



Bartosz Wojciech Wojciechowski*
Clinical and Forensic Psychology Department,
University of Silesia
Katowice
Poland

Content analysis algorithms: an innovative and accurate approach to statement veracity assessment

Key words: psychological content analysis, veracity assessment, statement validity assessment, decision algorithm

Introduction

There are four main areas of psychological research on the detection of deception: a) deceiver's personality traits; b) extralinguistic and nonverbal cues of deception; c) physiological cues; d) and the analysis of the qualitative aspects of the account (Sporer 1997). German forensic psychologists have developed qualitative criteria to analyse the content of statements and to assess their validity. Veracity assessment is based on an assumption called the Undeutsch hypothesis, maintaining that statements derived from a memory

* bartosz.wojciechowski@poczta.fm.

of a self-experienced event will differ qualitatively from statements that are based on a fantasy or suggestion (Undeutsch, 1967). On the basis of the work of Udo Undeutsch, Guenter Koehnken and Max Steller compiled a list of 19 criteria and described a procedure for evaluating the veracity of a statement, which led to development of a comprehensive method for evaluating witness statements – Statement Validity Assessment (SVA, Koehnken 2004). SVA focuses on specific content characteristics supporting the hypothesis that the account is based on genuine personal experience. SVA consists of three major elements: a structured interview, a criteria-based content analysis (CBCA) and the integration of all obtained case information into a final judgment as whether the account is based on genuine personal experience with the support of Validity Checklist (Koehnken 2004; Koehnken et al. 1995).

With regard to cognitive and motivational factors, it is assumed that the presence of several criteria indicates genuine experiences. Statements that are coherent and consistent (logical structure), with information that is not provided chronologically (unstructured production) and contains a significant amount of detail (quantity of detail), are more likely to be true. Statements are considered more likely to be truthful if they include references to time and space (contextual embeddings), descriptions of interactions (action – response – action...), reproduction of conversations (quotes of original speech), unexpected complications (unplanned changes in the course of events), unusual details (unusual or extraordinary detail), and superfluous details (peripheral contemporaneous events). Accurately reported and misunderstood details will also indicate truthfulness. Another possible indicators of truthfulness are reports of details that are not part of the allegation but are related to it (related external associations, accounts of a subjective mental state, attribution of perpetrator's mental state). Since truthful persons are not as concerned with impression management as deceivers, the following may occur in their statements: spontaneous corrections, admitting lack of memory, raising doubts about one's own testimony, self-deprecation, and pardoning the perpetrator. The last CBCA criterion is related to the characteristic features of the offence. It will be considered present, if key aspects and features of a given type of offence are presented. Description should be however counter-intuitive or discrepant to everyday knowledge (Koehnken 2004; Koehnken et al. 1995).

The Validity Checklist is used for CBCA. Validity Checklist refers to general categories of information to be evaluated: psychological traits, features of the interview, motivation, investigative questions and offence-specific ele-

ments. SVA evaluators consider the following issues: (a) appropriateness of language and knowledge (mental capability of the child) (b) appropriateness of affect shown by the interviewee, (c) interviewee's susceptibility to suggestion, (d) evidence of suggestive, leading, or coercive questioning, (e) overall adequacy of the interview, (f) motives to report, (g) context of the original disclosure or report, e.g. whether there are questionable elements in the context of the original disclosure, (h) pressures to report falsely, such as indications that others suggested, coached, pressured, or coerced the interviewee to make a false report, (i) consistency with the law of nature, i.e. whether the described events are unrealistic, (j) consistency with other statements, i.e. whether there are major elements of the statement that are inconsistent or contradicted by another statement made by the same interviewee, and (k) consistency with other evidence, e.g. presence of major elements in the statement that are contradicted by reliable physical evidence or other concrete evidence (Koehnken G. 2004; Koehnken G. et al. 1995).

There have been over 40 research papers testing the accuracy of statement analysis published in English or presented at conferences (Vrij 2005; Pezdek et al. 2004; Ruby, Brigham 1997; Buck, Warren, Betman, and Brigham 2002; Granhag, Stroemwall, Landstroem 2006; Vrij, Mann 2006). The average error rate of SVA judgments is estimated at 30%, both in experimental and laboratory studies (Vrij 2005; Vrij 2008). Research has demonstrated that content analysis scores are affected not only by the veracity of the statement but also by other factors, such as age, verbal proficiency and social skills of the interviewee, and the interview style of the interviewer (Vrij, Akehurst, Soukara, Bull 2004; Akehurst, Bull, Vrij, Koehnken 2004; Goedert, Gamer, Rill, Vossel 2005; Kapardis 2003). Some criteria were present more frequently in truthful statements, some SVA ratings were higher for false accounts, and none of the CBCA or Validity Checklist criteria proved its reliability in all the studies (Vrij 2008; Kapardis 2003). Moreover, several researchers have found that trained judges were better at distinguishing between truths and lies than nonprofessional evaluators, some found no training effect, and other found that training made judges worse at distinguishing between truths and lies (Vrij 2005, Akehurst, Bull, Vrij, Koehnken 2004; Blandon-Gitli, Pezdek, Lindsay, Hagen 2009; Rassin 2001). Aldert Vrij (2008) has pointed to another important limitation of SVA: the method is not a standardised instrument, there are no clear rules to determine the number of criteria that need to occur for a statement to be assessed as truthful, and there are no rules regarding the weight of the individual criteria. In consequence, SVA assessments are subjective.

The study

The major purpose of our study was to investigate whether SVA content criteria reliably and accurately discriminate between truthful and fabricated accounts, if there are different SVA content characteristics of true and false accounts, and to describe qualitative and quantitative features of the statements (Wojciechowski 2012).

Over 130 criminal cases were randomly selected from the register of four district and two regional courts, and thoroughly analysed with the use of files taxonomic scale. Information about crime, evidence material and witness(es) was collected. Testimonies were considered true if the courts of both first and second instance recognised them as reliable and there was additional evidence supporting the witnesses' accounts (such as physical traits of expert evidence). The testimonies were considered deceptive if the courts of the first and second instance found them unconvincing, the witness admitted to giving false evidence and was sentenced for perjury. The most important documents, such as interrogation protocols and expert evidence were photocopied. Witnesses were interviewed from one to seven times ($M = 2.51$, $sd = 1.19$), and the overall number of words ranged from 141 to 3146 ($M = 1,072$, $sd = 787.89$). False statements were longer than truthful accounts ($F(1.612) = 50.6493$, $p < 0.001$), but only in the case of 8.64% accounts, the variance of their length may be explained by the reliability of the statement ($\eta^2 = 0.08643$). There were no statistically significant effects for crime category, and the witness's gender, age, education level and status.

Forty-three forensic psychology graduates participated in the project on the voluntary basis as competent judges (raters). Content analysis was preceded by a 30-hour-long training. Participants read relevant books, research papers and detailed description of SVAs, and participated in lectures on rating methods and common rating errors. Extended criteria descriptions were presented, and each criterion was discussed. Raters rehearsed content analysis and identification of criteria in test transcribed statements. Rating scale for the Statement Validity Assessment was introduced, and raters had an opportunity to discuss cases of analysis and to compare their ratings with experts' ratings, and were given feedback. Each account was rated by two independent competent judges. Coding was carried out individually, and raters remained blind to the outcome of legal proceedings. 79 transcripts – 47 true and 32 false statements were rated with the SVA content criteria. Each criterion

was rated on a 6-point scale, where 1 indicated “absent” and 6 – “explicitly present”. Finishing the rating task, raters could judge the overall credibility of the account according to their subjective impression, irrespective of the SVA criteria ratings. Ratings were compared with the objective truth status, and correctness of classification decisions was assessed.

The results

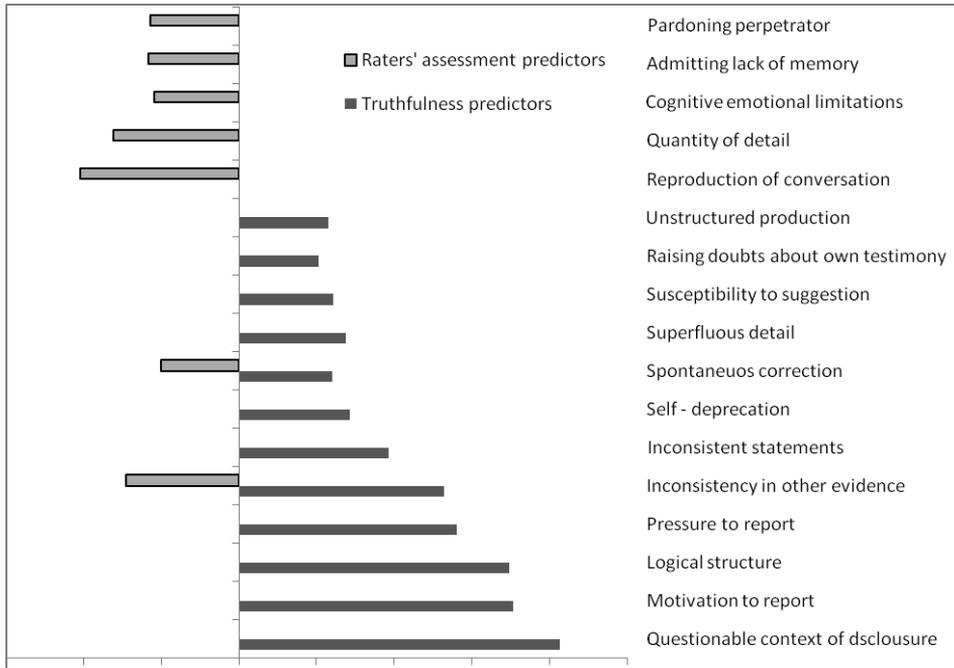
The use of SVA content criteria let raters classify properly 65,19% (103) of the cases, 64% of the false testimonies and 65.67% of truthful accounts. When judgments were based on the subjective impressions of the judges irrespective of the SVA ratings, 84.18% (133) of cases were properly classified, yet a truth bias was explicit (89.96% of truthful and 55.21% of false accounts).

To assess the effect of truth value on the SVA assessment, Kruskal-Wallis ANOVA was conducted. The mean overall score for all testimonies was 115.35 points (sd = 16.65), there was a statistically significant difference between two categories of accounts (83.13 and 63.20 respectively, $H = 3.9516$, $p = 0.0468$), true accounts received 116.05 points (sd = 15.67) and false ones – 111.56 points (sd = 21.13). However, only 15 out of 30 SVA criteria differentiate true and false accounts according to the methodological assumptions. The CBCA criteria 2, 3, 4, 8, 9, 12–18, and Validity Checklist items 1, 4 and 5 were evaluated higher for dishonest statements. Consistent with previous researches outcomes (Vrij 2008; Kapardis 2003; Rassin 2001), the presented results suggest that the false and true accounts could not be differentiated on the basis of the CBCA and Validity Checklist criteria, and the use of SVA reduces accuracy of veracity assessment.

In order to select SVA criteria reliably discriminating truthful and fabricated accounts and to narrow the list of predictor variables, a feature selection and variable screening tool was used. The analysis revealed that seven of the SVA content criteria are significantly related to the veracity of a statement at the value of $p < 0.05$; they are: reproduction of conversations ($\text{Chi}^2 = 20.4381$, $p = 0.0004$), quantity of detail ($\text{Chi}^2 = 16.1770$, $p = 0.0063$); inconsistency with other evidence ($\text{Chi}^2 = 14.5281$, $p = 0.0126$); spontaneous correction ($\text{Chi}^2 = 10.9737$, $p = 0.0269$); appropriateness of language and knowledge ($\text{Chi}^2 = 10.9036$, $p = 0.0277$); admitting lack of memory ($\text{Chi}^2 = 11.6637$, $p = 0.0397$) and pardoning the perpetrator ($\text{Chi}^2 = 11.4940$, $p = 0.0424$). Surprisingly, validity predictor variables did not correspond with rater decisions

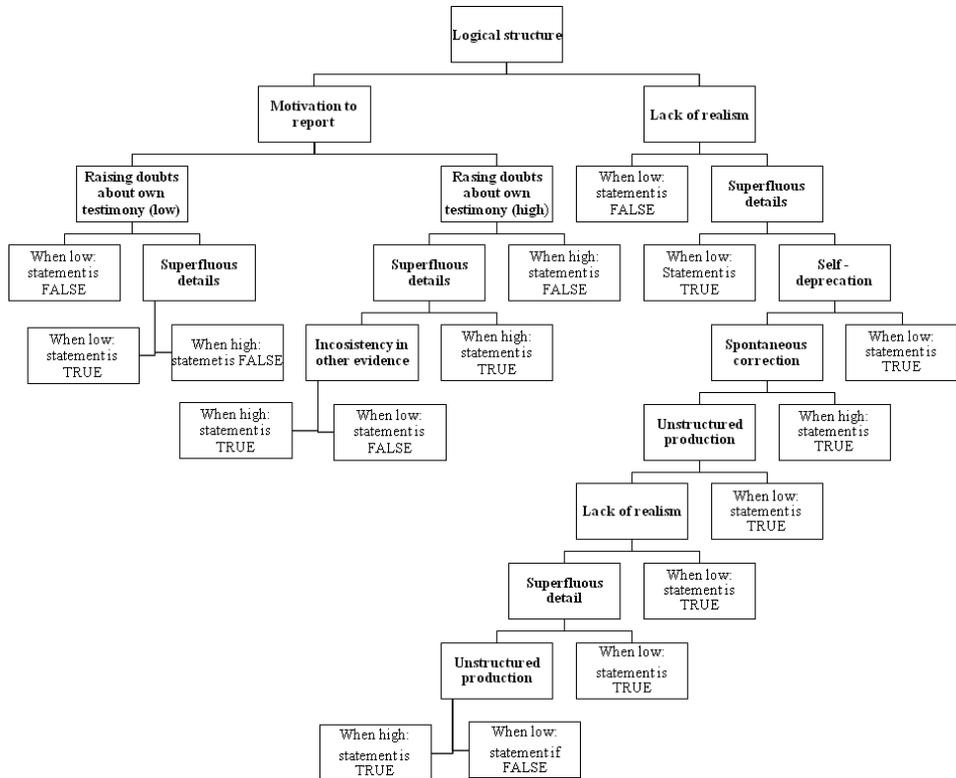
predictor variables. As presented in Figure 1, only two of account truthfulness predictors (spontaneous correction, and inconsistency with other evidence) influenced decisions made by the competent judges in the process of validity assessment.

Fig. 1. SVA veracity predictors and raters' decisions determinants



Preselected categorical variables, which are the best determinants of statement veracity, were used for recursive partitioning analysis, a non-parametric data mining algorithm for generating decision trees. Recursive partitioning creates a decision tree where observations are sorted into nodes, but compared to generalised linear models, no distribution assumptions are made about the data. Variables that best differentiate between the categories of account divide data into leaves of the tree or nodes, ordering the target variable. Decision tree is a high performance algorithm that groups accounts into two classes: true and false.

Fig. 2. Classification and decision tree for true and false statements



The screening of variables and focus on the most significant categorical statement value predictor – SVA content criteria organised into a classification tree – allow proper classification of all (100%) truthful accounts and 84% of false statements made by witnesses. Presented model not only improves accuracy of the ratings but may be deployed to make reliable and objective decisions on statement veracity. Classification tree offers rules regarding the weight of different criteria and the number of criteria that need to occur for assessing the statement as truthful. There are two steps of the veracity assessment with the use of the content analysis algorithm. First, the rater grades each of the content criteria. Second, the results of the individual criteria analysis are confronted with the decision tree. For example if the first CBCA criterion (“logical structure”) is rated 4 or lower on the 6-point scale used, “motivation to report” is rated between 1 and 4, and the criterion “rais-

ing doubts about own testimony” is rated 1, 2 or 3 points, this indicates that the statement is false. If, on the other hand, the first criterion (“logical structure”) were rated 4 or higher, “lack of realism” – 3 or higher, and “superfluous details” from 1 to 3 points, the analysed statement would be truthful.

Summary and discussion

There are three key findings in this study. First, that the results support the Undeutsch hypothesis claiming content differences between accounts of events that are based on true experience and accounts of events that are based on fantasy or invention. The results revealed that SVA content criteria, and reproduction of conversations, quantity of detail, inconsistency in other evidence, spontaneous corrections, appropriateness of language and knowledge, admitting lack of memory and pardoning perpetrator in particular reliably discriminate truthful and fabricated accounts.

The second key finding of this study is that the poor accuracy rates of the SVA may be explained by the lack of correspondence between the best validity predictors and the decisions predictors used by the raters. Feature selection and variable screening results showed that decisions made by the raters are not based on the results of content analysis including the best truthfulness predictors for the accounts. Only two SVA criteria of the 15 best veracity predictors (spontaneous correction and inconsistency with other evidence) are used by competent judges, and specific contents characteristics with poor discriminative power are the grounds for veracity assessment.

The third key finding of this study is that Statement Validity Assessment can enable accurate discrimination of truthful and false testimonies but decision algorithms in the form of classification trees should be applied. Assessment of the content criteria of particular SVAs must be followed by a formalised analysis employing rules that involve the weight of different criteria and number of criteria that need to occur for assessing the statement as true or false.

There is a considerable empirical support for the assumption that there are qualitative and quantitative differences between experience-based and fabricated statements. To improve the diagnostic strategy in specific cases, it is desirable to assess decision algorithms in further studies and to use of clas-

sification and decision trees to gather more information on how individual traits related to the quality of the statements can affect the accuracy of veracity assessment.

References

- Akehurst L., Bull R., Vrij A., Koehnken G., (2004), The effects of training professional groups and lay persons to use Criteria-Based Content Analysis to detect deception. *Applied Cognitive Psychology*, 18, 877.
- Blandon-Gitlin I., Pezdek K., Lindsay S.D., Hagen L. (2009), Criteria-Based Content Analysis of true and suggested accounts of events. *Applied Cognitive Psychology*, 23, 901.
- Buck J.A., Warren A.R., Betman S.I., Brigham J.C. (2002), Age differences in criteria-based content analyses scores in typical sexual abuse interviews. *Applied Developmental Psychology*, 23, 267.
- Goedert H.W., Gamer M., Rill H.R., 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.
- Granhag P.A., Stroemwall L.A., Landstroem S. (2008), Children recalling an event repeatedly: effects on RM and CBCA scores. *Legal and Criminological Psychology*, 11, 81–98.
- Kapardis A., (2003), *Psychology and law. A critical introduction*, Cambridge University Press, Cambridge.
- Koehnken G.(2004), Statement Validity Analysis and the “detection of truth” in P.A. Granhag, L.A. Stroemwall, eds. *The detection of deception in forensic contexts*, Cambridge University Press, Cambridge, 41.
- Koehnken G., Schimossek E., Aschermann E., Hoefler E., (1995), The Cognitive Interview and the assessment of the credibility of adults’ statements. *Journal of Applied Psychology*, 80, 6, 671.
- Pezdek K. (2004), Detecting deception in children: event familiarity affects Criterion-Based Content Analysis ratings. *Journal of Applied Psychology*, 89, 1, 119–126.
- Rassin E. (2001), Criteria-based content analysis: the less scientific road to truth. *Expert Evidence*, 7, 265.

Ruby C.L., Brigham J.C. (1997), The usefulness of the Criterion-Based Content Analysis technique in distinguishing between truthful and fabricated allegations: A critical review. *Psychology, Public Policy and Law*, 3, 705.

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.

Trankell A. (1972), Reliability of evidence, Beckmans, Stockholm.

Undeutsch U. (1967), Beurteilung der Glaubhaftigkeit von Zeugenaussagen, in: *Handbuch der Rechtspsychologie*, ed. Undeutsch U., Gettingen, Hogrefe, 26.

Vrij A., (2008), *Detecting lies and deceit: pitfalls and opportunities*. Wiley, Chichester.

Vrij A. (2005), Criteria-Based Content Analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy and Law*, 11, 1, 3.

Vrij A., Mann S., (2006), Criteria-Based Content Analysis: an empirical test of its underlying processes. *Psychology, Crime and Law*, 12, 337.

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, 2, 113.

Wojciechowski B.W. (2012). Ocena wiarygodności zeznań świadków dorosłych przy zastosowaniu psychologicznych i prawniczych kryteriów treściowych; unpublished doctoral dissertation, University of Silesia in Katowice.