Baselining as a Lie Detection Method

Aldert Vrij

University of Portsmouth (UK)

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

For entirely valid reasons practitioners urge researchers to develop within-subjects lie detection methods. This article highlights some problems with the existing and popular nonverbal baseline lie detection method and with the use of the Validity Checklist in SVA. The article further suggests how verbal within-subjects lie detection methods can be introduced in interview settings by implementing verbal lie detection interview protocols designed in recent years. The article concludes that these within-subjects methods do not provide the ultimate solution as clear cut-off scores cannot be established, but they are a step forward in verbal lie detection.
Baselining as a Lie Detection Method

In private talks with practitioners, they often stress the importance of ‘within-subjects’ lie detection tools. That is, they wish to make a decision about the veracity status of an interviewee by comparing different responses made by the same interviewee during a single interview. This request makes perfect sense. There are large individual differences in people’s behaviour, speech and physiological responses (DePaulo & Friedman, 1998). Some people typically make many movements, others do not; some people are eloquent, others are not; some people show large variations in physiological responses, others do not, and so on. Therefore, simple decision rules such as ‘He makes many movements, so he must be lying’ or ‘He does not say much, so he must be lying’ will not work. In physiological (polygraph) lie detection this is widely acknowledged and the two main polygraph tests, the Comparison Question Test (Raskin, 1982, 1986; Raskin & Honts, 2002) and the Concealed Information Test (Lykken, 1959, 1998; Verschuere, Ben-Shakhar, & Meijer, 2011) are both within-subjects tests. Within the polygraph world the debate is heated about which questions to ask to make an adequate within-subjects comparison (Ben-Shakhar, 2008; Grubin, 2008; Vrij, 2008), but the within-subjects comparisons are less prominent in the (non)verbal lie detection domain. In this article I will focus on within-subjects designs in (non)verbal lie detection. I will discuss some of the problems and possible solutions.

The problems

Nonverbal Baselining

Investigators trained in interrogation techniques such as the Reid technique are advised to examine a suspect’s natural, truthful, behaviour at the beginning of an
interview through small talk (Inbau, Reid, Buckley, & Jayne, 2013, p.140). Similarly, in the Improving Interpersonal Evaluations for Law Enforcement and National Security technique, investigators are instructed to observe the interviewees’ normal mode of behaviour (baseline behaviour) when asking non-threatening background questions (Frank, Yarbrough, & Ekman, 2006). Investigators are encouraged to take note of how expressive interviewees are; how much movement they show in their hands, feet, and head; what words they use; and what tone of voice they express. This behaviour is then used as a baseline comparison with the investigative part of the interview, whereby any difference in response between the baseline and investigative part of the interview could be interpreted as a sign of deceit.

This baseline lie detection technique has been described as one of the most striking misuses of psychological research (Moston & Engelberg, 1993). Fundamental differences exist between small talk and the investigative part of the interview. Small-talk conversations are low-stakes situations where the suspect’s 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 the suspect in case s/he will not be believed by the investigator. As a result, both guilty and innocent people 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).

In other words, when making baseline comparisons in the way investigators are taught to do, they fail to take situational factors into account, which is the notion that the same person behaves differently in different situations. This is a well-known error in
social perception, called the fundamental attribution error: the tendency to overlook the impact of situations when explaining someone’s responses (Ross, 1977). There are several situational factors, other than stakes, that affect someone’s behaviour. People react differently: (i) in a formal setting compared with an informal setting (Vrij, 2008); (ii) when they are accused of wrongdoing compared with when they are unchallenged (Vrij, 2006); and (iii) when interviewed by different people (Vrij & Winkel, 1991). In addition, behaviour is topic related. People will respond differently when discussing an embarrassing topic compared with a neutral topic (Kleinke, 1986) or a topic they care about compared with one that they have no interest in (Davis & Hadiks, 1995; Matarazzo, Wiens, Jackson, & Manaugh, 1970). Finally, behaviour can change over time. This can be either in the same interview (Buller & Burgoon, 1996; Burgoon, Buller, White, Afifi, & Buslig, 1999; Stiff, Corman, Krizek, & Snider, 1994; White & Burgoon, 2001) or between interviews held on more than one occasion (Granhag & Strömwall, 2002).

An example of how baseline comparisons can go wrong is a real life high-stakes deception study in which 23 verbal and non-verbal cues were observed, displayed by former US President Bill Clinton during his Grand Jury Testimony about his alleged relationship with Monica Lewinsky (Hirsch & Wolf, 2001). The authors compared a 23-minute segment of the videotape in which Clinton denied having committed various acts of a sexual nature with Monica Lewinsky with 11 minutes of the same testimony when he answered basic questions (his name, his attorney’s name, etc.). Significant differences were obtained for 19 cues. Additionally, the 23-minute segment was compared with a five-minute section of a fundraising speech to a sympathetic crowd. This time, 20 significant differences emerged. Based on these analyses, it cannot be concluded that
Clinton showed 19 or 20 cues to deceit, as Hirsch and Wolf suggested, because the comparisons made between the truthful and deceptive statements were unfair. It seems obvious that a person will show different behaviours when answering basic questions or addressing a crowd in a fundraising speech than when interviewed about an alleged affair.

Although the baseline lie detection method is frequently advocated in interrogation techniques and frequently used (Moston & Engelberg, 1993), it has received virtually no attention from researchers. I could only find one published study (Ewens, Vrij, Jang, & Jo, 2014). In that study, similar to the recommendations in the police literature, truthful baseline behaviour was established prior to the interview through a non-threatening question (‘You just read and signed an informed consent form, could you please tell me what you remember about it and what it said’?). The investigative part of the interview then followed in which the interviewee was aware that s/he would be assessed on whether s/he was lying. During the investigative part, interviewees either discussed the job that they had (truth tellers) or pretended to have (liars). Two target periods were introduced. Target period 1 was at the beginning of the interview, almost immediately after the truthful baseline was established, whereas Target period 2 occurred almost at the end of the interview. In Target 1 interviewees answered the question: ‘Please describe your place of work in as much detail as you can?’ and for Target period 2 they answered: ‘If you were training me to do your job for a day, what things would I need to know about it?’ The extent to which interviewees looked nervous, gave the impression to having to think hard, or attempted to control themselves was assessed. Both liars and truth tellers’ behavioural patterns differed to the same extent between the
baseline behaviour and the investigative parts of the interview, which means that the nonverbal baseline technique did not receive empirical support.

**Verbal Baselining**

Two verbal lie detection tools are used in the field, Statement Validity Assessment (SVA, Köhnken & Steller, 1988; Steller & Köhnken, 1989) and Scientific Content Analysis (SCAN, Sapir, 1987/2000; Vrij, 2008). In SCAN no attention is paid to the fact that individual differences in speech occur and no effort is made to control for such differences. In SVA the problem of individual differences is acknowledged. SVA consists of several phases including a criteria-based content analysis (CBCA) and a Validity Checklist phase. During the CBCA phase a trained coder rates a transcripts for the occurrence (presence or absence) of 19 different CBCA criteria, and calculates a CBCA score, which typically varies from 0 (no criterion is present) to 19 (all criteria are present). Truthful accounts are thought to be richer in quality than deceptive accounts and this is reflected in the number of criteria present: The more criteria present, the more likely it is that a statement is truthful. There is indeed empirical support for this assumption (Amado, Arce, & Fariña, 2015; Amado, Arce, Fariña, & Vilarino, 2016; Oberlader, Naefgen, Koppehele-Gossel, Quinten, Banse, & Schmidt, 2016; Vrij, 2005), yet there is a problem. CBCA scores are affected by factors other than veracity including age and cognitive development, suggestibility, and quality of the interview. In an attempt to control for these factors a Validity Checklist has been developed (Vrij, 2005, 2008). The Validity Checklist consists of factors thought to be worth examining because they may have affected the CBCA score. By systematically addressing each of the factors mentioned on the Validity Checklist, the SVA evaluator explores and considers
alternative interpretations of the CBCA outcomes. Each time an alternative interpretation is rejected, it strengthens the assumption that the CBCA score accurately reflects the veracity of the statement, whereas each time an alternative interpretation is thought to be plausible, the evaluator should consider whether the CBCA score accurately reflects the veracity of the statement.

Assessing the impact of the factors Validity Checklist factors on a CBCA score remains a difficult task. Some factors are difficult to measure, and even if factors can be measured it is often difficult to assess their exact impact. A factor difficult to measure is susceptibility to suggestion. Some interviewees are more prone to an interviewer’s suggestions than are others. The danger is that a suggestible person may be inclined to provide information that confirms the interviewer’s expectations but that, in fact, is inaccurate. Accordingly, Yuille (1988) and Landry and Brigham (1992) have recommended asking the interviewee a few misleading questions at the end of the interview to assess his or her susceptibility to suggestion. Because asking such questions about central information could harm the statement (it could contaminate someone’s memory; Loftus, 2005; Loftus & Palmer, 1974; Porter, Yuille, & Lehman, 1999), Yuille (1988) recommends focusing on peripheral information (e.g., “When you were with your sister, which friend was also there, Claire or Sarah?” when the interviewer is aware that there was no friend present). However, being restricted to asking questions about peripheral information is problematic because interviewees show more resistance to suggestibility for central aspects of an event than for peripheral aspects of an event (Dalton & Daneman, 2006; Goodman, Rudy, Bottoms, & Aman, 1990; Porter, Spencer, & Birt, 2003), and they are more resistant to suggestibility for stressful events, most
likely the central information, than for less stressful events, most likely the peripheral information (Davies, 1991; Porter & Peace, 2007). Therefore, insight into interviewees’ suggestibility for peripheral parts of the event cannot be effectively used to draw conclusions about their suggestibility for core events.

Some Validity Checklist factors are easier to measure, but once measured it is still difficult or perhaps even impossible, to determine their exact impact on CBCA scores. The quality of interviewing, for example, can be measured and often influences the amount of information that an interviewee reports (Griffiths & Milne, 2006; Oxburg, Ost, & Cherryman, 2012; Vrij, Hope, & Fisher, 2014). However, this does not mean that each individual interviewee will necessarily be affected by a particular interview style. And even if the interviewee was affected by the style of interviewing, it is often impossible to determine afterwards the precise impact this had on the quality of his/her statement. In other words, the precise impact of Validity Checklist factors on an individual statement can often not be determined, but only estimated. Obviously, mistakes in these estimations can occur.

A good illustration of the difficulty that SVA experts face in determining the exact impact of Validity Checklist factors on CBCA scores is the field study conducted by Lamers-Winkelman and Buffing (1996). In this study, raters were instructed to take the age of the child into account when calculating CBCA scores. Nevertheless, six criteria positively correlated with age. In other words, even after being instructed to correct CBCA scores for age, the results still showed age related effects with older children obtaining higher CBCA scores than younger children.

Given these difficulties in identifying the presence of Validity Checklist factors, and in examining the exact impact of these factors on CBCA scores, it is clear that the
Validity Checklist procedure is more subjective and less formalised than the CBCA procedure (Steller, 1989; Steller & Köhnken, 1989). It is therefore not surprising that if two experts disagree about the veracity of a statement, the most likely reason for the disagreement is that they disagree about the impact of Validity Checklist factors on that statement (Gumpert & Lindblad, 1999; Vrij, 2008). Neither it is therefore surprising that experts often ignore the outcomes of the Validity Checklist analysis. That is, SVA experts sometimes highlight the influence of Validity Checklist factors on statements in general, without discussing how these factors may have influenced the statement of the particular interviewee they were asked to assess. Also, experts sometimes indicate possible external influences on statements, but are then still inclined to rely upon the CBCA outcome, and tend to judge high-quality statements as truthful and low-quality statements as fabricated.

The solutions

Nonverbal Baselining

Given that nonverbal behaviours are affected by a variety of factors, discussed above, when investigators wish to compare a person’s deceptive nonverbal response with a truthful nonverbal response from the same person, they need to make sure that the deceptive and truthful responses are taken from the same interview setting; that the person talks about similar topics in the deceptive and truthful parts; and that these parts were discussed within a short period of time from each other. In that respect, Vrij (2008) refers to a comparable truth baseline. Comparable means that the baseline the investigator uses must be similar in content, context, stakes, and cognitive and emotional involvement to investigative questions (Vrij, 2008).
In a recent study, such a comparable truth baseline was empirically tested (Palena, Vrij, Caso, & Orthey, 2016). In the baseline part of the interview truth tellers and liars truthfully reported three tasks they just had conducted. After that, in the investigative part of the interview in which truth tellers and liars were aware that their responses would be assessed in terms of veracity, truth tellers reported three more, but similar, tasks they had conducted, whereas liars pretended to have carried out these three additional tasks. Movements of arms, hands and fingers were assessed. Results indicated that truth tellers and liars changed their behaviours from the baseline to the investigative part of the interview to the same extent. For truth tellers the tasks they described in the baseline and investigative parts of the interview were very similar. The main difference for them was that they were aware that their veracity status would be assessed in the investigative part of the interview and not in the baseline part of the interview. This means that it will be virtually impossible to establish a reliable nonverbal baseline outside the investigative part of the interview, because truth tellers are aware that their veracity status is not assessed outside the investigative part of the interview. The only solution left over is to introduce a truthful baseline during the investigative part of the interview, but that may be difficult to achieve in real life, because investigators often do not know with any degree of certainty which (if any) of the details an interviewee reports during the investigative part of the interview are truthful. I am therefore pessimistic about the development of a reliable nonverbal baseline lie detection method.

**Verbal Baseline**

I am more optimistic about the possibility of developing verbal baseline lie detection methods. In fact, several of them already exist, and I will briefly introduce them
here. In recent years researchers started to develop interview protocols that elicit or enhance verbal cues to deception. The most frequently researched methods are cognitive lie detection, which consists of three elements, imposing cognitive load, encouraging interviewees to say more, and asking unexpected questions (Vrij, 2015; Vrij, Fisher, & Blank, 2016; Vrij, Leal, Mann, Vernham, & Blankaert, 2015), the verifiability approach (Vrij & Nahari, 2016; Vrij, Taylor, & Picornell, 2016) and the Strategic Use of Evidence (Granhag & Hartwig, 2015; Hartwig, Granhag, & Luke, 2014). Each interview protocol contains within-subjects comparisons.

**Imposing cognitive load.** Imposing cognitive load is based on the well-established empirical finding that in interview settings lying is typically more mentally taxing than truth telling (e.g., fMRI research, Christ, Van Essen, Watson, Brubaker, & McDermott, 2009; Vrij & Ganis, 2014). Imposing cognitive load refers to investigators’ interventions aimed at making the interview setting mentally more difficult. Liars, who often require more cognitive resources than truth tellers in interviews settings, will have fewer cognitive resources left over. If cognitive demand is further raised, which could be achieved by making additional requests, liars may be less able than truth tellers to cope with these additional requests. One way to impose cognitive load is by asking interviewees to tell their stories in reverse order.

The reverse order technique can be used as a within-subjects lie detection tool. Interviewees are first asked in an open-ended question to describe in detail what they have experienced. Interviews typically answer this question in chronological time order as that is how they have experienced the event they discuss (truth tellers) or have rehearsed the story they discuss (liars). After this free recall, interviewees are invited to
report the event again, but this time in reverse order. This is a technique commonly used in interviews with cooperative witnesses (truth tellers) as it typically leads to new information (reminiscences) (Fisher & Geiselman, 1992; Fisher, Schreiber Compo, Rivard, & Hirn, 2014). This instruction invites truth tellers to think about the event again, but from a different perspective. This often leads to reminiscences. Liars have shown to include fewer reminiscences than truth tellers (Ewens, Vrij, Mann, & Leal, 2016; Shaw et al., 2014; Vrij, Leal, Mann, & Fisher, 2012), probably for two reasons. First, they may see this as a test (‘Can I report again what I just reported, but now in reverse order?’), because they are very much concerned with consistency, more so than truth tellers (Vrij, Fisher, Blank, Leal, & Mann, 2016). Liars typically believe that consistent statements will appear more credible, and, indeed, investigators often use this consistency heuristic (Strömwall, Granhag, & Hartwig, 2004; Vredeveldt, van Koppes, Granhag, 2014). Adding new information in the reverse order recall makes this recall less consistent with the initial chronological recall, so from a consistency perspective liars are unlikely to add new detail. Second, when liars are satisfied with their initial chronological recall, they will see no reason to add information when reporting the rehearsed story in reverse order, because this additional information may give investigators further leads to check the veracity of the statement. Apart from reminiscences, another verbal cue may distinguish truth tellers from liars. Reporting information in reverse order is mentally taxing, particularly for liars. The complexity of the task means that errors may occur in the form of contradictions between the chronological and reverse order control, particularly in liars (Vrij, Leal, Mann, & Fisher, 2012). In sum, in comparing the chronological statement with the reverse order statement, the investigator should pay attention to reminiscences.
and (if they occur) contradictions. Interviewees who add reminiscences are likely to be truthful, interviewees who do not add reminiscences are likely to be deceptive. Contradictions may provide further evidence that the interviewee was lying.

**Encouraging interviewees to say more.** If people in daily conversations are asked a question about their activities, they rarely provide all the information they know (Vrij, Fisher, & Hope, 2014). Instead, they are likely to provide a summary of their activities, highlighting some core issues (‘I did some shopping in the morning, and had a BBQ in the evening’). Interviewees realise that in interview settings they have to provide more information than that but they still have inadequate expectations about how much detail is expected from them (Fisher, 2010; Fisher, Milne, & Bull, 2011).

Investigators can alter the participants’ expectations about how much detail is required by providing them with a model answer, a detailed statement about an event unrelated to the topic of investigation. In a study in which the model statement was used, half of the interviewees who lied or told the truth about an insurance claim, listened to an audiotaped detail account of someone describing his experiences at a day attending Formula 2 motor racing. Interviewees who were exposed to this model statement gave twice as much information about their insurance claim than those who did not listen to the model statement (Leal, Vrij, Warmelink, Vernham, & Fisher, 2015).

The model statement technique allows investigators to make within-subjects comparisons. Start the interview with inviting the interviewee to report in as much detail as possible what s/he has experienced. After this initial recall, let him/her listen to a model statement and invite him/her again to report in as much detail as possible what s/he
has experienced. Unlike in the reverse order technique described above, liars will understand that more information is required from them after listening to the model statement. As a result, both truth tellers and liars will add more detail, but the type of detail they add is different. First, the additional detail sounds more plausible in truth tellers than in liars (Leal et al., 2015). Truth tellers can search their memories and add more detail to their story, whereas liars have to fabricate additional detail on the spot. The latter is mentally taxing and leads to reminiscences that do not sound as plausible as the truth tellers’ reminiscences. Second, truth tellers elaborate on the core and peripheral elements of their story, whereas liars mainly elaborate on the peripheral elements of their story. Liars prefer to avoid providing potentially incriminating information (Granhag & Hartwig, 2008) and a possible solution is to talk around the core event and add less relevant information instead. Thus, in this method investigators should judge the plausibility of the reminiscences (truth tellers are more plausible) and the extent to which the additional information is about the core event (truth tellers elaborate more about the core event).

**Asking unexpected questions.** A consistent finding in the deception literature is that liars prepare themselves for anticipated interviews by preparing possible answers to questions they expect to be asked (Hartwig, Granhag, & Strömwall, 2007). This strategy of preparing answers for possible questions makes sense. Planning makes lying easier and planned lies typically contain fewer cues to deceit than spontaneous lies (DePaulo et al., 2003). However, preparing for answers has a limitation. It will be fruitful only if liars correctly anticipate which questions will be asked. Investigators can exploit this limitation by asking questions that liars do not anticipate. Though liars can refuse to
answer unexpected questions by saying “I don’t know” or “I can’t remember,” such responses will create suspicion if these questions are about central aspects of the target event. A liar, therefore, has little option other than to fabricate a plausible answer on the spot, which is cognitively demanding.

The unanticipated questions approach can be used as a within-subjects technique. For liars, expected questions should be easier to answer than unexpected questions, because they can give their planned and rehearsed answers to the expected questions but they need to fabricate answers to the unexpected questions. The difference liars experience in cognitive load while answering these two sets of questions should become evident in their verbal responses. In contrast, truth tellers experience similar levels of cognitive load while answering expected and unexpected questions, and they should produce more comparable answers to the expected and unexpected questions than liars.

To date, in the unexpected-questions technique, within-subjects comparisons have been made in three different manners, each of which examines a different outcome measure (e.g., dependent variable). In a first set of studies, truth tellers went on their missions in small groups, whereas liars pretended to have carried out such missions. They prepared themselves for the interview in these small groups but were interviewed separately. In the interview they were asked a mixture of expected and unexpected questions, and the answers given by the different individuals belonging to a group were compared to one another (Roos af Hjelmsäter, Ohman, Granhag, & Vrij, 2014; Vrij et al., 2009). Liars showed less overlap in their answers to the unexpected questions than to the expected questions, whereas no difference emerged in truth tellers when comparing the expected and unexpected questions.
In a second set of studies, individual truth tellers and liars were asked a mixture of expected and unexpected questions. People, including interviewers, pay attention to detail when making veracity judgments, and the richer an account is perceived to be in detail, the more likely it is to be believed (Bell & Loftus, 1989). Liars are aware of this and are therefore keen to provide details to make an honest impression (Nahari, Vrij, & Fisher, 2012). As a result, liars may prepare a detailed alibi and report it as soon as the opportunity arises. This opportunity does arise when an expected question is asked. In contrast, liars will not have prepared answers for unexpected questions and may therefore struggle to generate detailed answers. For truth tellers, the difference in detail between expected and unexpected questions should be less pronounced. When asked about an event they have experienced, truth tellers will search their memories for details about that event and there is no reason why those details are less accessible for unexpected than for expected questions, as long as the expected and unexpected questions are both about core aspects of the event. Therefore, as research has shown, liars’ prepared answers to expected questions are more detailed than the truth tellers’ answers to expected questions, because it is easier to recall a planned story (e.g., liars) than to search memory for details (e.g., truth tellers). In contrast, liars’ answers to unexpected questions are less detailed than truth tellers’ answers to unexpected questions, because it may be more difficult to fabricate an answer (e.g., liars) than to search memory for details (e.g., truth tellers) (Knieps, Granhag, & Vrij, 2013; Lancaster, Vrij, Hope, & Waller, 2012; Warmelink, Vrij, Mann, Jundi, & Granhag, 2012).

In a third set of studies individual truth tellers and liars were interviewed and were asked the same unexpected question twice. The overlap in the two answers was examined
Baselining

When liars have not anticipated the question, they have to fabricate an answer on the spot. A liar’s memory of this fabricated answer may be more unstable than a truth teller’s actual memory of the event. Therefore, liars should contradict themselves more than truth tellers when they are asked the same unexpected question twice, and this may be particularly the case when the question is asked twice in different formats (e.g., verbal recall versus sketching). Truth tellers will have encoded the topic of investigation along more dimensions than will liars. As a result, compared with liars, truth tellers should be able to recall the event more flexibly (along more dimensions). For example, in Leins, Fisher, and Vrij’s (2012) experiment truthful participants had visited a room whereas deceptive participants had not. In the interview however, all participants claimed to have visited the room. Participants were asked to verbally recall the layout of the room twice, to sketch it twice, or to verbally recall it once and to sketch it once. Liars contradicted themselves more than truth tellers, but only in the verbal recall – drawing condition. In other words, for this within-subjects comparison to work it appears to be necessary that the interviewee answers the same question in different formats.

In sum, the unexpected questions technique can be executed in different ways, and each of the three ways in which it has been employed to date has been fruitful in terms of eliciting verbal cues to deceit and distinguishing between truth tellers and liars using within-subjects lie detection methods.

Verifiability Approach. Central to the Verifiability Approach are two assumptions. First, truth tellers typically include more detail into their accounts than liars (Vrij, 2008). Observers seem to be aware of this, as the richer an account is perceived to
be in detail, the more likely it is to be believed (Bell & Loftus, 1989). As a result, liars are inclined to provide many details to make an honest impression on observers (Nahari, Vrij, & Fisher, 2012). Second, liars prefer to avoid mentioning too many details out of fear that investigators can check such details and will discover that they are lying (Nahari et al., 2012). A strategy that compromises between these two conflicting motivations is to provide details that cannot be verified. Verifiable detail include: Activities i) carried out with or ii) witnessed by named persons or persons who can be identified by their description who the interviewer can consult; iii) activities that the interviewee believes may have been captured on CCTV; and iv) activities that lead a trace, such as activities that have been recorded through technology (using debit cards, mobile phones, or computers). Indeed, results of all six Verifiability Approach studies published to date revealed that liars include fewer verifiable details than truth tellers (Harvey, Vrij, Nahari, & Ludwig, 2015; Nahari & Vrij, 2014; Nahari, Vrij, & Fisher, 2014a, b; Nahari, Leal, Vrij, Warmelink, & Vernham, 2014; Vrij, Nahari, Isitt, & Leal, 2016). Interestingly, this verifiability effect becomes stronger if interviewees are asked at the beginning of the interview to incorporate (if possible) details that the investigator can check (Harvey, Vrij, Nahari, & Ludwig, 2015; Nahari, Vrij, & Fisher, 2014b; Vrij, Nahari, Isitt, & Leal, 2016). The within-subjects aspect of this method is that the proportion of verifiable details (verifiable details / verifiable and unverifiable details) is higher for truth tellers than for liars.

**Strategic Use of Evidence.** Liars and truth tellers enter interviews with different counter-interrogation strategies (Gran Hag & Hartwig, 2008). Liars are inclined to use avoidance strategies (e.g., in a free recall avoid mentioning where they were at a certain time) or
denial strategies (e.g., denying having been at a certain place at a certain time when asked directly) (Hartwig et al., 2007). In contrast, truth tellers are generally more forthcoming and “tell the truth like it happened” (e.g., Hartwig et al., 2007; Kassin, 2005). When investigators possess critical and possibly incriminating background information (evidence) in a case, they can exploit these differential truth tellers’ and liars’ strategies by introducing the available evidence during the interview in a strategic manner. When questions about the evidence are asked, liars use more avoidance strategies, whereas truth tellers use more forthcoming strategies (Granhag & Hartwig, 2008). The result is that truth tellers’ accounts are more consistent with the available evidence than liars’ accounts (Hartwig, Granhag, & Luke, 2014).

Apart from this statement – evidence inconsistency, the SUE technique also elicits within-statement inconsistency, which is a within-subjects examination. During SUE interviews liars may start to become aware that the investigator possesses a piece of evidence liars initially did not think the investigator possessed. Liars then tend to adjust their story somewhat in an effort to provide a plausible but innocent explanation for that piece of evidence. Research has shown that this results in liars contradicting what they said previously (within-statement inconsistency, Hartwig, Granhag, & Luke, 2014). For example, a liar who tries to hide the fact that he used a safe at a train station on a particular Sunday afternoon may initially deny to have been at the train station on that Sunday afternoon. However, after realising that the investigator may have evidence that he was at the train station on that Sunday afternoon, he may now adjust his story by saying that he was at the train station (albeit not to use a safe).

**Applying Verbal Baselining in Real Life**
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 verifiable than unverifiable details in their statements and liars always include more unverifiable than verifiable details in their statements (e.g. Nahari & Vrij, 2015; Nahari et al., 2014a). In that case a clear cut-off score can be established, but this does not happen in real life. All that can be concluded is that truth tellers typically include a higher proportion of verifiable details in their statements than liars. This still leaves practitioners with the following problem: When is the proportion of verifiable details high enough to decide that the interviewee is telling the truth? 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 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 verifiable and unverifiable details are included (more likely to be influenced by veracity).

**Conclusion**

For entirely valid reasons practitioners urge researchers to develop within-subjects lie detection methods. In this article I highlighted some problems with the existing and popular nonverbal baseline lie detection method and with the use of the Validity Checklist in SVA. I further highlighted how verbal within-subjects lie detection methods can be introduced in interview settings by implementing verbal lie detection
interview protocols designed in recent years. Although within-subject lie detection methods do not provide the ultimate solution, they can be seen as an improvement. Hopefully this article will make practitioners think about how to correctly use verbal within-subjects lie detection methods and will encourage researchers to pay more attention to the important issue of within-subjects verbal lie detection.
References


Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin, 100,* 78-100. DOI: 10.1037/0033-2909.100.1.78


DOI: 10.1111/lcrp.12017


DOI:10.1080/10439463.1993.9964670


