Scientific Content Analysis (SCAN) cannot distinguish between truthful and fabricated accounts of a negative event

Glynis Bogaard[1*], Ewout H. Meijer[1], Aldert Vrij[2], Harald Merckelbach[1]

[1]Maastricht University, Netherlands, [2]Psychology, University of Portsmouth, United Kingdom

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Author contribution statement

All authors have made substantial contributions to the conception and design of the work, GB acquired and analysed the data, and all authors interpreted the data and revised the article concerning its content and approved the current version.

Keywords

deception detection, scan, Scientific Content Analysis, Lie Detection, verbal cues, verbal credibility assessment

Abstract

Word count: 108

The Scientific Content Analysis (SCAN) is a verbal veracity assessment method that is currently used worldwide by investigative authorities. Yet, research investigating the accuracy of SCAN is scarce. The present study tested whether SCAN was able to accurately discriminate between true and fabricated statements. To this end, 117 participants were asked to write down one true and one fabricated statement about a recent negative event that happened in their lives. All statements were analyzed using 11 criteria derived from SCAN. Results indicated that SCAN was not able to correctly classify true and fabricated statements. Lacking empirical support, the application of SCAN in its current form should be discouraged.

Funding statement

Ethics statement

(Authors are required to state the ethical considerations of their study in the manuscript including for cases where the study was exempt from ethical approval procedures.)

Did the study presented in the manuscript involve human or animal subjects: Yes

Please state the full name of the ethics committee that approved the study. If the study was exempt from this requirement please state the reason below.
The study was approved by the standing ethical committee of the Faculty of Psychology and Neuroscience, Maastricht University.

Please detail the consent procedure used for human participants or for animal owners. If not applicable, please state this.
Upon arrival at the lab, all participants provided written informed consent

Please detail any additional considerations of the study in cases where vulnerable populations were involved, for example minors, persons with disabilities or endangered animal species. If not applicable, please state this.
Not applicable

Scientific Content Analysis (SCAN) cannot distinguish between truthful and fabricated accounts of a negative event

Glynis Bogaard[1], Ewout H. Meijer[1], Aldert Vrij[2] and Harald Merckelbach[1]

[1] *Clinical Psychological Science, Maastricht University, Maastricht, the Netherlands*
[2] *Psychology, University of Portsmouth, Portsmouth, the United Kingdom*

*Corresponding author. Email: Glynis.bogaard@maastrichtuniversity.nl

26 **Abstract**
27 The Scientific Content Analysis (SCAN) is a verbal veracity assessment method that
28 is currently used worldwide by investigative authorities. Yet, research investigating
29 the accuracy of SCAN is scarce. The present study tested whether SCAN was able to
30 accurately discriminate between true and fabricated statements. To this end, 117
31 participants were asked to write down one true and one fabricated statement about a
32 recent negative event that happened in their lives. All statements were analyzed using
33 11 criteria derived from SCAN. Results indicated that SCAN was not able to correctly
34 classify true and fabricated statements. Lacking empirical support, the application of
35 SCAN in its current form should be discouraged.

Introduction

Research has revealed that non-verbal cues (e.g., behavioural cues such as gaze aversion, sweating) are faint and differences between liars and truth tellers are small at best (DePaulo et al., 2003; Sporer and Schwandt, 2007). However, findings about verbal cues are less variable and are more strongly related to deception (Vrij, 2008; Bond and DePaulo, 2006; Vrij, 2008). Verbal cues (or content cues) are cues that can be found in the content and meaning of a statement, such as the number of details that are included in a story (e.g., he had a large spider tattoo in his neck). Indeed, lying has been shown to result in qualitative differences between deceptive and truthful language. As a result, various verbal credibility assessment tools have been developed that address these content criteria within statements. Although the exact content criteria included may differ depending on the method, the procedure is highly similar. The presence of the criteria within the statements is carefully checked, and based on the presence or absence of the various criteria, a conclusion is drawn about its truthfulness.

One example of such a content criterion is "quantity of details". In order to fulfil this criterion, a statement has to be rich in details, such as mentioning places (e.g., it happened in the kitchen), times (e.g., on Sunday evening at 8 p.m.), descriptions of people and objects (e.g., a tall man with bright blue eyes), etc. Additionally, deceit has been related to the use of fewer personal pronouns (e.g., using "the house" instead of "our house") and fewer negations (e.g., no, never, not), using less perceptual information (e.g., "I could smell the alcohol in his breath"), less details overall and shorter statements (Amado et al., 2015; Hauch et al., 2014; Masip et al., 2005; Newman et al., 2003). As mentioned previously, several methods have been developed to address these issues.

Two well-established credibility assessment tools that tap into such content differences are the Criteria Based Content Analysis (CBCA) and Reality Monitoring (RM). For CBCA, two theoretical assumptions have been presented by Köhnken (1996). First, lying is seen as more cognitively challenging that telling the truth. Second, liars are expected to be more concerned with impression management than truth tellers. More precisely, a first subset of CBCA criteria is included because they are deemed too difficult to fabricate (e.g., descriptions of interactions with the perpetrator). Hence, their presence in a statement indicates an actual experience. The remainder of the CBCA criteria are concerned with the way an interviewee presents his or her story. It is expected that liars are concerned with how they are viewed by others and therefore leave out information that can possibly damage their view of being an honest person (e.g., mentioning self-deprecating information). Consequently, a truthful person is more likely to include these criteria in their statement than a deceptive person. RM, in contrast, is derived from memory research and holds that memories of real events are obtained through sensory processes, making them more clear, sharp and vivid. Fabricated statements, on the other hand, are the result of fantasy and are usually more vague and less concrete (Johnson and Raye, 1981). Indeed, various studies reported supportive evidence for these methods. Their overall accuracy for detecting deceit varies around 70%, and is considerably higher than chance level (Amado et al., 2015; Johnson and Raye, 1981; Masip et al., 2005; Steller and Köhnken, 1989; Undeutsch, 1967; Vrij, 2005).

85 Despite the research showing above chance accuracy for CBCA and RM, their field
86 use seems limited. A third method - that is used by Law enforcement worldwide - is
87 Scientific Content Analysis (SCAN). SCAN was developed by former Israeli
88 polygraph examiner Avinoam Sapir (2005), who - based on his experience with
89 polygraph examinees - argues that people who tell the truth differ from liars in the
90 type of language they use. Based on these assumed differences, Sapir developed
91 criteria that, according to him, can assist in differentiating between true and fabricated
92 statements, but without reporting a theoretical foundation as to why these specific
93 criteria should differ. For example, SCAN includes the criterion "social introduction".
94 It is argued that people who are described in the statement should be introduced with
95 name and role (e.g., My friend, John). If a person leaves out information (e.g., We
96 stole the key), so leaving out the name, role or both, this indicates deception. Another
97 criterion is the "structure of the statement". According to SCAN, 20% of the
98 statement should consist of information that led up to the event, 50% should be about
99 the main event and 30% of the statement should be about what happened after the
100 event. The more the statement deviates from this structure, the higher the likelihood
101 that the statement is deceptive. Yet, in contrast to CBCA and RM, no theoretical
102 rationale is presented, and there is no evidence that these criteria are actually
103 diagnostic (Bogaard et al., 2014; Nahari et al., 2012; Vanderhallen et al., 2015).
104
105 Research about SCAN is scarce, although the method is used worldwide (e.g.,
106 Australia, Belgium, Canada, Israel, Mexico, UK, US, the Netherlands, Qatar,
107 Singapore, South Africa) and is also used by federal agencies, military law
108 enforcement, private corporations, and social services (retrieved from
109 www.lsiscan.com/id29.htm). Moreover, the third author asked during an investigative
110 interviewing seminar which lie detection tool was used by the practitioners in the
111 audience. These practitioners came from many different countries and the most
112 frequent answer was SCAN (Vrij, 2008). In a typical SCAN procedure, the examinee
113 is asked to write down "everything that happened" in a particular period of time, to
114 get a "pure version" of the facts (Sapir, 2005). This pure version is typically obtained
115 without the interviewer interrupting or influencing the examinee. Next, a SCAN
116 trained analyst investigates a copy of the handwritten statement, using several criteria
117 that are described throughout the SCAN manual (Sapir, 2005). Criteria that are
118 present within the written statements are highlighted according to a specific colour
119 scheme, circled or underlined. The presence of a specific criterion can either indicate
120 truthfulness or deception, depending on the criterion itself. This SCAN analysis is
121 then used to generate questions that could elucidate important details within the
122 statement, and/or to make a judgment of the veracity of the statement. Although
123 SCAN is used worldwide, it lacks a well-defined list of criteria, as well as a
124 standardized scoring system. Bogaard et al. (2014) has shown that 12 criteria
125 primarily drove SCAN in sexual abuse cases, largely overlapping with the criteria list
126 described in Vrij (2008). Only six published studies examined the validity of SCAN
127 (Bogaard et al., 2014; Driscoll, 1994; Nahari et al., 2012; Porter and Yuille, 1996;
128 Smith, 2001; Vanderhallen et al., 2015) of which only four were published in peer
129 reviewed journals. The two studies that were not published in peer reviewed journals
130 (Driscoll (1994) and Smith (2001)) were both field studies investigating suspect
131 statements.
132
133 Driscoll (1994) investigated 30 statements that were classified as either apparently
134 accurate or doubtful. With the help of SCAN, 84% of the statements could be

4

classified correctly. In the study of Smith, five groups of experts were asked to analyse 27 statements. These statements were previously classified by police officers as truthful, false or undecided. This classification was made on the basis of confessions and supportive evidence. Three groups consisted of SCAN trained officers that had minimal, moderate or extensive experience with using SCAN. The two other groups consisted of newly recruited officers and experienced officers. The first three groups used SCAN to analyse the statements, while the latter two groups judged the veracity of the statements without using SCAN. Overall, the SCAN groups correctly judged 78% of the statements, which was similar to the accuracy of the experienced officers. At first glance, these results seem to support SCAN. Yet, in both studies ground truth of the statements was unknown and statements were categorized as either truthful or doubtful without having hard evidence supporting this categorisation. Moreover, it cannot be excluded that the SCAN outcome influenced the course of the investigation, and therefore the confessions and supporting evidence that was gathered. A typical problem that can occur in such studies is that errors are systematically excluded from the sample. For example, if a statement is erroneously judged as truthful, no further investigation takes place. This means that no evidence will be found revealing that an error has been made, and such erroneous classifications are then excluded from the sample. This way of selecting the sample may therefore be biased to overestimate SCAN's accuracy (for more information see Iacono, 1991; Meijer et al., In press). Moreover, in Smith's study, it was unclear whether the three undecided statements were included in the reported analyses (Armistead, 2011).

The following four studies investigating SCAN were published in peer-reviewed journals. Porter and Yuille (1996) resolved the problem of ground truth by asking participants to commit a mock crime. However, they only investigated three SCAN criteria (i.e., unnecessary connectors, use of pronouns and structure of the statement), and results indicated no significant differences between true and fabricated statements concerning these criteria. Nahari et al. (2012) asked six independent raters to assess the presence of 13 SCAN criteria within various true and fabricated statements. Results showed that SCAN did not discriminate between truthful and fabricated statements, a conclusion that was also supported by Bogaard et al. (2014). In their study, participants were asked to write down one truthful and one fabricated autobiographical statement about a negative event that recently happened to them. Two raters indicated the presence of 12 SCAN criteria, but no significant differences emerged between truth tellers and liars. Vanderhallen et al. (2015), finally, asked SCAN trained police officers to classify four statements as either truthful or deceptive based on SCAN, and compared their accuracy to students and police officers who made this classification without the help of SCAN. The SCAN group had an average accuracy of 68%, police officers without SCAN 72%, and students 65%. The accuracy of the SCAN group did not significantly differ from the police officers who did not use SCAN. Consequently, from these results it was concluded that SCAN did not have an incremental value in detecting deceit.

Given that SCAN is used worldwide in police investigations, providing support, or the lack thereof, is not trivial (Meijer et al., 2009). Using a data set of 234 statements, the current study aimed at extending previous SCAN findings, and to investigate whether the different SCAN criteria can actually discriminate between truthful and fabricated statements. Although Nahari et al. (2012), Bogaard et al. (2014) and Vanderhallen et

185 al. (2015) investigated SCAN, Bogaard et al. mainly focused on the SCAN total
186 scores, and not on the separate criteria, or the accuracy of SCAN. Separate criteria
187 scores were reported, but their power was too low to make any conclusions from these
188 results. In contrast, Nahari et al. asked participants to perform a mock crime, meaning
189 that the statements that were analysed with SCAN were restricted to "false denials"
190 (i.e., people who performed the mock crime but lied about it). Moreover, in the study
191 of Vanderhallen et al. four statements on traffic accidents were used. The statements
192 included in our study are broader than false denials or traffic accidents, as we
193 requested participants to write about a negative autobiographical event. In this way,
194 participants not only reported false denials, but also false allegations (i.e., stating they
195 fell victim to a crime, while in fact they were not). Participants could report about
196 whatever they preferred, thereby including various topics, as would also be the case in
197 police investigations where SCAN is usually applied.
198
199 **Method**
200 *Participants*
201 All participants ($N = 117$) were first and second year health sciences students (i.e.,
202 Mental health or Psychology) of Maastricht University (37 men). The data of 85
203 participants were collected specifically for this study, while the remaining 32 came
204 from the control group of Bogaard et al. (2014). Instructions for these 2 datasets were
205 identical, and they were combined to increase power. We report the analysis for the
206 entire sample below, but also include the findings for the new dataset in appendix B.
207 Participants could choose whether they wanted to receive one course credit or a 7,5
208 Euro gift voucher for their participation. Approximately 50 students chose the gift
209 voucher over the course credit. All participants read and signed a letter of Informed
210 Consent before they took part in this study. Participants had a mean age of 21 years
211 ($SD = 2.35$). The experiment was approved by the appropriate standing ethical
212 committee.
213
214 *Procedure*
215 Upon arrival in the lab, participants were told that the study was about the accuracy of
216 verbal lie detection methods. Participants were asked to write about a truthful and a
217 fabricated event. The order in which participants wrote these statements was
218 randomized. Approximately half of the participants started with the truthful statement,
219 the other half started with the fabricated statement. For the truthful statement
220 participants received the following instruction: "For this study we ask you to think
221 about an event you actually experienced. More specifically, this event should be about
222 a recent negative experience; think about a financial, emotional or physical negative
223 event you've been through the past months." For the fabricated statement participants
224 received the following instruction: " For this study we ask you to think about an event
225 that you have not actually experienced. This event should be about a recent negative
226 experience; think about a financial, emotional or physical negative event you could
227 have been through the past months. This event should not be based on something that
228 actually happened to you or your friends or family. Please pretend as if this event took
229 place somewhere in the previous months. Although the story should be fabricated, the
230 statement should consist of a realistic scenario." After the instruction, participants had
231 the opportunity to think about a real and a fabricated story for a maximum of 5
232 minutes. Participants were assured that their stories would be treated confidentially
233 and anonymously. They were told that the length of the stories should be

234   approximately one written page (A4). No time limit was set for the production of the
235   statements.
236
237   *Statement coding*
238   After participants finished their stories, these were analysed by four raters. One rater
239   completed the three-day SCAN course. The other three raters received a 2-hour
240   training about SCAN, using the SCAN manual (Sapir, 2005), given by the SCAN
241   trained rater. Moreover, they received the appropriate pages of Vrij (2008a) about
242   SCAN (Chapter 10; 282-287). During the training all 12 criteria were discussed
243   separately and examples of the specific criteria were presented and discussed. Next,
244   raters received two practice statements of 1 page each, and were asked to analyze
245   these statements. After all raters analyzed these statements, their analyses were
246   discussed and questions they still had about SCAN were answered. When the training
247   was completed, raters started analyzing the statements.
248
249   Although the raters were not blind to the aim of the study, they were blind to the
250   veracity of the statements. The first author served as one of the raters, the other raters
251   were not otherwise involved in the study and were research assistants of the first
252   author. The rater who completed the original SCAN training scored all 234
253   statements, while the other three raters scored approximately 80 statements each. In
254   order to control for potential order effects, the sequence of the statements to be scored
255   was varied from rater to rater. Rater A scored all statements in the order of 1 to 234,
256   while the other raters scored the statements in the reverse order (rater B started from
257   79 to 1, rater C started from 157 to 80 and rater D from 234 to 158).
258
259   A total of 12 criteria (Vrij, 2008a) were coded within the statements. According to
260   SCAN, seven of these criteria indicate truthfulness: (1) denial of allegations, (2)
261   Social introductions, (3) Structure of the statement, (4) Emotions, (5) Objective and
262   subjective time, (6) First person singular, past tense, (7) Pronouns, while the
263   remaining five indicate deception: (8) Change in language (9) Spontaneous
264   corrections (10) Lack of conviction or memory (11) Out of sequence and extraneous,
265   (12) Missing information. See appendix A for a complete description of the different
266   criteria. All criteria that are expected to indicate truthfulness were scored on a 3-point
267   scale ranging from 0 (not present) to 2 (strongly present), while the 5 criteria that are
268   expected to indicate deception were scored in reverse, ranging from -2 (strongly
269   present) to 0 (not present). By using this scoring system, a higher score indicates a
270   higher likelihood that the statement is truthful and vice versa.
271
272   **Results**
273   *Inter-rater reliablility*
274   Inter-rater reliability was calculated by means of Cohen's Kappa for each of the 12
275   separate criteria. The Kappa values for the truthful statements varied from 0.60 to 1
276   with an average Cohen's Kappa of 0.77. The Kappa values for the fabricated
277   statements varied from 0.65 to 1, with an average kappa of 0.78. These results
278   indicated that there is high agreement between the raters (Landis and Koch, 1977).
279   Because variance was low for several criteria, Cohen's Kappa could give a distorted
280   image of the actual inter-rater reliability. Therefore, we also included inter-rater
281   agreement calculated by means of percentage agreement and its presence in the
282   statement. Therefore, we dichotomized the original data set with presence coded as 1
283   and absence as 0. High agreement was achieved for all SCAN criteria ranging from

284 80.34% to 100% with an average of 90.56%. The scoring of the three raters was
285 always compared to those of the rater that completed the SCAN training. As
286 reliability showed to be sufficient, this also showed that our 2-hour SCAN training
287 was sufficient to score the investigated SCAN criteria reliably.
288
289 *Data analysis*
290 Because the inter-rater reliability was high, we averaged the scores of the two raters
291 for each criterion. Due to the nature of our instructions (i.e., autobiographical
292 statements) the first criteria could not have been coded in the statements. As such, we
293 have left out "denial of allegations" in the following analysis. Next, we calculated the
294 sum scores for each statement by summing up the averaged scores of the separate
295 criteria. To investigate the discriminability of SCAN, we conducted several
296 Generalized Estimation Equation (GEE) analyses (see for example Burton et al.,
297 1998); one for each separate criterion. Moreover, we conducted a paired samples t-test
298 for the sum score, and a discriminant analysis to test SCAN's predictive power
299 concerning the veracity of statements.
300
301 *Number of words*
302 The length of the statements did not significantly differ between the true ($M = 265.42$;
303 $SD = 85.48$) and fabricated statements ($M = 261.86$; $SD = 88.12$) [$t(116) = 0.63$, $p = $
304 0.53, $d = 0.04$].
305
306 *SCAN criteria scores*
307 Table 1 shows the mean differences in each of the SCAN criteria as a function of
308 veracity. To analyze the separate criteria, we have dichotomized our data by recoding
309 presence as 1 (regardless of whether the score was a 1 or a 2) and absence as 0. Next,
310 we analyzed the data with GEE in order to investigate the differences between truthful
311 and fabricated statement for each of the separate criteria. Due to very low variability
312 of the criterion "pronouns" (i.e., it was present in almost all of the statements), this
313 criterion was left out of the analysis. To correct for multiple testing we used an alpha
314 level of .01. As Table 2 shows, only one criterion significantly differed between the
315 statements, namely "Change in language". Participants included more changes in
316 language in their fabricated statements compared to their truthful statements. This
317 criterion was present in 29 out of 117 fabricated statements (24.8%) and in 14 out of
318 117 true statements (12%). In Appendix B (Table B1) we have presented the results
319 of only the new data, and results showed again that "Change in language"
320 significantly differed between statements.
321
322 *SCAN sum scores*
323 Results indicated that there were no differences in SCAN sum scores between true (M
324 $= 5.33$; $SD = 2.10$) and fabricated ($M = 5.15$; $SD = 2.25$) statements [$t(116) = 0.77$, p
325 $= 0.44$, $d = 0.12$].
326
327 Lastly, we conducted a discriminant analysis to investigate whether the SCAN criteria
328 were able to predict veracity. As can be seen in Table 3, only one significant mean
329 difference was observed, and this was for "Change in language" ($p < 0.01$). The
330 discriminate function revealed a low association between veracity and SCAN criteria,
331 only accounting for 7.20 % of the variability. Closer analysis of the structure matrix
332 revealed that three criteria that had moderate discriminant loadings (i.e., Pearson
333 coefficients), these were – again - "Change in language" (0.664), "Structure of the

334 statement" (0.412) and "Social introduction" (-0.353). The uncorrected model resulted
335 in correct classification of 59% of the truth tellers, and 65% of the liars. The cross-
336 validated classification, however, showed that 49.60 % of the liars and 53 % of the
337 truth tellers were correctly classified, thereby showing that SCAN performed around
338 chance level. In Appendix B (Table B2), we have presented the results of only the
339 new data, and results showed to be similar. The uncorrected model resulted in a
340 correct classification of 63% of the truth tellers, and 58% of the liars. The cross-
341 validated classification showed that 50 % of the liars and 55 % of the truth tellers
342 were correctly classified, again showing that SCAN performed around chance level.
343
344 **Discussion**
345 In the current study, we failed to find support for SCAN as a lie detection method.
346 The total SCAN score did not significantly differ between true and fabricated
347 statements, so confirming previous results (Bogaard et al., 2014; Nahari et al., 2012).
348 Interestingly, for a subset of our data CBCA and RM sum scores were coded and did
349 discriminate between the truthful and fabricated statements (Bogaard et al., 2014). As
350 such, it seems that the absence of significant SCAN findings cannot be attributed to
351 the quality of the statements used in this study. Furthermore, we investigated the
352 separate SCAN criteria, and only one criterion "Change in language" significantly
353 differentiated between true and fabricated statements; participants changed their
354 language more in their fabricated statements compared to their truthful statements.
355
356 Interestingly, the criterion "Change in language" is not described in other verbal
357 credibility methods (e.g., CBCA, RM). Therefore, our findings concerning this
358 criterion are noteworthy. Sapir (2005) explained in his manual that especially words
359 describing family members (e.g., mother, father, dad, mom, etc.), people (e.g.,
360 someone, individual, man, guy, etc.), communication (e.g., told, spoke, talked, etc.),
361 transport (e.g., vehicle, car, truck, etc.) and weapons (e.g., gun, rifle, revolver, pistol,
362 etc.) should be investigated carefully. The idea is that such a change indicates
363 something has altered in the mind of the writer. When the events in the statements
364 justify this change it does not indicate deception per se, however, in all other cases
365 these changes indicate deceit. But what exactly is meant by a justification is not
366 described in the manual. Consequently, due to the absence of clear guidelines on
367 verifying whether a change is justified, the current study scored all changes in
368 language as a cue to deceit, and might therefore differ from how SCAN is used in
369 practice.
370
371 Both the analyses of the SCAN sum score and the discriminant analysis showed
372 SCAN did not perform above chance level. This chance level performance can be
373 understood when looking at various contradicting interpretations of its criteria
374 compared with CBCA. More precisely, both methods describe "spontaneous
375 corrections" and "lack of conviction or memory", but differ in their use. For CBCA
376 both criteria are interpreted as a sign of truthfulness, while for SCAN both criteria are
377 interpreted as a sign of deceit. Commonsensically, only one interpretation can be
378 correct. As CBCA is far more embedded in the scientific literature and has been
379 shown to detect deceit above chance level (Amado et al., 2015; Vrij, 2005), CBCA's
380 interpretations should be favored over SCAN. Also, SCAN does not consider criteria
381 involved in judging distinctive types of details. Both CBCA and RM consist of
382 various types of details that have to be checked. For example, with these methods it is
383 checked whether there is information in the statement about when (i.e., temporal

9

384 details) and where (i.e., spatial details) the event took place, about what the writer saw
385 during the event (i.e., visual details) and whether there were any other perceptual
386 details (i.e., smells, tastes, sensations, sounds). Research showed that especially these
387 types of criteria are significantly more present in truthful compared to fabricated
388 statements (DePaulo et al., 2003; Masip et al., 2005; Vrij, 2005).
389
390 Relatedly, recent meta-analytical research reveals that passively observing cues only
391 has a limited influence on our deception detection abilities, as most of these cues are
392 generally weak (Hartwig and Bond, 2011). The authors argue we should actively
393 increase the verbal and non-verbal differences between liars and truth tellers. Various
394 techniques have already been suggested, such as focusing on unanticipated questions
395 during the interrogation (Vrij et al., 2009), applying the Strategic Use of Evidence
396 technique (Granhag et al., 2007) or inducing cognitive load (Vrij et al., 2006; Vrij et
397 al., 2008; Vrij et al., 2011; Vrij et al., 2012). SCAN fails to actively influence the
398 information that is provided by the interviewee, which potentially contributes to its
399 chance performance.
400
401 Finally, users of SCAN may argue that the way SCAN is tested in laboratory studies
402 such as these, is far from how it is applied in the field, and that the results therefore do
403 not translate. However, the diagnostic value of SCAN and its criteria lies within its
404 capabilities of discriminating between truthful and fabricated statements. SCAN
405 makes no assumptions as to *why* or *when* these differences between truths and lies
406 occur, only that they occur. As such, also laboratory studies – for example where
407 participants are asked to fabricate a negative event – should be able to pick up such
408 differences, if they exist. Moreover, it has proven to be exceptionally difficult to test
409 the accuracy of SCAN in field studies as the reliability of SCAN has shown to be
410 extremely low (Bogaard et al., 2014; Vanderhallen et al., 2015). The only way to
411 control for this low reliability is to use a more standardized scoring system, as we
412 have done so in the current study. For example, as is mentioned previously, SCAN
413 does not consist of a fixed list of criteria, and the criteria are not scored on a scale. In
414 field studies, SCAN analysts write a report about the presence or absence of the
415 criteria, and on the basis of this report, they make a conclusion about the truthfulness
416 of the statement. As such, it is unclear how many criteria are actually taken into
417 consideration when making a judgment, and whether these criteria are weighed
418 equally.
419
420 In sum, SCAN has no empirical support to date, and fails to include criteria
421 investigating different types of details. Only one criterion showed potential for lie
422 detection research, but has to be investigated more thoroughly in order to overcome
423 the problems that are inherent to SCAN and its criteria (e.g., vague description,
424 ambiguous interpretation). As a result, we discourage the application of SCAN in its
425 current form.

426 **References**

427 Amado, B. G., R. Arce, and F. Fariña. "Undeutsch Hypothesis and Criteria Based
428   Content Analysis: A Meta-Analytic Review.". *The European Journal of*
429   *Psychology Applied to Legal Context* 7 (2015): 1-10.

430 Armistead, T. W. "Detecting Deception in Written Statements: The British Home
431   Office Study of Scientific Content Analysis (Scan).". *Policing: An*
432   *International Journal of Police Strategies & Management* 34 (2011): 588-605.

433 Bogaard, G., E. Meijer, and A. Vrij. "Using an Example Statement Increases
434   Information but Does Not Increase Accuracy of Cbca, Rm, and Scan." *Journal*
435   *of Investigative Psychology and Offender Profiling* 11 (2014): 151-63.

436 Bogaard, G., E. Meijer, A. Vrij, N. J. Broers, and H. Merckelbach. "Scan Is Largely
437   Driven by 12 Criteria: Results from Field Data.". *Psychology, Crime and Law*
438   20 (2014): 430-49.

439 Bond, C. F., and B. M. DePaulo. "Accuracy of Deception Judgments." *Personality*
440   *and Individual Differences* 10 (2006): 214-34.

441 Burton, P., L. Gurrin, and P. Sly. "Tutorial in Biostatistics. Extending the Simple
442   Linear Regression Model to Account for Correlated Responses: An
443   Introduction to Generalized Estimating Equations and Multi-Level Mixed
444   Modeling." *Statistics in medicine* 17 (1998): 1261-91.

445 DePaulo, B. M., J. J. Lindsay, B. E. Malone, L. Muhlenbruck, K. Charlton, and H.
446   Cooper. "Cues to Deception." *Psychological Bulletin* 129 (2003): 74-118.

447 Driscoll, L. "A Validity Assessment of Written Statements from Suspects in Criminal
448   Investigations Using the Scan Technique." *Police Studies* 4 (1994): 77-88.

449 Granhag, P. A., L. A. Strömwall, and M. Hartwig. "The Sue-Technique: The Way to
450   Interview to Detect Deception.". *Forensic Update* 88 (2007): 25-29.

451 Hartwig, M., and C. F. Bond. "Why Do Lie-Catchers Fail? A Lens Model Meta-
452   Analysis of Human Lie Judgments.". *Psychological Bulletin* 137 (2011): 643-
453   59.

454 Hauch, V., I. Blandón-Gitlin, J. Masip, and S. L. Sporer. "Are Computers Effective
455   Lie Detectors? A Meta-Analysis of Linguistic Cues to Deception.".
456   *Personality and Social Psychology Review* (2014): 1-36.

457 Iacono, W. G. . "Can We Determine the Accuracy of Polygraph Tests? ." In *Advances*
458   *in Psychophysiology*, edited by J. R. Jennings, P. K. Ackles and M. G. H.
459   Coles, 201-07. London: Jessica Kingsley Publishers, 1991.

460 Johnson, M. K., and C. L. Raye. "Reality Monitoring." *Psychological Review* 88
461   (1981): 67-85.

462 Köhnken, G. "Social Psychology and the Law." In *Applied Social Psychology*, edited
463   by G Semin and K. Fiedler, 257-82. London: Sage, 1996.

464 Landis, J. R., and G. G Koch. "The Measurement of Observer Agreement for
465   Categorical Data." *Biometrics* 33 (1977): 159-74.

466 Masip, J., A. L. Sporer, E. Garido, and C. Herrero. "The Detection of Deception with
467   the Reality Monitoring Approach: A Review of the Empirical Evidence."
468   *Psychology, Crime and Law* 11 (2005): 99-122.

469 Meijer, E.H., B. Verschuere, M. Gamer, H. Merckelbach, and G. Ben-Shakhar.
470   "Deception Detection with Behavioral, Autonomic, and Neural Measures:
471   Conceptual and Methodological Considerations That Warrant Modesty.".
472   *Psychophysiology* (In press).

473 Nahari, G., A. Vrij, and R. P. Fisher. "Does the Truth Come out in the Writing? Scan
474   as a Lie Detection Tool." *Law and Human Behavior* 36 (2012): 68-76.

475   Newman, M. L., J. W. Pennebaker, D. S. Berry, and J. M. Richards. "Lying Words:
476   Predicting Deception from Linguistics Styles." *Personality and Social*
477   *Psychology Bulletin* 29 (2003): 665-75.
478   Porter, S., and Y. C. Yuille. "The Language of Deceit: An Investigation of the Verbal
479   Clues to Deception in the Interrogation Context ". *Law and Human Behavior*
480   20 (1996): 443-58.
481   Sapir, A. *The Lsi Course on Scientific Content Analysis (Scan)*. Phoenix, AZ:
482   Laboratory for Scientific Interrogation, 2005.
483   Smith, N. "Reading between the Lines: An Evaluation of the Scientific Content
484   Analysis Technique (Scan)." *Police Research Series Paper 135* (2001): 1-42.
485   Sporer, S. L., and B. Schwandt. "Moderators of Nonverbal Indicators of Deception."
486   *Psychology, Public Policy, and Law* 13 (2007): 1-34.
487   Steller, M., and G. Köhnken. "Criteria Based Statement Analysis." In *Psychological*
488   *Methods in Criminal Investigation and Evidence*, edited by D C Raskin, 217-
489   45. New York: Springer 1989.
490   Undeutsch, U. "Beurteilung Der Glaubhaftigkeit Von Aussagen." In *Handbuch Der*
491   *Psychologie Vol 11: Forensische Psychologie*, edited by U. Undeutsch.
492   Göttingen, Germany: Hogrefe, 1967.
493   Vanderhallen, M., E. Jaspaert, and G. Vervaeke. "Scan as an Investigative Tool."
494   *Police Practice and Research* (2015): 1-15.
495   Vrij, A. "Criteria Based Content Analysis: A Qualitative Review of the First 37
496   Studies." *Psychology, Public Policy, and Law* 11 (2005): 3-41.
497   ———. *Detecting Lies and Deceit: Pitfalls and Opportunities*. Chichester: Wiley,
498   2008.
499   ———. "Nonverbal Dominance Versus Verbal Accuracy in Lie Detection: A Plea to
500   Change Police Practive." *Criminal Justice and Behavior* 35 (2008): 1323-36.
501   Vrij, A., R. Fisher, S. Mann, and S. Leal. "A Cognitive Load Approach to Lie
502   Detection." *Journal of Investigative Psychology and Offender Profiling* 5
503   (2008): 39-43.
504   ———. "Detecting Deception by Manipulating Cognitive Load." *Trends in Cognitive*
505   *Sciences* 10 (2006): 141-42.
506   Vrij, A., P. A. Granhag, S. Mann, and S. Leal. "Outsmarting the Liars: Toward a
507   Cognitive Lie Detection Approach." *Current Directions in Psychological*
508   *Science* 20 (2011): 28-32.
509   Vrij, A., S. Leal, P. A. Granhag, S. Mann, R. Fisher, J. Hillman, and K. Sperry.
510   "Outsmarting the Liars: The Benefit of Asking Unanticipated Questions." *Law*
511   *and Human Behavior* 33 (2009): 159-66.
512   Vrij, A., S. Leal, S. Mann, and R. Fisher. "Imposing Cognitive Load to Elicit Cues to
513   Deceit: Inducing the Reverse Order Technique Naturally." *Psychology, Crime*
514   *and Law* 18 (2012): 579-94.
515
516

Appendix A

SCAN criteria (derived from Vrij (2008a)).

(1) **Denial of allegations**: Refers to whether the examinee directly denies the allegation in the statement by stating "I did not…". This criterion assumes that a truthful person is more likely to directly deny his or her involvement in the act.

(2) **Social introduction**: Refers to how the persons described in the statement are introduced. People that are described within a statement should be introduced in an unambiguous way, usually by mentioning their name and role (e.g., My wife, Susan). Deviations from this type of introduction indicate deception.

(3) **Structure of the statement**: Refers to the balance of the statement. In a truthful statement 20% is used to describe activities leading up to the event, the next 50% to describe the actual event, and the final 30% to discuss what happened after the event.

(4) **Emotions**: Refers to where there are emotions described in the statement. Usually emotions should be described in the epilogue of the statement. When emotions are already included within the description of the prologue (before the actual event), this indicates deception. For example, "On Saturday something strange happened to me, I was really scared" (emotions before main event) or "when he was gone, I felt disgusted with myself" (emotions after the main event). The former would indicate deception, the latter truthfulness.

(5) **Objective and subjective time**: Refers to how different time periods are covered in the statement. Objective time refers to the actual duration of events described, whereas subjective time refers to the number of words used describing these events. On average a writer is expected to need three or four lines per hour when describing one day. Large deviations from this pace suggest deception.

(6) **First person singular, past tense**: Refers to the format in which a statement is written. This is called the test of commitment, and holds that a truthful person will write the statement in first person singular, past tense. Deviations from past tense or writing in the third person could indicate a lack of commitment and hence could indicate deception.

(7) **Pronouns**: Refers to the use of pronouns in the statement. When pronouns (e.g., "he", "mine", "my") are missing in the statement, or more pronouns are expected, this could suggest that the writer wants to distance him/herself from the statement. This indicates deception. For example, when a writer refers to his car as "the car" and never as "my car" this could mean he is being deceptive about what happened to the car.

(8) **Change in language**: Refers to the change of terminology or vocabulary in the statement. This is especially important for words that are related to categories such as family members, people, communication, transport or weapons. When a change of language is obvious in a statement (e.g., knife to blade) but no justification is given for such a change, this indicates deception. A change in language indicates that something has altered in the mind of the writer.*

(9) **Spontaneous corrections**: Refers to all corrections that are made within the statements. Before the writer starts with the statement s/he is instructed not to

566		cross anything out, and when the writer fails to follow this instruction, this
567		indicates deception.*
568	(10)	**Lack of conviction or memory**: Refers to when the writer is vague about
569		certain elements within the statement (e.g. "I think…", "I guess…") or when
570		the writer admits he or she has forgotten something (e.g., "I do not remember
571		how we got to the house"). Lack of memory indicates deceit.*
572	(11)	**Out of sequence and extraneous information**: Examines whether the
573		statement includes information that is given by the writer, but has no
574		apparent meaning for the reader or whether there is strange or irrelevant
575		information within the statement. Whether the information is seen as strange
576		or irrelevant depends on the statement itself. It is thought that by including
577		this type of information, the writer is distracting the reader to hide more
578		important information. This is seen as a sign of deception.*
579	(12)	**Missing information**: Refers to phrases in the statement that indicate some
580		information has been left out. For example, words such as "after a while",
581		"shortly thereafter", or "the next thing I remember" all indicate there is
582		information missing within the statement. This is especially relevant when
583		the writer is discussing the main event. Missing information during the main
584		event could indicate that the writer is deliberately hiding information, which
585		indicates the person is deceptive.*

Appendix B

Table B1. *Overview of parameters from the GEE analysis of new data.*

Criteria	Beta Estimate	SE	95% CI	Odds ratio
8. Change in language	-1.08	0.44	-1.94, -.23	**1.18**
2. Social introduction	-0.87	0.52	-1.37, -0.05	0.75
4. Emotions	0.73	0.43	-0.12, 1.57	0.53
3. Structure of statement	0.45	0.31	-0.16, 1.07	0.21
5. Objective and subjective time	-0.46	0.28	-1.01, 0.10	0.21
6. First pers sing. past tense	-0.31	0.54	-1.38, 0.75	0.10
10. Lack of conviction or memory	0.20	0.33	-0.46, 0.86	0.04
12. Missing information	-0.21	0.33	-0.87, 0.43	0.05
9. Spontaneous corrections	-0.21	0.29	-0.78, 0.35	0.04
11. Out of sequence and extraneous information	-0.08	0.36	-0.80, 0.63	0.01

Note. Significant difference between statement types, $p = 0.01$ is in bold.

Table B2. *Detailed overview of discriminant analysis coefficients derived from the new data.*

Criteria	Mean	SD	Structure matrix	Discriminant function coefficients
8. Change in language	-0.17	0.38	0.05	1.51
3. Structure of statement	0.64	0.60	0.32	0.80
4. Emotions	1.11	0.60	-0.24	0.74
9. Spontaneous corrections	-0.71	0.63	0.36	0.34
11. Out of sequence and extraneous information	-0.16	0.41	0.14	0.48
12. Missing information	-0.80	0.50	0.12	0.23
7. Pronouns	1.69	0.51	-0.07	0.22
6. First pers sing. past tense	1.59	0.63	0.10	-0.42
10. Lack of conviction or memory	-0.20	0.39	-0.34	0.13

5. Objective and subjective time	0.91	0.66	-0.19	-0.66
2. Social introduction	1.56	0.69	0.56	-0.66

Table 1. *Means, standard deviations and percentage present for each SCAN criterion as a function of veracity.*

SCAN criteria	True			Fabricated		
	Mean	SD	% present	Mean	SD	% present
1. Denial of allegations	0	0	0	0	0	0
2. Social introduction	1.26	.81	76.90	1.40	.71	87.20
3. Structure of the statement	.73	.60	67.50	.59	.60	56.40
4. Emotions	1.05	.62	83.80	.95	.65	76.10
5. Objective and Subjective time	.71	.65	62.40	.79	.65	69.20
6. First pers sing, past tense	1.59	.63	92.30	1.60	.60	94.00
7. Pronouns	1.68	.49	97.40	1.69	.50	97.40
8. Change in language	-.09	.27	12.00	-.23	.43	24.80
9. Spontaneous corrections	-.61	.62	56.40	-.64	.63	58.10
10. Lack of conviction or memory	-.16	.36	18.80	-.14	.33	16.40
11. Out of sequence and extraneous info	-.18	.38	21.40	-.20	.43	22.20
12. Missing information	-.64	.55	75.00	-.67	.52	67.50

Table 2. *Overview of parameters from the GEE analysis.*

Criteria	Beta Estimate	SE	95% CI	Odds ratio
8. Change in language	-0.89	0.36	-1.59, -.18	**0.79**
2. Social introduction	-0.713	0.34	-1.37, -0.05	0.51
4. Emotions	0.48	0.27	-0.06, 1.03	0.23
3. Structure of statement	0.47	0.26	-0.04, 0.99	0.22
5. Objective and subjective time	-0.31	0.21	-0.73, 0.12	0.10
6. First pers sing, past tense	-0.27	0.51	-1.26, 0.72	0.07
10. Lack of conviction or memory	0.24	0.31	-0.36, 0.85	0.06
12. Missing information	-0.12	0.21	-0.53, 0.30	0.01
9. Spontaneous corrections	-0.07	0.22	-0.50, 0.36	0.00
11. Out of sequence and extraneous information	-0.05	0.28	-0.60, 0.50	0.00

Note. Significant difference between statement types, $p = 0.01$ is in bold.

Table 3. *Detailed overview of discriminant analysis coefficients.*

Criteria	Mean	SD	Structure matrix	Discriminant function coefficients
8. Change in language	-0.16	0.37	0.66	1.82
3. Structure of statement	0.66	0.60	0.41	0.79
4. Emotions	1.00	0.63	0.29	0.67
9. Spontaneous corrections	-0.62	0.62	0.07	0.23
11. Out of sequence and extraneous information	-0.19	0.40	0.10	0.20
12. Missing information	-0.65	0.53	0.12	0.12
7. Pronouns	1.69	0.49	-0.03	-0.09
6. First pers sing. past tense	1.59	0.61	-0.04	-0.14
10. Lack of conviction or memory	-0.15	0.35	-0.11	-0.21
5. Objective and subjective time	0.75	0.65	-0.23	-0.43
2. Social introduction	1.33	0.77	-0.35	-0.53