Observers' performance at evaluating truthfulness when provided with Comparable Truth or

Small Talk Baselines

Letizia Caso[1], Nicola Palena

Department of Human and Social Sciences, University of Bergamo

Aldert Vrij

Department of Psychology, University of Portsmouth

Augusto Gnisci

Department of Psychology, University of Campania Luigi Vanvitelli

Ethical standards

Ethical approval

All procedures performed in studies involving human participants were in accordance with the

ethical standards of the institutional and/or national research committee (the Department of

[1] Correspondence concerning this article should be addressed to: Prof. Letizia Caso, Department of Human and Social Sciences, University of Bergamo. Piazzale Sant'Agostino, 2, 24129 Bergamo (Italy) Email: letizia.caso@unibg.it

Running Head: Observers' accuracy in baseline interviews

Human and Social Sciences of University of Bergamo) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent

Informed consent was obtained from all individual participants included in the study.

Abstract

Research has shown that the Comparable Truth Baseline Technique outperforms the Small Talk with respect to the elicitation of cues to deception. However, their impact on observers' accuracy has not been evaluated yet. In this experiment, participants ($N = 74$) watched ten interviews where senders either lied or told the truth about a set of tasks. Half of the interviews were conducted with a Comparable Truth Baseline, the other half with a Small Talk Baseline. As predicted, results showed that observers in the Comparable Truth Baseline condition outperformed the participants in the Small Talk Baseline condition in terms of total accuracy rates. The paper sheds light on the impact of the two baseline techniques in distinguishing truth tellers from liars and discourages the use of a small talk baseline. It also provides insights for future studies.

Keywords: veracity assessment, baseline technique, observers' accuracy, comparable truth, small talk, deception detection, interviewing techniques

Observers' performance at evaluating truthfulness when provided with Comparable Truth or

Small Talk Baselines

Introduction

Research has shown that people's accuracy in assessing truthfulness and deceit is low (Bond & DePaulo, 2006; Vrij, 2008). Bond and DePaulo (2006) meta-analysis including almost 25,000 observers found a 54% accuracy rate, whereby 50% could be achieved by chance. In addition, accuracy of deception judgments are unrelated to confidence in one's own judgements (DePaulo, Charlton, Cooper, Lindsay, & Muhlenbruck, 1997), suggesting that people have no insight into their own competence. Additionally, people from whom one should expect a higher accuracy, such as police officers, are not more accurate than laypeople (Bond & DePaulo, 2006).

To explain the low accuracy rates, it has been argued that cues to deception are faint and unreliable (Vrij, 2008). DePaulo and colleagues (2003), in one of the most comprehensive meta-analyses on cues to deception, have found that effect sizes for behavioural and verbal differences between truth and lie telling are small, with an average effect size of Cohen's $d = .25$ for the most diagnostic cues (Vrij & Granhag, 2012). Consequently, it should not be surprising that observers' accuracy is low, as having weak cues to rely on when making judgments makes the judgment itself difficult (Hartwig & Bond, 2011). It is for this reason that academics are now focusing on interviewing techniques aiming at enhancing differences between truth and lie telling (Granhag & Vrij, 2010; Vrij, 2014; Vrij & Granhag, 2012). The rationale behind this is that if truth tellers and liars do not differ much, it is desirable to find a way to make such differences more evident. Academics have explored several approaches that indeed augmented differences between truth tellers and liars, including the Strategic Use of Evidence (Granhag & Hartwig, 2015) and cognitive credibility assessment (Vrij, 2015; Vrij, Fisher, & Blank, 2017).

There is an alternative interviewing strategy often proposed within police forces: The baseline technique (Ewens, Vrij, Jang, & Jo, 2014; Vrij, 2016). The underlying idea is that since there are interpersonal differences in behaviour, an interviewer should start by obtaining a

4

behavioural baseline of the interviewee through observing the responses in chitchat while answering trivial questions. Then, while the interview progresses, the interviewer should compare this baseline behaviour with interviewees' responses to target questions (questions related to the topic under investigation). If any difference arises, then one may conclude that the interviewee is lying (Frank, Yarbrough, & Ekman, 2006).

The use of the baseline technique in this manner is problematic. People's nonverbal behaviour is related to stakes and topic of conversation (Vrij, 2008), and in terms of stakes and topic of conversation the baseline response and the target response are not comparable to each other. The result is that both liars and truth tellers are likely to change their behaviour when baseline and target responses are compared (Caso, Maricchiolo, Bonaiuto, Vrij, & Mann, 2006; Gnisci, Caso e Vrij, 2010; Moston & Engelberg, 1993). Ewens, Vrij, Jang and Jo (2014) empirically tested the efficacy of this type of baselining, which they labelled "Small Talk". When interviewing their participants, they started by asking an initial small talk baseline question, followed by several target questions. It was found that both truth tellers and liars changed their behaviour during the interview. Therefore, they concluded that the small talk baseline in not an effective strategy to detect deception.

Ewens and colleagues (2014) also noted that there is a different type of baseline which may work better, the comparable truth baseline. They described it as follows: "*Comparable means that the baseline the investigator uses must be similar in content, context, stakes, and cognitive and emotional involvement to investigative questions*" (Vrij, 2016, p. 1114). Vrij and Mann (2001) gave a real life example of this type of baseline. They compared several behaviours displayed by a suspect in a murder case in different phases of the police interrogation. The suspects was interrogated about his activities during the day of the murder and provided a detailed answer covering the entire day. Police investigated his whereabouts and could only verify his morning activities. Eventually, it became known that he had met the victim in the afternoon and killed her later. Vrij and Mann (2001) analysed the videotaped interrogation and found that the suspect did

5

show a difference in his behaviour when he discussed his activities in the morning (truth) compared to the afternoon and evening (lie). Their results therefore gave credit to the potential effectiveness of the comparable truth baseline.

There is to date only one experimental study in which comparable truth and small talk baselines were compared (Palena, Caso, Vrij, & Orthey, 2018). In the comparable truth baseline condition, the baseline question referred to three tasks that the participants performed. All participants had to answer the baseline question truthfully. Then, in the target phase of the interview, half of the participants told the truth and the other half lied about three additional, but similar, tasks. In the small talk baseline condition, the baseline question referred to personal information about the interviewee (e.g.: the last year spent as a student). Again, all participants had to answer the baseline question truthfully. The target questions referred to the same three additional tasks as for participants in the comparable truth baseline condition. Again, half of the participants responded to this target question truthfully, whereas the other half lied. The authors compared the baseline and target responses in terms of similarity for both nonverbal and verbal behaviour. Truth tellers' levels of similarity did not differ from those of liars in the small talk baseline condition, but the truth tellers' levels of similarity were higher than those of liars in the comparable truth baseline condition, but only for spatial details. This study thus replicated Ewens et al.'s (2014) study that a small talk baseline does not work. Second, it showed that nonverbal baselining is problematic, but that verbal baselining may be more effective.

However, verbal baselining is not straightforward either as speech content is affected by the topic of conversation (Vrij, 2008). For example, in Ewens' et al (2014) study, the baseline concerned the informed consent form, whereas the target phase concerned the actual/pretended job. Consequently, the speech content of the two phases is expected to be different regardless of veracity, as the interviewee talks about two different topics.

Palena and colleagues (2018) did not test whether the differences in speech content would be clear to lay observers. Bond and DePaulo's meta-analysis (2006) underlined that when observers

were previously exposed to senders' truthful baselines, their accuracy improved. However, such previous exposure cannot be considered a baseline obtained through strategic questioning. Rather, since observers had simply the opportunity to become familiar with the sender, it was possibly the result of "*baseline familiarity*" (Bond & DePaulo, 2006; Brandt, Miller, & Hocking, 1980, 1982). Feeley, deTurck and Young (1995) provided their participants with zero, one, two or four truthful baseline exposures, and found that there was a positive linear relationship between the amount of familiarity with the sender and observers' accuracy. Results about the positive effects of baseline familiarity are important in personal and intimate relationships but baseline familiarity may be difficult to achieve during police investigations, as suspects and investigators are often strangers to each other and there is often the need to interview the suspect as soon as possible. Hence, it is more likely that police officers will try to obtain baselines with specific questioning rather than through increased familiarity.

Based on the literature presented above and expecting better efficacy of the comparable truth over the small talk baseline, we made the following predictions.

H1: Observers in the comparable truth baseline condition will reach higher overall accuracy rates than those in the small talk baseline condition.

H2: Observers in the comparable truth baseline condition will also reach higher accuracy rates for truth (Hypothesis 2a) and lie detection (Hypothesis 2b).

Material and methods

Participants

Seventy-four participants (56 women and 18 men), between 19 and 51 years of age ($M = 25.67$; $SD = 6.14$), took part in the experiment. One of them was excluded from the analyses as the person did not followed the instructions. This left us with a total sample of seventy-three participants. Thirty-seven participants were allocated to the comparable truth condition, whereas the remaining thirty-six were allocated to the small talk condition.

Design

The experiment was based on a two-group comparison. The type of baseline (Comparable Truth vs. Small Talk) was the between-subjects factor. Total, truth and lie accuracy rates, and d-prime and β values were the dependent variables.

Stimulus material

Twenty video stimuli were used for this experiment. All senders portrayed in the videos performed a mission that consisted of a series of tasks. The mission started with the participant receiving an envelop from the experimenter, which contained a PC password. Then, the experimenter left the room. As soon as the participant logged onto the PC only one file, a word document named "Read Me", appeared on the desktop. Such file informed the participant that s/he had now to look for CD-ROM that was placed inside the only backpack available in the room. The word document ended informing the participant to watch the video file that was recorded in such CD-ROM. This showed a mean telling to the participant to look again into the backpack to search for a key. The man also explained that the key served to open a safe-deposit box that was in the room. Once the participant opened the box, s/he found additional instruction that asked them to send an e-mail to a specific address and to exit the room and wait for a person to come. This person was actually a confederate and the meeting served to split the mission into two subsets. Everything that happened before the meeting was the comparable truth baseline section. Everything that happened after the meeting was the target section of the interview. The confederate gave a newspaper to the participant, and informed him/her that it contained further instruction. Then, the confederate left the room. The instruction guided the participant to a room adjacent to the first one and informed them that they had to look for a pen-drive, which was attached on a coat hook. Once found, the participant had to take it and switch it with another pen-drive, which was hidden inside a book at the bottom of the wardrobe located in the room. The instruction also told the participant to leave the newspaper with the instruction near that book and to keep the first pen-drive with them for the rest of the experiment. The participant then came back to the first room waited for the interviewer. Of the twenty videos, ten were used for the comparable truth condition. Here, for the

baseline questioning, the participants reported everything that happened before the meeting with the confederate. For the target questioning, the participant reported everything that happened after the meeting with the confederate. The remaining ten videos were used for the small talk condition. Here, the baseline questioning consisted of participants providing personal information (e.g.: describe their last year as a worker), whereas the target questioning was identical to that for participants in the comparable truth condition. That is, also participants in the small talk condition had to report everything that happened after the meeting with the confederate for the target questioning. All senders were honest in the baseline phase, whereas half of them were instructed to tell the truth and the other half to lie when responding to the target questioning.

Each video sequence was produced as follows: a black screen with a white text indicating "Baseline" appeared and lasted for three seconds. Then, the first sender appeared on the screen and started answering to the baseline question. Once s/he finished, a second black screen with the text "Target" appeared, also lasting three seconds. Then the first sender started answering the target question. Once s/he finished answering, another black screen appeared, and lasted for 30 seconds. In this time-window, participants had to make their veracity decisions. When the 30 seconds expired, a high frequency sound warned the participant that the time to evaluate the first sender was over and that the second sender was going to appear on the screen. This sequence was repeated until the end of the 10 videos. Participants expressed their veracity decisions answering the following question: "*Do you believe the interviewee was…*" . The answer alternatives were "*Telling the truth*" and *"Lying"*. Senders' veracity status was counterbalanced and each ten targets tape consisted of five truth tellers and five liars.

Procedure

Upon the arrival, participants met the experimenter and were briefed about the aim of the study. They were informed that they were going to watch some interviews that were divided into two sections: A baseline section and a target section. Observers where informed that the senders were always honest in the baseline section whereas they may have been either telling the truth or

lying in the target section. Observers were asked to decide whether each sender was telling the truth or lying in the target section. They were told that they needed to pay attention to (non)verbal deviations from the baseline to make their decision. No training was offered. Assignment to the two baseline conditions was random. The participants were also informed that there were more answer sheets (twenty) than the actual amount of stimuli and also that truth and lie telling may be balanced in different amounts. Such instructions were given to prevent participants making decisions based on balancing expectations rather than lie detection task decisions. The participant also read and signed a consent form and were offered an additional point for a university exam if they performed well. Eventually, all participants received the point, regardless. Once they confirmed they understood the instructions, they were left alone in a room with a computer and the answer sheets. They were also instructed to exit the room once they completed their task. They then started watching the stimuli. Each participant was shown the tapes individually.

Results

A *t*-test with type of Baseline as factor and overall accuracy rates as the dependent variable revealed that participants in the comparable truth condition were more accurate [$M = 56.49$; $SD = 18.74$, 95% CI (50.24, 62.73)] than those in the small talk condition [$M = 47.41$; $SD = 18.62$, 95% CI (41.11, 53.71)], $t(71) = 2.076$, $p = .042$, Cohen's $d = .49$, supporting Hypothesis 1. One *t*-test, again with type of Baseline as factor, showed no difference between conditions for truth accuracy, $t(71) = 1.200$, $p = .22$, Cohen's $d = .28$ but showed a difference for lie accuracy, $t(71) = 1.990$, $p = .050$, Cohen's $d = .47$, with the Comparable Truth baseline condition resulting in a higher lie accuracy rate than the Small Talk baseline condition (Table 1). Therefore, Hypothesis 2a was not supported but Hypothesis 2b received support.

Enter Table 1 about here

Further analyses (see Table 2) showed that in the *comparable truth baseline* condition the total accuracy rate was significantly above chance. No other accuracy rate differed from chance and only lie accuracy rate differed from the accuracy rates found in Bond and DePaulo's (2006) meta-

analysis. In the *small talk baseline* condition, none of the accuracy rates differed from chance and the total and truth accuracy rates were significantly lower than those found in Bond and DePaulo's (2006) meta-analysis.

Enter Table 2 about here

Signal Detection Analyses

It has been suggested to use Signal Detection Theory to analyse the accuracy of deception judgments in more detail (Jupe, Akehurst, Vernham, & Allen, 2016; Meissner & Kassin, 2002). Therefore, participants' performance was assessed via discrimination accuracy, using d-prime values, and responding bias, using β values. The former is a measure of sensitivity expressed in standard deviations units (Stanislaw & Todorov, 1999). Values of 0 indicate an inability to distinguish between the signal and noise, in our case liars from truth tellers. Values greater than 0 indicate that the observers are indeed able to distinguish truthful from lying senders.

The β value is a measure of response bias (Stanislaw & Todorov, 1999) whereby values of 1 indicate no response bias. On the other hand, values greater than 1 indicate a truth bias and values below 1indicate a lie bias.

A first t-test with baseline condition as factor and d-prime values as the dependent variable showed that observers in the comparable truth baseline [$M = .35$; $SD = 1.00$, 95% CI (.02, .68)] were better than those in the small talk baseline [$M = -.13$; $SD = .99$, 95% CI (-.47, .20)] at discriminating truth tellers from liars, $t(71) = 2.079$, $p = .040$, Cohen's $d = .49$. This, again, supports Hypothesis 1.

A between-subjects t-test on β values showed that participants in the comparable truth condition [$M = 1.27$; $SD = .44$, 95% CI (1.12, 1.42)] did not differ from those in the small talk condition [$M = 1.21$; $SD = .48$, 95% CI (1.04, 1.37)] in their response bias, $t(71) = .59$, $p = .55$, Cohen's $d = .14$. Additionally, β values for participants in both baseline conditions were significantly greater than 1, indicating that all participants were truth biased, regardless of the type of Baseline exposure (Table 3).

Enter Table 3 about here

Discussion

In this experiment, we found that observers in the comparable truth condition were more accurate in distinguishing truth tellers from liars than observers in the small talk condition. In addition, observers in the comparable truth condition performed significantly better than chance levels whereas observers in the small talk condition did not. This shows the benefit of using a comparable truth baseline compared to a small talk baseline. Our results discourage the use of a small talk baseline technique, the technique used by practitioners (Inbau, Reid, Buckley, & Jayne, 2013) and advocated by some academics (Frank et al., 2006).

It should be noted that, although better than the small talk baseline, the comparable truth baseline is still not ready to be implemented in real life. It needs to be improved and should focus on verbal content only (Vrij, 2016). Verbal content is more diagnostic than nonverbal behaviour (Bond & DePaulo, 2006) and it may be easier to establish verbal than nonverbal baselines as Palena et al's (2018) study suggests. Future efforts can improve the comparable truth baseline technique integrating it with the techniques used for strategic questioning. For example, the interviewer can start the interview with the baseline approach and then employ strategic questioning using the already available within-subjects comparisons techniques reported in Vrij (2016), such as the reverse order technique (Ewens, Vrij, Mann, & Leal, 2016; Vrij, Leal, Mann, & Fisher, 2012) and the verifiability approach (Nahari, Vrij, & Fisher, 2014; Vrij, Nahari, Isitt, & Leal, 2016). In the reverse order technique, the interviewee is first asked to report the story in chronological order. Then, s/he is asked to report the same story from the end to the beginning, which typically results in truth tellers to report more reminiscences but fewer contradictions than liars (Vrij, 2016). According to the verifiability approach liars try not to report details which can be verified by an investigator. Consequently, the proportion of verifiable details (verifiable detail/(verifiable and unverifiable details combined) is higher for truth tellers than for liars (Nahari et al., 2014; Vrij, 2016).

12

Another possibility is to create a (semi)structured interview protocol which incorporates a baseline technique. A good example is the Assessment Criteria Indicative of Deception (ACID, Colwell et al., 2007). This protocol starts with a baseline question coupled with rapport building and is based on the concept of Differential Recall Enhancement (DRE), whereby it is thought that the use of mnemonics techniques and forced choice questions will result in truth tellers reporting more new details than liars. Colwell et al. (2007, 2013) used a small talk baseline (the last meal someone had or the first day of a semester). Future studies may explore how the ACID technique performs when a comparable truth baseline is used.

Future studies can also account for some of the limitations present in this study. First, the target event represented here is of low stakes. Increase in stakes makes a small talk baseline even less effective, as emotions experienced by the interviewee during the target response such as fear (of being caught, for the liar and of not being believed, for the truth teller), may result in more pronounced differences between the two phases of the interview, regardless of sender's veracity (Ewens et al., 2014; Palena et al., 2018). Second, our participants did not receive any form of training in truth/lie detection. Therefore, their accuracy may benefit from training, as they would focus on more effective cues to truth/deception. For example, a recent meta-analysis (Hauch, Sporer, Michael, & Meissner, 2016) explored to what degree different types of training (e.g.: in verbal content, nonverbal behaviour, paralinguistic cues, etc.) affected credibility assessments. Their results showed that training did improve credibility assessments, particularly training focusing on speech content. Third, our sample consisted of university students. It may be worthwhile to explore how professionals (e.g. police officers) perform when they are exposed to a comparable truth baseline.

References

Bond Jr, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and social psychology Review*, *10*(3), 214-234. DOI: 10.1207/s15327957pspr1003_2.

Brandt, D. R., Miller, G. R., & Hocking, J. E. (1980). The truth-deception attribution: Effects of familiarity on the ability of observers to detect deception. *Human Communication Research*, *6*(2), 99-110. DOI: 10.1111/j.1468-2958.1980.tb00130.x.

Brandt, D. R., Miller, G. R., & Hocking, J. E. (1982). Familiarity and lie detection: A replication and extension. *Western Journal of Communication (includes Communication Reports)*, *46*(3), 276-290. DOI: 10.1080/10570318209374086.

Caso L., Maricchiolo F., Bonaiuto M., Vrij A., Mann S. (2006). The impact of deception and suspicion on different hand movements. *Journal of Nonverbal behavior*, *30*(1), 1-19. DOI: 10.1007/s10919-005-0001-z.

Colwell, K., Hiscock-Anisman, C., & Fede, J. (2013). Assessment criteria indicative of deception: An example of the new paradigm of differential recall enhancement. In B. S. Cooper, D. Griesel, & M. Ternes (Eds.), *Applied issues in investigative interviewing, eyewitness memory, and credibility assessment* (pp. 259-291). New York: Springer.

Colwell, K., Hiscock-Anisman, C. K., Memon, A., Taylor, L., & Prewett, J. (2007). Assessment Criteria Indicative of Deception (ACID): An integrated system of investigative interviewing and detecting deception. *Journal of Investigative Psychology and Offender Profiling, 4*, 167–180. DOI: 10.1002/jip.73.

DePaulo, B. M., Charlton, K., Cooper, H., Lindsay, J. J., & Muhlenbruck, L. (1997). The accuracy-confidence correlation in the detection of deception. *Personality and Social Psychology Review*, *1*(4), 346-357. DOI: 10.1207/s15327957pspr0104_5.

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., Jang, M., & Jo, E. (2014). Drop the small talk when establishing baseline behaviour in interviews. *Journal of Investigative Psychology and Offender Profiling*, *11*(3), 244-252. DOI: 10.1002/jip.1414.

Ewens, S., Vrij, A., Mann, S., & Leal, S. (2016). Using the reverse order technique with non-native speakers or through an interpreter. *Applied Cognitive Psychology, 30,* 242–249. DOI:10.1002/acp.3196.

Feeley, T. H., deTurck, M. A., & Young, M. J. (1995). Baseline familiarity in lie detection. *Communication Research Reports*, *12*(2), 160-169. DOI: 10.1080/08824099509362052.

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.

Gnisci A., Caso L., Vrij A. (2010). Have you made up your story? The effect of suspicion and liars' strategies on reality monitoring. *Applied cognitive psychology*, *24*(6), 762-773. DOI: 10.1002/acp.1584.

Granhag, P. A., & Hartwig, M. (2015). The Strategic Use of Evidence technique: A conceptual overview. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Detecting Deception: Current Challenges and Cognitive Approaches* (pp. 231-251). Chichester: John Wiley & Sons.

Granhag, P. A., & Vrij, A. (2010). Interviewing to Detect Deception. In P. A. Granhag (Eds.), *Forensic psychology in context: Nordic and international approaches* (pp. 75-93). Collumpton: Willian.

Hartwig, M., & Bond Jr, C. F. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bullettin, 137*(4), 643-459. DOI: 10.1037/a0023589.

Hartwig, M., & Bond, C. F. (2014). Lie detection from multiple cues: A meta-analysis. *Applied Cognitive Psychology*, *28*(5), 661-676. DOI: 10.1002/acp.3052.

Hauch, V., Sporer, S. L., Michael, S. W., & Meissner, C. A. (2016). Does training improve the detection of deception? A meta-analysis. *Communication Research*, *43*(3), 283-343. DOI: 10.1177/0093650214534974.

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

Jupe, L., Akehurst, L., Vernham, Z., & Allen, J. (2016). Teenage offenders' ability to detect deception in their peers. *Applied Cognitive Psychology*, *30*(3), 401-408. DOI: 10.1002/acp.3214.

Meissner, C. A., & Kassin, S. M. (2002). "He's guilty!": Investigator bias in judgments of truth and deception. *Law and human behavior*, *26*(5), 469.

Moston, S., & Engelberg, T. (1993). Police questioning techniques in tape recorded interviews with criminal suspects. *Policing and Society: An International Journal of Research and Policy, 3*, 223–237. DOI: 10.1080/10439463.1993.9964670.

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.

Palena, N., Caso, L., Vrij, A., & Orthey, R. (2018). Deception detection through small talk and comparable truth baselines. *Journal of Investigative Psychology and Offender Profiling, 1-9*. DOI: 10.1002/jip.1495

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior research methods, instruments, & computers*, *31*(1), 137-149. DOI: 10.3758/BF03207704.

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

Vrij, A. (2014). Interviewing to detect deception. *European Psychologist*, *19*(3), 184-194. DOI: 10.1027/1016-9040/a000201.

Vrij, A. (2015). A cognitive approach to lie detection. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Detecting Deception: Current Challenges and Cognitive Approaches* (pp. 205-230). Chichester: John Wiley & Sons.

Vrij, A. (2016). Baselining as a Lie Detection Method. *Applied Cognitive Psychology*, *30*(6), 1112-1119. DOI: 10.1002/acp.3288.

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

Vrij, A., & Granhag, P. A. (2012). Eliciting cues to deception and truth: What matters are the questions asked. *Journal of Applied Research in Memory and Cognition*, *1*(2), 110-117. DOI: 10.1016/j.jarmac.2012.02.004.

Vrij, A., Leal, S., Mann, S., & Fisher, R. (2012). Imposing cognitive load to elicit cues to deceit: Inducing the reverse order technique naturally. *Psychology, Crime & Law, 18*, 579–594. DOI:10.1080/1068316X2010.515987

Vrij, A., & Mann, S. (2001). Telling and detecting lies in a high-stake situation: the case of a convicted murderer. *Applied Cognitive Psychology, 15*, 187–203. doi: 10.1002/1099-0720(200103/04)15:2<187::AID-ACP696>3.0.CO;2-A.

Vrij, A., Nahari, G., Isitt, R., & Leal, S. (2016). Using the verifiability lie detection approach in an insurance claim setting. *Journal of Investigative Psychology and Offender Profiling*, *13*(3), 183-197. DOI:10.1002/jip.1458.

Table 1. *Means, Standard Deviation and Confidence Intervals for Truth and Lie Accuracy according to the Baseline Condition*

	Comparable Truth			Small Talk		
	M	SD	CI	M	SD	CI
Truth accuracy	57.84	30.10	47.80, 67.88	49.44	29.66	39.41, 59.48
Lie accuracy	55.14	20.22	48.49, 61.88	45.42	21.49	38.14, 52.69

Table 2. *One Sample t-tests Comparing Observers' Accuracy Rates with a Meta Analysis Average Scores[1] and with Chance in each Baseline Condition.*

Comparable Truth Baseline

	Test value: meta-analyses average scores			Test value: chance (50%)		
Accuracy	t	p	d	t	p	d
Total	.80	.42	.13	2.105	.04	.34
Truth	.64	.52	-.10	1.584	.12	.25
Lie	2.447	.02	.39	1.544	.13	.25

Small Talk Baseline

	Test value: meta-analyses average scores			Test value: chance (50%)		
Accuracy	t	p	d	t	p	d
Total	-2.124	.04	-.34	-.83	.40	-.13
Truth	-2.338	.02	-.38	-.11	.91	-.01
Lie	-.44	.66	-.07	-1.279	.20	-.21

[1]Bond & DePaulo (2006). Total accuracy, 54%. Truth accuracy, 61%. Lie accuracy, 47%.

Table 3. *One Sample t-tests Exploring Observers' Sensitivity and Response Bias in each Baseline Condition*

	Comparable Truth			Small Talk		
	t	p	d	t	p	d
d'	2.129	.04	.34	-.817	.42	-.13
β	3.737	.001	.60	2.629	.01	.43