999; Fisher & Cutler, 1995). Research has found that judges, lawyers and police believe that truthful statements are more consistent and more detailed over time than deceptive reports, but are generally unable to identify what constitutes adequate consistency (e.g. Strömwall & Granhag, 2003). In addition, consistency is used as a cue to justify decisions that a person is being either deceptive or truthful (Granhag & Strömwall, 2001).

The current study focused on the consistency of true and false trauma narratives over time. While past research suggests that inconsistencies undermine perceived credibility, there is little consensus on how truthful and deceptive claims of trauma will vary over time (e.g. Erdmann, Volbert, & Bohm, 2004; Granhag et al., 2006; Masip, Sporer, Garrido, & Herrero, 2005). There are three main possibilities regarding the consistency of deceptive statements. One view is that fabricated or deceptive claims should be less consistent over time relative to genuine experiences, especially in terms of the salient or central aspects of events (e.g. persons, actions) (e.g. Greuel, 1992). Wagenaar, van Koppen, and Crombag (1993) argued that truth can be inferred from level of consistency in consecutive statements. Deceptive statements are less likely to contain relevant and detailed information reported in a consistent manner. In a series of studies on memory for lying, Polage (2004) found that deceptive items were less believable over time due to inconsistencies—which she calls ‘fabrication deflation’. Studies on imagination inflation (i.e. repeated imagination and rehearsal; Mazzoni & Memon, 2003) support the view that
inflated details could ‘deflate’ consistency. Some studies suggest that the more time one has to fine-tune a lie, the more ‘fantastic’ it may become in successive retellings (Johnson, Hashtroudi, & Lindsay, 1993), which in turn leads to inconsistencies across reports. Alternatively, perhaps truthful and deceptive statements are equally consistent over time and change at similar rates (e.g. Granhag & Strömwall, 2000; Granhag, Strömwall, & Jonsson, 2003). This rests on the idea that in order for fabricated claims to be successful, the liar must have a good memory and engage in more rehearsal of deceptive details. Because of the cognitive load of such a task (e.g. Vrij, 2004), he/she may evidence more fluctuation or a greater drop in consistency between initial recalls but this will level off as the story gets solidified over successive retellings. With genuine experiences, the malleability of memory also is associated with variations in details reported over time, corresponding to a decrease in consistency (e.g. Loftus, 2003). For example, Granhag and Strömwall (2002) evaluated verbal and nonverbal cues to deception and found that levels of consistency for truthful and fabricated claims were relatively equivalent for information reported across interviews (over 11 days), indicating that both decrease over time. The final view is that deceptive statements may be retained more consistently than truthful reports. One possible mechanism is that liars are focused on remembering what they have said in previous statements, whereas truth-tellers base their reports on their memory for the event and may show more variation (Granhag & Strömwall, 2000). Deceptive claims of trauma could be based on a schema or stereotype of traumatic victimization and could appear more stable over time due to the repeated activation of a ‘trauma-script’ (e.g. Anderson, Cohen, & Taylor, 2000).

THE PRESENT STUDY

Little research has examined the characteristics and consistency of real and fabricated traumatic memory reports over an extended period of time. We predicted that the narratives of truthful and fabricated traumas would be qualitatively different according to the MAP and the Emotional Memory Survey (EMS). It was anticipated that deceptive reports would have discernable characteristics, both other-rated (e.g. less detail, less contextual information) and self-rated (e.g. lower vividness, clarity) that would betray their lack of veracity. Specifically, truthful narratives would be more detailed, emotional, contextually embedded and plausible relative to fabricated narratives. On the other hand, fabricated narratives of trauma may be more coherent and have a rehearsed quality as a result of increased cognitive efforts to ‘keep the story straight’. In addition, it was anticipated that memories for both types of trauma will decrease in consistency over time, but that the rate of decay for fabricated claims would be greater.

METHOD

Participants

Participants (N = 291; 58 males, 233 females) were undergraduates with a mean age of 19.64 years (SD = 2.38) at Time 1. Of this sample, 252 participants were successfully re-interviewed at Time 2 (M = 91.44 days; SD = 12.77; rate of attrition = 13.4%). Finally, 181 participants were involved in the third phase of the experiment (M = 72.13 days after Time
Participants received course credit or monetary compensation for their involvement. Criteria to participation were that participants had experienced a moderate to severe traumatic event (self-selected due to the subjective nature of trauma) after the age of 16. All events were assessed using the Revised Impact of Event Scale (IES-R; Weiss & Marmar, 1997) and those that scored above 20 (arbitrary cut-off to eliminate ‘mildly negative’ experiences) were included in the sample \( (M = 46.55, SD = 16.75; \text{see Peace, Porter, } & \text{Cook, 2010, for discussion of this trauma symptom data). Participants reported genuine traumatic events according to the following categories: sexual/physical violence (11.8%), theft/robbery/mugging (4.9%), serious accident/injury (21.7%), death of a loved one (22.3%), serious medical condition (9.4%), relationship breakdown/divorce (15.0%) and other (including natural disasters, getting lost, extreme failure; 14.9%). Distribution of events reported at Times 1, 2 and 3, did not significantly differ (Friedman’s \( \chi^2 = .65 \text{ with df } = 2, p > .05 \)), indicating that attrition was not due to trauma type\(^1\). A large initial sample was used to account for attrition due to the longitudinal design and difficulties in participant retention (five attempts at re-contact were made before dropping).

**Measures**

*Instructions and Written Narratives*

Participants were provided questionnaires instructing them to report on a real and false traumatic event (counterbalanced) in as much detail as possible. For the fabricated traumas, an additional instruction was provided for participants to complete the written description and questionnaires as if the ‘fake trauma’ had really happened to them (at any point since the age of 16), and to make their responses as convincing as possible for future assessors. Participants also were advised their reports would be shown to legal professionals to judge their credibility in training to create added incentive for participants to generate plausible fabricated claims (see Porter et al., 2007, for further information).

*Memory Assessment Procedure (MAP)*

The MAP includes a series of narrative ratings that involved numerical tallies (e.g. amount of detail, emotional components, contextual details, number of words) and ratings on a seven-point Likert scale (e.g. coherence, relevance, plausibility). These features were made by trained coders blind to veracity condition. Previous research has established high inter-rater reliability for all criteria (e.g. Porter et al., 2007: \( r_s = .82–1.0 \)).

*Emotional Memory Survey (EMS)*

The EMS (see Porter et al., 2007) includes features of memory self-reported by participants. All features (e.g. vividness/clarity, quality, sensory details, confidence, credibility) were rated on a seven-point Likert scale for both narratives in order to address RM features of truthful and fabricated narratives. Both the MAP and EMS have been used extensively in research evaluating the characteristics of memories for traumatic and non-traumatic experiences (e.g. Peace & Porter, 2004; Peace, Porter, & ten Brinke, 2008; Porter & Peace, 2007; Woodworth, Porter, ten Brinke, Doucette, Peace, & Campbell, 2009).

\(^1\)There were no significant differences on major demographic, traumatic impact, memory and consistency variables across participants who remained in the study relative to those who did not (all \( ps > .05 \)).
Consistency Questionnaire (CQ)

This questionnaire was modified from previous research (see Peace & Porter, 2004; Porter & Peace, 2007) and consisted of details from 23 factual categories (e.g. time of day, number of people at the scene, event duration, what the participant was wearing, etc.) that were applicable to both truthful and fabricated traumatic experiences. Participants provided responses to each of these questions following provision of both their truthful and fabricated narratives, which were later rated by trained coders. Each category was rated as inconsistent (0), mainly inconsistent (1), mainly consistent (2) or fully consistent (3) across the three interviews, to generate consistency scores for each memory feature. Due to the longitudinal design of this study, comparisons were made between Time 1 and 2 (T1–T2), Times 2 and 3 (T2–T3), and across the extended interval of Times 1 to 3 (T1–T3). Total consistency scores were calculated for each memory type by summing the numerical ratings across the 23 event categories. This scheme has a high level of inter-rater reliability; consistency scores for the two raters were highly correlated for both narrative types ($r > .95$, $p < .001$) and across each consistency item ($r = .86$ to .98, $ps < .005$).

Procedure

All potential participants who had experienced a traumatic event at some point after the age of 16 were screened using the IES-R and included in the sample if they scored over 20 (arbitrary cut-off indicating moderate trauma). Participants then received instructions indicating they were to provide a narrative of this genuine traumatic event, as well as provide a deceptive account of trauma (counterbalanced) that occurred within the same time frame. Following each narrative, they completed the EMS (self-rated memory features) and the CQ (self-report answers to a series of direct questions), with approximately 40–45 minutes between provision of the two narratives. This procedure was identical at Times 2 and 3, and took approximately 1.5–2 hours for each phase. Coders were trained on the MAP criteria and consistency coding scheme, and provided rating forms on these measures for each narrative. Coders were blind to narrative condition and inter-rater reliability was assessed. This study used a 2 (narrative type: Truthful, fabricated) × 3 (testing time: Initial, 3 months, 6 months) within subjects repeated measures multivariate design.

RESULTS

Differences in truthful and fabricated trauma narratives

The means for other-rated (e.g. amount of detail, coherence, plausibility) and self-rated (e.g. vividness, sensory detail) features are presented in Tables 1 and 2 (respectively). A 2 (narrative type) × 3 (time) MANOVA on other-rated narrative properties revealed a main effect of narrative type, Wilks’ lambda ($\lambda$) = .49, $F(8,166) = 21.32$, $p < .001$, $\eta^2 = .51$, and $r = .77$–.99 and total consistency scores ($r > .95$) at each time. Paired t-test analyses revealed no significant mean differences across all measures.

2An inter-coder reliability check was conducted on the MAP and consistency coding schemes for 20% of the sample by a trained coder who was blind to the hypotheses of the study and the narrative condition of the reports. Bivariate correlations indicated highly reliable associations between coders for the eight memory features ($r = .77$–.99) and total consistency scores ($r > .95$) at each time. Paired t-test analyses revealed no significant mean differences across all measures.
Table 1. Means (SDs) for other-rated MAP features for truthful and fabricated narratives over Times 1, 2 and 3 (including marginal means and standard deviations)

<table>
<thead>
<tr>
<th>MAP narrative features</th>
<th>Truthful traumas</th>
<th>Fabricated traumas</th>
<th>Overall effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time 1</td>
<td>Time 2</td>
<td>Time 3</td>
</tr>
<tr>
<td>Amount of detail $^a$</td>
<td>211.59 (129.18)</td>
<td>158.28 (91.17)</td>
<td>154.11 (93.18)</td>
</tr>
<tr>
<td>Emotional (self) $^a$</td>
<td>4.48 (4.87)</td>
<td>3.40 (4.78)</td>
<td>1.92 (2.46)</td>
</tr>
<tr>
<td>Emotional (other) $^a$</td>
<td>1.20 (2.34)</td>
<td>1.23 (2.32)</td>
<td>.73 (.120)</td>
</tr>
<tr>
<td>Time &amp; place details $^a$</td>
<td>18.72 (12.80)</td>
<td>16.56 (11.60)</td>
<td>14.77 (9.99)</td>
</tr>
<tr>
<td>Coherence $^b$</td>
<td>6.21 (.81)</td>
<td>5.98 (1.10)</td>
<td>6.14 (.90)</td>
</tr>
<tr>
<td>Relevant details $^a$</td>
<td>6.02 (.98)</td>
<td>5.79 (1.06)</td>
<td>5.93 (1.04)</td>
</tr>
<tr>
<td>Plausibility $^a$</td>
<td>6.07 (.80)</td>
<td>6.00 (.94)</td>
<td>6.04 (.78)</td>
</tr>
<tr>
<td>Word count $^a$</td>
<td>389.72 (217.02)</td>
<td>307.23 (179.57)</td>
<td>289.48 (172.96)</td>
</tr>
</tbody>
</table>

$^a$main effect of narrative type.
$^b$main effect of time.
Table 2. Means (SDs) for self-rated EMS features for truthful and fabricated narratives over Times 1, 2 and 3 (including marginal means and standard deviations)

<table>
<thead>
<tr>
<th>EMS narrative features</th>
<th>Truthful traumas</th>
<th>Fabricated traumas</th>
<th>Overall effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time 1 n = 290</td>
<td>Time 2 n = 250</td>
<td>Time 3 n = 179</td>
</tr>
<tr>
<td>Vividness/ clarity abc</td>
<td>5.52 (1.46)</td>
<td>5.21 (1.43)</td>
<td>5.07 (1.50)</td>
</tr>
<tr>
<td>Memory quality abc</td>
<td>5.48 (1.60)</td>
<td>5.27 (1.47)</td>
<td>5.17 (1.48)</td>
</tr>
<tr>
<td>Coherence abc</td>
<td>5.33 (1.42)</td>
<td>5.16 (1.40)</td>
<td>5.10 (1.44)</td>
</tr>
<tr>
<td>Sensory details b</td>
<td>4.15 (1.81)</td>
<td>3.96 (1.62)</td>
<td>4.06 (1.61)</td>
</tr>
<tr>
<td>Confidence in memory abc</td>
<td>6.07 (1.28)</td>
<td>5.67 (1.35)</td>
<td>5.30 (1.50)</td>
</tr>
<tr>
<td>Credibility ab</td>
<td>5.93 (1.32)</td>
<td>5.41 (1.43)</td>
<td>5.35 (1.45)</td>
</tr>
</tbody>
</table>

*a* main effect of narrative type.

*b* main effect of time.

*c* interaction between narrative type and time.
a main effect of time, $\lambda = .42, F(16,158) = 13.46, p < .001, \eta^2 = .58$, with no significant narrative type × time interaction. Overall, truthful narratives were associated with significantly more detail, $F(1,173) = 48.54, p < .001$, emotional components relating to the self, $F(1,173) = 20.44, p < .001$, and others, $F(1,173) = 12.07, p < .01$, time and place details, $F(1,173) = 42.95, p < .001$, relevant details, $F(1,173) = 11.92, p < .01$, higher levels of plausibility, $F(1,173) = 64.37, p < .001$, and a greater number of words, $F(1,173) = 53.85, p < .001$, relative to fabricated narratives (see Table 1 for individual, marginal means and Cohen’s $d$ effect sizes). Truthful (TT) and fabricated (FT) narratives did not differ on their level of coherence ($p > .05$). All other-rated characteristics also decreased over time for both TTs and FTs (all $ps < .05$). Pair-wise comparisons indicated a significant narrative type × time interaction for plausibility, $F(2,346) = 4.39, p < .05, \eta^2 = .03$, suggesting that lower ratings over time were accounted for by the reduction of plausibility in fabricated narratives but not truthful narratives (see Figure 1).

Similar to the above analysis, a 2 (narrative type) × 3 (time) MANOVA on self-rated memory features yielded an overall main effect of narrative type, $\lambda = .64, F(6, 154) = 14.35, p < .001, \eta^2 = .36$, a main effect of time, $\lambda = .55, F(12,148) = 9.94, p < .001, \eta^2 = .45$ and a narrative type × time interaction, $\lambda = .83, F(12,148) = 2.6, p < .01, \eta^2 = .17$. Truthful narratives were rated as being more vivid, of higher quality, more coherent, higher in confidence in the accuracy of the narrative, and greater perceived credibility, relative to fabricated narratives (all $ps < .001$; see Table 2 for means and $d$-values). Univariate analyses of these memory characteristics yielded significant decreases over time for vividness, quality, coherence, sensory details, confidence and credibility when collapsed over narrative type (all $ps < .001$). There was a narrative type × time interaction.
interaction for each (see Table 2 for individual and marginal means). This effect indicated that self-reported vividness, quality, coherence and confidence (all \( p < .05 \)) decreased in a small but significant linear fashion for truthful traumas across the testing times relative to fabricated narratives (which decreased to a greater degree from T1 to T2 and remained relatively stable between T2 and T3).

**Consistency of real and fabricated traumas over time**

The consistency of truthful and fabricated narratives over time was evaluated using total and item scores from the Consistency Questionnaire. Mean scores are reported for all time interval comparisons in Table 3. The mean level of consistency (derived from summing the scores from the 23 item categories) over the extended Time 1 to 3 (T1–T3) interval was 41.23 (SD = 7.83) for truthful narratives, relative to 35.02 (SD = 6.81) for fabricated narratives. To investigate the consistency of truthful and fabricated narratives of trauma, a 2 (narrative type) × 3 (time) MANOVA with total consistency scores as the dependent measure was conducted. There were main effects of narrative type, \( \lambda = .67, F(1,96) = 46.36, p < .001 \), \( \eta^2 = .33 \), and time, \( \lambda = .41, F(2,95) = 67.35, p < .001 \), \( \eta^2 = .59 \), as well as a narrative type × time interaction, \( \lambda = .91, F(2,95) = 4.52, p < .05 \), \( \eta^2 = .09 \). Overall, TTs (\( M = 43.15, SE = .64 \)) were more consistent than FTs (\( M = 38.17, SE = .60 \)). Pair-wise comparisons indicated that scores (across truthful and fabricated narratives) differed at each of the time intervals (all \( p s < .01 \); see Table 3 for means and effect sizes). The interaction indicated that these findings resulted from increases in consistency scores between T2–T3 relative to both T1–T2 and T1–T3 for both narrative types, with a slightly greater increase for FTs. An additional MANOVA comparing consistency scores for both narrative types across T1–T2 and T1–T3 intervals indicated that both memory types decreased over time, \( \lambda = .93, F(1,96) = 7.01, p < .01 \), and that TTs were more consistent overall relative to FTs, \( \lambda = .67, F(1,96) = 47.27, p < .001 \) (see Figure 2). The memory type × time interaction was not significant.

To examine these differences further, mean scores for each of the 23 consistency categories were examined (see Table 3) as a function of time and narrative type (using the initial and extended interval). The MANOVA yielded a main effect of narrative type only, \( \lambda = .33, F(23,74) = 6.67, p < .001 \), \( \eta^2 = .67 \), where the overall mean item score for truthful traumas was 1.88 (SE = .03) relative to 1.67 (SE = .03) for fabricated events (Cohen’s \( d = .70 \)). Memory differences were revealed for the following consistency categories: Clothes, time of day, day of the week, calendar date, smells, sensations, age at time of event and physical injury (all \( p s < .05 \)). The mean level of consistency was higher in TTs relative to FTs for each of these categories. These data indicate that consistency was a relatively static characteristic of individual participants over time (within narrative types).

**DISCUSSION**

False allegations hold enormous harmful consequences for both accused parties and society in general. Yet, little research has addressed how to better identify fabricated reports of traumatic victimization by complainants, or their evolution over time. One of the main hypotheses in this study was that truthful and fabricated narratives of trauma would be qualitatively different. Our findings supported one of our main predictions: Truthful narratives contained more detail overall, more emotional components, more contextual

Table 3. Mean item and total scores (and standard deviations) for truthful (TT) and fabricated (FT) traumas across Times 1 and 2, Times 2 and 3 and the overall interval of Times 1 to 3 on the Consistency Questionnaire.

<table>
<thead>
<tr>
<th>Consistency items</th>
<th>Time 1–Time 2</th>
<th>Time 2–Time 3</th>
<th>Time 1–Time 3</th>
<th>Marginal means (SE)</th>
<th>Overall effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TT (n = 249)</td>
<td>TT (n = 179)</td>
<td>TT (n = 178)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clothes</td>
<td>2.04 (.95)</td>
<td>1.58 (.92)</td>
<td>2.47 (.95)</td>
<td>2.12 (.95)</td>
<td>2.29 (.95)</td>
</tr>
<tr>
<td>Weather</td>
<td>1.76 (.95)</td>
<td>1.42 (.92)</td>
<td>1.88 (.94)</td>
<td>1.70 (.94)</td>
<td>1.78 (.94)</td>
</tr>
<tr>
<td>Location</td>
<td>1.99 (.99)</td>
<td>1.71 (.98)</td>
<td>2.26 (.98)</td>
<td>1.98 (.98)</td>
<td>1.97 (.98)</td>
</tr>
<tr>
<td>Time of day</td>
<td>1.70 (1.12)</td>
<td>1.53 (1.15)</td>
<td>1.99 (1.05)</td>
<td>1.83 (1.14)</td>
<td>1.82 (.88)</td>
</tr>
<tr>
<td>Emotions before</td>
<td>1.39 (1.20)</td>
<td>1.39 (1.20)</td>
<td>1.65 (1.15)</td>
<td>1.45 (1.23)</td>
<td>1.45 (.88)</td>
</tr>
<tr>
<td>Number people</td>
<td>1.52 (1.37)</td>
<td>1.65 (1.41)</td>
<td>1.83 (1.21)</td>
<td>1.46 (1.23)</td>
<td>1.75 (.88)</td>
</tr>
<tr>
<td>Day of week</td>
<td>2.07 (1.12)</td>
<td>1.75 (1.30)</td>
<td>2.26 (1.07)</td>
<td>2.06 (1.14)</td>
<td>2.18 (.88)</td>
</tr>
<tr>
<td>Duration</td>
<td>1.31 (1.19)</td>
<td>1.20 (1.15)</td>
<td>1.58 (1.20)</td>
<td>1.21 (1.21)</td>
<td>1.36 (.98)</td>
</tr>
<tr>
<td>Personal items</td>
<td>2.06 (1.18)</td>
<td>2.04 (1.05)</td>
<td>2.11 (1.13)</td>
<td>2.09 (.12)</td>
<td>2.07 (.88)</td>
</tr>
<tr>
<td>Emotions during</td>
<td>1.49 (1.13)</td>
<td>1.60 (1.10)</td>
<td>1.67 (1.04)</td>
<td>1.43 (1.11)</td>
<td>1.53 (.88)</td>
</tr>
<tr>
<td>Calendar date</td>
<td>1.89 (.97)</td>
<td>1.13 (1.02)</td>
<td>2.03 (.98)</td>
<td>1.94 (1.00)</td>
<td>1.98 (.78)</td>
</tr>
<tr>
<td>Smells</td>
<td>2.06 (1.22)</td>
<td>1.52 (1.31)</td>
<td>2.26 (1.10)</td>
<td>1.90 (1.23)</td>
<td>2.13 (.88)</td>
</tr>
<tr>
<td>Visual details</td>
<td>1.28 (1.22)</td>
<td>1.04 (1.18)</td>
<td>1.67 (1.17)</td>
<td>1.30 (1.21)</td>
<td>1.37 (.98)</td>
</tr>
<tr>
<td>Landmarks</td>
<td>1.24 (1.27)</td>
<td>1.24 (1.26)</td>
<td>1.55 (1.25)</td>
<td>1.03 (1.21)</td>
<td>1.29 (.88)</td>
</tr>
<tr>
<td>Emotions after</td>
<td>1.32 (1.13)</td>
<td>1.25 (1.14)</td>
<td>1.50 (1.06)</td>
<td>1.27 (1.12)</td>
<td>1.35 (.88)</td>
</tr>
<tr>
<td>Weapons</td>
<td>2.76 (.78)</td>
<td>2.54 (1.01)</td>
<td>2.61 (.93)</td>
<td>2.49 (.07)</td>
<td>2.62 (.78)</td>
</tr>
<tr>
<td>Noises</td>
<td>1.53 (1.26)</td>
<td>1.20 (1.25)</td>
<td>1.81 (1.21)</td>
<td>1.45 (1.26)</td>
<td>1.60 (.98)</td>
</tr>
<tr>
<td>Sensations</td>
<td>1.48 (1.19)</td>
<td>1.20 (1.19)</td>
<td>1.71 (1.17)</td>
<td>1.49 (1.19)</td>
<td>1.61 (.88)</td>
</tr>
<tr>
<td>Others involved</td>
<td>2.28 (1.09)</td>
<td>2.36 (1.08)</td>
<td>2.52 (1.88)</td>
<td>2.34 (1.02)</td>
<td>2.42 (.78)</td>
</tr>
<tr>
<td>Age at time</td>
<td>2.34 (1.17)</td>
<td>1.55 (1.47)</td>
<td>2.49 (1.04)</td>
<td>2.50 (1.07)</td>
<td>2.53 (.78)</td>
</tr>
<tr>
<td>Actions before</td>
<td>1.53 (1.31)</td>
<td>1.57 (1.30)</td>
<td>1.97 (1.23)</td>
<td>1.54 (1.32)</td>
<td>1.78 (.08)</td>
</tr>
<tr>
<td>Physical injury</td>
<td>2.63 (.83)</td>
<td>1.97 (1.26)</td>
<td>2.53 (.97)</td>
<td>2.40 (1.07)</td>
<td>2.56 (.69)</td>
</tr>
<tr>
<td>Actions after</td>
<td>1.50 (1.21)</td>
<td>1.38 (1.20)</td>
<td>1.76 (1.20)</td>
<td>1.49 (1.17)</td>
<td>1.68 (.88)</td>
</tr>
<tr>
<td>Overall item consistency</td>
<td>1.83 (.33)</td>
<td>1.59 (2.92)</td>
<td>2.02 (2.92)</td>
<td>1.79 (.34)</td>
<td>1.88 (.30)</td>
</tr>
<tr>
<td>Total mean consistency</td>
<td>42.08 (7.50)</td>
<td>36.64 (6.64)</td>
<td>46.43 (6.72)</td>
<td>42.71 (8.00)</td>
<td>43.25 (6.30)</td>
</tr>
</tbody>
</table>

*a main effect of narrative type
(time and place) information, more relevant details, a greater number of words and were rated as more plausible relative to fabricated statements. These results corroborate research using both CBCA and RM procedures for assessing the credibility of a statement (e.g. Granhag & Vrij, 2005; Masip et al., 2005; Vrij, 2005). Although we predicted that fabricated narratives would be rated by others as more coherent than truthful narratives, the results did not bear this out.

We also examined the self-reported features of trauma narratives. As stated by DePaulo et al. (2003): ‘Liars usually make an effort to seem credible; truth tellers more often take their credibility for granted’ (p. 78). Our prediction was in line with this conclusion, but was not supported by our results. Truthful narratives were rated as vivid, having a higher overall memory quality, more coherent and perceived by the participant as more credible relative to fabricated narratives. It is likely that this pattern resulted from participants’ personal knowledge of the veracity of the event which was revealed in their subjective ratings, whether intentional or not. This reflects meta-memory processes (such as ‘feelings of knowing’) and the surrounding context of recall (i.e. ‘I know I was never sexually assaulted’) when participants are required to make assessments of their own memory features (e.g. Johnson et al., 1988; Kelly, Carroll, & Mazzoni, 2002; Sporer, 2008). In their review, DePaulo et al. (2003) also argued that liars have more difficulty convincingly embracing their deceptive claims due to moral qualms, personal investment and lack of knowledge and experience in the circumstances being lied about. Studies based on real and imagined events using RM techniques have frequently found that participants rate their own truthful experiences as more realistic, memorable, believable and detailed (e.g. Arbuthnott, Geelen, & Kealy, 2002; Kealy et al., 2006; Sporer, 1997). In addition, the traumatic nature of the events in this study may have impeded participants’ abilities to respond as if the fabricated trauma had happened. Thus, these overall findings support the theoretical basis of RM decisions when applied to self-reported assessments.
Differences in the subjective features of narratives were especially noteworthy between Times 1 and 2, and the overall Time 1 to 3 interval. Higher overall ratings associated with truthful narratives (as indicated above) gradually decreased over the 6-month interval. However, truthful and fabricated narratives did not differ between Times 2 and 3. While subjective ratings of memory features declined in a more linear fashion for truthful traumas, ratings for fabricated traumas dropped more (in general) between Times 1 and 2 and then leveled off between Times 2 and 3 (refer back to Table 2 for means on these subjectively reported narrative properties). The only exception to this pattern was for sensory details; fabricated narratives were rated as having more sensory components initially, but ratings were equivalent to truthful narratives at Times 2 and 3. In contrast to these results, memory assessments using RM and/or CBCA techniques have found that sensory details (both other- and self-rated) are more common in truthful relative to fabricated narratives (see Masip et al., 2005). Perhaps fabrication inflation of sensory components may have resulted from information commonly portrayed in popular crime-oriented television programs (e.g. Crime Scene Investigation; Law & Order) where investigators focus on sounds, smells and sensations surrounding criminal victimization when questioning victims and witnesses. Based on the collective findings, it appears that self-reported features of fabricated narratives become more similar to truthful narratives (and more difficult to distinguish) with repeated retellings (e.g. Granhag, Strömwall, & Olsson, 2001; Granhag & Strömwall, 2002; Suengas & Johnson, 1988). Vrij (2004) argued that memory ratings may fluctuate to a greater extent during initial recall attempts, but these become more stable over time as the deceptive statement is solidified with successive retellings.

Due to the legal significance of testimonial consistency in perceptions of credibility (e.g. Bell & Loftus, 1989; Brewer et al., 1999), narrative changes over time are a critical, but understudied, component of veracity determinations (e.g. Granhag & Strömwall, 2001). It was predicted that truthful narratives of trauma would maintain higher levels of consistency relative to fabricated narratives. In addition, the effects of rehearsal or retelling were predicted to influence fabricated narratives to a greater extent, such that initial decreases may be evident but level off as details become consolidated in memory. The results supported the former prediction, but not the latter. Overall, truthful narratives were recalled more consistently than fabricated narratives, and both decreased in consistency over time. The lack of any interaction when evaluating consistency over the 6-month interval indicated that fabricated narratives did not stabilize over time due to rehearsal and memory consolidation, but continued to decline in the same manner as truthful narratives. Although total mean consistency scores for Times 1–2 and Times 1–3 were different between narratives, truthful and fabricated traumas changed only a small degree over time (approximately 1–1.5 points on the consistency scale). While professionals and laypersons alike often perceive the ‘truth’ as being steadfast and deception as more variable (e.g. King, 2006), the results of this study do not confirm this idea with respect to direct questions about an event. Studies have reported that fabricated narratives of trauma are less consistent relative to truthful narratives (e.g. Mazzoni & Memon, 2003; Polage, 2004), but few (if any) have found similar rates of change in narratives. These results can be explained by considering memory for genuine versus imagined (or deceptive) events. As discussed previously, there are differences in recalling a genuine memory versus retelling of a story (e.g. Dudukovic, Marsh, & Tversky, 2004; Granhag et al., 2006; Lampinen, Odegard, & Bullington, 2003; Sporer & Sharman, 2006; Suengas & Johnson, 1988). Truthful traumas are based in reality, with a detailed memory that has been encoded, stored and recalled. As

evidenced in this study, memories of truthful trauma were highly detailed in terms of contextual and perceptual information. On the other hand, fabricated narratives have little or no basis in reality, thus perceptual and contextual information must be generated and is unlikely to leave a lasting ‘impression’ in memory. Recollection of genuine events is derived from the memory of the original experience, whereas recollection of fabrication involves remembering a prior episode of being deceptive. More detailed and consistent reports are associated more so with truthful claims due to difficulties in embellishing a fabricated report with details generated solely from the imagination and maintaining this over time (Brewer et al., 1999; Vrij, 2008).

In addition, the superior recall of trauma relative to other types of emotional experiences has been repeatedly demonstrated in the literature (e.g. Alexander et al., 2005; Porter & Peace, 2007). While both narratives elicited in the present study were traumatic in nature, a genuinely experienced traumatic event is more personally salient and would be well retained in memory relative to one that was imagined (e.g. Pezdek, 2003). However, truthful traumas would not be fixed and indelible but exhibit minor changes in consistency over time due to the natural variations in memory—such as remembering another detail about the event or temporarily forgetting a detail previously reported (e.g. Loftus, 2003). On the other hand, false allegations of trauma may be distinctive in their invention, but are not directly experienced and are more difficult to maintain consistently over time. The design of this study afforded participants the opportunity to rehearse and retell their fabricated narratives on multiple occasions which likely preserved the gist of their deceptive claims (such as the main details). Analysis of specific consistency items yielded no effect of time (i.e. ratings did not change significantly overall when collapsed across narrative type) and no interaction effects. Overall, truthful narratives were more consistent in details pertaining to clothing, time of day, day of the week, calendar date, smells, sensations, age at the time of the event and physical injuries, relative to fabricated claims. These categories reflect more factual details that genuine victims may naturally recall, and alleged victims may use to ‘anchor’ their story (to the best of their ability) over time. Further research evaluating the consistency of the narratives themselves and the specific details mentioned at Times 1, 2 and 3 (rather than responses to direct questions on the CQ) may yield a different pattern of results and show more variation in narratives over time.

**CONCLUSION**

The present longitudinal study demonstrated with a large non-clinical sample that the features of truthful and fabricated narratives of traumatic victimization differ as a function of narrative type and time. In general, truthful traumas contained more detail, contextual information, emotional details and were considered more plausible relative to fabricated traumas. The details of truthful narratives also were more consistent than fabricated narratives over a 6-month interval, although levels of consistency decreased over time for both honest reports and lies. Ultimately, ‘there is no single piece of information that invariably marks a memory as an accurate reflection of the past’ (Johnson & Raye, 1998, p. 137). As such, the use of global and holistic assessments of memories and narratives are critical. Future research should continue to address various factors associated with deception detection in truthful and fabricated narratives of victimization (e.g. individual differences), and the validity of content-based techniques. While further research on truthful and fabricated narratives of trauma is certainly needed, the current findings provide
useful information for pinpointing possible false allegations and differentiating fact from fiction.

ACKNOWLEDGEMENTS

The authors thank the Social Sciences and Humanities Research Council of Canada (SSHRC) for providing funding throughout the duration of this project. Special acknowledgement goes to Naomi Doucette, Leanne ten Brinke and Allison Eisner for their research assistance. They also thank all of the participants who were involved in this research and the volunteers that assisted in coding the data.

REFERENCES


Truthful vs. fabricated trauma narratives


