Within-subjects verbal lie detection measures:

A comparison between total detail and proportion of complications

Aldert Vrij

Sharon Leal

Louise Jupe

Adam Harvey

Department of Psychology, University of Portsmouth

Author notes

This work was funded by the Centre for Research and Evidence on Security Threats (ESRC Award: ES/N009614/1)

Correspondence concerning this article should be addressed to Aldert Vrij, Department of Psychology, University of Portsmouth, King Henry Building, King Henry 1 Street, PO1 2DY, Hants, United Kingdom. Email: aldert.vrij@port.ac.uk

Within-subjects verbal lie detection measures

Abstract

Background. We examined whether the verbal cue, proportion of complications, was a more diagnostic cue to deceit than the amount of information provided (e.g. total number of details).

Method. In the experiment, 53 participants were interviewed. Truth tellers (n = 27) discussed a trip they had made during the last twelve months; liars (n = 26) fabricated a story about such a trip. The interview consisted of an initial recall followed by a model statement (a detailed account of an experience unrelated to the topic of investigation) followed by a post model statement recall. The key dependent variables were the amount of information provided and the proportion of all statements that were complications.

Results. The proportion of complications was significantly higher amongst truth tellers than amongst liars, but only in the post model statement recall. The amount of information provided did not discriminate truth tellers from liars in either the initial or post model statement recall.

Conclusion. The proportion of complications is a more diagnostic cue to deceit than the amount of information provided as it takes the differential verbal strategies of truth tellers and liars into account.

Keywords: within-subjects measures, information-gathering, verbal deception detection

Within-subjects verbal lie detection measures:

A comparison between total detail and proportion of complications

In each investigative interview two aims are central: To obtain as much information as possible and to determine whether the provided information is true or false (Loftus, 2011). Regarding the second aim, verbal deception research has shown that the total amount of information provided is amongst the most diagnostic cues to deception, with truth tellers typically reporting more details than liars (Amado, Arce, & Fariña, 2015; DePaulo et al., 2003; Vrij, 2008). In this article, we tested the hypothesis that another verbal deception cue, the proportion of complications, would be more successful in distinguishing truth tellers from liars than the total amount of information, because this variable takes the differential verbal strategies of truth tellers and liars more into account and is more suitable for within-subject assessments.

Total amount of information

In their meta-analysis of Criteria-Based Content Analysis, a list of 19 different verbal criteria which includes the criterion 'total details' (e.g. units of information), the criterion 'total details' emerged as the most diagnostic cue to deceit, d = .77 (Amado et al., 2015). Truth tellers typically provide more details than liars because liars lack the imagination to fabricate details that sound plausible, a conclusion derived from the CBCA literature (Köhnken, 2004; Leal, Vrij, Warmelink, Vernham, & Fisher, 2015), or because they are unwilling to provide many details out of fear that those details give leads to investigators that they are lying (Nahari, Vrij, & Fisher, 2014).

The total details variable is difficult to use in real life. The difficulty investigators face is that it is impossible to determine how much detail someone should provide to be judged as truthful (Nahari & Vrij, 2014, 2015; Vrij, 2016) This of course depends on the context, events rich in detail can be described in much more detail than events poor in detail

(Vrij, 2008); and it also depends on individual differences, some people are more eloquent than others (Merckelbach, 2004; Nahari, & Pazuelo, 2015). One way around this issue is to introduce a within-subjects comparison, that is, to compare one part of a statement with another part of a statement. At least two different within-subjects comparisons can be made. First, someone can examine the proportion of a specific type of detail within a statement, such as the proportion of details that can be verified (verifiable details divided by all details), as used in the Verifiability Approach (Nahari, Vrij, & Fisher, 2014). In this comparison, it does not matter anymore how detailed a statement is, thereby taking away the possible effects of the type of event and individual differences, all that matters is the percentage of details that can be verified. A second comparison is to introduce two phases in an interview and to compare the statements of both phases. In this case, again, it does not matter anymore how detailed a statement is, all that matters is the difference in detail between the two phases. In this experiment, we combined these two within-subjects comparisons.

Proportion of complications

Total amount of information is a generic measure which does not take into account all types of detail someone can report. In the present experiment, following Vrij et al. (2017b), we broke down this generic measure into components that we believe are more sensitive to the different verbal strategies that truth tellers and liars use. In brief, we expected truth tellers to provide stories that include non-essential details that make the story more complex (complications). By comparison, we expected liars to provide details that are based on common knowledge, or to justify why they cannot provide certain types of information (self-handicapping strategies). As a result, the proportion of complications (complications / (complications + common knowledge details + self-handicapping strategies)) should be higher in truth tellers than in liars.[1]

Within-subjects verbal lie detection measures

According to the CBCA literature a complication is a reported activity or event that was not expected or planned ("The air conditioning was not working properly in the hotel") (Steller & Köhnken, 1989; Vrij et al., 2017b). We prefer a slightly more inclusive definition: "an occurrence that makes a situation more difficult than necessary". Information which is considered to be a complication according to the CBCA definition is also a complication according to this definition, but this definition includes more information as complications. For example, if someone says that when driving from their hometown A to their holiday destination B, they also paid a short visit to town C, this short visit to C is considered a complication according to our definition but not according to the CBCA definition. Complications are one of the 19 criteria that appears on the CBCA list of criteria, and are more likely to occur in truthful statements than in deceptive statements (Amado et al., 2015; Vrij, 2008; Vrij et al., 2017b). Making up complications requires imagination, but liars may not have adequate imagination to do so (Köhnken, 2004; Vrij, 2008). In addition, research examining liars' interview strategies showed that liars prefer to keep their stories simple (Hartwig, Granhag, & Strömwall, 2007), but adding complications makes the story more complex.

Common knowledge details refer to strongly invoked stereotypical information about events ("We visited the Louvre museum where was saw the Mona Lisa") (Vrij et al., 2017b). Liars are more likely to include common knowledge details in their statements than truth tellers (Sporer, 2016; Volbert & Steller, 2014; Vrij et al., 2017b). Truth tellers have personal experiences of an event and are likely to report such unique experiences (DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996). When they do so the statement is no longer scripted. If liars do not have personal experiences of the event they report, they then will draw upon general knowledge to construe the event (Sporer, 2016). In case liars do have personal

experiences of the event, they may not report them due to their desire to keep their stories simple.

Self-handicapping strategies refer to explicit or implicit justifications as to why someone is not able to provide information ("I can't remember; it was a while ago when this happened", "Nothing unexpected happened; I am a very organised person"; "I fell asleep in the bus") (Vrij et al., 2017b). Liars are more likely to include self-handicapping strategies in their statements than truth tellers (Vrij et al., 2017b). For liars, who are inclined to keep stories simple, not having to provide information is an attractive strategy. However, liars are also concerned about their credibility and believe that admitting lack of knowledge and/or memory appears suspicious (Ruby & Brigham, 1998). A potential solution is to provide a justification for the inability to provide information. Note that the justification is crucial for a self-handicapping strategy: the statement 'I can't remember' itself is not a self-handicapping strategy, it is called admitting lack of memory and part of the Criteria-Based Content Analysis tool (Amado et al., 2015). Also, note that the justification does not have to be made explicit. The example "I fell asleep in the bus" is an implicit justification for not being able to provide information.

Examining complications, common knowledge details and self-handicapping strategies has two advantages compared to examining total details. First, the three separate measures should be more effective in discriminating truth tellers from liars than the generic measure of total details. Both truth tellers and liars must provide some details in interviews, with truth tellers perhaps providing somewhat more total details than liars. However, the overall set of details includes many types of detail including complications, common knowledge details and self-handicapping strategies. Truth tellers are more likely to report complications, whereas liars are more likely to report common knowledge details and self-handicapping strategies. Based on this the *proportion of complications* (complications /

Within-subjects verbal lie detection measures

(complications + common knowledge details + self-handicapping strategies)) -which represents the proportion of cues to truthfulness- can be calculated. This proportion of complications score exploits the different verbal strategies that truth tellers and liars use – it should be higher in truth tellers- and should therefore be a more diagnostic cue to deceit than the generic measure total details. Second, the proportion of complications is a within-subjects measure, which are preferred by practitioners (Vrij, 2016) and scholars (Nahari & Pazuelo, 2015; Nahari & Vrij, 2014; Nahari, in press), because it controls, at least to some extent, for differences in context and personal characteristics.

Two phases interview

In the present experiment, the interview consisted of two phases. In phase 1, the initial recall phase, interviewees (truth tellers and liars) were invited to report their alleged experiences (an alleged trip) in as much detail as possible. This was followed by a second phase, the post model statement recall phase, in which they were again invited to report in as much detail as possible all they could remember. In between these two phases, they listened to a model statement, a detailed account of an experience unrelated to the topic of investigation (Leal, Vrij, Warmelink, Vernham, & Fisher, 2015).

A model statement raises the expectations amongst interviewees how much information is required (Ewens et al., 2016) and therefore typically results in more detailed answers (Bogaard, Meijer, & Vrij, 2014; Ewens et al., 2016; Vrij et al., 2017b). Because both truth tellers and liars will realise that they are expected to provide more detailed answers after having listened to a model statement, they are both likely to report more details after listening to a model statement (Bogaard et al., 2014; Ewens et al., 2016; Vrij et al., 2017b). However, the type of detail they report may differ (Vrij et al., 2017b). Truth tellers are more likely to add complications to their stories than liars. Complications are often parts of an account that truth tellers may have thought were not essential and therefore worthwhile to report. The

raised expectations to provide more details may encourage truth tellers to include more complications than liars. Common knowledge details may also become a more diagnostic cue to deceit after listening to a model statement. The urge to provide more details may encourage truth tellers to decide to describe separate events in more detail, which reduces the likelihood that they sound scripted. The same urge to provide more details may encourage liars to describe more events, and these will probably discussed in a more scripted manner. Also, self-handicapping strategies may become more diagnostic after being exposed to a model statement. Rather than providing more details, a model statement may encourage liars to say why they cannot provide more details, which would mean introducing a self-handicapping strategy.

Vrij et al. (2017b) found support for this reasoning regarding complications, with truth tellers –but not liars- reporting more complications after listening to a model statement. As a result, the proportion of complications was higher for truth tellers than for liars, particularly in the condition where a model statement was present. However, in Vrij et al. (2017b), interviewees only gave one statement compared to two statements in the current experiment. In other words, in the present experiment we examined the impact of a model statement in a within-subjects design, which Vrij et al. (2017b) did not do.

Hypotheses

Truth tellers will report more total details, more complications, fewer common knowledge details and fewer self-handicapping strategies than liars (Hypothesis 1). From Hypothesis 1 follows that the proportion of complications (e.g., proportion of cues to truthfulness) will be higher in truth tellers than in liars (Hypothesis 2). Truth tellers will report more total details, more complications and fewer common knowledge details in Phase 2, post model statement recall phase, than in Phase 1, initial recall phase (Hypothesis 3). Liars will report more total details, more common knowledge details and more self-

handicapping strategies in Phase 2 than in Phase 1 (Hypothesis 4). From Hypotheses 2 and 3 it follows that the difference between truth tellers and liars in the proportion of complications will be more pronounced in Phase 2 than in Phase 1 of the interview (Hypothesis 5).

Method

Participants

A total of 53 participants (14 males and 39 females) took part in the study. Their average age was $M = 21.70$ years ($SD = 5.14$).

Procedure

The Procedure was derived from Vrij et al. (2017a, b). Participants were recruited via an advert on the university intranets and advertisement leaflets. The advert explained that the experiment would require participants to tell the truth or lie about a trip away that they may (or may not) have taken within the last year. We decided upon last year so that truth tellers would still remember many details about their trip and liars could not easily say "I can't remember" when answering the questions. On arrival to the laboratory, participants received a participant information sheet and signed an informed consent form. Both truth tellers and liars then completed a selection form on which they wrote down which cities they had visited during the last twelve months for at least two nights. They also wrote down whether they had ever visited these cities before and had ever lived there. For truth tellers ($n = 27$), the experimenter selected a city where the participant had stayed during the last twelve months for at least two nights without having been there before and without having lived there. For liars ($n = 26$), the experimenter selected a city chosen for a truth teller after checking that the liar had never been before to that city and had never lived there.

Truth tellers and liars were then left with a computer with internet access and were told they had twenty minutes to prepare for their interview, or to let the experimenter know if they were ready before that time. In order to avoid participants feeling that they had to

prepare in a rush we ensured that they were given plenty of time to prepare. All participants were told that they were allowed to make notes while doing their research. Truth tellers and liars were told that it was important to be convincing because, by means of a sanction, if they did not appear convincing they would be asked to write a statement about what they told the interviewer in the interview. In total 19 cities were discussed amongst the participants during their interviews. In a pre-interview questionnaire, the truth tellers and liars rated their thoroughness of preparation via three items: (1) shallow to (7) thorough; (1) insufficient to (7) sufficient; and (1) poor to (7) good. The answers to the three questions were averaged (Cronbach's alpha = .80) and the variable is called 'preparation thoroughness'. They were also asked whether they thought they were given enough time to prepare themselves with the following question: 'Do you think the amount of time you were given to prepare was: (1) insufficient to (7) sufficient.

The interview.

Prior to the interview the experimenter told the interviewer about which city to interview the participant. The interviewer was unaware of the veracity status of the participant. To make the interviewee feel comfortable and to avoid floor effects in establishing rapport interviewees were offered a glass of water from the interviewer, as offering something helps rapport building (reciprocation principle, Cialdini, 2007).

The female interviewer started by saying: "I understand from my colleague that you have visited _____". This was followed by the question: "Could you please tell me in as much detail as possible everything you did when you were at _____ from the moment you arrived to the moment you left?"

After the answer the interviewer said that the she wanted the participant to report the trip again but that she now wanted to give the participant an idea how much detail she wanted to hear by playing a model statement. An audiotaped model statement was then played. It was

Within-subjects verbal lie detection measures

1.30 minutes long detailed account of someone attending a Formula 2 motor racing event (Leal et al., 2015). The account was a spontaneous, unscripted, recall of an event truly experienced by the person. After the model statement the interviewee was again asked to tell in as much detail as possible everything s/he did when you were at _____ from the moment s/he arrived to the moment s/he left?

The interviews were audio recorded and subsequently transcribed.

Post-interview questionnaire.

After the interview, participants completed a post-interview questionnaire in which they indicated how motivated they were to perform well during the interview (measured on a 5-point Likert scale from 1 = not at all motivated to 5 = very motivated).

Coding

All coders were blind to the hypotheses and Veracity condition.

Detail.

The coders were taught the coding scheme by the first author who has more than twenty years of experience in coding details. A coder first read the transcripts and coded each detail in the interview, whereby a detail was defined as a unit of information. To give an example, the following answer contains ten details: "I have an aunt there, ah – living in a house close to the lake. We went in my truck, ah, five of us, also my sister's boyfriend and my boyfriend". Each detail in the actual trip section and each detail in the planning of the trip section were coded only once; thus repetitions were not coded. A second coder coded a random sample of 20 transcripts. Inter-rater reliability between the two coders, using the two-way random effects model measuring consistency, was satisfactory (Single Measures ICC = .86).

The coders were also taught how to code complications, common knowledge details and self-handicapping strategies by the first author. One coder coded all complications,

Within-subjects verbal lie detection measures

common knowledge details and self-handicapping strategies that appeared in the transcripts. Complications are occurrences that makes a situation more difficult than necessary. Three examples of complications are: i) …"she was meant to get a sirloin and I was meant to get a rump but she wanted hers medium rare and they did it the wrong way round and when we tried to complain they didn't like it" ii)… "when we got on to the M23 there was a lot of traffic there, I'm not sure what was causing the hold-up but yeah took a bit longer than expected to get there". and iii)… "I remember my en-suite the toilet wouldn't flush properly, so we had to call maintenance for them to try to sort it out".

Common knowledge details refer to strongly invoked stereotypical knowledge about events. Three examples of common knowledge details are: i)… "we visited the haunted house and we went to London Eye" ii)… "we just went sightseeing to Bath Abbey and then just looked around there" and iii)… "yeah it was wonderful sightseeing. We went to the Colosseum".

Self-handicapping strategies refer to explicit or implicit justifications as to why someone is not able to provide information. Three examples of self-handicapping strategies are: i) "I'm not sure exactly what shops we went in because it was quite a while ago", ii) "And then we just all sort of fell asleep in the car on the way back home" and iii) "We got there around the afternoon-ish and we looked around. And we went home after that because we were really tired because it's quite tiring looking around and stuff". (Examples 1 and 3 are explicit justifications and example 2 is an implicit justification.)

A second coder coded a random sample of 20 transcripts. Inter-rater reliability between the two coders, using the two-way random effects model measuring consistency, was high for complications (Average Measures, Intraclass correlation coefficient, ICC = .87) and for common knowledge details (Average Measures ICC = .87), and satisfactory for self-handicapping strategies (Average Measures ICC = .65). Disagreements were resolved

between the two coders. All disagreements occurred because one coder missed a cue. To calculate the proportion of complications a total score was computed (number of complications + number of common knowledge details + number of self-handicapping strategies) and the number of complications was divided by this total score. Scores > .50 indicate that the participants reported more complications than common knowledge details and self-handicapping strategies (when summed), whereas scores < .50 indicate that the participants reported more common knowledge details and self-handicapping strategies (when summed) than complications.

Results

Motivation and preparation

Participants reported to have been motivated to perform well during the interview (M = 4.33, SD = .65), with no significant difference in motivation between truth tellers and liars, $F(1, 51) = .08$, $p = .778$, $d = .08$. Neither did significant differences emerge between truth tellers and liars in terms of preparation thoroughness, $F(1, 51) = 2.71$, $p = .106$, $d = .46$ and preparation time, $F(1, 51) = 1.92$, $p = .172$, $d = .38$.

Hypothesis Testing

A 2 (Veracity) X 2 (Phase) MANOVA was conducted. Veracity was a between-subjects factor and Phase a within-subjects factor. The five variables listed in Table 1 were the dependent variables.[2] At a multivariate level the analysis revealed significant main effects for Veracity, $F(5, 47) = 6.34$, $p < .001$, *partial eta*^{2} = .40), and Phase, $F(5, 47) = 11.62$, $p < .001$, *partial eta*^{2} = .55. The Veracity x Phase interaction effect was also significant, $F(5, 47) = 4.25$, $p =. 003$, *partial eta*^{2} = .31.

Table 1 about here

The univariate Veracity main effects are presented in Table 1. Compared to liars, truth tellers reported significantly more complications and fewer self-handicapping strategies.

Within-subjects verbal lie detection measures

However, truth tellers and liars did not differ from each other in the total number of details or common knowledge details they reported. Thus, Hypothesis 1 is supported for the variables complications and self-handicapping strategies, but we failed to find evidence for a difference regarding the variables total details and common knowledge details. In addition, the proportion of complications was significantly higher for truth tellers than for liars, supporting Hypothesis 2.

The univariate Phase main effects are also presented in Table 1. Participants reported more details and more complications in Phase 2 than in Phase 1. The proportion of complications was significantly higher in Phase 2 than in Phase 1. At a univariate level, significant Veracity x Phase interaction effects emerged for complications, $F(1, 51) = 11.60$, $p = .001$, *partial eta*^{2} = .45, common knowledge details, $F(1, 51) = 5.75$, $p = .020$, *partial eta*^{2} = .10, and proportion of complications, $F(1, 51) = 6.68$, $p = .013$, *partial eta*^{2} = .12. The effects for total details, $F(1, 51) = 1.87$, $p = .177$, *partial eta*^{2} = .04 and self-handicapping strategies, $F(1, 51) = 1.29$, $p = .262$, *partial eta*^{2} = .03, were not significant.

Table 2 about here

Table 2 presents the results of simple effect analyses in which we compared the difference in detail between Phases 1 and 2 for truth tellers and liars separately. Despite the non-significant results for total details and self-handicapping strategies we also report the simple effect analyses for those two variables to obtain a complete picture.

Truth tellers reported more total details, more complications, and fewer common knowledge details in Phase 2 than in Phase 1. This supports Hypothesis 3. Liars reported more details in Phase 2 than in Phase 1, whereas the other effects were not significant. This support Hypothesis 4 regarding total details but we failed to find evidence for a difference regarding common knowledge details.

Within-subjects verbal lie detection measures

Table 2 also shows that truth tellers obtained a significantly higher proportion of complications score in Phase 2 than in Phase 1, whereas the proportion of complication score did not change for liars between Phases 1 and 2. This suggests that the proportion of complications score between truth tellers and liars was more pronounced in Phase 2 than in Phase 1, as predicted in Hypothesis 5. To statistically test Hypothesis 5, we first compared truth tellers' and liars' proportion of complications scores in Phase 1 as well as in Phase 2. In Phase 1 the proportion of complications score for truth tellers (M = .57, SD = .40, *95% CI* [.41,.73]) and liars (M = .48, SD = .39, *95% CI* [.33,.64]) were similar to each other, $F(1, 51) = 0.64$, p = .427, d = .23 (-.31,.77). In contrast, in Phase 2 the proportion of complications score for truth tellers (M = .84, SD = .26, *95% CI* [.72,.96]) was significantly higher than the proportion of complications score for liars (M = .51, SD = .36, *95% CI* [.39,.64]), $F(1, 51) = 14.62$, $p < .001$, d = 1.05 (.63,1.44). The effect size was also far more substantial in Phase 2 than in Phase 1. Second, we compared truth tellers' and liars' proportion of complications scores in Phase 2 and included the Phase 1 complication score as a covariate. The covariate was significant, $F(1, 50) = 22.47$, $p < .001$, $\text{eta}^2 = .31$. More importantly, the effect for Phase 2 was still significant, $F(1, 50) = 16.01$, $p < .001$, $\text{eta}^2 = .24$. These findings support Hypothesis 5. We employed in Phase 2 the decision rule 'Interviewees with a proportion of complication score > .50 are truth tellers, the others are liars'. This resulted in correct classifications of 20 out of 27 truth tellers (74%) and 17 out of 26 liars (65%). By comparison, we compared truth tellers and liars in terms of reported total number of details in Phases 1 and 2. In neither phase a Veracity effect emerged: Phase 1, $F(1, 51) = 0.50$, p = .483, d = .19 (-.35,.73) and Phase 2, $F(1, 51) = 1.90$, p = .174, d = .38 (-.17,.92). In sum, the proportion of complications Veracity effect was significant in Phase 2, but not in Phase 1 and remained significant in Phase 2 even when we controlled for the proportion of complications score in Phase 1. The total number of details Veracity effect was not significant in Phases 1 or 2. This means that

proportion of complications was a more diagnostic cue to deceit than total number of details, at least in Phase 2.

Discussion

We found that the proportion of complications was a more diagnostic cue to deceit than total number of details, but only in Phase 2 of the interview, the post model statement recall phase. This finding emerged because truth tellers and liars responded differently to the model statement. Both truth tellers and liars added details to their stories and, as a result, total number of details did not differ between the two groups. However, the type of detail they added was different. Unlike liars, truth tellers added more complications to their stories and made their stories less scripted.

That a model statement makes the proportion of complications cue more diagnostic is a replication of Vrij et al. (2017b). In that experiment the model statement was a between-subjects factor, in the current experiment we showed that the same effect occurs when a model statement is introduced as a within-subjects factor. This adds practical value to the finding because practitioners prefer within-subjects designs (Vrij, 2016). We introduced a model statement between Phases 1 and 2 because a model statement has shown to encourage interviewees to add details to their accounts. However, there is no theoretical reason why the present findings will only occur when a model statement is used. In theory, any intervention that encourages interviewees to say more should yield this effect. Several ways to encourage interviewees to say more have been examined to date, including the use of drawings (Mattison, Dando, & Ormerod, 2015; Vrij et al., 2017a), the introduction of a supportive second interviewer (Mann et al., 2013) and mimicking the nonverbal behaviour of the interviewee (Shaw et al., 2015).

Although we reasoned that examining complications, common knowledge details and self-handicapping strategies would be more effective for lie detection purposes than

examining total details, we did not expect that the variable total details could not discriminate truth tellers from liars at all. The absence of a Veracity effect regarding total details is unusual in deception research (Vrij, 2008). One obvious explanation for the absence of a Veracity effect is that our coding system was not sensitive enough to pick up differences in detail between truth tellers and liars. We think this is unlikely. We have used the same coding scheme in studies before and often found differences in detail between truth tellers and liars (Ewens et al., 2016; Harvey, Vrij, Leal, Hope, & Mann, 2017; Vrij et al., 2008).

It is more likely that a combination of factors caused the absence of a Veracity effect for details. First, interviewees discussed an alleged trip. Of course, many liars have made trips in their lives before and probably could use those experiences to make up a detailed fabricated trip (Leins, Fisher, & Ross, 2013). Second, liars were allowed to thoroughly prepare themselves for the interview which may have contributed to being able to provide detailed stories. Third, the interview protocol was very basic, as the interviewees were asked only one question. Research has shown that more sophisticated theory-based interview protocols are typically required to obtain differences between truth teller and liars (Vrij, Blank, & Fisher, 2018; Vrij, Fisher, & Blank, 2017; Vrij, Fisher, & Blank, Leal, & Mann, 2016).

An important point to consider is whether the proportion of complications will always be a more diagnostic cue to deceit than number of details or whether that depends on the scenario. This is an empirical question worth considering. We do not rule out that it will be situation-related. For example, for proportions of complications to become effective, complications should occur and we think this unlikely to happen in very short encounters.

In this article, the within-subjects measure we used was a comparison between two phases of the interview. In real life practitioners often use a different within-subjects measure. Typically, prior to the interview, someone's truthful responses during small-talk are

observed and these responses are then compared with responses during the actual interview. Different responses in both parts could then be indicative of deceit (Frank, Yarbrough, & Ekman, 2006; Inbau, Reid, Buckley, & Jayne, 2013). This type of baselining has been described as one of the most striking misuses of psychological research (Moston & Engelberg, 1993; Palena, Caso, Vrij, & Orthey, 2018). Fundamental differences exist between small-talk and the investigative part of the interview. For example, small-talk conversations are low-stakes situations where the interviewees' responses are unlikely to have any negative consequences. In contrast, the investigative part of the interview is a high-stakes situation that can have negative consequences for interviewees in case they will not be believed by the investigator. As a result, both guilty and innocent interviewees are likely to exhibit different behaviours during small talk compared with the actual interview (Hartwig & Bond, 2014), and this 'apple–orange' comparison will be prone to incorrect judgements (Moston & Engelberg, 1993). See Vrij (2016) for an overview of baselining.

Applying verbal baselining in real life is challenging. Baselining methods would be most effective if truth tellers and liars display truly different response patterns, for example, if truth tellers always include more complications in their statements than common knowledge details and self-handicapping strategies combined and if liars always include more common knowledge details and self-handicapping strategies combined in their statements than complications. In that case a clear cut-off score can be established, but this does not happen in real life. The current experiment showed that the decision rule 'Interviewees with a proportion of complication score > .50 are truth tellers, the others are liars' resulted in correct classifications of 74% of truth tellers and 65% of liars. Yet, the benefit of using within-subject measures compared to between-subjects measures still exists. If just 'amount of detail' is considered, the problem arises that the amount of detail will not only be affected by veracity but also by individual differences in being eloquent (eloquent people may provide

more details) or preparedness (well prepared answers are likely to be longer than spontaneous answers). Those additional factors play a lesser role in within-subjects comparisons. That is, it is no longer relevant how detailed an answer is (which is largely influenced by being eloquent and prepared), but it becomes relevant how many complications, common knowledge details and self-handicapping strategies are included (more likely to be influenced by veracity).

Acknowledgements

This work was funded by the Centre for Research and Evidence on Security Threats (ESRC Award: ES/N009614/1)

References

Amado, B. G., Arce, R., Fariña, F. (2015). Undeutsch hypothesis and Criteria Based Content Analysis: A meta-analytic review. *The European Journal of Psychology Applied to Legal Context, 7,* 3-12. Doi:10.1016/j.ejpal.2014.11.002

Bogaard, G., Meijer, E. H., & Vrij, A. (2014). Using an example statement increases information but does not increase accuracy of CBCA, RM, and SCAN. *Journal of Investigative Psychology and Offender Profiling, 11,* 151-163. Doi: 10.1002/jip.1409

Cialdini, R. B. (2007) *Influence: The psychology of persuasion*. New York: William Morrow and Company.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology, 70,* 979-995. Doi: 10.1037/0022-3514.70.5.979

DePaulo, B. M., Lindsay, J. L., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129,* 74–118. Doi: 10.1037/0033-2909.129.1.74

Ewens, S., Vrij, A., Leal, S., Mann, S., Jo, E., Shaboltas, A., Ivanova, M., Granskaya, J., & Houston, K. (2016). Using the model statement to elicit information and cues to deceit from native speakers, non-native speakers and those talking through an interpreter. *Applied Cognitive Psychology, 30,* 854-862. Doi: 10.1002/acp.3270

Frank, M. G., Yarbrough, J. D., & Ekman, P. (2006). Investigative interviewing and the detection of deception. In T. Williamson (Ed.), *Investigative interviewing: Rights, research and regulation* (pp. 229-255). Cullompton, Devon: Willan Publishing.

Hartwig, M., & Bond, C. F. (2014). Lie detection from multiple cues: A meta-analysis. *Applied Cognitive Psychology, 28,* 661-667. Doi: 10.1002/acp.3052.

Within-subjects verbal lie detection measures

Hartwig, M., Granhag, P. A., & Strömwall, L. (2007). Guilty and innocent suspects' strategies during interrogations. *Psychology, Crime, & Law*, *13*, 213–227. Doi: 10.1080/10683160600750264.

Harvey, A., Vrij, A., Leal, S., Hope, L., & Mann, S. (2017). Deception and decay: Verbal lie detection as a function of delay and encoding quality. *Journal of Applied Research in Memory and Cognition***,** *6*, 306-318. Doi: 10.1016/j. jarmac.2017.04.002

Inbau, F. E., Reid, J. E., Buckley, J. P., & Jayne, B. C. (2013). *Criminal interrogation and confessions, 5th edition*. Burlington, MA: Jones & Bartlett Learning.

Köhnken, G. (2004). Statement Validity Analysis and the 'detection of the truth'. In P. A. Granhag & L. A. Strömwall (Eds.), *Deception detection in forensic contexts* (pp. 41-63). Cambridge, England: Cambridge University Press.

Leal, S., Vrij, A., Warmelink, L., Vernham, Z., & Fisher, R. (2015). You cannot hide your telephone lies: Providing a model statement as an aid to detect deception in insurance telephone calls. *Legal and Criminological Psychology*, *20*, 129-146. Doi: 10.1111/lcrp.12017

Leins, D., Fisher, R. P., & Ross, S. J. (2013). Exploring liars' strategies for creating deceptive reports. *Legal and Criminological Psychology, 18*, 141-151. Doi: 10.1111/j.2044-8333.2011.02041.x

Loftus, E. F. (2011). Intelligence gathering post-9/11. *American Psychologist, 66*, 532-541. Doi: 10.1037/a0024614

Mann, S., Vrij, A., Shaw, D., Leal, S., Ewens, S., Hillman, J., Granhag, P. A., & Fisher, R. P. (2013). Two heads are better than one? How to effectively use two interviewers to elicit cues to deception. *Legal and Criminological Psychology, 18*, 324-340. Doi: 10.1111/j.2044-8333.2012.02055.x

Within-subjects verbal lie detection measures

Mattison, M. C. L., Dando, C. J., & Ormerod, T. C. (2015). Sketching to remember: Episodic free recall task support for child witnesses and victims with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 45, 1751-1765. Doi: 10.1007/s10803-014-2335-z

Merckelbach, H. (2004). Telling a good story: Fantasy proneness and the quality of fabricated memories. *Personality and Individual Differences, 37,* 1371–1382. Doi: 10.1016/j.paid.2004.01.007

Moston, S. J., & Engelberg, T. (1993). Police questioning techniques in tape recorded interviews with criminal suspects. *Policing and Society, 6,* 61-75. Doi: 10.1080/10439463.1993.9964670

Nahari, G. (in press). Advances in lie detection: Limitations and potential for investigating allegations of abuse. In R. Burnett (Ed.), *Wrongful Allegations of Sexual and Child Abuse*. Oxford: Oxford University Press.

Nahari, G., & Pazuelo, M. (2015). Telling a convincing story: Richness in detail as a function of gender and priming. *Journal of Applied Research in Memory and Cognition, 4* , 363-367. Doi:10.1016/j.jarmac.2015.08.005

Nahari, G., & Vrij, A. (2014). Are you as good as me at telling a story? Individual differences in interpersonal reality-monitoring. *Psychology, Crime and Law, 20,* 573 – 583. Doi:10.1080/1068316X.2013.793771

Nahari, G. & Vrij, A. (2015). Systematic Errors (Biases) in Applying Verbal Lie Detection Tools: Richness in Detail as a Test Case. *Crime Psychology Review, 1,* 98-107. Doi: 10.1080/23744006.2016.1158509

Nahari, G., Vrij, A., & Fisher, R. P. (2014). Exploiting liars' verbal strategies by examining the verifiability of details. *Legal and Criminological Psychology, 19,* 227-239. Doi: 10.1111/j.2044-8333.2012.02069.x

Within-subjects verbal lie detection measures

Palena, N., Caso, L., Vrij, A., & Orthey, R. (2018). Detecting deception through small Talk and comparable truth baselines. *Journal of Investigative Psychology and Offender Profiling*. Doi: 10.1002/jip.1495

Ruby, C. L., & Brigham, J. C. (1998). Can Criteria-Based Content Analysis distinguish between true and false statements of African-American speakers? *Law and Human Behavior*, *22*, 369-388. Doi: 10.1023/A:1025766825429

Shaw, D. J., Vrij, A., Leal, S., Mann, S., Hillman, J., Granhag, P. A., & Fisher, R. P. (2015). Mimicry and investigative interviewing: Using deliberate mimicry to elicit information and cues to deceit. *Journal of Investigative Psychology and Offender Profiling*, *12*, 217-230**.** Doi: 10.1002/jip.1438

Sporer, S. L. (2016). Deception and cognitive load: Expanding our horizon with a working memory model. *Frontiers in Psychology: Hypothesis and Theory*, *7*, article 420. Doi: 10.3389/fpsyg.2016.00420.

Steller, M., & Köhnken, G. (1989). Criteria-Based Content Analysis. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217-245). New York: Springer-Verlag.

Volbert, R., & Steller, M. (2014). Is this testimony truthful, fabricated, or based on false memory? Credibility assessment 25 years after Steller and Köhnken (1989). *European Psychologist*, 19, 207-220. Doi: 10.1027/1016-9040/a000200.

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities, second edition*. Chichester: John Wiley and Sons.

Vrij, A. (2016). Baselining as a lie detection method. *Applied Cognitive Psychology*, *30*, 1112-1119. Doi: 10.1002/acp.3288

Within-subjects verbal lie detection measures

Vrij, A., Blank, H., & Fisher, R. (2018). A re-analysis that supports our main results: A reply to Levine et al. *Legal and Criminological Psychology*, *23*, 20-23. Doi: 10.1111/lcrp.12121

Vrij, A., Fisher, R., Blank, H. (2017). A cognitive approach to lie detection: A meta-analysis. *Legal and Criminological Psychology*, *22*, 1-21. Doi: 10.1111/lcrp.12088

Vrij, A., Fisher, R., Blank, H., Leal, S., & Mann, S. (2016). A cognitive approach to elicit nonverbal and verbal cues of deceit. In J. W. van Prooijen & P. A. M. van Lange (Eds.), *Cheating, corruption, and concealment: The roots of dishonest behavior* (pp. 284-310). Cambridge, England: Cambridge University Press.

Vrij, A., Leal, S., Fisher, R. P., Mann, S., Dalton, G. Jo, E., Shaboltas, A., Khaleeva, M., Granskaya, J., & Houston, K. (2017a). Sketching as a technique to elicit information and cues to deceit in interpreter-based interviews. *Journal of Applied Research in Memory and Cognition*. Doi: 10.1016/j.jrarmac.2017.11.001

Vrij, A., Leal, S., Mann, S., Dalton, G. Jo, E., Shaboltas, A., Khaleeva, M., Granskaya, J., & Houston, K. (2017b). Using the Model Statement to elicit information and cues to deceit in interpreter-based interviews. *Acta Psychologica*, *177*, 44-53. Doi: 10.1016/j.actpsy.2017.04.011

Vrij, A., Mann, S., Fisher, R., Leal, S., Milne, B., & Bull, R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. *Law and Human Behavior*, *32*, 253-265. Doi: 10.1007/s10979-007-9103-y.

Within-subjects verbal lie detection measures

Table 1.

Total detail, complications, common knowledge details, self-handicapping strategies and proportion of complications as a function of Veracity or Phase.

	Truth	**Lie**	***F***	***p***	**Cohen's *d***
	***M* (*SD*) 95% CI**	***M* (*SD*) 95% CI**			**[95% CI]**
Number of details	137.44 (119.39) (101.03,173.86)	108.15 (57.47) (71.04,145.27)	1.28	. 263	0.31 (-.24,.85)
Number of complications	8.41 (10.91) (5.22,11.59)	2.62 (3.85) (-0.63,5.86)	6.53	.014	0.70 (.14,1.25)
Number of common knowledge details	0.78 (1.31) (-0.29,1.84)	1.85 (3.70) (0.76,2.93)	2.00	.164	0.39 (.16,.93)
Number of self-handicapping strategies	.04 (0.19) (-0.16,0.24)	0.54 (0.76) (0.32,0.76)	11.01	.002	0.91 (.33,1.46)
Proportion of complications	0.71 (0.28) (0.58,0.83)	0.50 (0.35) (0.39,0.64)	5.85	.019	0.62 (.06,1.16)
	Phase 1 (initial recall)	**Phase 2 (post model statement recall)**			
	***M* (*SD*) 95% CI**	***M* (*SD*) 95% CI**			
Number of details	47.28 (44.25) (34.94,59.74)	75.79 (55.03) (60.55,90.65)	41.79	< .001	0.57 (.17,.95)
Number of complications	1.83 (3.74) (0.80,2.83)	3.74 (5.30) (2.34,5.05)	25.61	< .001	0.42 (.03,.72)
Number of common knowledge details	0.68 (1.25) (0.33,1.03)	0.63 (1.71) (0.17,1.09)	0.11	.739	0.03 (-.35,.41)
Number of self-handicapping strategies	0.15 (0.36) (0.06,0.25)	0.13 (0.34) (0.04,0.23)	0.16	.693	0.06 (-.48,.59)

Within-subjects verbal lie detection measures

Proportion of complications	0.53 (0.39) (0.43,0.65)	0.68 (0.35) (0.59,0.77)	10.71	.002	0.37 (.02,.75)

Table 2.

Total detail, complications, common knowledge details, self-handicapping strategies and proportion of complications as a function of Veracity.

	Phase 1 (initial recall)	**Phase 2 (post model statement recall)**	***F***	***p***	**Cohen's *d***
	***M* (*SD*) 95% CI**	***M* (*SD*) 95% CI**			**[95% CI]**
Truth tellers					
Number of details	51.52 (58.70) (34.34,68.70)	85.93 (65.62) (64.85,107.01)	25.58	< .001	0.55 (.01,1.09)
Number of complications	2.63 (4.88) (1.21,4.05)	5.78 (6.46) (3.88,7.68)	22.49	< .001	0.55 (.00,1.09)
Number of common knowledge details	0.59 (1.01) (0.11,1.08)	0.19 (0.48) (-0.46,0.83)	5.68	.025	0.51 (-.04,1.04)
Number of self-handicapping strategies	0.00 (0.00) (0.00,0.00)	0.04 (0.19) (-0.04,0.11)	1.00	.327	0.30 (.24,.83)
Proportion of complications	0.57 (0.40) (0.41,0.73)	0.84 (0.26) (0.72,0.96)	13.52	.001	.80 (.23,1.34)
Liars					
Number of details	42.88 (21.20) (25.38,60.39)	65.27 (39.93) (43.79,86.75)	16.58	< .001	0.70 (.13,1.25)
Number of complications	1.00 (1.70) (-0.45,2.45)	1.62 (2.42) (-0.32,3.55)	3.84	.061	0.30 (.25,.84)

Within-subjects verbal lie detection measures

Number of common knowledge details	0.77 (1.48) (0.27,1.27)	1.08 (2.33) (0.42,1.73)	1.56	.224	0.16 (-.39,0.70)
Number of self-handicapping strategies	0.31 (0.47) (0.12,0.50)	0.23 (0.43) (0.10,0.36)	0.65	.425	0.18 (-0.21,.56)
Proportion of complications	0.48 (0.39) (0.33,0.64)	0.51 (0.36) (0.39,0.64)	0.33	.568	0.08 (.-.47,.62)

[1] Note that many more types of detail than complications, common knowledge details and self-handicapping strategies exist, and that total details is not the equivalent of complications + common knowledge details + self-handicapping strategies. See the Method section for more information.

[2] The correlation between number of complications and the proportion of complications score was $r = .558$, indicating that multicollinearity is not a concern.