

Fantasy Proneness as a Confounder of Verbal Lie Detection Tools

KAREN SCHELLEMAN-OFFERMANS^{1,*} and HARALD MERCKELBACH²

¹*Faculty of Health, Medicine, and Life Sciences, Maastricht University, Maastricht, The Netherlands*

²*Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands*

Abstract

This study investigated whether high fantasy-prone individuals have superior storytelling abilities. It also explored whether this trait is related to specific linguistic features (i.e. self-references, cognitive complexity, and emotional words). Participants high (n = 30) and low (n = 30) on a fantasy proneness scale were instructed to write down a true and a fabricated story about an aversive situation in which they had been the victim. Stories were then examined using two verbal lie detection approaches: criteria-based content analysis (CBCA) and linguistic inquiry and word count (LIWC). Irrespective of the truth status of the stories, independent observers rated stories of high fantasy-prone individuals as being richer in all nine CBCA elements than those of low fantasy-prone individuals. Furthermore, overall, high fantasy-prone people used more self-references in their stories compared with low fantasy-prone individuals. High fantasy prones' fabricated stories scored higher on various truth indices than authentic stories of low fantasy prones. Thus, high fantasy-prone people are good in creating a sense of authenticity, even when they fabricate stories. Forensic experts should bear this in mind when they employ verbal lie detection tools. Copyright © 2010 John Wiley & Sons, Ltd.

Key words: fantasy proneness; linguistic inquiry and word count; criteria-based content analysis; verbal lie detection; false allegations

INTRODUCTION

Stories of witnesses and defendants play a critical role in the legal arena. To the extent that these stories make a trustworthy impression on police, expert witnesses, and judges, they may provide legal facts with an interpretation. In this way, stories may become intimately related to legal evidence (Bennett, 1992; Pennington & Hastie, 1986). It is not surprising, then, that in the forensic literature, many methods have been proposed to

*Correspondence to: Karen Schelleman-Offermans, Faculty of Health, Medicine, and Life Sciences, Maastricht University, Maastricht, The Netherlands.
E-mail: karen.offermans@maastrichtuniversity.nl

discriminate between true and false statements on the basis of their narrative and linguistic features (see, for reviews, Vrij, 2008; Zhou, Burgoon, Nunamaker, & Twitchell, 2004). The current study focused on two of them: the criteria-based content analysis (CBCA; Steller & Köhnken, 1989) and linguistic inquiry and word count (LIWC; Newman, Pennebaker, Berry, & Richards, 2003; see also Pennebaker & King, 1999).

The CBCA is based on the so-called Undeutsch hypothesis, named after the German psychologist Udo Undeutsch (1989). Referring to his extensive forensic experience, Undeutsch argued that descriptions of authentic autobiographical events differ from fabricated stories on a number of narrative dimensions such as richness in details, reproduction of conversations, and spontaneous corrections. According to Undeutsch, truthful statements have more of these features than fabricated statements. Steller and Köhnken (1989) took the Undeutsch hypothesis as the starting point for developing the CBCA. Originally, the CBCA was only employed as a forensic tool for evaluating the veracity of child witness accounts, but nowadays, it is also used to evaluate adults' written statements, both in the research context and in the legal domain (e.g. Vrij, Akehurst, Soukara, & Bull, 2004). Vrij (2008) critically reviewed 6 field studies and 25 laboratory studies that looked at the extent to which CBCA features are present in truthful and false statements. The author concluded that the 'CBCA criteria emerged more frequently in truthful than in deceptive statements regardless of the experimental research paradigm that was used, that is actual involvement, watching a video or staged event, or statements derived from memory' (Vrij, 2008, p. 228). Vrij also noted that the error rate of CBCA assessments is too high to make them accurate enough to be used in court. Nevertheless, the CBCA is employed as a forensic tool on a wide scale by German and Dutch expert witnesses.

CBCA evaluations are context sensitive. For example, experts can only identify a conversational part (i.e. reproduction of conversations; see Table 2) when they understand the entire text. Another approach is to look for general linguistic markers that can be counted, irrespective of the semantic context. Such approach lends itself to automatic categorisation of text cues (e.g. Zhou *et al.*, 2004). The best example of this approach is LIWC, a software program that analyses written texts on a word-by-word basis. It is based on work by Pennebaker and King (1999) and Newman *et al.* (2003), who identified three linguistic markers of deception. Thus, whilst the CBCA is focused on truthfulness, LIWC markers are believed to indicate fabrication. More specifically, Newman *et al.* (2003) found evidence that compared with true stories, fabricated stories are characterised by (1) fewer markers of cognitive complexity; (2) less self-references; and (3) more negative emotion words. The first characteristic—lack of cognitive complexity—is consistent with the well-validated idea that in general, lying itself is cognitively demanding (Vrij, 2008). To compensate for this, liars would use simpler language (e.g. simple verbs such as 'walk' and 'go') than truth tellers (Newman *et al.*, 2003). Also, relative to truth tellers, liars would less frequently employ exclusive words such as 'but' and 'except', because unlike truth tellers, they are not able to indicate what did and did *not* happen (Newman *et al.*, 2003; but see Hancock, Curry, Goorha, & Woodworth, 2008).

The second marker of lying (i.e. less self-references) has been documented in several studies (e.g. Burgoon, Buller, Guerrero, Afifi, & Feldman, 1996; Ebesu & Miller, 1994; but see: DePaulo, Lindsay, Malone, Muhlenbruck, Charlton, & Cooper, 2003). Compared with truth tellers, liars employ less self-references, such as 'I' and 'mine'. Several authors have argued that this has to do with the fact that liars are not personally involved in the story, which means that they did not really experienced the thoughts, feelings, and emotions that they try to describe (Newman *et al.*, 2003; Vrij, 2008). Another reason that has

been proposed in the literature is that liars want to distance themselves from their lie, because they do not want to take responsibility for their behaviour, or feel guilty or ashamed about lying (Newman *et al.*, 2003; Vrij, 2008). This moral explanation might also account for the third marker of deception that has been observed in previous studies (Stirman & Pennebaker, 2001; Vrij, 2008; but see: Hancock *et al.*, 2008), namely, liars' relatively frequent use of negative emotion words.

In a systematic study by Newman *et al.* (2003), LIWC was used in five different experimental settings to evaluate truthful and fabricated statements of undergraduate students. In the first three settings (videotaped, typed, and handwritten statements), students were asked to present true and false views on abortion. Analyses showed that LIWC parameters identified truth tellers and liars above chance level. In the fourth setting, students were asked to explain in front of a camera their positive feelings about two people they disliked and two people they liked. In the fifth setting, some participants were involved in a mock crime and were instructed to deny their involvement during a follow-up interview. Others were truly innocent participants who also denied involvement in the mock crime. In these more complex settings, LIWC parameters were not successful in discriminating between liars or truth tellers. However, across settings, the overall hit rate of the combined LIWC parameters was 67% against 52% for judges. Nevertheless, the Newman *et al.*'s (2003) study demonstrates that context matters, i.e. that LIWC parameters are more effective in a simple setting (e.g. talking about a single but highly emotional topic like abortion) than in a complex setting (e.g. responding to an accusation).

Bond and Lee (2005) performed LIWC analyses on truthful and fabricated comments of prisoners about video clips. These researchers found an overall hit rate of 69.9%, indicating that the three LIWC markers possess discriminative power when they are combined to diagnose the truthfulness of statements. On the whole, then, there is evidence to support the notion that the LIWC word count approach might have potential as a forensic tool, although its discriminatory power seems to be context dependent for reasons that are ill understood (e.g. Hancock *et al.*, 2008).

There is an extensive literature on the constraints of verbal lie detection tools such as the CBCA. This literature makes clear, for example, that CBCA dimensions are sensitive to coaching and social skills (e.g. Vrij *et al.*, 2004), and that their ability to discriminate between truthful and false accounts in which people really believe is seriously restricted (Blandon-Gitlin, Pezdek, Lindsay, & Hagen, 2009). Also, much depends on whether verbal accounts have been obtained with information gathering or accusatory interviews. Accusatory interviews produce statements that are less suitable for CBCA assessments for the simple reason that they are shorter in length than those obtained with information gathering interviews (Vrij, Mann, Kristen, & Fisher, 2007).

An issue that has received less attention is how personality traits are related to the verbal features tapped by CBCA and LIWC. People differ in their storytelling ability, and one trait that is intimately related to this ability is fantasy proneness (Merckelbach, Horselenberg, & Muris, 2001a; Sánchez-Bernardos & Avia, 2004, 2006). Individuals who score high on fantasy proneness exhibit an extensive involvement in fantasy, vivid imagery, and daydreaming (Wilson & Barber, 1983). Based on structured interviews with fantasy-prone people, Wilson and Barber (1983, p. 352) concluded that their fantasies are similar to a good movie: 'In the same way as a good movie, the fantasy can be fun and exciting and can be experienced as vividly and realistically as any other aspect of life. They can experience anything in fantasy—people can be seen and heard to speak; food can be smelled and tasted; sensations such as touch, heat, and cold can be felt; and emotions such as fright

and joy can be experienced—and when immersed in fantasy, they do not ask whether their experiences are real.’

Given their control over imagery, one would expect that people scoring high on fantasy proneness have a talent to tell stories that are detailed and complex, and that contain dialogues and self-references, i.e. stories rich in CBCA characteristics and low on LIWC markers of deception. One would also expect that this pattern emerges not only when high fantasy prones describe imagined experiences, but also when they describe things that really happened. Preliminary support for these predictions comes from two studies (Merckelbach, 2004). In these studies, undergraduates high and low in fantasy proneness were asked to write down true and fabricated stories, which were then given to raters who evaluated the stories using the CBCA. Raters consistently judged stories of high fantasy-prone participants to be richer in CBCA elements than those of low fantasy-prone participants. A limitation of these studies was the relatively low intercorrelation between CBCA ratings of different judges. Also, sample sizes in both studies were relatively small ($n \leq 38$), and analyses of the story material did not include the three LIWC parameters discussed above.

The current study tried to replicate the basic finding of Merckelbach (2004) that high fantasy-prone people are able to produce true and fabricated stories that are richer in CBCA dimensions than those produced by low fantasy-prone participants. In addition, our study looked at the three LIWC parameters. Thus, we expected that high fantasy-prone individuals would use more markers of cognitive complexity, more self-references, and less negative emotion words than low fantasy-prone individuals, even when they write down stories that are fabricated (Newman *et al.*, 2003). If such associations with fantasy proneness do, indeed, occur, this is important information for researchers and practitioners who employ instruments such as CBCA and LIWC to evaluate the truth status of statements.

METHOD

Participants

Participants were selected on the basis of testing sessions in which a fantasy proneness questionnaire (i.e. the Creative Experiences Questionnaire [CEQ]; see below) was administered to first-year psychology, drama, and art students ($n = 351$). Those with high fantasy proneness scores (CEQ > 12; top 15% of the distribution) were invited to participate in a follow-up study, as were those with low fantasy proneness scores (CEQ < 6; fifth decile of the distribution).

Students who could be contacted and who were willing to come to the laboratory were invited to participate in a study on storytelling with two test sessions, separated by a week. Participants were not informed about their fantasy proneness status. The final sample consisted of 60 participants: 30 in the high fantasy-prone group and 30 in the low fantasy-prone group. The details of the two groups are given in Table 1. The high fantasy-prone group mainly consisted of art and drama students, whilst the low fantasy-prone group consisted of psychology students (Fisher exact $p < 0.001$). The groups did not differ with respect to their gender distribution (Fisher exact $p = 0.19$), but high fantasy-prone participants were significantly older than low fantasy-prone participants ($t [58] = 2.11, p < 0.05$). Fantasy proneness is intimately linked to dissociative experiences (e.g. feelings of de-realisation, absorption, and identity confusion; e.g. Merckelbach, 2004). As a check on participant selection, we asked participants to complete the C version of the Dissociative

Table 1. Characteristics of high and low fantasy-prone groups

	High fantasy prone	Low fantasy prone
Number of psychology students included	3	30
Number of drama students included	2	0
Number of art students included	25	0
Number of men included	12	18
Mean age	22.5 (2.9)	21.1 (2.2)
Mean CEQ score pre-test	15.0 (1.9)	3.0 (1.3)
Mean CEQ score post-test	14.9 (3.2)	3.9 (1.7)
Mean DES-C score	2.7 (1.2)	5.5 (1.1)

Note. Standard deviations are shown in parentheses.

CEQ, Creative Experiences Questionnaire; DES-C, C version of the Dissociative Experiences Scale.

Experiences Scale (DES-C; Wright & Loftus, 1999; see below). Replicating Merckelbach (2004), high fantasy prones had higher DES-C scores than low fantasy prones ($t [58] = 25.18, p < 0.01$; Cohen's $d = 2.47$).

The experiment was approved by the standing ethical committee of the Faculty of Psychology and Neuroscience. Psychology students who participated received course credits for their participation.

Measures

CEQ

The CEQ (Cronbach's $\alpha = 0.79$; Merckelbach *et al.*, 2001a; Sánchez-Bernardos & Avia, 2004, 2006) is a self-report measure of fantasy proneness. It comprises 25 yes–no items that were derived from extensive case descriptions of fantasy proneness provided by Wilson and Barber (1983). None of the CEQ items allude to lying or deceiving. It is important to stress this point because if the CEQ would include such items, it would be circular to study fabricated stories of high fantasy prones. Some CEQ items refer to the developmental antecedents of fantasy proneness. Other items have to do with intense elaboration of and profound involvement in fantasy and daydreaming. Still, others pertain to the concomitants and consequences of fantasising. Sample items are: 'In general, I spend at least half of the day fantasizing or daydreaming'; 'My fantasies are so vivid that they are like a good movie'; and 'I tend to confuse my fantasies with memories of real events'. Yes answers are summed to obtain a total score (range 0–25), with higher CEQ scores indicating higher levels of fantasy proneness. CEQ's internal and test–retest reliabilities are sound, and the scale correlates strongly with concurrent measures of fantasy proneness (e.g. $r = 0.77$ with the Inventory of Childhood Memories and Imaginations; see Merckelbach, Wiers, Horselenberg, & Wessel, 2001b). Merckelbach *et al.* (2001a) reported that the CEQ does not correlate with age ($r = 0.01$).

In the current study, participants completed the CEQ twice: during the initial mass testing session (pre-test) and—to ensure that group assignment was correct—after the second experimental session (post-test). Test–retest stability was $r = 0.92$ ($p < 0.01$).

DES-C

As an integrity check on our participant selection, we also asked participants to complete the C version of the DES (Wright & Loftus, 1999). This version lists 28 phenomena that

are considered typical manifestations of dissociation (e.g. missing part of a conversation, talking out loud to oneself when alone, not recognising friends or family members, looking at the world through a fog). Respondents are asked how often they have these experiences compared with other people. They are instructed to tick one of the 11 horizontally presented boxes, with the extreme boxes having the labels 'much less than others' and 'much more than others'. Responses are scored on a 0 ('much less than others') to 10 ('much more than others') scale. Total DES-C scores are obtained by averaging across the 28 items.

CBCA

CBCA (Steller & Köhnken, 1989; Vrij, 2005, 2008) was originally designed to evaluate statements of child witnesses in sexual abuse cases. Although it consists of 19 criteria, the current study used only those criteria that have received strong empirical support (e.g. Vrij, 2005) and that are suitable for the laboratory context. Thus, criteria that have to do with typical child witness issues (e.g. details of an event that are misunderstood) were excluded. In our study, true and fabricated stories were evaluated against the following nine criteria (see also Merckelbach, 2004): logical structure, quantity of relevant details, contextual embedding, descriptions of interactions, reproduction of speech, unusual details, superfluous details, referral to own subjective experience, and attribution of the perpetrator's mental state. The reader is referred to Vrij (2008) for a detailed description of these criteria.

Using four-point scales (0 = criterion is not present; 3 = criterion is strongly present), two independent judges evaluated to what extent each of the nine criteria were present in true and fabricated stories. Judges were familiar with the CBCA literature and had been performing CBCA ratings for a previous study. Scores were summed and averaged across the judges (range 0–36). Inter-rater correlations were 0.81 for true stories and 0.85 for fabricated stories (both $ps < 0.01$). Note, in passing, that these rs are higher than those in the Merckelbach (2004) study, where inter-rater correlations were in the order of 0.55.

LIWC

LIWC (Pennebaker, Francis, & Booth, 2001) is a software program that analyses written statements word by word and stores them into several word categories. The number of words in each category (e.g. negative emotion words) is counted, adjusted for the number of total words used, and is expressed as frequency per 100 words. Thus, LIWC can provide a quantitative summary of the linguistic features in a statement, thereby correcting for the length of the statement. Our study made use of a Dutch dictionary database that was developed by researchers of Utrecht University (Zijlstra, Middendorp, Meerveld, & Geenen, 2005). Their psychometric analyses indicate that the Dutch LIWC version possesses adequate reliability in terms of internal consistency (Cronbach's $\alpha > 0.70$) and test-retest stability (e.g. rs of 0.75 and 0.88 for self-references and negative emotion words, respectively). In the current study, LIWC analysis focused on the three parameters that have been found to be related to fabrication (i.e. markers of cognitive complexity, self-reference, and negative emotion words). More specifically, for cognitive complexity, LIWC counted the rates of simple motion verbs (e.g. 'go'; 'take') and exclusive words (e.g. 'but'; 'although'), adjusted for the total number of words by reporting rates per 100

words. Frequencies of self-references (e.g. 'I'; 'my') and negative words (e.g. 'hate'; 'death') were also counted and calculated per 100 words. LIWC parameters did not correlate significantly with one another, with the possible exception of self-references and negative emotion words in true stories ($r = 0.32, p = 0.02$).¹

Procedure

Participants were told beforehand that the study was about stories and that there were two test sessions. However, we did not inform them beforehand that they had to fabricate stories and that we had selected them on the basis of their fantasy proneness scores. The second test session was at least 1 week after the first one. At the start of both sessions, participants were given written instructions that asked them to think for a while about a real story or a fabricated story, in which they personally suffered from other people's actions. Instructions emphasised that the stories they wrote down would be treated confidentially and anonymously. Next, they were asked to write down the story using a text program on a computer. They were told that the length of the story had to be approximately 400 words. Participants were given unrestricted time to prepare and type the story. For the fabricated stories, the instructions stressed that participants should write down a realistic scenario that, however, never had played a role in their true life. Half of the participants within each group started with the fabricated story in the first test session and wrote the real story during the second session. The other half had the reversed order. At the end of session 1, participants were not informed about the type of story that they had to write during session 2. Appendix A shows two illustrative fragments of stories. After participants had written down their stories in the second session, they were asked to complete the CEQ and the DES-C.

The printed versions of the true and fabricated stories were scored on CBCA dimensions by two blind (i.e. blind as to fantasy proneness status) and independent observers (a 24-year-old psychology student and a 52-year-old psychiatric nurse) in a random order. As said before, both judges had extensive experience with the CBCA in a research context. Next, the stylistic features of the stories (motion verbs, exclusive words, self-references, and negative emotions) were scored using LIWC.

Statistical analyses

We first tested with 2 (order: true–fabricated versus fabricated–true) $\times 2$ (group: high versus low fantasy proneness) $\times 2$ (truth status: true versus fabricated stories) analyses of variance (ANOVAs) performed on CBCA and LIWC parameters whether there were any main effects of or interaction effects with order. In these ANOVAs, order and group were between-subject factors, whilst truth status was a within-subject factor. Next, 2 (group) $\times 2$ (truth status) ANOVAs with repeated measurements on the second factor were carried out for total CBCA ratings averaged across judges, for separate CBCA dimensions, and for separate LIWC parameters. We also conducted specific follow-up *t*-tests to evaluate whether fabricated stories of high fantasy-prone participants differ from true stories of low fantasy-prone participants.

¹This correlation is non-significant when Bonferroni corrections are used.

RESULTS

Topics of the stories were very diverse, and there were no obvious thematic differences between true and false stories of high fantasy-prone and those of low fantasy-prones.² Stories were about how participants had been the victim of gossip, bullying, accidents caused by others, robbery, threat, scam, cheating, sexual assault, vandalism, and so on. True stories of high and low fantasy-prone participants did not differ in length, means being 543 (standard deviation [SD] = 212) and 472 (SD = 109) words ($t [58] = 1.62$, $p = 0.11$). Neither did fabricated stories of high and low fantasy-prones, means being 528 (SD = 154) and 480 (SD = 114) words ($t [58] = 1.38$, $p = 0.17$).

A series of 2 (order) \times 2 (group) \times 2 (truth status) repeated measures ANOVAs yielded no significant main effects of or interaction effects with order, all F s (1, 56) < 3.02 , all p s > 0.09 . Therefore, we did not include this factor in our further analyses. Figure 1 shows mean total CBCA scores for the stories of high and low fantasy-prone participants. A 2 (group) \times 2 (truth status) repeated measures ANOVA revealed a main effect of group ($F [1, 58] = 37.50$, $p < 0.01$; $\eta^2 p = 0.39$), indicating that overall, stories of high fantasy-prone participants were given higher total CBCA ratings than those of low fantasy-prone participants. Also, a main effect of truth status emerged ($F [1, 58] = 8.49$, $p < 0.01$; $\eta^2 p = 0.13$). As can be seen in Figure 1, true stories received higher CBCA scorings than fabricated stories. The interaction term failed to reach significance ($F [1, 58] = 0.21$,

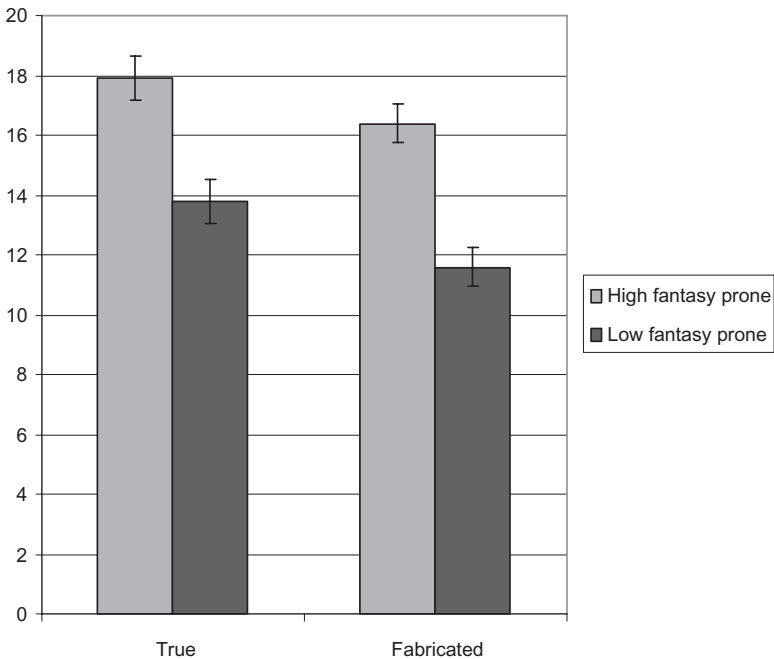


Figure 1. Criteria-based content analysis ratings for true and fabricated stories of high fantasy-prone ($n = 30$) and low fantasy-prone ($n = 30$) participants.

²A complete list of story themes and their frequencies across conditions and groups can be obtained from the first author.

$p = 0.65$). However, fabricated stories of high fantasy prones did receive higher total CBCA ratings than true stories of low fantasy prones, means being 16.4 (SD = 3.4) and 13.8 (SD = 3.3), respectively ($t [58] = 3.01, p < 0.01$; Cohen's $d = 0.79$). As age did not correlate with total CBCA ratings for true or fabricated stories (both $r_s < 0.23$, both $p_s > 0.07$), there was no point in conducting follow-up analyses with age as a covariate.

Table 2 shows averaged ratings for separate CBCA dimensions. A series of 2 (group) \times 2 (truth status) repeated measures ANOVAs were conducted to explore to what extent fantasy proneness is a relevant background variable for individual dimensions. The main effect of group reached significance for all CBCA dimensions, with $F_s (1, 58)$ ranging from 6.48 to 46.07 (all $p_s < 0.015$, all $\eta^2 p_s > 0.10$). As can be seen in Table 2, relative to low fantasy prones, high fantasy prones scored higher on all CBCA dimensions. However, the main effect of truth status attained significance for only three CBCA dimensions: logical structure ($F [1, 58] = 5.63, p < 0.05; \eta^2 p = 0.02$), contextual embedding ($F [1, 58] = 4.03, p < 0.05; \eta^2 p = 0.049$), and attribution of the perpetrator's state ($F [1, 58] = 11.65, p < 0.01; \eta^2 p = 0.001$). For these dimensions, true stories received higher scores than fabricated stories. None of the interaction terms were significant, with all $F_s (1, 58) < 1.0$.

Table 3 shows the mean frequencies of motion verbs, exclusive words, self-references, and negative emotions per 100 words in stories of high and low fantasy prones. As these LIWC parameters were not significantly correlated with each other (cf. supra), separate 2

Table 2. Mean ratings (0–3) of CBCA dimensions in high and low fantasy prones' stories

CBCA parameters	High fantasy prone		Low fantasy prone	
	True	Fabricated	True	Fabricated
Logical structure	2.7 (0.4)	2.6 (0.4)	2.5 (0.4)	2.2 (0.7)
Quantity of relevant details	2.7 (0.4)	2.7 (0.4)	2.3 (0.5)	2.2 (0.7)
Contextual embedding	2.3 (0.6)	2.2 (0.7)	1.8 (0.7)	1.5 (0.8)
Descriptions of interactions	2.3 (0.8)	2.1 (0.8)	2.0 (0.7)	1.6 (1.0)
Reproduction of speech	1.2 (1.2)	0.7 (0.7)	0.4 (0.9)	0.3 (0.8)
Unusual details	2.1 (0.7)	2.0 (0.8)	1.5 (0.7)	1.4 (0.9)
Superfluous details	1.1 (0.9)	1.0 (0.9)	0.3 (0.4)	0.5 (0.8)
Referral to own subjective experience	2.5 (0.6)	2.4 (0.5)	1.9 (0.8)	1.8 (0.8)
Attribution of the perpetrator's mental state	1.7 (1.0)	1.0 (0.8)	1.0 (0.9)	0.7 (0.8)

Note. Standard deviations are shown in parentheses.

CBCA, criteria-based content analysis.

Table 3. Mean frequencies of motion verbs, exclusive words, self-references, and negative emotions (per 100 words) in high and low fantasy prones' stories

LIWC parameters	High fantasy prone		Low fantasy prone	
	True	Fabricated	True	Fabricated
Motion words	2.4 (1.0)	2.6 (1.1)	3.0 (1.4)	2.8 (1.3)
Exclusives	4.3 (1.1)	3.8 (1.2)	5.2 (1.0)	4.1 (1.1)
Self-references	7.9 (2.9)	8.4 (2.2)	6.4 (2.7)	6.9 (2.7)
Negative emotions	1.5 (0.8)	1.5 (0.8)	1.4 (0.7)	1.6 (0.8)

Note. Standard deviations are shown in parentheses.

LIWC, linguistic inquiry and word count.

(group) \times 2 (truth status) repeated measures ANOVAs were conducted. For cognitive complexity of stories, we carried out two ANOVAs: one for motion verbs and another for exclusive words. As to motion verbs, the main effect of group reached borderline significance, with high fantasy prones using fewer simple motion words than low fantasy prones ($F [1, 58] = 3.00, p = 0.08; \eta^2 p = 0.05$). However, the ANOVA revealed no significant difference between true and fabricated stories ($F [1, 58] < 1.0$). Neither was the interaction term significant ($F [1, 58] < 1.0$). Although Table 3 suggests that high fantasy prones' fabricated stories contained fewer simple motion words than low fantasy prones' true stories, this difference was not convincing ($t [58] = 1.23, p = 0.16$; Cohen's $d = 0.31$).

As for exclusive words, a main effect of group emerged ($F [1, 58] = 7.57, p < 0.01; \eta^2 p = 0.12$). Contrary to expectation, high fantasy prones employed significantly fewer exclusive words in their true and fabricated stories than low fantasy prones. There was also a main effect of truth status ($F [1, 58] = 15.19, p < 0.01; \eta^2 p = 0.21$), indicating that true and fabricated stories differed in the frequency of exclusive words. That is, fabricated stories contained less exclusives compared with true stories (see Table 3). The interaction effect of group by truth status did not attain significance ($F [1, 58] = 2.11, p = 0.15$). Most importantly, high fantasy prones' fabricated stories contained significantly fewer exclusive words, compared with true stories of low fantasy-prone participants ($t [58] = 4.91, p < 0.01$; Cohen's $d = 1.26$).

High fantasy prones employed significantly more self-references in their true and fabricated stories compared with low fantasy prones, as indicated by a main effect of group ($F [1, 58] = 9.22, p < 0.01; \eta^2 p = 0.14$). However, no significant main effect of truth status was found ($F [1, 58] < 1.0$), which indicates that overall, true and fabricated stories did not differ with regard to this LIWC parameter. As well, the interaction term failed to reach significance ($F [1, 58] < 1.0$). Yet, high fantasy prones' fabricated stories relied on more self-references than low fantasy prones' true stories ($t [58] = 3.14, p < 0.01$; Cohen's $d = 0.81$).

An ANOVA performed on negative emotion words did not reveal a significant main effect of group ($F [1, 58] < 1.0$), meaning that high and low fantasy prones did not differ in their use of negative emotion words. Neither was there a main effect of truth status ($F [1, 58] < 1.0$), implying that true stories did not differ from fabricated stories in the frequency of negative emotion words used. The interaction effect was also non-significant ($F [1, 58] < 1.0$).

None of the LIWC parameters correlated with age (all $r_s < 0.13$; all $p_s > 0.34$). Therefore, we did not conduct follow-up analyses with age as covariate.

DISCUSSION

In the legal context, assessment of veracity is often based on written transcripts of statements provided by eyewitnesses, victims, or suspects. One widely used tool for this type of assessment is the CBCA. Ruby and Brigham (1997, p. 723) argued that 'a person who is good at telling stories would likely to be judged by the CBCA as more truthful than someone who is not good at storytelling'. The current data as well as those of Merckelbach (2004) clearly demonstrate that Ruby and Brigham's concern is justified. In fact, effect sizes found in the current study suggest that fantasy proneness is a more important determinant of CBCA ratings than is truth status.

Why is it important to know that people high on fantasy proneness more often spontaneously use CBCA dimensions than people low in fantasy proneness? Although their prevalence may be low, false claims of sexual abuse and sexual harassments do exist (Hamilton, Feldman, & Cunnien, 2008). Typically, they arise against a background of child custody or divorce disputes. This is also the context in which the CBCA is often used as a practical tool in Germany and the Netherlands. As our study demonstrates that each CBCA dimension is sensitive to fantasy proneness, we would argue that forensic experts employing the CBCA should take into account the fantasy proneness levels of the individuals whose statements they are going to evaluate. We feel that there is now sufficient empirical justification to consider fantasy proneness a potential confounder of the CBCA.

Our study extends that of Merckelbach (2004) by showing that people high and low on fantasy proneness differ with regard to several LIWC parameters identified by Newman *et al.* (2003). Fabricated stories of high fantasy prones were found to be richer in self-references compared with true stories of low fantasy prones. This shows that high fantasy prones are superior in pretending to be personally involved in the stories that they tell. Also, relative to low fantasy prones, high fantasy prones employed *fewer* 'but's', 'however's', and other exclusive words when fabricating stories. Admittedly, this pattern conflicts with Newman *et al.*'s (2003) idea that exclusive words reflect narrative complexity, which in turn is an indication of truthfulness. However, it may well be the case that high fantasy prones intuitively understand that exclusive words may impress as powerless speech (Boccaccini, 2002), because they function to restrict what is claimed. Clearly, this issue warrants further study. In more general terms, our results suggest that it might be worthwhile to conduct a more systematic analysis of the links between fantasy proneness and Newman *et al.*'s (2003) LIWC parameters. The current study may have underestimated these links because it relied on a homogenous sample of intelligent participants who were instructed to write down stories of approximately 400 words. These features might have introduced a restriction of range on the LIWC parameters, thereby lowering the changes to detect robust associations amongst truth status, fantasy proneness, and LIWC characteristics.

Are the group differences in narrative features that we found driven by the low fantasy prones scoring low on fantasy proneness or the high fantasy prones scoring high on fantasy proneness? The mean CEQ score of our low fantasy-prone group was 3.4 (SD = 1.3), which is in the lower part of the CEQ distribution (Merckelbach *et al.*, 2001a). However, this CEQ score is above those of the two low fantasy-prone subsamples in the Merckelbach (2004) study (M = 2.71, SD = 1.20; and M = 2.27, SD = 0.83, respectively; both $t_s > 1.70$, both $ps > 0.09$). Therefore, our impression is that the group differences in narrative characteristics that we found are related to high fantasy proneness.

There are several limitations of the current study that merit comment. To begin with, we instructed our participants to write a true and a fabricated story, but we had no possibility to test the stories against the ground truth. Also, true and fabricated stories did not always concern high-stake situations. Research summarised by Vrij (2008) shows that it might be important to control for this feature. Thus, it is entirely possible that differences in narrative ability between high and low fantasy prones become particularly pronounced when they are instructed to fabricate stories about high-stake situations. Furthermore, the current study focused on two verbal assessment tools: the CBCA dimensions and the LIWC parameters. An alternative method is the reality monitoring approach. This approach heavily relies on memory research (Sporer, 1997) and has been shown to be a useful alternative to the CBCA (e.g. Vrij *et al.*, 2004). Future research on fantasy

prone as a confounder in verbal lie detection should take these limitations into account.

In sum, the current study found that stories of high fantasy prones are richer in CBCA dimensions, contain more self-references, and use less exclusive words than stories of low fantasy prones. This is also the case when one compares fabricated stories of high fantasy prones with authentic stories of low fantasy prones. Apparently, high fantasy prones intuitively understand how to tell stories that are compelling and persuasive. The obvious explanation for this talent is that they can draw on their imaginative abilities, which allow them to describe fantasies as real events (Wilson & Barber, 1983). The practical implication of this is *not* that statements of fantasy-prone people can never be trusted. Rather, our study suggests that it is important to measure this trait. Furthermore, a good strategy to tackle the confounding influence of fantasy proneness is to adopt Vrij's (2008) comparable truth technique. Basically, this technique compares the qualities of statements against a baseline measurement that consists of a known truthful response (i.e. a story about an event that is known to have happened). In this way, differences in fantasy proneness can be taken into account.

REFERENCES

- Bennett, W. L. (1992). Legal fiction: Telling stories and doing justice. In M. J. McLaughlin, S. J. Cody, & S. J. Read (Eds.), *Explaining one's self to others* (pp. 149–165). Hillsdale, NJ: Erlbaum.
- Blandon-Gitlin, I., Pezdek, K., Lindsay, D. S., & Hagen, L. (2009). Criteria-based content analysis of true and suggested accounts of events. *Applied Cognitive Psychology, 23*, 901–917.
- Boccaccini, M. T. (2002). What do we really know about witness preparation? *Behavioral Sciences and the Law, 20*, 161–189.
- Bond, G. D., & Lee, A. Y. (2005). Language of lies in prison: Linguistic classification of prisoners' truthful and deceptive natural language. *Applied Cognitive Psychology, 19*, 313–329.
- Burgoon, J. K., Buller, D. B., Guerrero, L. K., Afifi, W. A., & Feldman, C. M. (1996). Interpersonal deception. XII. Information management dimensions underlying deceptive and truthful messages. *Communication Monographs, 63*, 50–69.
- DePaulo, B. M., Lindsay, J. L., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*, 741–758.
- Ebesu, A. S., & Miller, M. D. (1994). Verbal and nonverbal behaviors as a function of deception type. *Journal of Language and Social Psychology, 13*, 418–442.
- Hamilton, J. C., Feldman, M. D., & Cunnien, A. J. (2008). Factitious disorder in medical and psychiatric practices. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 128–144). New York: Guilford.
- Hancock, J. T., Curry, L. E., Goorha, S., & Woodworth, M. (2008). On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes, 45*, 1–23.
- Merckelbach, H. (2004). Telling a good story: Fantasy proneness and the quality of fabricating memories. *Personality and Individual Differences, 37*, 1371–1382.
- Merckelbach, H., Horselenberg, R., & Muris, P. (2001a). The Creative Experiences Questionnaire (CEQ): A brief self-report measure of fantasy proneness. *Personality and Individual Differences, 31*, 987–995.
- Merckelbach, H., Wiers, R., Horselenberg, R., & Wessel, I. (2001b). Effects of retrieving childhood events on metamemory depend on the questions you ask. *British Journal of Clinical Psychology, 40*, 215–240.
- Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin, 29*, 665–675.
- Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count (LIWC): LIWC 2001 manual*. Mahwah, NJ: Erlbaum.

- Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, *77*, 1296–1312.
- Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology*, *51*, 651–672.
- Ruby, C. L., & Brigham, J. C. (1997). The usefulness of the criteria-based content analysis technique in distinguishing between truthful and fabricated allegations. *Psychology, Public Policy, and Law*, *3*, 705–737.
- Sánchez-Bernardos, M. L., & Avia, M. D. (2004). Personality correlates of fantasy proneness among adolescents. *Personality and Individual Differences*, *37*, 1069–1079.
- Sánchez-Bernardos, M. L., & Avia, M. D. (2006). The relationship between fantasy proneness and schizotypy in adolescents. *Journal of Nervous and Mental Disease*, *194*, 411–414.
- Sporer, S. L. (1997). The less travelled road to truth: Verbal cues in deception detection in accounts of fabricated and self-experienced events. *Applied Cognitive Psychology*, *11*, 373–397.
- Steller, M., & Köhnken, G. (1989). Criteria-based statement analysis. In D. C. Raskin (Ed.), *Psychological methods in criminal investigations and evidence* (pp. 217–245). New York: Springer.
- Stirman, S. W., & Pennebaker, J. W. (2001). Word use in poetry of suicidal and nonsuicidal poets. *Psychosomatic Medicine*, *73*, 517–522.
- Undeutsch, U. (1989). The development of the statement reality analysis. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 101–119). Dordrecht, The Netherlands: Kluwer.
- Vrij, A. (2005). Criteria-based content analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy and Law*, *11*, 3–41.
- Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities* (2nd ed.). Chichester, UK: John Wiley.
- Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2004). Let me inform you how to tell a convincing story: CBCA and reality monitoring scores as a function of age, coaching, and deception. *Canadian Journal of Behavioural Science*, *36*, 113–126.
- Vrij, A., Mann, S., Kristen, S., & Fisher, R. P. (2007). Cues to deception and ability to detect lies as a function of police interview styles. *Law and Human Behavior*, *31*, 499–518.
- Wilson, S. C., & Barber, T. X. (1983). Fantasy prone personality: Implications for understanding imagery, hypnosis, and parapsychological phenomena. In A. A. Sheikh (Ed.), *Imagery: Current research and application* (pp. 340–387). New York: Wiley.
- Wright, D. B., & Loftus, E. F. (1999). Measuring dissociation: Comparison of alternative forms of the dissociative experiences scale. *American Journal of Psychology*, *112*, 497–519.
- Zhou, L., Burgoon, J. K., Nunamaker, J. F., & Twitchell, D. (2004). Automating linguistics-based cues for detecting deception in asynchronous computer-mediated communications. *Group Decision and Negotiation*, *13*, 81–106.
- Zijlstra, H. M., Middendorp, H., Meerveld, T., & Geenen, R. (2005). Validiteit van de Nederlandse versie van de [Validity of the Dutch version of] Linguistic Inquiry and Word Count (LIWC). *Nederlands Tijdschrift voor de Psychologie [Dutch Journal of Psychology]*, *60*, 55–63.

APPENDIX A. FRAGMENTS FROM TRUE STORY OF LOW FANTASY-PRONE AND FABRICATED STORY OF HIGH FANTASY-PRONE PARTICIPANT³

True story (first five sentences)

When I was about 21 years old, I had a girlfriend for already 3 years. The relationship went well for a long time and we had a lot of fun together. When we were together for about a year, I think it actually was 1.5 years, we broke up for a while, because she wasn't sure if she wanted to continue the relationship. After about 3 months, we tried to get back

³For illustrative purposes, we show fragments of stories with the highest CBCA ratings. Translations of the complete stories can be obtained from the first author.

together again and that actually went pretty good. Until the day she announced some bad news.

(CBCA rating and number of exclusives and self-references per 100 words: 20.5, 6.5, and 8.9, respectively)

Fabricated story (first five sentences)

When I was 19, I had a relationship with a boy that was in my class. In the beginning, it was really great. He could and was willing to help me with a lot of my school courses. I had a bit of troubles with language courses, while he really had a talent for languages. About four days each week, he was at my place to make homework together and on top of that we saw each other every weekend while going out at night.

(CBCA rating and number of exclusives and self-references per 100 words: 23.5, 5.5, and 7.2, respectively)