Police accuracy in truth/lie detection when judging baseline interviews

Letizia Caso, PhD[1]

Nicola Palena, MSc[2]

Elga Carlessi, MSc[3]

Aldert Vrij, PhD[4]

[1]Department of Human and Social Sciences, University of Bergamo. Associate Professor in Psychology and Law

[2]Department of Human and Social Sciences, University of Bergamo. PhD Student

[3]Department of Human and Social Sciences, University of Bergamo, MSc Student / Bergamo Police

[4]Department of Psychology, University of Portsmouth. Full Professor in Applied Social Psychology

Ethical standards

Declaration of conflicts of interest

Author A [Letizia Caso] has declared no conflicts of interest

[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

Abstract

Research has shown that a Comparable Truth Baseline approach elicits more cues to deception and results in higher accuracy rates than a Small Talk Baseline. Past research focused on accuracy rates obtained by laypeople. In the current experiment we examined whether the Comparable Truth Baseline also has a positive effect on law enforcement personnel accuracy. In this study, ninety-five police officers judged ten interviews, whereby half of the senders told the truth and the other half lied about a mock undercover mission. Half of the interviews included only questioning about the event under investigation, whereas the other half also included questioning aimed at creating a Comparable Truth Baseline. Observers did not differ in their total and truth accuracy, but those who watched interviews with a Comparable Truth Baseline obtained higher lie detection accuracy rates than those who watched interviews without the Baseline questioning. Signal detection analyses showed that this effect could be attributed to a decreased response bias in the Comparable Truth Baseline condition.

Police accuracy in truth/lie detection when judging baseline interviews

Introduction

Cues to Deception and Interviewing Techniques

Detecting deception is a flourishing research area, from both a theoretical (Bond, Levine, & Hartwig, 2015; Caso, Maricchiolo, Livi, Vrij, & Palena, 2018; Vrij, Hartwig, & Granhag, 2018; Walczyk, Harris, Duck, & Mulay, 2014) and applied perspective (Vrij & Fisher, 2016). Research initially focused on searching for reliable cues to deceit that liars display spontaneously (see Vrij [2008] for a comprehensive review of this research). This research has proven to be largely unsuccessful: Cues to deception in this setting are typically faint and unreliable (DePaulo et al., 2003; Vrij, 2008).

Due the paucity of support for the search for reliable cues to deception that liars display spontaneously, scholars have started to examine whether such cues can be elicited or enhanced through specific interviewing protocols (Vrij & Granhag, 2012; Vrij, 2014). Several interview protocols have emerged. In the Strategic Use of Evidence technique the available evidence is presented to interviewees in such a way that it elicits within-statement and between-statements inconsistencies in liars and elicits admissions from them (Granhag & Hartwig, 2008, 2015; May, Granhag, & Tekin, 2017). The Cognitive Credibility Assessment approach builds on the research findings that lying is usually more mentally taxing than telling the truth (Vrij, Fisher, & Blank 2017). This approach is based on the idea that questions can be asked, or instructions can be given, that elicit different mental processes in truth tellers and liars (Vrij, 2015; Vrij, Fisher, Blank, Leal, & Mann, 2016). Another approach, the Assessment Criteria Indicative of Deception technique, is a (semi)structured interview which builds on the memory-enhancement techniques (mnemonics) which are part of the Cognitive Interview (Fisher & Geiselman, 1992). Such memory-enhancing techniques have a different impact on truth tellers and liars when considering cues such as response length, amount of details and coherence (Colwell et al., 2009; Colwell, Hiscock-Anisman, Memon,

Taylor, & Prewett, 2007). This so-called interviewing to detect deception approach has proven to be successful and several of the techniques proposed and tested in this area are ready to be implemented in real life (Vrij & Fisher, 2016; Vrij, 2018).

Observers' Lie Detection Accuracy

Another line of research examines observers' accuracy at evaluating interviewees credibility. The most comprehensive meta-analysis published to date showed that people – laypersons and professionals alike- are poor at this task. The average accuracy, obtained by over 24,000 observers, was 54%, which is only just above the level of chance (50%) (Bond & DePaulo, 2006). However, the studies presented in the meta-analysis did not account for the effect of strategic interviewing, which is worth exploring. Luke et al. (2016) found that training law enforcements in the SUE technique increased their accuracy rates up to 65%. Similarly, another study found that police officers trained in the Cognitive Credibility Assessment approach asked more effective questions and improved their accuracy rates up to 74% (Vrij, Leal, Mann, Vernham, & Brankaert, 2015). Finally, Colwell et al. (2009) found that observers trained in the Assessment Criteria Indicative of Deception technique obtained accuracy rates up to 77%. These pictures show strong potential for training practitioners into strategic interviewing.

The Baseline Approach

Another interview technique is based on the rationale that if the observer has previous knowledge of sender's truthful behavior, this can be used as a truthful baseline to inform observer's decisions (see Vrij, 2008). Indeed, research has supported this claim, as it was found that being familiar with the sender increases observer's performance (Brandt, Miller, & Hocking, 1980, 1982). Feeley, de Turk, and Young (1995) found a positive linear relationship between the level of familiarity and observers' performance; that is, the more the observer was familiar with senders' truthful behavior, the more s/he was accurate in detecting deception. However, research also shows that familiarity only works when the baseline is truthful. For example, Garrido and Masip (2001)

provided observers with a baseline behavior of the sender, which could be either truthful or deceitful. The authors found that observers benefited from the baseline exposure only when this was truthful.

The type of baselining reported above is different from the one that one can expect in investigative interviewing. The suspect and the interviewer are often strangers; therefore, such *familiarity baseline* is difficult to obtain (if possible at all) during investigations. Instead, investigators must obtain the baseline during the first phases of the interview. There have been some suggestions on how to create a baseline *on the spot*.

Initially, it was suggested to create such a baseline by asking neutral, non-threatening questions (Frank, Yarbrough, & Ekman, 2006) but both theoretical reasons and experimental results have revealed that this method is ineffective (Ewens, Vrij, Yang, & Yo, 2014; Moston & Engelberg, 1993; Palena, Caso, Vrij, & Orthey, 2018). This approach is thought to be a striking misuse of psychological theory (Moston & Engelberg, 1993; Vrij, Leal, Jupe, & Harvey, 2018). The problem is that the baseline and target (which concerns the event under investigation) sections are fundamentally different in such an approach. For example, the stakes are higher in the target than in the baseline section and interviewee's engagement is often different between the two sections. As a result, not only liars but also truth tellers change their behaviour when the baseline and target responses are compared (Ewens et al., 2014; Palena et al., 2018; Vrij, 2016). Therefore, both theoretical explanations (Moston & Engelberg, 1993; Vrij, 2016; Vrij, Leal, Jupe, & Harvey, 2018) and experimental findings (Caso, Palena, Vrij, & Gnisci, 2019; Ewens, Vrij, Yang, & Yo, 2014) support the assumption that a baseline created through small talk (i.e., Small Talk Baseline) is ineffective.

Vrij (2016) states that a different baselining approach, the creation of a comparable truth baseline (CTB), may be more fruitful. In the CTB approach factors such as engagement, emotion, cognitive load and context are kept comparable between the baseline and the target sections of the

interview. To keep such factors comparable, baseline questioning should pertain to the same topic of that discussed when posing question concerning the event under investigation. Similarly, questions should not be posed in a way that creates different emotional arousal. Two recent studies found that using a CTB, liars changed their behaviour and speech more than truth tellers (Palena et al., 2018) and that observers provided with a CTB obtained high accuracy rates than those provided with a small talk baseline (Caso et al., 2019).

The main aim of the present study was to explore whether a CTB would also be effective with law enforcement personnel as observers. Involving law enforcement personnel as participants in studies is important as it may make it more likely that they will endorse the findings and, if the findings are positive, start to use the techniques themselves (Vrij & Granhag, 2012).

We decided to compare the CTB approach to a "no baseline" rather than to a Small Talk Baseline for two main reasons. First, we wanted to test the hypothesis that having a comparable baseline behaviour of the interviewee makes the observers more accurate than not having such reference. This builds on the idea that accuracy is increased because a comparable baseline reduces the effect of interpersonal differences (Vrij, 2016). Second, we did not want to compare the CTB with the small talk baseline because: i) research has already shown that such an approach does not work because truth tellers and liars appear equally deceptive (Ewens et al., 2014; Palena et al., 2018); and ii) research with laypersons has shown that observers provided with a CTB outperform those provided with a Small Talk baseline (Caso et al., 2019) This is unlikely to change with practitioners, as the inefficacy of the Small Talk baseline approach depends on its underpinnings rather than on the observer: Both truth tellers and liars display differences between baseline and target periods with this approach (Ewens et al., 2014). Therefore, no reliable cue to deception appears and neither laypersons nor practitioners can benefit from the small talk baseline approach.

Building on previous results on objective cues to deception elicited with a CTB (Palena et al., 2018) and on the rationale presented above, we expected that practitioners in the CTB condition

would achieve higher total (Hypothesis 1), truth (Hypothesis 2) and lie (Hypothesis 3) accuracy rates than practitioners in the no-baseline condition.

Materials and Method

Participants

A total of 95 practitioners (88 men and seven women) took part in the experiment. Of them, 42 belonged to the state police, 28 to the financial and economic crimes police, and 25 to the Italian Military police (Carabineri). Age ranged from 28 to 58, with a mean of $M = 45.39$ ($SD = 6.71$). Professional experience ranged from five to 38 years, with a mean of $M = 23.84$ ($SD = 7.82$). One participant was excluded from the analyses because he did not follow the instructions.

Design

The experiment utilised a 2 (Baseline: no baseline vs. CTB, between-subjects) by 2 (Veracity: Truth tellers vs. Liars, within-subjects) mixed design. For the factor Baseline, observers in the no-baseline condition just saw suspects being questioned about the event under investigation. On the other hand, observers in the CTB condition saw interviews where suspects were also questioned about an event other than -but comparable to- that under investigation. For the factor Veracity, half of the senders told the truth, whereas the other half lied. For the CTB condition, all senders truthfully reported the additional event, which served as the CTB. The dependent variables were the three accuracy rates obtained by the observers: total accuracy, truth accuracy and lie accuracy. Accuracy rates were obtained using the formula $\frac{number\ of\ correct\ answers}{total\ number\ of\ possible\ answers} \cdot 100$.

Procedure

The first and the second author made an appointment with the prosecutor (who is the person in charge for recruiting participants for any research project with practitioners) and presented him an outline of the study. The importance of conducting research with practitioners was stressed. The prosecutor was interested in the project and proceeded to contact high rank officials belonging to

the three organisations (State Police, Economic and Financial Police, and Military Police), who were informed about the research goals. High rank officers then provided the authors with a list of participants from the three police organisations. The participants were then contacted to take part in the study. Data were collected in three different places, one for each police organisation. Each participant took part in the experiment individually. Upon arrival, s/he was welcomed and briefed about the aim of the study. They were informed that they were going to watch a series of interviews with different people and then decide for each interviewee whether s/he was lying or telling the truth. They were not informed about the number of interviews they would see and about the truth/lie telling ratio in the interviews to avoid these aspects affecting their decisions. In total, each judge evaluated every sender (ten in total).

Participants in the no baseline condition were not given further information. Participants in the CTB condition were informed that interviews were split in two sections. The first section was the baseline, the second was the target section of the interview. They were also informed that senders were always honest in the baseline section and therefore they had to take a decision about sender's (dis)honesty only regarding the target section. They were invited to examine deviations from the baseline to inform their decisions on the rationale that the more a sender changed his/her behaviour and speech between the two phases, the more likely it was that s/he was lying. Information about which behaviours or speech patterns might be indicative of deceit was not provided. The participants were not informed about the aims of the study (comparing accuracy of practitioners in the two experimental conditions).

In the experiment, veracity decisions were made answering the dichotomous question "Do you think the interviewee was…"Lying" or "Telling the truth". After the experimenter felt confident that the participant understood the instruction, s/he was left alone to carry out the lie detection test, which lasted on average about 30 minutes. Once the participant had finished the test, they were thanked and debriefed.

Stimulus material

Senders depicted in the videoclips had performed a mock undercover mission. The videoclips were obtained from a previous study (reference not reported for blind review). The mock crime started with participants receiving an envelope from the experimenter that contained the password for accessing a laptop. Once logged in, the sender read a Word document file that informed him/her to look for a CD-ROM in a backpack that was placed in the room. The CD-ROM depicted a man informing the sender to look for a key in the same backpack, which would open a safe deposit box placed near the window in the same room. The safe deposit box contained further written instructions, asking the participant to send an email to a specific address and to wait for a person to collect them from the room. Everything until that moment was part of the CTB, whereas everything happened that after this point was part of the target event, the event under investigation.

After a short while the person (a confederate) arrived and gave a newspaper to the sender, informing him/her to read it for further instructions. These instructions informed the sender that s/he had to go to an adjacent room and to look for a USB stick hidden behind a coat hook. The sender had to take that USB stick and put it in place of a second USB stick that was hidden in a book placed in a wardrobe. The sender then had to leave the newspaper next to the book and keep the second USB (the one found in the book) until the end of the experiment. Once these tasks were completed, the sender had to come back to the first room and wait for the interviewer.

We used ten senders in total. For the CTB, the observer watched both the baseline and target sections, which was composed by the sender answering a free recall baseline question and a free recall target question. All ten senders answered the baseline question truthfully, whereas for the target question, five of them told the truth and five lied. The veracity status was counterbalanced. For the condition without the baseline, the observers only saw the target question and answers.

The ten clips in the "no baseline" condition lasted 76.70 seconds on average (SD = 17.36), those in the CTB lasted 170.10 seconds on average (SD = 20.51). This difference is due to the presence of the baseline.

Results

Total Accuracy

To test Hypothesis 1, an ANOVA was carried out with Baseline (no baseline vs. CTB) as factor, and the total accuracy rate as dependent variable. No significant effect appeared for the Baseline factor, $F(1, 92)$ = 1.554, p = .21, d = -0.26 [-0.66, 0.15], post-hoc achieved power .24. Mean accuracy for participants in the no baseline condition (M = 49.58; DS = 13.67; 95% CI [45.61, 53.55]) was similar to that of participants in the CTB (M = 53.26; DS = 14.91; 95% CI [48.83, 57.69]). Hypothesis 1 was thus rejected.

Truth Accuracy

Preliminary tests assessing ANOVA assumptions showed that homoscedasticity was not respected, $F(1, 92)$ = 7.110, p = .01. Consequently, we run a Mann-Whitney U test to test Hypothesis 2. The truth accuracy for participants in the no-baseline condition (M = 60.00; SD = 17.01; 95% CI [55.06; 64.93]) did not differ from that of participants in the CTB condition (M = 51.73; SD = 22.93; 95% CI [44.93, 58.55]), U = 863.50, p = .058, post-hoc achieved power .48. Therefore, Hypothesis 2 was also rejected.

Lie Accuracy

To test Hypothesis 3, an ANOVA was carried out with Baseline as factor and lie accuracy as the dependent variable. The Baseline effect was significant, $F(1, 92)$ = 17.16, p < .001, d = -0.85 [-1.27, -.43], post-hoc achieved power .98. Supporting Hypothesis 3, observers in the CTB condition (M = 54.78; SD = 20.41, 95% CI [48.72, 60.84]) outperformed those in the no baseline condition (M = 39.16; SD = 15.95, 95% CI [34.53, 43.80]). In addition, more observers (63%) in the CTB

condition than in the control condition (25%) obtained an accuracy rate of at least 60%, χ^2 (1, $N =$ 94) = 13.82, $p < .001$, Cramer's V = .38. Figure 1 depicts a density plot of such values.

enter Figure 1 about here

Signal detection analyses.

Our results supported Hypothesis 3. However, lie detection accuracy for participants in the CTB condition did not differ from chance, $t(45) = 1.589$, $p = .12$, $BF_{10} = 0.513$, $d = 0.23$ [-0.06, 0.52], post-hoc achieved power .33. This makes an alternative explanation possible. Since with a baseline approach (including the CTB) both truth tellers and liars appear to behave and speak differently between the two phases of the interview (Palena et al., 2018; Vrij, 2016), the increased lie accuracy may be partly due to a lie bias. We tested this possibility by exploring participants' response bias. Although historically the β value was the preferred measure for bias, some scholars have suggested to use c instead of β (see Stanislaw & Todorov, 1999) as the former is less biased than the latter. c is described as deviations, in standard deviation units, from the neutral point (where neither answer is preferred, i.e., there is no bias), which is set at 0. If the c value is greater than 0, there is a bias toward responding "no" (in our context, "truth teller"); if the value is less than 0 there is a bias toward responding "yes" (in our context, "liar"). A one-sample t-test with c as the dependent variable and 0 as test score, was significant for participants in the no baseline condition, $t(47) = 7.803$, $p < .001$, $d = 1.13$ [0.76, 1.49], post-hoc achieved power 1, but not significant for participants in the CTB condition, $t(45) = -.555$, $p = .58$, $d = -0.08$ [-0.37, 0.21], $BF_{01} = 5.404$, post-hoc achieved power .08. An independent sample t-test with Baseline (no baseline vs. CTB) as the factor and c as the dependent variable showed that participants in the "no baseline" condition obtained higher scores ($M = 0.29$; $SD = 0.26$, 95% CI [0.22, 0.37]) than those in the CTB condition ($M = -0.05$; $SD = 0.61$, 95% CI [-0.23, 0.13]), $t(60.62) = 3.547$, $p = .001$, $d = 0.73$ [0.31, 1.14], post-hoc achieved power 0.94. Results for the c score therefore suggest that participants in the "no

baseline" condition were truth biased, whereas those in the CTB displayed no bias. Table 1 displays the rates of truth and lie judgments in each condition.

enter Table 1 about here

Given the response bias results, we also analysed *d'* scores, which is a measure of sensitivity reported in standard deviation units (Stanislaw & Todorov, 1999). A *d'* score of 0 indicates an inability to distinguish between the two stimuli (truth telling vs. lying), whereas scores greater than 0 indicate that participants were able to make such a distinction. Neither participants in the "no baseline" condition (*M* = -0.002; *SD* = 0.82, 95% CI [-0.24, 0.23]), $t(47) = -0.021$, $p = .98$, $BF_{01} = 6.376$, post-hoc achieved power .05, nor those in the CTB condition (*M* = 0.24; *SD* = 0.99, 95% CI [-0.05, 0.53]), $t(45) = 1.620$, $p = .11$, $BF_{01} = 1.864$, post-hoc achieved power .36, were able to discriminate truth tellers from liars. Considering the analyses on response bias and sensitivity, we can conclude that the difference in accuracy between the two conditions was driven by the difference in response bias.

Discussion

In this experiment, we tested Italian police officers' ability to evaluate credibility when provided with a CTB compared to when no baseline was presented. We found support for Hypothesis 3, as officers in the CTB condition (54.78%) outperformed those in the no-baseline condition (39.16%) in terms of lie accuracy. However, further analyses showed that this was the result of officers in the CTB condition being less biased rather than being more accurate.

Our results are in alignment with the Truth Default Theory (Levine, 2014), which predicts that observers usually tend to believe others, unless "deceptive triggers" appear and make the observer considering the possibility of deception. In our context, such "deceptive triggers" may originate from behavioural differences that appear- for both truth tellers and liars- between the baseline and the target phase of the interview. However, the idea of humans having a cognitive

default (truth) bias has been questioned (see Street, 2015). Future studies should try to disentangle this issue.

The baseline approach comes with several issues. First, truth tellers may be classified as liars because they also often change their behaviour throughout the interview. Second, the problems with the CTB are also shown in Palena et al. (2018), who found that it only worked for one cue (spatial details) when comparisons between baseline and target periods were made. The authors noted that their result was likely due to the task performed by the interviewees, which was mainly spatial. Indeed, participants in Palena's et al. (2018) study had to commit a mock crime which requested them to explore different rooms and interact with several objects, making the statements rich in spatial details. This makes clear that the cues to be used with the CTB approach are tightly connected to the content of the story itself.

Third, obtaining a CTB in laboratory settings is easy as the experimenters exert full control over baseline veracity. However, in real life it may be difficult to obtain a CTB which is really truthful and comparable, as a ground truth is often missing. And, in case the baseline is a lie, it loses its efficacy (Garrido & Masip, 2001).

In sum, the positive result for CTB concerning lie accuracy was the result of reduced probability of guessing truth, and no differences were found for total and truth accuracy. This shows no real positive effect for using a CTB. Previous research on the effectiveness of various interview techniques reached better results than those we found in the current experiment. This may be partly due to the fact that such techniques are more active approaches -the interviewer conducts the interview actively- whereas the baseline approach is more passive. The only thing the interviewer has to do is to create the baseline. It has already been suggested that active strategic interviewing is more effective for deception detection (Vrij & Granhag, 2012; Vrij, 2014).

Limitations

There were some limitations in the present study. First, our stimulus material was relatively low-stakes. This may have affected the results, although it is not straightforward to predict how. However, baseline research relies in part on behavioural patterns and such patterns are affected by stakes (Hartwig & Bond, 2014). It is therefore important to start examining the baseline technique also in higher-stakes situations. Second, observers only watched ten senders, which may not represent variations in senders' performance adequately. Third, it appeared that police officers participating in this study do not use Small Talk baselining or any other type of baselining in their daily practice. Given the problems associated with baselining, this cannot be considered a bad thing. Rather, they reported that when they interview real suspects, they tend to base their decision on available evidence Yet, we did not provide our participants with evidence.

The authors have declared no conflict of interests

The authors declare that all measures and conditions have been reported, that one participant was excluded from the analyses because he did not follow the instructions, and that the sample size was determined in order to have at least 30 participants per cell, accepting any further available participant.

References

Bond, C. F. Jr., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review, 10*(3), 214–234. doi:10.1207/s15327957pspr1003_2.

Bond, C. F., Jr., Levine, T. R., & Hartwig, M. (2015). New findings in non-verbal lie detection. In P. A. Granhag, A. Vrij, & B. Vershuere (Eds.), *Detecting deception: Current challenges and cognitive approaches* (pp. 37-58). Chichester, UK: John Wiley.

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., Livi, S., Vrij, A., & Palena, N. (2018). Factors affecting observers' accuracy when assessing credibility: The effect of the interaction between media, senders' competence and veracity. *The Spanish Journal of Psychology*.

Caso, L., Palena, N., Vrij, A., & Gnisci, A. (2019). Observers' performance at evaluating truthfulness when provided with Comparable Truth or Small Talk Baselines. *Psychiatry, Psychology, and Law.*

Colwell, K., Hiscock-Anisman, C. K., Memon, A., Colwell, L. H., Taylor, L., & Woods, D. (2009). Training in assessment criteria indicative of deception to improve credibility judgments. *Journal of Forensic Psychology Practice, 9*(3), 199–207. doi:10.1080/15228930902810078.

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*(3), 167–180. doi:10.1002/jip.73.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*(1), 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.

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.

Fisher, R. P., & Geiselman, R. E. (1992). *Memory enhancing techniques for investigative interviewing: The cognitive interview*. Charles C. Thomas Publisher.

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.

Garrido, E., & Masip, J. (2001). Previous exposure to the sender's behavior and accuracy at judging credibility. In R. Roesch, R. R. Corrado, & R. J. Dempster (Eds.), *Psychology in the courts: International advances in knowledge*, (pp. 271-287). London: Routledge.

Granhag, P. A., & Hartwig, M. (2008). A new theoretical perspective on deception detection: On the psychology of instrumental mind reading. *Psychology, Crime & Law, 14*(3), 189–200. doi:10.1080/10683160701645181.

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. 231–252). Chichester: John Wiley and Sons.

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

Levine, T. R. (2014). Truth-Default Theory (TDT): A Theory of Human Deception and Deception Detection. *Journal of Language and Social Psychology, 33*(4), 378–392. doi:10.1177/0261927X14535916.

Luke, T. J., Hartwig, M., Joseph, E., Brimbal, L., Chan, G., Dawson, E., ... & Granhag, P. A. (2016). Training in the Strategic Use of Evidence technique: Improving deception detection accuracy of American law enforcement officers. *Journal of Police and Criminal Psychology*, *31*(4), 270-278.

May, L., Granhag, P. A., & Tekin, S. (2017). Interviewing Suspects in Denial: On How Different Evidence Disclosure Modes Affect the Elicitation of New Critical Information. *Frontiers in psychology*, *8*, 1154. doi: 10.3389/fpsyg.2017.01154.

Meissner, C. A., & Kassin, S. M. (2002). "He's guilty!": Investigator Bias in Judgments of Truth and Deception. *Law and Human Behavior, 26*(5), 469-480. doi:10.1023/A:1020278620751.

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.

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*, *15*(2), 124-132. 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.

Street, C. N. (2015). ALIED: Humans as adaptive lie detectors. *Journal of Applied Research in Memory and Cognition*, *4*(4), 335-343. doi: 10.1016/j.jarmac.2015.06.002.

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

Vrij, A. (2014). Interviewing to detect deception. *European Psychologist, 19*, 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 and Sons.

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

Vrij, A. (2018). Verbal Lie Detection Tools From an Applied Perspective, In J. P. Rosenfeld (Eds.), *Detecting Concealed Information and Deception* (pp. 297-327). London, UK: Elsevier Academic Press.

Vrij, A., Fisher, R. P., Blank, H., Leal, S., & Mann, S. (2016). A cognitive approach to elicit verbal and nonverbal cues to deceit. In J.W. Van Prooijen, & P. A. M. Van Lange (Eds.), *Cheating, corruption, and concealment: The roots of dishonesty* (pp. 284-302). Cambridge: Cambridge University press.

Vrij, A., & Fisher, R. P. (2016). Which lie detection tools are ready for use in the criminal justice system?. *Journal of Applied Research in Memory and Cognition*, *5*(3), 302-307. doi: 10.1016/j.jarmac.2016.06.014.

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: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., Hartwig, M., & Granhag, P. A. (2018). Reading lies: Nonverbal communication and deception. *Annual Review of Psychology*.

Vrij, A., Leal, S., Jupe, L., & Harvey, A. (2018). Within-subjects verbal lie detection measures: A comparison between total detail and proportion of complications. *Legal and Criminological Psychology, 23*(2), 265-279. doi: 10.1111/lcrp.12126.

Vrij, A., Leal, S., Mann, S., Vernham, Z., & Brankaert, F. (2015). Translating theory into practice: Evaluating a cognitive lie detection training workshop. *Journal of Applied Research in Memory and Cognition*, *4*(2), 110-120. doi:10.1016/j.jarmac.2015.02.002.

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., . . . Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review, 25*(1), 58-76. doi:10.3758/s13423-017-1323-7.

Walczyk, J. J., Harris, L. L., Duck, T. K., & Mulay, D. (2014). A social-cognitive framework for understanding serious lies: Activation-decision-construction-action theory. *New Ideas in Psychology*, *34*, 22-36. doi: 10.1016/j.newideapsych.2014.03.001.

Figure 1. *Density plot for Lie Accuracy according to the Baseline condition.*
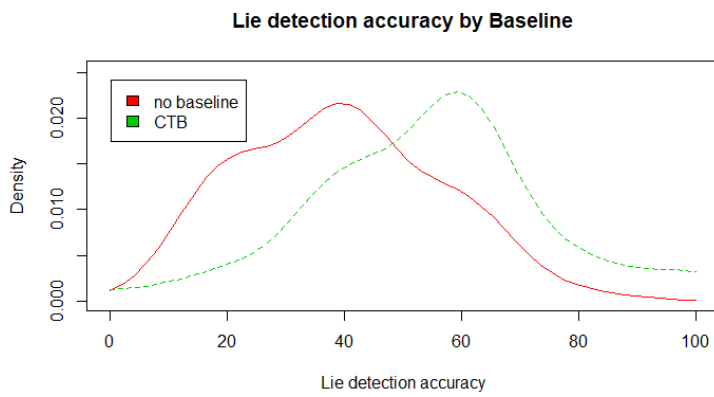
Table 1. *Means and Standard Deviations for rates of truth and lie judgments in each condition.*

	Baseline	Decides Truth	Decides Lie
Mean	No Baseline	60.42	39.58
	Comparable truth baseline	48.48	51.52
Standard deviation	No Baseline	9.22	9.22
	Comparable truth baseline	15.77	15.77