

Bij hun opsporingswerk maken politie-
mensen steeds vaker gebruik van
verbale analysemethoden. Daarbij
kijkt de rechercheur naar allerlei ken-
merken van het verhaal dat de ver-
dachte of getuige vertelt om zo de ge-
loofwaardigheid ervan te kunnen
peilen. Neem de verdachte die zegt: 'ik
kwam op mijn werk en daarna was het
geld weg.' Volgens zo'n methode is deze
vage chronologie een veeg teken.
Glynis Bogaard, Ewout Meijer, Aldert
Vrij en Harald Merckelbach kijken
daar toch anders tegen aan.

VERBALE ANALYSEMETHODEN

LEUGENAARS PRATEN ANDERS

In 1995 werd de Amerikaanse Susan Smith tot een levenslange gevangenisstraf veroordeeld voor de moord op haar twee kinderen. Ze had de kinderen slapend in haar auto laten zitten, de handrem er vanaf gehaald en de auto zo het water in laten rijden. Om haar daad verborgen te houden, vertelde Smith de politie aanvankelijk dat haar auto was gestolen door een man met een donkere huidkleur, terwijl haar kinderen er nog in zaten.

In de verklaring die Smith tegenover de politie aflegde, was het opvallend dat ze in de verleden tijd over haar kinderen sprak: 'My children wanted me. They needed me. And now I can't help them.' Bij sommige onderzoekers ging de alarmbel af. Je zou immers verwachten dat een moeder in de tegenwoordige tijd over haar vermiste kinderen spreekt. Tenzij ze meer weet natuurlijk. Hier leek het erop dat de moeder wist dat de kinderen haar niet meer nodig hadden. Haar man – die in de veronderstelling verkeerde dat ze nog in leven waren – sprak wél in de tegenwoordige tijd als het over zijn kinderen ging.

VERBALE ANALYSEMETHODEN In dit voorbeeld maten de onderzoekers de betrouwbaarheid van de verdachte af aan hoe zij zich verbaal uitdrukte. Forensische toepassingen die van dit principe gebruik maken – de zogenaamde verbale analysemethoden – winnen aan terrein. Zo zijn Nederlandse en Belgische opsporingsambtenaren enthousiast over de zogenaamde *Scientific Content Analysis* (SCAN) en passen allerlei getuige-deskundigen de *Criteria Based Content Analysis* (CBCA) toe in zedenzaken.

Hoe goed is het timmermansoog van experts als zij een onderscheid moeten maken tussen ware en gelogen verklaringen? Eén van ons (Vrij, 2008; zie ook Bond & DePaulo, 2006) zette nog niet zo lang geleden 39 studies daarover op een rij. Het bleek dat experts gemiddeld slechts 56% van de verklaringen juist beoordelen. Het enige verschil tussen experts en leken is dat experts vaker hun kundigheid om leugens te ontmaskeren overschatten (Meissner & Kassin, 2002). Niettemin voelen ook experts de behoefte aan hulpmiddelen, dat wil zeggen methoden die de nauwkeu-

righeid van experts vergroten als zij moeten oordelen over het waarheidsgehalte van verklaringen. Dit artikel gaat over verbale analysemethoden en of zij zo'n hulpmiddel kunnen zijn. Om op onze conclusie vooruit te lopen: ze zetten maar in beperkte mate zoden aan de dijk. Hieronder leggen we uit waarom we deze sombere conclusie trekken.

HOE NAUWKEURIG IS DE CRITERIA BASED CONTENT ANALYSIS?

De belangstelling voor verbale analysemethoden dateert uit de jaren vijftig, toen de Keulse hoogleraar psychologie Udo Undeutsch de *Reality Criteria* ontwikkelde. Tijdens zijn vele interviews met kinderen die zeiden seksueel misbruikt te zijn, meende Undeutsch te zien dat ware verklaringen qua vorm en inhoud verschillen van gelogen verklaringen (Undeutsch, 1967). Undeutsch maakte een lijst met aandachtspunten die konden worden gebruikt om te bepalen hoe geloofwaardig de verklaringen van jeugdige getuigen zijn. Deze punten werden later door de Duitse psychologen Steller en Koehnken (1989) verfijnd tot een min of meer formele procedure, *Statement Validity Assessment* (sva) genaamd.¹

De Criteria Based Content Analysis (CBCA) vormt het kernstuk van deze procedure. CBCA bestaat eruit dat de expert een getuigenverklaring toetst tegen de criteria die zijn weergegeven in Tabel 1. Ware verklaringen zouden meer en sterker aan deze criteria beantwoorden dan gelogen verklaringen. Volgens CBCA zegt bijvoorbeeld de hoeveelheid details die wordt beschreven in een verklaring iets over de waarachtigheid ervan. Hetzelfde geldt voor beschrijvingen van conversaties (toen riep hij: 'naar de slaapkamer, nu!') en emoties ('ik werd daar erg bang van'). De aanwezigheid van dit type detail zou dus op een ware verklaring duiden.

De bedenkers van CBCA vonden het belangrijk dat de verklaring die de getuige tijdens het verhoor aflegt, wordt opgenomen en letterlijk wordt uitgetypt. Deskundigen die CBCA willen gebruiken, hebben op deze manier inzicht in hoe de getuige of de verdachte zich tijdens de ondervraging uitdrukte (Steller & Koehnken, 1989). De criteria van CBCA worden doorgaans gescoord aan de hand van een driepuntschaal; een criterium krijgt een 0 toebedeeld indien de verklaring er niet aan voldoet, een 1 indien zij in zekere mate eraan voldoet en een 2 indien zij sterk eraan voldoet (Vrij, 2008). De scores worden vervolgens gesommeerd. Hoe hoger deze somscore, hoe authentiekere de verklaring zou zijn.

Hoe nauwkeurig is CBCA? In nogal wat veldstudies en laboratoriumexperimenten stond die vraag centraal. Veldstudies maken gebruik van de verklaringen van echte getuigen

of verdachten. In laboratoriumstudies krijgen proefpersonen de opdracht om te liegen of de waarheid te spreken over bepaalde gebeurtenissen. Vrij (2005) zette 37 veld- en laboratoriumstudies op een rij, en daaruit bleek dat CBCA in staat is om 55% tot 90% van de verklaringen juist te classificeren. De gemiddelde nauwkeurigheid voor ware en valse verklaring is nagenoeg gelijk en varieert rond de 70%. In elf van de twaalf nader onderzochte studies kwamen de criteria van de CBCA vaker voor in ware dan in gelogen verklaringen (Vrij, 2005).

Veldonderzoek naar CBCA werd uitgevoerd door de Amerikaanse onderzoeker Lamb en zijn medewerkers (1997). Zij analyseerden 98 verklaringen van vermoedelijk misbruikte kinderen in de leeftijdscategorie van 4 tot 12 jaar. Afhankelijk van het juridisch bewijs – zoals de aanwezigheid van schaamhaar van de verdachte op het slachtoffer – werden hun verklaringen in twee groepen verdeeld. Als er aanvullend bewijs bestond, werden de verklaringen ingedeeld in de 'waarschijnlijke' groep; alle overige verklaringen werden ingedeeld in de 'onwaarschijnlijke' groep. Verklaringen werden gescoord op de mate waarin ze voldeden aan veertien criteria van CBCA. Ieder criterium kreeg een 0 toebedeeld indien de verklaring er niet aan voldeed en een 1 indien dat wel het geval was. De maximale score was zodoende 14. De 'waarschijnlijke' groep

TABEL 1. CRITERIA BASED CONTENT ANALYSIS CRITERIA

1.	Logische structuur
2.	Ongestructureerde productie
3.	Hoeveelheid details
4.	Contextuele omkadering
5.	Beschrijving interactie
6.	Reproductie van spraak
7.	Onverwachte complicaties
8.	Ongewone details
9.	Overbodige details
10.	Nauwkeurige, verkeerd geïnterpreteerde details
11.	Samenhangende, externe associaties
12.	Beschrijvingen eigen mentale toestand
13.	Attributies mentale toestand dader
14.	Spontane correcties
15.	Toegeven geheugenfalen
16.	Twijfelen over eigen verklaring
17.	Zelfafkeuring
18.	Excuses voor het gedrag van de beschuldigde uiten
19.	Details kenmerkend voor het misdrijf

De verklaringen van fantasierijke mensen krijgen een hoge CBCA-score, ongeacht of ze waar zijn of bij elkaar gelogen

haalde hogere scores (gemiddeld 6.7 criteria per verklaring) dan de 'onwaarschijnlijke' groep (gemiddeld 4.9 criteria per verklaring). Dit verschil was significant, maar de lezer ziet ook wel dat het niet om een verpletterend effect gaat.

Ofschoon onderzoek duidelijk maakt dat waar-onwaar-classificaties van CBCA boven kansniveau uitkomen, is er toch ook sprake van een aanzienlijke foutenmarge. Het punt is dat een CBCA-score niet alleen wordt beïnvloed door de (on)waarheid van de verklaring, maar ook door factoren die meer te maken hebben met de persoon van de getuige of de verdachte. Te denken valt aan leeftijd, intelligentie, verbale ontwikkeling en fantasierijkheid (Merckelbach, 2004; Santtila, Roppola, Runtti & Niemi, 2000; Vrij, 2005). Hoe ouder de persoon en hoe groter zijn verbale en cognitieve vermogens, des te hoger de CBCA-score. Ook blijken de verklaringen van fantasierijke mensen een relatief hoge CBCA-score te halen, ongeacht of ze waar of bij elkaar gelogen zijn (Schelleman-Offermans & Merckelbach, 2010).

CBCA-experts werpen vaak tegen dat ze niet alleen maar volstaan met het toetsen van de verklaringen aan de criteria, maar dat zij zich ook een oordeel vormen over de persoon van de verdachte of de getuige en wat daarover zoal in het strafdossier te vinden is. Sterker nog, zeggen de experts, we beginnen ons werk met zo'n soort inventarisatie. Of dit tot voordeel van hun werkwijze strekt, valt echter te betwijfelen. Aangenomen mag worden dat zulke dossierinformatie de CBCA-scores vertroebelt. Neem – bij wijze van parallel – het onderzoek van Elaad en collega's (1994). Zij lieten een uitdraai van een leugendetector aan experts zien. De ene groep van experts kreeg te horen dat de uitdraai van een verdachte was die net had bekend. De andere groep hoorde dat de uitdraai weliswaar van een verdachte was, maar dat ondertussen een andere persoon had bekend. Vervolgens kregen beide groepen de opdracht om te turen naar de uitdraai en te zoeken naar verdachte tekens. De eerste groep zag beduidend meer verdachte aanwijzingen in de uitdraai dan de tweede groep.

Dit brengt ons bij de betrouwbaarheid van CBCA. De interbeoordelaarsbetrouwbaarheid verwijst naar de mate waarin de scores van twee of meer onafhankelijke beoordelaars met

elkaar overeenkomen. Een goede interbeoordelaarsbetrouwbaarheid is belangrijk omdat een diagnostisch instrument niet te veel afhankelijk mag zijn van de expert die het toepast. Om dat voor de CBCA te onderzoeken lieten Anson, Golding, en Gully (1993) geoefende beoordelaars 23 video's zien van kinderen die verklaarden dat ze seksueel misbruikt waren. Hun verklaringen werden beoordeeld aan de hand van CBCA. De interbeoordelaarscoëfficiënt voor de totale CBCA-score bleek goed tot zeer goed. Deze coëfficiënt lag beduidend hoger dan coëfficiënten voor de afzonderlijke CBCA-criteria, die varieerden van slecht tot goed. Gelijksortige resultaten werden gemeld door Horowitz en collega's (1997) en Gödert en collega's (2005). In beide studies onderzochten drie getrainde beoordelaars ongeveer honderd geschreven verklaringen van slachtoffers die vermoedelijk seksueel misbruikt waren. Opnieuw bleken de interbeoordelaarscoëfficiënten voor de totale CBCA redelijk goed. Vrij (2008) bracht de resultaten van 39 CBCA-studies in kaart en kwam tot dezelfde conclusie.

Uit deze resultaten valt te concluderen dat de interbeoordelaarsbetrouwbaarheid van de totale CBCA in orde is, maar dat die voor de individuele criteria toch vaak te wensen overlaat. Deze vaststelling is niet van praktisch belang ontbloeit, aangezien deskundigen zich in de rechtszaal geregeld op individuele criteria beroepen. Het is bijvoorbeeld niet uitzonderlijk om een deskundige te horen zeggen dat de verklaring van het kind geloofwaardig moet zijn, omdat het dialogen beschrijft (Anson et al., 1993; Gödert et al., 2005; Horowitz et al., 1997).

CBCA IN DE PRAKTIJK Hoewel in vakartikelen richtlijnen worden beschreven voor het toepassen van CBCA, wijkt het gebruik ervan in de praktijk nogal eens af. Van Nierop en collega's (2006) bestudeerden aan de hand van interviews met deskundigen en hun pro-justitia-rapportages het gebruik van CBCA in de praktijk. Uit hun inventarisatie bleek dat experts uiteenlopende criteria van de CBCA belangrijk vinden. Eén deskundige gebruikte bijvoorbeeld slechts zes van de negentien criteria. Vaak stelden de deskundigen

SCAN lijkt niets toe te voegen aan het timmermansoog van ervaren onderzoekers

de criteria eigenhandig bij. Een andere deskundige paste bijvoorbeeld vier van de criteria naar eigen inzicht aan. Zulke amendementen konden zelfs zo'n vorm aannemen dat ze niet meer waren terug te voeren op de oorspronkelijke CBCA-lijst. Bovendien bleek geregeld dat de criteria niet puntsgewijs werden langsgelopen, zoals de richtlijnen van de CBCA voorschrijven, maar slechts beschrijvend werden weergegeven. Anders dan de wetenschappelijke literatuur doet vermoeden, blijkt het gebruik van CBCA in de praktijk dus niet gestandaardiseerd te zijn.

De inventarisatie van Van Nierop en haar collega's laat ook een groot verschil zien in de wijze waarop deskundigen hun conclusies weergeven. Slechts een enkeling betreft kenmerken als suggestibiliteit en cognitieve capaciteit in zijn analyses. Volgens Rassin en collega's (1997) is dit te wijten aan het ontbreken van systematisch onderzoek: daardoor bestaat er ruimte voor het naar eigen inzicht inzetten van CBCA.

Deze indruk wordt bevestigd in de Zweedse veldstudie van Gumpert en Lindblad (1999). Zij onderzochten hoe Zweedse deskundigen de CBCA bij een tiental zaken hadden gebruikt. De deskundigen bleken allerlei criteria van CBCA verkeerd te gebruiken. Zo kwam het voor dat de afwezigheid van emoties (beschrijving van de eigen mentale toestand) door de ene deskundige als indicatie van leugenachtigheid werd beschouwd, terwijl een andere deskundige hier geen veeg teken in zag. Het probleem dat zich hier voordoet, heeft te maken met een foute interpretatie van CBCA. Bij CBCA is de aanwezigheid van criteria een indicatie van waarachtigheid, maar de afwezigheid ervan is *geen* indicatie van liegen. De deskundigen gebruikten ook criteria die niet zijn terug te voeren op CBCA. Zo rapporteerde één deskundige dat het gaandeweg bijstellen van het verhaal een aanwijzing voor leugenachtigheid was, terwijl dit geen criterium van CBCA is. Daarnaast verschilden deskundigen ook in de weging die ze gaven aan de diverse criteria. Omdat deze weging bepaalt welke criteria zwaarder tellen bij het trekken van een conclu-

sie, kan dit voor eenzelfde verklaring uiteenlopende conclusies opleveren.

Hoewel CBCA beschikt over een gestandaardiseerde set van criteria, een redelijke nauwkeurigheid en een acceptabele interbeoordelaarsbetrouwbaarheid, lijkt het gebruik in de praktijk toch te wensen over te laten. Wetenschappelijk onderzoek laat weliswaar zien dat CBCA boven kansniveau presteert, maar ad-hoc-aanpassingen van de methode door praktiserende deskundigen doen ons twijfelen aan de soliditeit van de methode.

HOE NAUWKEURIG IS REALITY MONITORING?

Een andere vaak onderzochte verbale analysemethode is *Reality Monitoring* (RM). De methode werd ooit ontwikkeld om een onderscheid te maken tussen echte herinneringen en pseudo-herinneringen. De gedachte daarbij was dat echte herinneringen ontstaan vanuit de waarneming en daarom meer sensorische en affectieve informatie zullen bevatten dan fictieve herinneringen (Johnson & Raye, 1981). Echte herinneringen zullen daarom veel details over geur, smaak, gevoel en geluid in zich dragen en deze details kunnen binnen een afgebakende context worden geplaatst. Zo zullen herinneringen aan uw laatste vakantie waarschijnlijk redelijk precieze beelden, geuren en gevoelens omvatten. Pseudo-herinneringen vereisen daarentegen meer cognitieve inspanning ('cognitieve operaties'). Het gaat dan om redeneringen en conclusies van het type: 'Volgens mij moet het die dag koud zijn geweest want ik had een sjaal om' en 'Ik heb de deur vast wel dichtgedaan want zo iets vergeet ik niet.'

Een aantal onderzoekers heeft gekeken of RM ook bruikbaar kan zijn om een onderscheid te maken tussen ware en onware verklaringen. De Duitse psycholoog Sporer (1997) beschreef acht punten waarop echte en valse verklaringen van elkaar zouden verschillen. Zeven daarvan komen volgens hem vaker voor bij ware verklaringen dan bij onware verkla-

TABEL 2. REALITY MONITORING CRITERIA

1.	Helderheid
2.	Perceptuele informatie
3.	Ruimtelijke informatie
4.	Temporele informatie
5.	Affect
6.	Reconstructie
7.	Realisme
8.	Cognitieve Operaties

ringen, terwijl het criterium 'cognitieve operaties' vaker zou voorkomen bij onware verklaringen (Zie Tabel 2; Johnson & Raye, 1981; Vrij, Mann & Leal, 2008). Ook RM-criteria worden doorgaans gescoord aan de hand van een 3-puntsschaal. Vervolgens kan dan een totale RM-score worden berekend.³ Ook hier geldt: hoe hoger deze score, hoe authentieker de verklaring.

Omdat RM oorspronkelijk uit het geheugenonderzoek stamt, blijft het de vraag of de methode wel toegepast kan worden als hulpmiddel om de waarachtigheid van verklaringen te beoordelen. Met dat in het achterhoofd bekeek Vrij (2008) de gemiddelde nauwkeurigheid van RM in zeven laboratoriumstudies en drie veldstudies. Hij vond dat het

jarigen en een groep studenten de opdracht om een ware of gelogen verklaring te produceren over een recente gebeurtenis. De gebeurtenis bestond uit het spelen van 'vier op een rij' in een kamer met een schoolbord waarop allerlei informatie stond. Tijdens het spel kwam iemand de kamer binnen, maakte een opmerking over het spel en veegde vervolgens de informatie op het schoolbord uit. De verkregen verklaringen werden geanalyseerd met behulp van de CBCA en RM. De RM-methode bleek niet in staat om de verklaringen van de jongste kinderen boven kansniveau te beoordelen. Met behulp van CBCA kon echter 69% van de verklaringen in deze groep correct worden geclassificeerd. Bij de verklaringen van studenten bleek de RM-methode het juist beter te doen

De ene deskundige beschouwt de afwezigheid van emoties als indicatie van leugenachtigheid, een andere ziet hier geen veeg teken in

percentage correcte classificaties op basis van RM varieert tussen 61% en 83%, met een gemiddelde van 69% (Vrij, 2008). Dat is vergelijkbaar met de prestaties van de CBCA. Wanneer de criteria afzonderlijk op hun verdiensten werden beoordeeld, vond hij dat de contextuele criteria (wanneer en waar speelde het zich allemaal af?) meer diagnostisch gewicht hebben dan de andere criteria (Masip, Sporer, Garido & Herrero, 2005; Vrij, 2008). De interbeoordelaarsbetrouwbaarheid van de totale RM score is – net als bij CBCA – goed.

Hoewel CBCA en RM veel raakvlakken hebben – het CBCA-criterium *contextuele omkadering* doelt bijvoorbeeld op hetzelfde als de twee RM-criteria *ruimtelijke informatie* en *temporele informatie* – zijn er ook duidelijke verschillen. CBCA werd oorspronkelijk ontwikkeld voor het beoordelen van de verklaringen van jonge kinderen, terwijl RM haar wortels in het geheugenonderzoek heeft. Vrij en collega's (Vrij, Akehurst, Soukara & Bull, 2004) onderzochten of RM bruikbaar is bij het beoordelen van verklaringen van kinderen jonger dan tien jaar. In hun studie kregen zesjarigen, elfjarigen, vijftien-

(77% versus 62% correcte classificatie). In de twee middelste leeftijdsgroepen was de nauwkeurigheid van RM en CBCA vergelijkbaar.

De RM-methode is niet alleen leeftijdsgevoelig, maar blijkt ook onbruikbaar bij verklaringen die gaan over oude herinneringen. Dat komt waarschijnlijk omdat hoe ouder een herinnering is, hoe meer cognitieve operaties er nodig zijn om haar vlot te trekken (Roediger, 1996). Het omgekeerde geldt trouwens ook: hoe ouder een gebeurtenis waarover vaak is gefantaseerd, hoe makkelijker deze kan worden opgeroepen (Manzanero & Digest, 1996).

Omdat RM slechts over acht criteria beschikt, neemt deze methode beduidend minder tijd in beslag dan de CBCA. Een ander voordeel van RM is dat deze techniek niet ligt ingebed in een quasi-protocol dat voorschrijft dat de expert eerst het dossier moet raadplegen. Daarmee worden verwachtingseffecten van het *seek-and-you-will-find*-type voorkomen. Het zou interessant zijn om de nauwkeurigheid van een gecombineerde CBCA/RM-methode te onderzoeken. De

criteria van beide methoden richten zich – althans voor een deel – op andere aspecten van de verklaringen. De CBCA kijkt vooral naar cognitieve en motivationele factoren, terwijl de RM-methode vooral kijkt naar aspecten van het geheugen. Op grond van psychometrische overwegingen – hoe meer verschillende items, hoe beter het instrument – zou men verwachten dat een CBCA/RM-combinatie een hoger percentage treffers oplevert. Dat is ook wat Vrij en zijn collega's (2004) vonden in de eerder aangehaalde studie. Zij vergeleken de nauwkeurigheid van CBCA, RM en non-verbale gedragingen zoals spreekfouten, hand- en voetbewegingen, en oogcontact. Wanneer deze drie benaderingen werden gecombineerd, was de nauwkeurigheid in elke leeftijdsgroep groter dan de nauwkeurigheid van elke methode afzonderlijk (69% versus 61% juiste classificatie). Deze resultaten doen vermoeden dat

de nauwkeurigheid toeneemt wanneer RM en CBCA worden gecombineerd (zie ook Colwell, Hiscock-Anisman, Memon, Taylor & Prewett, 2008).

SCIENTIFIC CONTENT ANALYSIS (SCAN) Anders dan de naam doet vermoeden is de *Scientific Content Analysis* (SCAN; Sapir, 1996) niet ontstaan vanuit wetenschappelijke inzichten, maar vanuit de ervaring die haar geestelijke vader, Avinoam Sapir, als speurder opdeed. Sapir is een voormalig Israëlich luitenant en ook polygrafist. Een polygrafist is iemand die een leugendetector bedient en de uitkomsten ervan interpreteert. Sapir merkte tijdens zijn werkzaamheden als polygrafist dat verdachten die volgens hem logen en verdachten die volgens hem de waarheid spraken er een ander taalgebruik op nahielden. Op basis van deze observa-

TABEL 3. SCIENTIFIC CONTENT ANALYSIS CRITERIA NAAR VRIJ (2008)

Kenmerk	Voorbeeld (moord)	Beschrijving kenmerk
Ontkennen van beschuldiging	Ik heb het niet gedaan!	Letterlijk schuld ontkennen duidt op onschuld
Sociale introductie	Mijn vrouw, Yvonne	Een goede introductie bestaat uit rol en naam
Spontane correcties	Het was om 8 uur 's morgens, nee half negen 's morgens	Spontane correctie duidt op misleiding
Gebrek aan herinneringen of betrokkenheid met de gebeurtenis	Ik denk dat ik haar ongeveer om half negen gevonden heb, maar ik weet het niet zeker.	Vaag taalgebruik of gebrek aan herinnering duiden op misleiding
Structuur van de verklaring		Een verklaring bestaat uit 20% proloog, 50% hoofdzaak en 30% epiloog. Afwijkingen duiden op misleiding
Plaats van emoties	Toen ik haar zo vond, was ik heel bang	Emoties moeten na de hoofdzaak beschreven worden.
Objectieve en subjectieve tijd		De gemiddelde schrijfsnelheid is 3 regels per objectief uur. Afwijkingen hiervan duiden op misleiding.
Disproportionele of buiten-contextuele vermeldingen	Ik ging toevallig de keuken binnen omdat ik daar mijn beurs had laten liggen	Al de uitleg (omdat, want, daarom...) die spontaan gegeven wordt, geeft gevoelige punten in een verklaring aan
Ontbrekende informatie	Een uur later vond ik haar dood	Missende tijd duidt op misleiding. Wat gebeurde er tijdens dat uur?
Eerste persoon enkelvoud, verleden tijd	Ik liep naar de keuken en zie haar daar liggen.	Afwijken van eerste persoon enkelvoud verleden tijd duidt op misleiding
Voornaamwoorden	Zag haar liggen en verliet het huis	Missende voornaamwoorden duiden op afstand en misleiding
Verandering in taalgebruik	Mijn vriend wilde praten. Ik zei tegen Robert dat ik niet wilde.	Verandering van taal duidt op een verandering in realiteit.

De nauwkeurigheid neemt waarschijnlijk toe wanneer twee of drie methoden worden gecombineerd

ties ontwikkelde hij de SCAN-methode, die hij wereldwijd in meerdaagse cursussen onderwijst. Inmiddels wordt SCAN in landen als België, Israël, Mexico, Singapore, Zuid-Afrika, Nederland, Canada, het Verenigd Koninkrijk en de Verenigde Staten toegepast (Bockstaele, 2008; Vrij, 2008), en de claims zijn niet van de lucht. Zo lezen we op de website⁴ dat: 'SCAN will solve every case for you quickly and easily. You only need the subject's own words, given of his/her own free will' en dat 'Everyone wants to tell you everything – even the people who commit crimes. As an interrogator you need the tools to read this information clearly – and SCAN is the tool that will give you the truth.' Ze komen niet alleen van Sapir zelf. Zo merken twee Belgische magistraten op: 'Kortom, SCAN kan zonder enige twijfel bijdragen tot de waarheidsvinding' (Sabbe & de Smet, 2008, p.165).

SCAN gaat uit van een handgeschreven verklaring. Deze verklaring dient zoveel mogelijk de woorden van de betrokkene weer te geven, en mag niet het resultaat zijn van een uitgebreide ondervraging door de politie. De verklaring wordt vervolgens geanalyseerd met behulp van de SCAN-criteria (zie Tabel 3). De lijst van SCAN-criteria werd in de loop van de tijd voortdurend aangepast en tot op heden bestaat er geen gestandaardiseerde set. Zoals de Belgische hoofdcommissaris Bockstaele (2008; p. 136) zegt, is het scoren van de criteria 'afhankelijk van de inventiviteit en kwaliteit van de analist.' Dat klinkt sympathiek, maar het zet de deur voor willekeurige interpretaties wagenwijd open.

Ondanks de grootschalige inzet van SCAN hebben wij slechts drie studies kunnen vinden waarin de SCAN-techniek op haar merites werd beoordeeld. Slecht één studie verscheen in een serieus wetenschappelijk tijdschrift. Het gaat dan om het laboratoriumonderzoek van de Canadese psychologen Porter en Yuille (1996). In hun experiment kregen proefpersonen de opdracht om wel of niet een geënceneerd misdrijf te plegen. Hierna legden ze een mondelinge verklaring af die later op schrift werd gesteld. De verklaringen werden aan de hand van slechts drie SCAN-criteria geanalyseerd (structuur van de verklaring; missende informatie; eerste persoon enkelvoud, verleden tijd). Ware en gelogen verklaringen verschilden hierin niet (Porter & Yuille, 1996).

Driscoll (1994) probeerde met behulp van een veldstudie de kwaliteit van de SCAN vast te stellen. Hij vond dat op basis van tien SCAN-criteria ware en valse verklaringen boven kansniveau correct geclassificeerd konden worden. Dat klinkt bemoedigend, maar allerlei methodologische tekortkomingen maken zijn studie problematisch. Zo kon voor geen van de verklaringen met enige mate van zekerheid

worden vastgesteld of ze ook echt waar of onwaar waren. Studies waarin deze zogeheten *ground truth* ontbreekt, zijn natuurlijk beperkt in de wetenschappelijke bewijskracht die ze hebben (zie bijvoorbeeld Iacono, 1991).

Een tweede veldstudie wordt beschreven in het onderzoeksrapport van Smith (2001) voor het British Home Office, de Engelse overheidsdienst die onder andere verantwoordelijk is voor immigratiecontrole, veiligheid en orde. In dit rapport werden 27 verklaringen geanalyseerd door vijf groepen die zich van SCAN bedienden, waarbij iedere groep uit twee à drie beoordelaars bestond. Drie groepen omvatten experts die in wisselende mate geoefend waren in SCAN. Het ging dan om beginnende, geregelde, en ervaren SCAN-gebruikers. Deze groepen slaagden erin minimaal 80% van de ware en 75% van de gelogen verklaringen juist te beoordelen. Maar de twee groepen zonder enige SCAN-training – en deze groepen bestonden uit ervaren en debuterende rechers – presteerden ook boven kansniveau. Sterker nog: de groep ervaren rechers zonder SCAN-training verschilde qua prestatie niet van de drie SCAN-groepen met training. SCAN lijkt dus niets toe te voegen aan het timmermansoog van ervaren rechers. Ook bij deze resultaten moet een voorbehoud worden gemaakt. Het onderzoek kampt met hetzelfde methodologische euvel als dat van Driscoll: de *ground truth* is onbekend. Opvallend in de tweede veldstudie was bovendien dat er een lage interbeoordelaarsbetrouwbaarheid gevonden werd voor de afzonderlijke SCAN-criteria. Beoordelaars kwamen vaak wel tot dezelfde conclusies, maar baseerden deze conclusie op verschillende criteria. SCAN zet dus blijkbaar niet aan tot een gestandaardiseerde toepassing.

Samengevat kan gezegd worden dat de wetenschappelijke onderbouwing van SCAN mager is, om niet te zeggen afwezig. Systematisch onderzoek ontbreekt, evenals standaardisatie van SCAN zelf. Dit laatste betekent dat een analyse van verklaringen langs de lijnen van SCAN helemaal afhangt van de vorm die een expert er zelf aan geeft. Het is opmerkelijk dat

Noten

1. SVA wordt momenteel gebruikt in allerlei West Europese landen waaronder Oostenrijk, Duitsland, Zweden, Zwitserland en Nederland (Vrij, 2008). SVA bestaat uit vier fases; analyse van dossierinformatie; interview met de getuige; Criteria-Based Content Analysis (CBCA) van zijn of haar verklaring; en een evaluatie van de CBCA-analyse met behulp van de Validity Checklist. Fase 3, de CBCA, vormt zoals gezegd het kernstuk van SVA.
2. Interbeoordelaarsbetrouwbaarheid wordt op uiteenlopende manieren berekend zoals: proportie overeenkomst, correlatie of Cohens kappa's. Ongeacht de methode duidt een score van minder dan .50 op een ondermaatse betrouwbaarheid, .50 op een matige betrouwbaarheid, 0.60 tot .75 op een goede betrouwbaarheid en hoger dan .75 op een zeer goede betrouwbaarheid (Vrij, 2005).
3. Het criterium 'cognitieve operaties' krijgt echter een omgekeerde waarde toegewezen, omdat dit criterium wordt gezien als een aanwijzing voor leugenachtigheid.
4. Zie www.lsiscan.com.
5. Voor een ander voorbeeld van een geautomatiseerd systeem zie Zhou, Burgoon, Nunamaker & Twitchell (2004).

Summary

Verbal veracity assessment tools

G. Bogaard, E.H. Meijer, A. Vrij, H. Merckelbach
Verbal lie detection tools generate a great deal of interest, not only from researchers, but also from police. In this article, we discuss three verbal lie detection methods: Criteria Based Content Analysis (CBCA), Reality Monitoring (RM), and Scientific Content Analysis (SCAN). Especially SCAN is a popular technique that is widely used by the police in various countries. However, in contrast to CBCA and RM, SCAN does not consist of a clear and coherent set of criteria, and so far, there is no solid scientific support for its ability to discriminate between true and false statements. SCAN has also been criticised for its lack of standardisation and low reliability. For these reasons, it is surprising that SCAN is employed on such a wide scale practice.

ondanks het gebrek aan empirische steun er toch op ruime schaal van deze methode gebruik wordt gemaakt. Grootschalige toepassing van de SCAN binnen de politionele opsporing is op haar minst voorbarig.

TOT SLOT CBCA, RM en SCAN zijn gebaseerd op het idee dat ware en gelogen verklaringen te onderscheiden zijn op basis van hun inhoudelijke en stilistische kenmerken. Uit onderzoek naar de CBCA en RM blijkt dat beide instrumenten inderdaad in staat zijn om boven kansniveau ware van gelogen verklaringen te onderscheiden. Dat is bemoedigend. Terughoudendheid is niettemin geboden. Zo wordt de CBCA in de praktijk vaak weinig gestandaardiseerd toegepast en dat brengt problemen met zich mee. Zo specificeert geen van de methoden een afkappunt. Welke score volstaat om te besluiten dat een verklaring waar moet zijn, is aan de beoordeelaar zelf en aan het gewicht dat hij of zij toekent aan bepaalde criteria. RM, daarentegen, blijkt momenteel helemaal niet toegepast te worden in de praktijk en beperkt zich dus enkel tot gebruik in de wetenschap. Redenen hiervoor zijn echter onbekend. Voor de SCAN ontbreekt deugdelijk wetenschappelijk onderzoek. Ook lijkt de betrouwbaarheid van SCAN laag, vanwege haar gebrek aan standaardisatie. De conclusie kan geen andere zijn dan dat van de drie besproken methoden de SCAN het minst onderbouwd is.

Grootschalige toepassing van SCAN is vanwege het ontbreken van degelijk onderzoek voorbarig. De CBCA en RM kunnen wél bijdragen aan de politionele opsporing. Maar daarbij zouden gebruikers niet uit het oog mogen verliezen dat ook deze methoden een foutmarge hebben die in de tientallen procenten belooft. Het is dus een riskante zaak hieraan bewijs voor de rechtszaal te onttelen.

Toekomstig onderzoek zou zich moeten richten op de psychometrische kenmerken van SCAN. Daarnaast is het interessant om te onderzoeken of het mogelijk is om CBCA, RM en wellicht criteria van SCAN in één systeem te integreren.

Mw drs. G. Bogaard is als promovenda verbonden aan het Department klinische Psychologie, sectie Forensische Psychologie van de Universiteit Maastricht.

E-mailadres: glynis.bogaard@maastrichtuniversity.nl

Dr. E.H. Meijer en prof.dr. H. Merckelbach zijn verbonden aan dezelfde sectie, respectievelijk als postdoctoraal onderzoeker en als hoogleraar psychologie.

Prof.dr. A. Vrij is hoogleraar toegepaste sociale psychologie aan de Universiteit van Portsmouth.

Literatuur

- Anson, D.A., Golding, S.L. & Gully, K.J. (1993). Child sexual abuse allegations: reliability of criteria based content analysis. *Law and Human Behavior*, 17, 331-341.
- Bockstaele, M. (2008). Scientific Content Analysis (SCAN). Een nuttig instrument bij verhoren? In L. Smets & A. Vrij (red.), *Het analyseren van de geloofwaardigheid van verhoren: Het gebruik van leugendetectiemethoden* (p. 105-156). Brussel: Politeia.
- Bond, C.F. & DePaulo, B.M. (2006). Accuracy of deception judgments. *Personality and Individual Differences*, 10, 214-234.
- Colwell, K., Hiscock-Anisman, C.K., Memon, A., Taylor, L. & Prewett, J. (2008). Assessment criteria indicative of deception (ACID). An integrated system of investigative interviewing and detecting deception. *Journal of Investigative Psychology and Offender Profiling*, 4, 167-180.
- Driscoll, L. (1994). A validity assessment of written statements from suspects in criminal investigations using the SCAN technique. *Police Studies*, 4, 77-88.
- Elaad, E., Ginton, A. & Ben-Shakhar, G. (1994). The effects of prior expectations and outcome knowledge on polygraph examiners' decisions. *Journal of Behavioral Decision Making*, 7, 272-292.
- Gödert, H.W., Gamer, M., Rill, H. & 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.
- Gumpert, C.H. & Lindblad, F. (1999). Expert testimony on child sexual abuse. A qualitative study of the Swedish approach to statement analysis. *Expert Evidence*, 7, 279-314.
- Horowitz, S.W., Lamb, M.E., Esplin, P.W., Boychuk, T.D., Krispin, O. & Reiter-Lavery, L. (1997). Reliability of criteria-based content analysis of child witness statements. *Legal and Criminological Psychology*, 2, 11-21.
- Iacono, W.G. (1991). Can we determine the accuracy of polygraph tests? In J.R. Jennings, P.K. Ackles & M.G.H. Coles (Eds.), *Advances in Psychophysiology* (Vol. 4, p. 201-207). London: Jessica Kingsley Publishers.
- Johnson, M.K. & Raye, C.L. (1981). Reality Monitoring. *Psychological Review*, 88, 67-85.
- Lamb, M.E., Sternberg, K.J., Esplin, P.W., Hershkowitz, I., Orbach, Y. & Hovav, M. (1997). Criterion-Based Content Analysis: a field validation study. *Child Abuse & Neglect*, 21, 255-264.
- Manzanero, A.L. & Digest, M. (1996). Effects of preparation on internal and external memories. In G. Davies, S. Lloyd-Bostock, M. McMurrin & C. Wilson (Eds.), *Psychology, Law, and Criminal Justice: International developments in research and practice* (p. 56-63). Berlin, Germany: de Gruyter.
- Masip, J., Sporer, A.L., Garido, E. & Herrero, C. (2005). The detection of deception with the reality monitoring approach: a review of the empirical evidence. *Psychology, Crime and Law*, 11, 99-122.
- Meissner, C.A. & Kassin, S.M. (2002). 'He's guilty!' Investigator bias in judgments of truth and deception. *Law and Human Behavior*, 26, 469-480.
- Merckelbach, H. (2004). Telling a good story: fantasy proneness and the quality of fabricated memories. *Personality and Individual Differences*, 37, 1371-1382.
- Nierop, N.M. van, Eshof, P. van de & Brandt, C. (2006). De beoordeling van geloofwaardigheid in zedenzaken. Theorie en praktijk. *Nederlands Juristenblad*, 43, 2456-2464.
- 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.
- Rassin, E., Merckelbach, H. & Crombag, H.F.M. (1997). De Criteria Based Content Analysis (CBCA) als instrument om de geloofwaardigheid van getuigenverklaringen te bepalen. Een kritische notie. *Nederlands Juristenblad*, 42, 1923-1929.
- Roediger, H.L. (1996). Memory illusions. *Journal of Memory and Language*, 35, 76-100.
- Sabbe, J. & Smet, K.D. de (2008). SCAN: revolutionaire leugenontmaskering of steuntje in de rug voor de waarheidsvinding? In L. Smets & A. Vrij (red.), *Het analyseren van de geloofwaardigheid van verhoren: het gebruik van leugendetectiemethoden*. Brussel: Politeia.
- Santtila, P., Roppola, H., Runtti, M. & Niemi, P. (2000). Assessment of child witness statements using Criteria-Based Content Analysis (CBCA): the effects of age, verbal ability, and interviewer's emotional style. *Psychology, Crime and Law*, 6, 159-179.
- Sapir, A. (1996). *The LSI Course on Scientific Content Analysis (SCAN)*. Phoenix, AZ: Laboratory for Scientific Interrogation.
- Schelleman-Offermans, K. & Merckelbach, H. (2010). Fantasy proneness as a confounder of verbal lie detection tools. *Journal of Investigative Psychology and Offender Profiling*, 7, 247-260.
- Smith, N. (2001). Reading between the lines. An evaluation of the Scientific Content Analysis technique (SCAN). *Police Research Series Paper 135*.
- 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. & Koehnken, G. (1989). Criteria Based Content Analysis. In D.C. Rassin (Ed.), *Psychological Methods in Criminal Investigation and Evidence* (p. 217-245). New York: Springer Publishing Company.
- Undeutsch, U. (1967). Beurteilung der Glaubhaftigkeit von Aussagen. In U. Undeutsch (Hrsg.), *Handbuch der Psychologie Vol 11: Forensische Psychologie*. Göttingen, Germany: Hogrefe.
- 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*. Chichester: Wiley.
- Vrij, A., Akehurst, L., Soukara, S. & Bull, R. (2004). Detecting deceit via analysis of verbal and nonverbal behavior in children and adults. *Human Communication Research*, 30, 8-41.
- Vrij, A., Mann, S. & Leal, S. (2008). Reality Monitoring. In L. Smets & A. Vrij (red.), *Het analyseren van de geloofwaardigheid van verhoren: het gebruik van leugendetectiemethoden* (p. 81-84). Brussel: Politeia.
- Zhou, L., Burgoon, J.K., Nunamaker, J.F. & Twitcheil, D. (2004). Automating linguistics-based cues for detecting deception in text-based asynchronous computer mediated communication. *Group Decision and Negotiation*, 13, 81-106.

‘Het scoren van de criteria van SCAN is afhankelijk van de inventiviteit en de kwaliteit van de analist’