



Offenders' uncoerced false confessions: A new application of statement analysis?

Rebecca M. Willén* and Leif A. Strömwall

University of Gothenburg, Sweden

Purpose. Deception detection research has mainly studied denials and distortions given by students. This study examined true and false confessions as told by offenders. It was hypothesized that the statement analytic techniques Criteria-Based Content Analysis (CBCA) and Reality Monitoring (RM) would discriminate truths and lies.

Methods. Truthful and deceptive confessions to crime were given by 30 offenders (both women and men) in a within-subject design. The participants were in prison at the time of data collection, and told the truth about a crime they had committed and been sentenced for. In addition, they made up a lie about a different crime after a few minutes of preparation. The transcribed statements were scored for CBCA and RM criteria.

Results. Results showed that neither total CBCA nor total RM scores differentiated between lies and truths. Some individual CBCA criteria, however, showed differences: more *self-deprecations* and *doubts about own testimony* in the told lies, and more *unexpected complications* in the truths.

Conclusions. The results are discussed in relation to statement analysis of offenders' accounts, individual CBCA criteria, as well as the development of criminal experience and familiarity with the event and setting. Implications for triers of fact and suggestions for future research are considered.

In forensic interview contexts, it is crucial to detect deception in order to find the perpetrator or to narrow down the number of suspects. A meta-analysis by Bond and DePaulo (2006) showed that people typically reach an overall accuracy of about 54% when not using specialized techniques for detecting deception. This was also the case for supposed lie detection experts such as police detectives. Hence, there is a need for some kind of tool to improve the correctness of veracity judgements. One such tool is statement analysis in which the content of a person's statement is analysed, for example, according to certain criteria (Vrij, 2008). Research on statement analysis has focused on two criteria-based techniques: Criteria-Based Content Analysis (CBCA) and Reality Monitoring (RM). Previous research on statement analysis has mainly concerned

Parts of this research were presented at the 20th conference of the European Association of Psychology and Law, Gothenburg, Sweden, June 2010.

*Correspondence should be addressed to Rebecca Willén, Department of Psychology, University of Gothenburg, PO Box 500, SE 405 30 Gothenburg, Sweden (e-mail: Rebecca.Willen@psy.gu.se).

accounts by witnesses and plaintiffs (Vrij, 2008). The possibilities of detecting false confessions through the use of statement analysis have so far not been investigated. The aim of this study was to examine if CBCA and/or RM can be useful in differentiating offenders' true and false confessions.

Criteria-based content analysis and reality monitoring

Statement Validity Assessment (SVA) is a comprehensive tool for analysing children's statements in cases concerning sexual abuse. It was originally developed in Germany in the 1950s and has been used in Germany ever since (Köhnken, 2004). CBCA is a core feature in this technique. The foundation of CBCA is the so-called Undeutsch hypothesis: there are qualitative and quantitative differences between true and deceptive accounts (Undeutsch, 1982). CBCA consists of 19 criteria; all are supposed to occur more frequently in truthful than in deceptive statements. These criteria are supposed to occur due to either motivational or cognitive reasons. Liars are assumed to be more concerned about impression management than truth tellers. Therefore, they would not undermine the account by including a certain type of details (motivation); for instance, liars would not admit *lack of memory* since it could undermine their credibility. Other criteria are supposed to not be used by liars because the criteria are too cognitively difficult to fabricate (cognition); for example, reports of *unexpected complications* (Köhnken, 2004). The cognitive criteria in general tend to receive more support than the motivational criteria (Vrij, 2008).

Quantity of details is a criterion that has received extensive empirical support; truthful accounts are usually richer in detail than deceptive accounts (Vrij, 2008). Three more criteria have received strong support: *unstructured production*, *contextual embeddings*, and *reproduction of conversations*. These criteria are all more often found in truths than in lies (Vrij, 2008).

Self-deprecations belong to the motivational cluster of criteria, and this is the only CBCA criterion that has still not received any support at all. In the few studies in which a difference has been found, it has been in the direction not predicted; *self-deprecations* occurred more frequently in lies (Vrij, 2008). Quandt (2010) suggests that *self-deprecations* is susceptible to which type of event the statement concerns: individuals lying about forensic events want to avoid *self-deprecations* while deceptive accounts about non-forensic events tend to include *self-deprecations* (see also Niehaus, Krause, & Schmidke, 2005).

CBCA was originally developed specifically to analyse children's accounts in sexual abuse cases, however, some researchers have suggested that it might be useful also when it concerns adults and other types of crime (see for instance, Porter & Yuille, 1996; Steller & Köhnken, 1989). Porter and Yuille (1995) argued that the Undeutsch hypothesis (as well as RM theory that is described below) regarding differences in quantity and quality between truths and lies should be valid also in accounts by suspects. In line with this reasoning, Yuille and Cutshall (1989) rated the occurrence of some CBCA criteria in a confession of a real-life suspect (a breaking-and-entering during which three people were killed). The outcome of this analysis showed that the criteria could be used to correctly identify the confession as true. Yuille and Cutshall therefore expressed optimism regarding using CBCA, and statement analysis in general, on suspects' accounts and they pushed for more such research. However, such research has been very rare.

RM is, unlike CBCA, originally based on psychological theory and was first introduced by Johnson and Raye (1981). They suggested that there are two types of memory sources:

internal and external. When imagining an event, one would have to internally create a picture of something not experienced. This internally created picture would then be the memory source. A memory of something actually experienced, on the other hand, would be based on an external source: the memory would have been derived through the five senses. According to RM theory, statements containing real memories should therefore among other things include more *spatial, temporal, and perceptual details*. Imagined events, on the other hand, would contain more details related to cognitive processes (e.g., ‘I suppose it was early in the morning since my colleagues had not yet arrived’). Johnson and Raye called this type of information *cognitive operations*. RM was initially introduced as a model for people judging their own memories as real or false. Alonso-Quecuty (1992) suggested that RM might also be useful as a tool for lie detection. Sporer (1997, 2004) continued the development of RM as a deception detection tool.

There are still few published studies examining RM. However, *visual, auditory, temporal and spatial details*, and *realism* have all received support; they are found more often in truthful than in deceptive statements (Masip, Sporer, Garrido, & Herrero, 2005). *Cognitive operations* is the only lie criterion; it is supposed to occur more frequently in lies than in truths. The criterion has received weak support (Masip *et al.*, 2005). Vrij (2008) suggests that the reason for this might be that it has been operationalized differently by different researchers.

Both RM (Masip *et al.*, 2005) and CBCA (Vrij, 2008) seem to have stronger discriminative potential when used as a whole set of criteria compared to the discriminative power of single criteria (in the present study we will examine both). The discriminative ability is, according to Vrij (2008), similar for both techniques: an overall accuracy rate of about 70%. RM seems to be a somewhat stronger tool than CBCA when they are both used on the same statements (e.g., Strömwall, Bengtsson, Leander, & Granhag, 2004). Compared with CBCA, RM has the advantage of being easier to use, easier to teach and rests on solid theory. RM might come to be a useful tool for practitioners in the future, but more research is needed (Masip *et al.*, 2005; Vrij, 2008).

Statement analysis of suspects’ statements

Research on statement analysis has examined the two techniques applied to statements by witnesses and plaintiffs. Only a few studies have tested CBCA or RM on suspects’ accounts. ‘Suspects’ in such research have usually been undergraduates acting as suspects of a mock crime, or suspects of a non-criminal but not allowed act. The present study recruited participants with authentic criminal experience.

The general finding for CBCA on suspects’ accounts are results in line with previous research on accounts by witnesses and plaintiffs; for instance higher total CBCA scores in truthful than in deceptive statements (Caso, Vrij, Mann, & De Leo, 2006; Gödert, Gamer, Rill, & Vossel, 2005; Vrij, Akehurst, Soukara, & Bull, 2004a; Vrij, Akehurst, Soukara, & Bull, 2004b; Vrij & Mann, 2006).

Contrary results were found in the study by Caso *et al.* (2006): there were more *conversations* and more *accounts of subjective mental state* in lies than in truths. Lee, Klaver, and Hart (2008) also found a result contrary to the prediction: *spontaneous corrections* occurred more often in lies than in truths.

Even fewer studies have examined RM on statements by suspects. The general finding was higher total RM scores in truths than in lies (Vrij *et al.*, 2004a; Vrij *et al.*, 2004b; Vrij, Mann, Kristen, & Fisher, 2007). However, Bond and Lee (2005) found that deceptive

statements contained more *spatial details* than true statements, a finding that is contrary to what is expected according to RM theory.

Two of the above-mentioned studies (Bond & Lee, 2005; Lee *et al.*, 2008) have employed offenders as participants. As already mentioned, both reported some results in the non-predicted direction. Porter and ten Brinke (2010) pointed out the need for more research on CBCA with experienced and skilled deceivers, such as criminal offenders. The current study employs offenders as participants and as in the study by Lee *et al.* (2008), they provide statements about crime. This study is the first to use both CBCA and RM on offenders' truthful and false confessions about committed crimes. A false confession (in this study) is an account in which the participant claims to have committed a crime he or she in fact has not committed at all.

This study

The purpose of this study was to examine if CBCA and/or RM would have potential in distinguishing offenders' true and false confessions. In line with Undeutsch's hypothesis (1982) and RM theory (Johnson & Raye, 1981) as well as recent empirical studies (as reviewed in Vrij, 2008), it was hypothesized that there would be higher CBCA scores (Hypothesis 1) and higher RM scores (Hypothesis 2) in truthful confessions than in false.

Method

Participants

Thirty offenders serving time in one of three Swedish prisons participated: 9 women ($M = 39.33$ years, $SD = 10.45$) and 21 men ($M = 32.00$ years, $SD = 8.00$). All prisons were classified as low-security prisons. The participants' experience of police interviews ranged from 4 to over 200, with an average (median) of having been in 23 police interviews ($M = 47.97$, $SD = 56.49$). Time served in prison ranged from 1 month to 15 years ($Md = 22$ months, $M = 42.38$, $SD = 47.60$). The men had a criminal history mostly including violence, drug offences, and theft. The typical crime experiences of the women were drug offences, theft, and fraud. Participation was voluntary and each participant received compensation (equivalent to just over 6 EUR). The participants were informed both verbally and in writing about anonymity and ethics.

Questionnaire

The participants answered a short questionnaire in which questions about their interview experience were answered by specifying the number of interviews experienced. An open question about crime experience allowed the participants to describe in their own words what types of crime they had committed (during their lifetime). There were also questions about gender, age, and time served in prison.

Procedure

During a first meeting, the participants were informed (verbally as well as in writing) about the study and its procedure. Participants were instructed to tell the truth about a crime s/he had committed and had been convicted for (i.e., give a truthful confession) in one interview, and in another interview tell a lie about a crime that s/he had never

committed (i.e., falsely confess). The truthful confession was not confined to the reason for the present imprisonment. It was emphasized that the truthful statement must be all true (i.e., not contain distortions of any kind) and that the false statement must be all false (i.e., not contain any true parts).

Each participant was interviewed twice and each interview lasted about 5–10 min. The order of truthful and deceptive interviews was counterbalanced. Before each interview, the participant received instructions from a research assistant who told the participant whether s/he was going to lie or tell the truth in the upcoming interview. The participants did not know the topic of their deceptive statement until they were given it from the assistant who selected it from a long pre-prepared list of crimes. The selection of topics was carried out with regard to what type of crime the participant chose to tell the truth about (e.g., due to ethical reasons participants were never given a violent lie scenario if their truth did not involve any violence at all). In addition, it was regarded whether the topic of the truthful account concerned one specific event or a series of events (e.g., when the truth concerned an isolated incident the participant would not receive a lie scenario that concerned a series of events). The participants received 5–10 min to plan their lie.

Each interview started with the interviewer (always the same person) pointing out that she did not know if the participant was going to lie or tell the truth in that particular interview. The participant was reminded to give each statement in a convincing manner irrespective of the objective veracity status. After a free recall, the interviewer used at least two and never more than five (depending on how much was said in the free recall) open questions (e.g., ‘Do you remember anything more?’) in order to elicit a complete statement. All interviews were audiotaped. These 60 interviews have also served as the subject of other analyses in a different study (Willén & Strömwall, in press).

Ground truth

Ground truth (i.e., objective truth status) was not objectively established in the present study. However, participants received strict information to give one truthful and one false statement. At least 12 times did participants reject the first lie scenario handed out to him or her due to our instruction to not accept lie scenarios that were close to an event they had experienced in real life. That is, many participants acted upon our instruction and consequently to some extent provided a ground truth.

Our conviction that the participants did follow the instructions is strengthened by the fact that judicial consequences (e.g., detention, trial, and sentence) were described in 18 truthful accounts (out of 30), as compared to only in six of the deceptive accounts. Such consequences were in all six cases described in both the participant’s statements. Hence, in no case did a participant mention judicial consequences in his/her deceptive statement only.

Coding procedure

The interviews were transcribed and anonymized. Judicial consequences of the crime were italicized prior coding and not considered in the analyses. A 5-point rating scale was used, as suggested by Köhnken (2004), for the CBCA and RM ratings, in order to use a more sensitive measure than the standard 0-1-2 scoring. Zero indicated that the criterion was not at all present or fulfilled, and 4 indicated that the criterion was present or fulfilled to a very high degree.

Table 1. Inter-rater reliability for CBCA and RM criteria

CBCA criterion	κ_w	r	RM criterion	κ_w	r
1. Logical structure	.77	.80**	Visual details	.73	.73**
2. Unstructured production	.91	.92**	Audio details	.77	.78**
3. Quantity of details	.77	.81**	Smell	1.00	1.00**
4. Contextual embeddings	.92	.94**	Taste ^a	-	-
5. Interactions	.84	.86**	Physical sensations	1.00	1.00**
6. Conversations	1.00	1.00**	Affective details	.62	.67**
7. Unexpected complications	.75	.77**	Spatial details	.54	.56*
8. Unusual details	1.00	1.00**	Temporal details	.69	.76**
9. Superfluous details	.55	.64*	Cognitive operations	.61	.65**
10. Details misunderstood ^a	-	-	Clarity	.65	.66**
11. External associations	.77	.77**	Reconstruction	.43	.52*
12. Subjective mental state	.96	.96**	Realism	.56	.61*
13. Other's mental state	.59	.61*			
14. Spontaneous corrections	.89	.90**			
15. Lack of memory	.83	.83**			
16. Doubts about testimony	.71	.74**			
17. Self-deprecations	.94	.94**			
18. Pardoning the other ^a	-	-			
19. Crime-specific details	.66	.69**			

Note. κ_w = weighted kappa; r = Pearson correlations. ^aThe criterion did not occur in any of the 60 statements. * $p < .05$; ** $p < .01$.

Training

We created a manual for each technique (CBCA and RM) in which each criterion was described and supplemented with examples. We trained two coders in CBCA and two other coders in RM. All coders were psychology students without criminal and/or working experience relevant to criminality or to the coding procedure in this study. The training included core readings, followed by ratings of several example statements (carried out by the coders individually). None of the example statements originated from the 30 offenders included in this study, and were not part of the upcoming reliability assessment. The coders had discussions with the experiment leader at several training sessions that lasted about 6 h in total (for each technique). The coders were blind to the actual veracity of the training statements.

Inter-rater reliability

Fifteen statements were randomly selected for the reliability assessment that was checked with both Cohen's weighted kappa and Pearson correlations (see Table 1 for inter-rater data). The obtained overall agreement was excellent (according to Gödert *et al.*, 2005, coefficients above .75 is excellent) for CBCA; $r = .91$, $p < .01$ ($\kappa_w = .91$), as well as for RM; $r = .90$, $p < .01$ ($\kappa_w = .90$). Once acceptable reliability had been obtained, one of the CBCA coders and one of the RM coders rated the remaining statements according to the procedure agreed upon.

Data preparation and initial analyses

Two CBCA criteria were never present: *details misunderstood* and *pardoning the other*. Neither was the RM criterion *taste* ever present. Furthermore, the RM criteria *smell* and *physical sensations* only occurred once each. These five criteria were therefore excluded from all further analyses. The *cognitive operations* scores were reversed in all analyses (i.e., high scores indicates truth).

Dependent variables

Total CBCA and total RM scores were calculated for truths and lies separately, resulting in two sum scores for each participant since we employed a within-subjects design. These scores were used to test the techniques' overall ability to discriminate truths and lies. The scores for the individual CBCA and RM criteria were used as dependent variables in subsequent analyses aimed at examining the differentiating power at the criteria-specific level.

Covariates

Initial analyses indicated that both Interview experience and Gender correlated with total CBCA and total RM scores (*rs* ranging from .09 to .53). Since we had not made any predictions about these variables but wanted as precise error terms as possible, we included Gender and Interview experience as covariates in the analyses.¹

Results**Order effects**

To find out if the order in which the participants had told their truth and lie had any effects on the scores, we conducted two MANOVAs separately for CBCA and RM criteria, with Order and Veracity as independent variables. On the multivariate level there were no significant effects for CBCA, $F(17,12) = 0.93$, $p = .57$, $\eta_p^2 = .57$, or for RM, $F(9,20) = 1.02$, $p = .46$, $\eta_p^2 = .31$. None of the multivariate interactions were significant, both $ps > .06$. Hence, there were no differences in CBCA or RM scores that could be attributed to the order in which the participants told their statements.

CBCA analyses**Total scores**

In order to test if the whole set of CBCA criteria could distinguish truths and lies, a repeated measures ANCOVA was conducted. The total truth and lie scores were the repeated factor; Gender and Interview experience the covariates. The test was not significant, $F(1,27) = 1.58$, $p = .22$, $\eta_p^2 = .06$. Hence, Hypothesis 1 received no support.

¹Running the analyses for total scores without the covariates included showed the same overall results: no significant differences between total truth and total lie scores for either CBCA, $F(1, 29) = 1.50$, $p = .23$, $\eta^2 = .05$, or RM, $F(1, 29) = 0.06$, $p = .81$, $\eta^2 = .002$.

Table 2. CBCA scores as a function of veracity

Criterion	Truth		Lie	
	M	SD	M	SD
1. Logical structure	2.70	0.65	2.93	0.45
2. Unstructured production	2.57	0.68	2.47	0.82
3. Quantity of details	2.57	0.68	2.83	0.53
4. Contextual embeddings	0.77	0.90	1.20	0.89
5. Interactions	1.23	1.17	1.03	1.07
6. Conversations	0.77	1.25	0.63	1.25
7. Unexpected complications	0.80	1.03	0.30	0.65
8. Unusual details	0.10	0.40	0.13	0.35
9. Superfluous details	1.20	1.06	1.17	1.09
11. External associations	0.47	0.86	0.37	0.61
12. Subjective mental state	1.10	1.19	1.37	1.19
13. Other's mental state	0.40	0.68	0.30	0.53
14. Spontaneous corrections	0.27	0.58	0.37	0.81
15. Lack of memory	0.57	0.77	0.47	0.90
16. Doubts about testimony	0.20	0.48	0.50	0.86
17. Self-deprecations	0.10	0.40	0.70	0.92
19. Crime-specific details	1.03	1.25	1.20	1.13
Total CBCA score ^a	16.83	6.01	17.97	5.53

Note. The *M*'s are corrected for the influence of the covariates. ^aTotal scores could range from 0 to 68.

Individual criteria

A repeated measures ANCOVA was conducted with Veracity as the repeated factor, the 17 individual CBCA criteria as dependent variables, and Gender and Interview experience as covariates. A significant multivariate effect of Veracity was found when Gender and Interview experience were controlled for, $F(17,11) = 2.81, p < .05, \eta_p^2 = .81$. This effect indicates that for at least some individual criteria, differences between truths and lies emerged. Pairwise Bonferroni-corrected comparisons showed significant differences for *unexpected complications*, $p < .05, d = 0.44$, *doubts about own testimony*, $p < .05, d = -0.40$, and *self-deprecations*, $p < .001, d = -0.78$. As can be seen in Table 2, there were more *unexpected complications* in truths than in lies, but more *doubts about own testimony* and *self-deprecations* in lies than in truths. In fact, only two participants were found to have more *self-deprecations* in truthful accounts (than in deceptive accounts), while the opposite was found for 14 participants.

RM analyses

Total scores

In order to test if the whole set of RM criteria could distinguish truths and lies, we conducted a repeated measures ANCOVA. The total truth and lie scores were the repeated factor; Gender and Interview experience were the covariates. The test was not significant, $F(1,27) = 0.06, p = .80, \eta_p^2 = .002$. Hypothesis 2 was thus not supported.

Table 3. RM scores as a function of veracity

Criterion	Truth		Lie	
	M	SD	M	SD
Visual details	2.50	0.68	2.47	0.63
Audio details	0.90	1.16	0.77	0.90
Affective details	0.47	0.51	0.60	0.50
Spatial details	2.00	0.79	1.90	0.80
Temporal details	1.83	0.70	2.00	0.64
Cognitive operations ^a	2.27	0.91	2.30	0.92
Clarity	2.20	0.66	2.27	0.69
Reconstruction	2.50	0.73	2.60	0.62
Realism	2.50	0.73	2.43	0.73
Total RM score ^b	17.17	4.21	17.33	3.36

Note. The *M*'s are corrected for the influence of the covariates. ^aThis is reversed; high scores indicate truth. ^bTotal scores could range from 0 to 36.

Individual criteria

A repeated measures ANCOVA was conducted with Veracity as the repeated factor, the 9 RM criteria as dependent variables, and Gender and Interview experience as covariates. There was no significant multivariate effect of Veracity, $F(9,19) = 0.64$, $p = .75$, $\eta_p^2 = .23$, or any significant effects at the individual criteria level. See Table 3 for descriptive statistics.

Discussion

The main purpose of the present study was to examine if CBCA and/or RM could be useful in discriminating offenders' true and false confessions. Previous research has not, with only a few exceptions, explored the applicability of these techniques on statements by offenders.

CBCA and RM

It was hypothesized that there would be higher CBCA and RM scores in truthful statements than in false. None of the hypotheses were supported. Neither CBCA nor RM distinguished between truthful and false statements when using the total scores. Previous studies in which the techniques have been compared have shown RM to be a somewhat better technique (Strömwall *et al.*, 2004; Vrij, 2008). In this study we have to conclude that neither of them was useful in assessing veracity, although some individual CBCA criteria did differentiate between the true and false confessions (these criteria are further discussed below). However, our findings do not dismiss SVA/CBCA and/or RM as techniques for assessing veracity. In a situation such as the one examined in the present study - criminally experienced individuals providing true and false confessions - the techniques did not function as well as research has shown them to do for other situations. In our view, the overall finding points to the need of the development of scientifically based assessment techniques designed specifically for criminally experienced suspects' accounts, unlike the techniques tested in the present study.

Our unpredicted results can be interpreted as in line with previous research on the effect of familiarity. In two studies, Pezdek *et al.* (2004) and Blandon-Gitlin, Pezdek, Rogers, and Brodie (2005) have shown how CBCA scores are affected by the level of familiarity the participants (children) had with the event they were interviewed about. The explanation given is that familiar events are likely to contain more information and have a more coherent structure than descriptions of unfamiliar events (Blandon-Gitlin *et al.*, 2005). In the present study, most participants had committed more than one crime during their lifetime, were more or less familiar with the interrogation setting and all reported to have lied in previous police interviews. Arguably, the level of familiarity might have affected many of the participants' deceptive accounts to such an extent that neither CBCA nor RM could separate truths from lies. Further research could examine the effects of event familiarity for criminally experienced individuals.

Another possible explanation to why neither tools functioned as we expected may be due to a memory effect: the participants were recounting truthful events that had taken place at least several months and sometimes years before the interview. The false events, on the other hand, were invented just prior to the interviews. Previous CBCA and RM research seem to not have tested truthful accounts given such a long time after the to-be-remembered event (Masip *et al.*, 2005; Vrij, 2008). However, research has shown that prolonged retention intervals in general have a negative effect on memory recall (Wixted & Ebbesen, 1991). It is therefore possible that long retention intervals made some of the truthful accounts in this study contain less CBCA and RM information.

The results of this study were not as predicted, but perhaps not surprising. For reasons outlined above (event familiarity and memory effects) as well as the fact that neither CBCA nor RM is specifically developed to detect offenders' truths and lies, the findings are not astounding. However, only by testing the techniques it is possible to conclude that neither of them appears to be suited for offenders' statements.

Despite the nil results for the total scores, some differences found at the individual criteria level emanated. Three CBCA criteria distinguished truthful and false confessions: *self-deprecations*, *doubts about own testimony*, and *unexpected complications*. The two first-mentioned criteria are motivational based and occurred more frequently in deceptive accounts whereas the latter criterion belongs to the cognitive cluster and occurred more often in truthful accounts. These findings are somewhat in line with previous research since the motivational criteria have been shown to be less useful as truth indicators than the cognitive criteria (Vrij, 2008).

Almost half of all the false confessions in our study contained *self-deprecations* to some degree. In line with our finding, Ruby and Brigham (1998) found more *self-deprecations* in lies than in truths. Perhaps more interesting though is a finding by Appleby, Hasel, and Kassin (2011) who analysed the content of 20 proven (real life) false confessions. They found that eight of the suspects expressed remorse in their false confession, and five of them apologized. Translated into suspects' confessions (from plaintiffs' statements), remorse and apologies can be seen as related to *self-deprecations* since it in both cases concerns admitting one's own wrongdoing and/or taking at least some responsibility for the crime. Appleby *et al.* (2011) did not include a comparison group (i.e., truthful confessions), however, their results are in line with ours regarding this specific content and we did have a comparison group. *Self-deprecation* has still not received any support as truth verifying (Vrij, 2008). Perhaps future research will show that it in some contexts can be more useful as a lie criterion.

In the review of all published CBCA studies by Vrij (2008), *doubts about own testimony* were never found to be more salient in lies than in truths. However, the

criterion has in previous research not received strong support in the predicted direction either. Our results further question the validity of this particular criterion. The present study is arguably the first in which more *doubts about own testimony* have been found in lies than in truths. One possible explanation for this is that we employed unusual participants. The offenders may have realized in previous police interviews that raising doubts about their own memory may be a useful tactic, if used with finesse, to come out as convincing.

Unexpected complications distinguished true and false statements in the predicted direction. Lee *et al.* (2008) made the same finding when analysing offenders' accounts. In the review by Vrij (2008) however, *unexpected complications* received support only in five of 22 studies (the study by Lee *et al.* was not included). Further research is needed to find out if offenders include *unexpected complications* more often than other participants, and if the criterion has potential as a discriminating criterion in a reversed manner (i.e., more often found in false accounts).

One frequently used criterion for credibility assessment in courts (May & Wierda, 2002; Schelin, 2007) and asylum procedures (Kagan, 2003) is *quantity of details*. In a study by Granhag, Andersson, Strömwall, and Hartwig (2004), offenders reported to believe that deceptive statements are generally more detailed than truthful ones. In line with this, several participants in the present study told the experiment leader that one has to give a very detailed statement in order to create a truthful impression (these data are reported elsewhere, Strömwall & Willén, 2011). *Quantity of details* is a criterion with strong empirical support (Vrij, 2008), and as mentioned above, it is widely used in practise. Despite this, *quantity of details* was not a useful criterion in separating true and false confessions in our study. Our null finding for this criterion is somewhat supported by another finding in the study by Appleby *et al.* (2011): the real-life false confessions they analysed were indeed very detailed. Because of the wide use of the criterion among judicial decision-makers, it is crucial to explore it further in statements told by criminally experienced individuals.

The offenders in the present study told so convincing lies that neither CBCA nor RM could pinpoint them. In the course of the development of their criminal lifestyle, offenders might have gained experience in telling credible lies. This is in line with what Vrij and Semin (1996) and Granhag *et al.* (2004) found when investigating prisoners' beliefs about cues to deception. The prisoners were found to hold more correct beliefs than for instance prison personnel, students, and police detectives. It is likely that the offenders, in line with their more correct beliefs, had developed a kind of expertise in telling convincing lies in police interviews.

Limitations

Ground truth was not objectively established in the present study, although participants received strict information and to some extent themselves provided a ground truth. As Vrij (2008) has pointed out, establishment of ground truth is essential when carrying out deception research. One possibility could have been to have asked the participants for permission to compare their truthful statements with protocols from the court. However, this would have required interferences with the anonymity promised and that was considered to decrease the number of offenders willing to participate. Furthermore, a possible consequence of such a procedure is that the participants would not tell the truth as it happened if they had given a sweetened version of the crime in the court. They may then stick to the sweetened version, which would result in 'truths' that were

not all true. The lack of ground truth is nonetheless a limitation and in future research of this kind one should give priority to the establishment of ground truth.

In addition, there was a small sample size in this study since it was not easy to gain access to criminally experienced persons willing to participate. In spite of the small sample, we argue that this study contributes more to our understanding of how lies and truths in criminal cases are told, and therefore can be detected, than yet another study using non-criminally experienced undergraduate students.

Implications and future directions

Statement analysis is presently used foremost to assess reliability of plaintiffs' statements (Köhnken, 2004). There is a need for an empirically based and scientifically evaluated technique for analysing suspects' accounts. This would further strengthen the legal security since in difficult cases (e.g., sexual abuse or domestic violence) statements from both parties could be properly analysed and assessed.

However, the status of the interviewee (suspect/witness/plaintiff) may sometimes be of less importance when trying to detect deceit than criminal experience. For instance, individuals involved in criminal activity are themselves in high risk of becoming crime victims and witnesses (Rock, 2007). That is, plaintiffs and witnesses can be at least as criminally experienced as suspects.

This study suggests that analysis of accounts given by offenders may give rise to different results than analysis of statements by non-offenders. More research employing criminally experienced plaintiffs and witnesses, as well as suspects, is certainly needed.

We found that false confessions made voluntary by offenders were difficult to detect, but also that there were differences in quality and quantity on some individual criteria that did distinguish the truthful and deceptive statements. Hence, our results suggest that further development and research could make it possible to differentiate offenders' true and false confessions through a scientifically based technique for statement analysis.

Acknowledgements

The authors would like to thank David Mendes-Lopes, Erik Nilsson, Henrik Lind, and Robin Mörck for coding the data; and Maria Bing for assistance during the data collection. The offenders who participated in the study, and the prison management and personnel are especially acknowledged for great collaboration. The authors thank two anonymous reviewers for helpful comments and suggestions.

References

- Alonso-Quecuty, M. L. (1992). Deception detection and reality monitoring: A new answer to an old question? In F. Lösel, D. Bender, & T. Bliesener (Eds.), *Psychology and law: International perspectives* (pp. 328–332). Berlin: Walter de Gruyter.
- Appleby, S. C., Hasel, L. E., & Kassin, S. M. (2011). *Police-induced confessions: An empirical analysis of their content and impact*. Manuscript submitted for publication.
- Blandon-Gitlin, I., Pezdek, K., Rogers, M., & Brodie, L. (2005). Detecting deception in children: An experimental study of the effect of event familiarity on CBCA ratings. *Law and Human Behavior, 29*, 187–197.

- Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgements. *Personality and Social Psychology Review, 10*, 214–234.
- 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.
- Caso, L., Vrij, A., Mann, S., & De Leo, G. (2006). Deceptive responses: The impact of verbal and non-verbal countermeasures. *Legal and Criminological Psychology, 11*, 99–111.
- Granhag, P. A., Andersson, L. O., Strömwall, L. A., & Hartwig, M. (2004). Imprisoned knowledge: Criminals' beliefs about deception. *Legal and Criminological Psychology, 9*, 103–119.
- Gödert, H. W., Gamer, M., Rill, H-G., & Vossel, G. (2005). Statement validity assessment: Inter-rater reliability of criteria-based content analysis in the mock-crime paradigm. *Legal and Criminological Psychology, 10*, 225–245.
- Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*, 67–85.
- Kagan, M. (2003). Is truth in the eye of the beholder? Objective credibility assessment in refugee status determination. *Georgetown Immigration Law Journal, 17*, 367–381.
- Köhnken, G. (2004). Statement validity analysis and the 'detection of the truth'. In P. A. Granhag & L. A. Strömwall (Eds.), *Detection of deception in forensic contexts* (pp. 41–63). Cambridge: Cambridge University Press.
- Lee, Z., Klaver, J. R., & Hart, S. D. (2008). Psychopathy and verbal indicators of deception in offenders. *Psychology, Crime & Law, 14*, 73–84.
- Masip, J., Sporer, S. L., Garrido, E., & Herrero, C. (2005). The detection of deception with the reality monitoring approach: A review of the empirical evidence. *Psychology, Crime & Law, 11*, 99–122.
- May, R., & Wierda, M. (2002). *International criminal evidence*. Ardsley: Transnational Publishers, Inc.
- Niehaus, S., Krause, A., & Schmidke, J. (2005). Täuschungsstrategien bei der Schilderung von Sexualstraftaten [Deception strategies in the description of sex crimes]. *Zeitschrift für Sozialpsychologie, 36*, 175–187.
- Pezdek, K., Morrow, A., Blandon-Gitlin, I., Goodman, G. S., Quas, J. A., Saywitz, K. J., . . . Brodie, L. (2004). Detecting deception in children: Event familiarity affects criterion-based content analysis ratings. *Journal of Applied Psychology, 89*, 119–126.
- Porter, S., & ten Brinke, L. (2010). The truth about lies: What works in detecting high-stakes deception? *Legal and Criminological Psychology, 15*, 57–75.
- Porter, S., & Yuille, J. C. (1995). Credibility assessment of criminal suspects through statement analysis. *Psychology, Crime & Law, 1*, 319–331.
- Porter, S., & Yuille, J. C. (1996). The language of deceit: An investigation of the verbal clues to deception in the interrogation context. *Law and Human Behavior, 20*, 443–458.
- Quandte, S. (2010). *It's always best to speak the truth, unless you are a good liar: Children's strategies when telling the truth and lying*. Unpublished Ph.D. thesis. Department of Psychology, University of Portsmouth.
- Rock, P. (2007). Theoretical perspectives on victimisation. In S. Walklate (Ed.), *Handbook of victims and victimology* (pp. 37–61). Cullompton, UK: Willan Publishing.
- Ruby, C. L., & Brigham, J. C. (1998). Can criteria-based content analysis distinguish between true and false statements of African-American speakers? *Law and Human Behavior, 22*, 369–388.
- Schelin, L. (2007). *Bevisvärdering av utsagor i brottmål [Evidentiary assessment of statements in criminal cases]*. Stockholm: Norstedts Juridik.
- 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.
- Sporer, S. L. (2004). Reality monitoring and detection of deception. In P. A. Granhag & L. A. Strömwall (Eds.), *Detection of deception in forensic contexts* (pp. 64–102). Cambridge: Cambridge University Press.
- Steller, M., & Köhnken, G. (1989). Criteria-based content analysis. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217–245). New York: Springer.

- Strömwall, L. A., Bengtsson, L., Leander, L., & Granhag, P. A. (2004). Assessing children's statements: The impact of a repeated experience on CBCA and RM ratings. *Applied Cognitive Psychology, 18*, 653-668.
- Strömwall, L. A., & Willén, R. M. (2011). *Inside criminal minds: Offenders' strategies when lying*. Manuscript submitted for publication.
- Undeutsch, U. (1982). Statement reality analysis. In A. Trankell (Ed.), *Reconstructing the past: The role of psychologists in criminal trials* (pp. 27-56). Stockholm: Norstedts.
- Vrij, A. (2008). *Detecting lies and deceit. Pitfalls and opportunities*. (2nd ed.). Chichester: Wiley.
- Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2004a). 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., Akehurst, L., Soukara, S., & Bull, R. (2004b). Detecting deceit via analyses of verbal and nonverbal behavior in children and adults. *Human Communication Research, 30*, 8-41.
- Vrij, A., & Mann, S. (2006). Criteria-based content analysis: An empirical test of its underlying processes. *Psychology, Crime & Law, 12*, 337-349.
- 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*, 499-519.
- Vrij, A., & Semin, G. R. (1996). Lie experts' beliefs about nonverbal indicators of deception. *Journal of Nonverbal Behavior, 20*, 65-80.
- Willén, R. M., & Strömwall, L. A. (in press). Offenders' lies and truths: An evaluation of the Supreme Court of Sweden's criteria for credibility assessment. *Psychology, Crime and Law*.
- Wixted, J. T., & Ebbesen, E. B. (1991). On the form of forgetting. *Psychological Science, 2*, 409-415.
- Yuille, J. C., & Cutshall, J. (1989). Analysis of statements of victims, witnesses and suspects. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 175-191). London: Kluwer Academic Publishers.

Received 22 October 2010; revised version received 25 April 2011