

SCAN :

Scientific? Content Analysis

Drs. Ewout Meijer
Faculteit der Psychologie, Universiteit Maastricht

Prof. dr. Aldert Vrij
Faculty of Psychology, University of Portsmouth

Prof. dr. Harald Merckelbach,
Faculteit der Psychologie, Universiteit Maastricht

Definitieve versie, aangeboden aan:
Frits Vlek
Politie & Wetenschap
29 april 2008

1. Vraag

Scientific Content Analysis (SCAN) is een veel gebruikte leugendetectie methode. Zij wordt wereldwijd gebruikt, onder andere door politie en leger in landen als de Verenigde Staten, Canada, België, Israël, Mexico, Singapore, Zuid Afrika en het Verenigd Koninkrijk . Inmiddels zijn er ook in Nederland 40 opsporingsbeambten die in het gebruik van SCAN zijn getraind. Deze recente opkomst van SCAN in Nederland is gedeeltelijk terug te voeren op het werk van Thinka Bethlem en Janny Oosting. In 2005 schreven zij ter afronding van hun studie Tactisch Management aan de Politieacademie hun afstudeerscriptie over leugendetectie. Daarin omschrijven ze onder andere verschillende verbale analysemethoden en concludeerden ze dat de Nederlandse Politie kansen laat liggen op dit vlak. De afstudeerscriptie heeft ertoe geleid dat er in de regio's Amsterdam-Amstelland en Noord- en Oost- Gelderland een pilot heeft gelopen, waarin SCAN is toegepast binnen lopende onderzoeken. In totaal werd bij 118 personen SCAN afgenomen. Op basis van de gegevens uit deze pilot concludeerden de auteurs dat *'verbale analysemethode SCAN kan bijdragen aan de waarheidsvinding in opsporingsonderzoeken.'* (Bethlem e.a., 2006, p. 20) De auteurs bevelen dan ook aan om SCAN in te zetten als hulpmiddel in de opsporing. Om dit op een verantwoorde manier te kunnen doen verdient het volgens de auteurs de voorkeur om eerst wetenschappelijk onderzoek uit te (laten) voeren naar de methode.

Dit rapport is een eerste aanzet daartoe. Het betreft een literatuurstudie naar de wetenschappelijke merites van SCAN. Het beantwoordt de volgende vraag: zijn er in de wetenschappelijke literatuur aanknopingspunten te vinden voor de werkzaamheid van SCAN, en zo ja, wat zijn hierbij de cruciale elementen.

2. SCAN

SCAN is een methode om geschreven verklaringen te beoordelen op waarheidsgehalte en volledigheid. De methode is ontworpen door Avinoman Sapir, en kan toegepast worden op verklaringen van verdachten, slachtoffers en ooggetuigen. Volgens Sapir is de achterliggende rationale dat ware verklaringen uit het geheugen worden opgehaald, terwijl verzonnen of gelogen verklaringen vanuit de fantasie en verbeelding worden opgebouwd. Gelogen verklaringen zouden om deze reden over andere taalkundige kenmerken beschikken dan ware verklaringen, en zo van elkaar te onderscheiden zijn (Sapir, 1996). SCAN hoort hiermee tot de familie verbale analyse methoden, een familie waartoe ook de *Statement Validity Analysis* (SVA), en *Reality Monitoring* (RM) horen.

Startpunt van SCAN is een handgeschreven verklaring. Het is belangrijk dat deze verklaring de eigen woorden van de persoon weergeeft, en dus wordt geproduceerd voordat de persoon door opsporingsambtenaren wordt ondervraagd. Vervolgens wordt deze verklaring beoordeeld op de aan of afwezigheid van een aantal criteria. De aanwezigheid van deze criteria zou duiden op een onware verklaring. Een aantal van deze criteria worden vermeld in Smith (2001) en Vrij (2008). Zo zou bijvoorbeeld het afwijken van het schrijven in de eerste persoon enkelvoud, verleden tijd ('het huis gaat dan in vlammen op', in plaats van 'ik zag dat het huis in vlammen opging') een aanwijzing zijn voor bedrog, net als afwijkend gebruik of weglaten van persoonlijk voornaamwoorden ('parkeerde de auto' in plaats van 'ik parkeerde de auto' en 'mijn vrouw en ik gingen' in plaats van 'wij gingen'). Ook zou een ware verklaring een andere structuur hebben dan een gelogen verklaring. Een ware verklaring zou voor 20% bestaan uit een proloog, voor

50% uit het incident zelf, en voor 30% uit wat er na het incident gebeurde.¹ Hoe groter de afwijking van deze structuur, hoe groter de kans dat een verklaring onwaar is.

De bevindingen van de SCAN analyse worden in een rapport weergegeven. Dit rapport identificeert delen van de verklaring die verduidelijking behoeven en waar in een vervolgvraag op kan worden doorgevraagd. Op deze manier wordt de uitkomst van SCAN gebruikt om richting te geven aan verdere verhoren, meer dan als leugendetectietest.²

¹ Voor een vollediger lijst met criteria zie bijlage 1.

² Dit is de werkwijze van de Politie Amsterdam-Amstelland. Volgens de handleiding van de SCAN workshop is de methode weldegelijk geschikt als leugendetectie methode. Op de eerste pagina van het workbook is te lezen:

Dear student,

During the week you will be introduced to SCAN (Scientific Content Analysis), a very powerful technique to analyze statements in order to find out what the person actually said, what the person didn't say, and if the person is truthful or not. (Sapir, 1996, p.3)

3 Wetenschappelijke verankering

Een goede wetenschappelijke test moet aan een aantal criteria voldoen. De belangrijkste drie criteria zijn betrouwbaarheid, criterium validiteit en constructvaliditeit. Betrouwbaarheid betreft de mate waarin een test onder gelijkblijvende condities dezelfde uitkomst oplevert. In het geval van SCAN bijvoorbeeld of twee verschillende analysten op basis van dezelfde verklaring tot dezelfde conclusies komen. Criterium validiteit betreft de mate waarin een test meet wat die beoogt te meten. In het geval van SCAN gaat het dus om hoe goed de methode ware en onware verklaringen van elkaar kan onderscheiden. Construct validiteit, ten slotte, verwijst naar de theoretische verankering. Op basis van welke psychologische theorie verwacht men dat de methode zal werken.

Tot op heden zijn er drie studies die SCAN op haar wetenschappelijke merites hebben trachten te beoordelen. Slechts één hiervan is in een *peer reviewed* tijdschrift gepubliceerd. Dit betreft het onderzoek van Porter and Yuille uit 1996. Deze auteurs lieten proefpersonen een geënceneerd misdrijf plegen, en onderzochten vervolgens hun verklaringen op basis van drie van de SCAN criteria (structuur, missende informatie en eerste persoon enkelvoud verleden tijd). Deze criteria bleken niet te discrimineren tussen ware en onware verklaringen. Problematisch bij deze studie was echter dat de verklaringen niet volgens de SCAN voorkeurswijze tot stand gekomen. Het waren mondelinge verklaringen die pas daarna op schrift waren gesteld.

Naast de laboratoriumstudie van Porter and Yuille bestaan er ook twee veldstudies naar de werking van SCAN. Deze studies gebruiken verklaringen uit dossiers waarvan later is vastgesteld of zij waar of onwaar waren, bijvoorbeeld aan de hand van een bekentenis van de verdachte. Driscoll (1994) analyseerde 30 geschreven verklaringen

met behulp van SCAN. Deze verklaringen waren aan de hand van later verzamelde bewijsmiddelen ingedeeld als *apparently accurate* (AA) of *doubtful* (D). Van de 11 AA verhalen classificeerde SCAN er 8 (73%) correct, en van de 19 D verhalen 18 (95%).

De meest uitgebreide studie naar SCAN is uitgevoerd door Nicky Smith, verbonden aan de Britse Home Office (Smith, 2001). In deze studie beoordeelde vijf verschillende groepen rechercheurs 27 verklaringen. Drie van de vijf groepen bestonden uit in SCAN getrainde rechercheurs met verschillende ervaringsniveau's. De andere twee groepen bestonden uit rechercheurs zonder SCAN training; één groep uit ervaren, en één groep uit nieuwe rechercheurs. Vier van de groepen bestonden uit drie personen en één uit twee. Iedere groep kwam gezamenlijk tot één conclusie. Alle vijf groepen presteerde boven kansniveau. De drie SCAN groepen verschilden onderling niet van elkaar, en classificeerde ten minste 80% van de ware verklaringen en 75% van de onware verklaringen correct. Hiermee deden de SCANers het beter dan de onervaren rechercheurs zonder SCAN training, maar even goed als de ervaren rechercheurs zonder SCAN training. SCAN presteerde dus weliswaar boven kansniveau, het voegde niets toe aan het oordeel van ervaren rechercheurs.

Op het eerste gezicht lijken de resultaten van de veldstudies van Driscoll (1994) en Smith (2001) het gebruik van SCAN te steunen. Beide studies kampen echter met een groot methodologisch probleem dat samenhangt met selectie van de verklaringen. Al deze verklaringen kwamen uit echte dossiers, en de SCAN uitslag heeft invloed gehad op het verloop van het verhoor en onderzoek. Er is in deze dossiers dus eerder verder gerechercheerd als de SCAN analyse daar aanleiding toe gaf. De kans dat er vervolgens bewijsmiddelen gevonden worden is daarmee groter naarmate men op basis van SCAN

concludeert de verklaring onwaar is. Verzamelt men nu alle zaken waarin er bewijs is om de SCAN uitkomst hieraan te staven, zal dit in hoge mate overeen komen. Informatie over het aantal vals negatieve (onware verklaring wordt als waar aangemerkt) uitkomsten ontbreekt echter. Bij deze SCAN uitslag wordt er immers niet verder gerechercheerd, en zullen er dus ook geen bewijsmiddelen worden gevonden. Hetzelfde geldt voor de vals positieve (ware verklaring wordt als onwaar aangemerkt) uitkomsten. Bij deze SCAN uitslag wordt weliswaar verder gerechercheerd, maar geen bewijs gevonden, omdat dit er niet is. Wegens het ontbreken van bewijs zullen deze verklaringen nooit in de te onderzoeken steekproef worden opgenomen. De selectie van de verklaringen is dus sterk bevooroordeeld ten opzichte van correcte SCAN uitkomsten. De uitkomst van zulke studies zullen dan ook een overschatting zijn van de werkelijke nauwkeurigheid (zie ook Iacono, 1991). Om deze reden moeten de uitkomsten van de studies van Driscoll (1994) en Smith (2001) dan ook met grote voorzichtigheid geïnterpreteerd worden.

Eén andere bevinding uit het onderzoek van Smith (2001) is nog het vermelden waard. Er was weinig overeenstemming tussen de criteria die de verschillende SCAN groepen gebruikten om tot hun oordeel te komen. Alhoewel ze wel vaak tot dezelfde uitslag kwamen, baseerden zij dat op verschillende criteria. Een lage (interbeoordelaar) betrouwbaarheid dus, extra opmerkelijk omdat de groepen al uit meerdere personen bestonden die gezamenlijk tot één oordeel kwamen.

Concluderend kan gezegd worden dat de wetenschappelijke verankering van SCAN ontbreekt. Onderzoek van acceptabel niveau ontbreekt, en de enige houdbare conclusie op basis ervan lijkt dat SCAN weinig meerwaarde heeft ten opzichte van het oordeel van ervaren rechercheurs. De betrouwbaarheid is laag, waarschijnlijk doordat

standaardisatie ontbreekt. Verder is het onduidelijk hoe de verschillende SCAN criteria zich tot het theoretisch kader te verhouden. Zo lijkt bijvoorbeeld een criterium als 'ontbreken van ontkenning van aantijgingen' weinig te maken te hebben met het door Sapir gegeven theoretisch kader, namelijk dat verklaringen afkomstig uit het geheugen andere kenmerken bezitten dan verklaringen afkomstig uit de fantasie. Het ontbreekt SCAN in ieder geval aan drie belangrijke psychometrische eigenschappen, namelijk betrouwbaarheid, criterium- en construct validiteit.

4. Andere verbale analyse methoden

Naast SCAN zijn er nog twee verbale analyse methoden. Dit zijn Statement Validity Analysis / Criteria Based Content Analysis (SVA / CBCA) en Reality Monitoring (RM).

SVA verwijst naar een procedure om uitgetypte versies van een semi-gestructureerd verhoor op waarheid te beoordelen. De methode vindt zijn oorsprong in de jaren 60 (Undeutsch, 1967) en wordt tegenwoordig op grote schaal gebruikt in landen als Zweden en Duitsland. Oorspronkelijk is de methode ontworpen om aangiftes van seksueel geweld door kinderen te beoordelen, maar tegenwoordig wordt ze ook gebruikt om verklaringen van volwassen getuigen en verdachten te beoordelen.

In zijn huidige vorm is de belangrijkste component van SVA de CBCA, een lijst met negentien criteria (Kohnken en Steller, 1988; zie bijlage 1). Aanwezigheid van deze criteria in een uitgetypte versie van een verhoor duidt op waarheid. CBCA is gebaseerd op de zogenaamde Undeutsch hypothese: een verklaring afkomstig uit het geheugen verschilt in vorm en inhoud van een verzonnen verklaring (Steller, 1989). Vrij (2008) heeft de verschillende studies naar de nauwkeurigheid van CBCA op een rijtje gezet. De nauwkeurigheid varieerde van 54% tot 90% en was gemiddeld 70%. Op basis hiervan oordeelt Vrij dat CBCA in de politionele opsporing gebruikt kan worden.

Opvallend is dat een aantal van de CBCA criteria overlapt met de SCAN criteria. CBCA onderzoek laat zien dat ware verklaringen meer emoties bevatten, een criterium dat SCAN ook bevat. CBCA onderzoek laat echter ook zien dat op een aantal SCAN criteria geen verschil, of zelf het tegenovergestelde gevonden wordt. Zo voorpelt SCAN bijvoorbeeld dat spontane correcties duiden op een onware verklaring, terwijl CBCA

onderzoek laat zien dat in onware verklaringen juist minder spontane correcties worden gemaakt.

De tweede verbale analyse methode is Reality Monitoring (RM). Dit concept verwees oorspronkelijk naar de cognitieve processen die een rol spelen bij het bepalen of iemand een situatie daadwerkelijk heeft meegemaakt of zich dit slechts heeft ingebeeld (Johnson en Raye, 1981). Achterliggende rationale is dat herinneringen aan situaties die iemand heeft meegemaakt scherp en levendig zijn en bijvoorbeeld veel sensorische en contextuele informatie bevatten (ik weet nog dat het heel koud was en de wind in mijn gezicht striemde). Situaties die zijn ingebeeld zouden deze kenmerken ontberen, en juist over veel cognitieve operaties beschikken (volgens mij had ik mijn jas aan dus het zal wel koud geweest zijn). Met behulp van RM hoopte men oorspronkelijk om valse van echte herinneringen te kunnen onderscheiden, iets wat overigens maar matig lukte (Schooler, Gerhard en Loftus 1986; Porter, Yuille en Lehman 1999).

Sinds begin jaren 90 is RM ook onderzocht op het vermogen om ware van onware verklaringen te onderscheiden. Ware verklaringen komen immers voort uit herinneringen aan situaties die zijn meegemaakt, terwijl onware verklaringen uit ingebeeld situatie voort kunnen komen. Ware en onware verklaringen zouden om deze reden op basis van de RM criteria van elkaar verschillen. Deze RM criteria zijn weergegeven in bijlage 1. De nauwkeurigheid gevonden in een achttal studies varieert van 61% tot 82% met een gemiddelde van 69% (Vrij, 2008; zie ook Masip e.a., 2005). Hiermee verschilt RM wat betreft nauwkeurigheid niet veel van CBCA.

Een aantal van RM criteria overlapt met criteria van CBCA. Zo komt bijvoorbeeld het CBCA criterium *contextual embeddings* overeen met de RM criteria

spatial information en *temporal information*. De methoden lijken elkaar echter ook aan te kunnen vullen: CBCA bevat veel cognitieve en motivationele criteria, terwijl RM zich voornamelijk richt op het geheugen. Onderzoek naar het combineren van de twee lijkt daarmee zeer relevant (Sporer, 2004; zie ook Sporer, 1997; Vrij e.a., 2004).

5. Conclusies

SCAN, SVA(CBCA) en RM zijn verbale analyse methoden die grote overlap vertonen. Allen zijn gebaseerd op de hypothese dat ware verklaringen uit het geheugen komen en daarom over andere kenmerken beschikken dan onware verklaringen, die uit de fantasie afkomstig zijn. Van deze drie methoden is SCAN wetenschappelijk het minst onderbouwd. Er is geen duidelijke, samenhangende set van criteria, en het is onduidelijk hoe de criteria zich verhouden tot het theoretisch kader. SCAN lijkt vooral gebaseerd te zijn op de intuïtie van de bedenker Sapir, en onderzoek van kwalitatief acceptabel niveau ontbreekt.

Voor de werkzaamheid van CBCA en RM bestaat een redelijke verzameling empirisch bewijs. Dit betreft voornamelijk laboratorium onderzoek, en kwalitatief hoogstaande veldstudies ontbreken. Zowel CBCA als RM presteren boven kansniveau. Op basis hiervan kunnen ze worden ingezet om richting te geven aan politionele opsporing. Gezien de aanzienlijke foutenmarge moeten de uitkomsten wel met grote voorzichtigheid geïnterpreteerd worden.

6 Aanbevelingen

Voordat de Nederlandse politie besluit om het gebruik van verbale analyse methoden voort te zetten en of uit te bereiden dient er wetenschappelijk onderzoek plaats te vinden.

Dit onderzoek zou de volgende twee vragen moeten beantwoorden:

- 1) Wat zijn de psychometrische kenmerken van SCAN
- 2) Hoe verhouden SCAN, CBCA en RM zich tot elkaar, en zijn deze mogelijk te combineren

Om de continuïteit van het onderzoek te waarborgen verdient het de voorkeur dit onderzoek in de vorm van een AiO positie vorm te geven. Deze positie zal zich op twee lijnen concentreren. Als eerste zullen in Maastricht verschillende laboratoriumstudies worden uitgevoerd. Parallel hieraan zal in samenwerking met politie Amsterdam-Amstelland veldstudies worden uitgevoerd. Het uiteindelijke doel is om tot de best mogelijke, wetenschappelijk verankerde verbale analyse methoden te komen. Want zoals Sapir zelf in zijn *workbook* meldt: *'If a technique doesn't work, we shouldn't use it'* (Sapir, 1996, p. 23).

Referenties

- Bethlem, T., Oosting, J., & Teernstra, M. (2006). Eindrapportage waarheidsvinding via woordanalyse. Unpublished document.
- Driscoll, L. N. (1994). A validity assessment of written statements from suspects in criminal investigations using the SCAN technique. *Police Studies*, 17, 77-88.
- 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, pp. 201-207). London: Jessica Kingsley Publishers.
- Johnson, M. K., & Raye, C. L. (1981). Reality Monitoring. *Psychological Review*, 88, 67-85.
- Köhnken, G., & Steller, M. (1988). The evaluation of the credibility of child witness statements in German procedural system. In G. Davies & J. Drinkwater (Eds.), *The child witness: Do the courts abuse children?* (pp. 37-45). Leicester, UK: British Psychological Society.
- 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 and Law*, 11, 99-122.
- 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, 443-459.
- Porter, S., Yuille, J. C., & Lehman, D. R. (1999). The nature of real, implanted, and fabricated memories for emotional childhood events: implications for the recovered memory debate. *Law and Human Behavior*, 23, 517-537.

- Sapir, A. (1996). *The LSI course on scientific content analysis (SCAN)*. Phoenix, AZ: Laboratory for Scientific Interrogation.
- Schooler, J. W., Gerhard, D., & Loftus, E. F. (1986). Qualities of the unreal. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 171-181.
- Smith, N. (2001). *Reading between the lines: An evaluation of the scientific content analysis technique (SCAN)*. London: UK Home Office, Research, Development and Statistics Directorate.
- Sporer, S. L. (1997). The less traveled 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 the detection of deception. In L. A. Stroemwall & P. A. Granhag (Eds.), *The detection of deception in forensic context*. Cambridge: Cambridge University Press.
- Steller, M. (1989). Recent developments in statement analysis. In J. C. Yuille (Ed.), *Credibility Assessment* (pp. 135-154). Deventer, The Netherlands: Kluwer.
- Undeutsch, U. (1967). Beurteilung der Glaubhaftigkeit von Aussagen. In U. Undeutsch (Ed.), *Handbuch der Psychologie* (pp. 26-181). Göttingen, Germany: Hogrefe.
- Vrij, A. (2008). *Detecting lies and deceit. Pitfalls and opportunities*. Chichester: Wiley.
- Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2004). Let me inform you how to tell a convincing story: CBCA and Reality Monitoring scores as a function of age, coaching and deception. *Canadian Journal of Behavioral Science*, 36, 113-126.

Bijlage 1: SCAN, CBCA en RM criteria (volgens Vrij, 2008)

SCAN criteria:

1. Denial of allegations
2. Social introduction
3. Spontaneous correction
4. Lack of conviction or memory
5. Structure of the statement
6. Emotions
7. Objective and subjective time
8. Out of sequence and extraneous information
9. Missing information
10. First person singular, past tense
11. Pronouns
12. Change in language

CBCA criteria:

1. Logical structure
2. Unstructured production
3. Quantity of details
4. Contextual embedding
5. Description of interactions
6. Reproduction of conversation
7. Unexpected complications during the incident
8. Unusual details
9. Superfluous details
10. Accurately reported details misunderstood
11. Related external associations
12. Accounts of subjective mental state
13. Attribution of perpetrator's mental state
14. Spontaneous corrections or additions
15. Admitting lack of memory
16. Raising doubts about one's own testimony
17. Self-deprecation
18. Pardoning the perpetrator
19. Details characteristic of the offence

RM criteria:

1. Clarity
2. Perceptual information
3. Spatial information
4. Temporal information
5. Affect
6. Reconstructability of the story
7. Realism
8. Cognitive operations