Drop the Small Talk when Establishing Baseline Behaviour in Interviews

Sarah Ewens
Aldert Vrij
University of Portsmouth

Minhwan Jang
Eunkyung Jo
Hallym University

Author Note
Sarah Ewens, Aldert Vrij, Department of Psychology, University of Portsmouth; Minhwan Jang, Eunkyung Jo, Department of Psychology, Hallym University.

Correspondence concerning this article should be addressed to Sarah Ewens, Department of Psychology, University of Portsmouth, King Henry Building, King Henry 1 Street, PO1 2DY, Hants, United Kingdom. Email sarah.ewens@port.ac.uk
Abstract

The present experiment investigated the behavioural patterns of interviewees when comparing their baseline behaviour, prior to the interview, to their behaviour during the investigative interview. Similar to what has been advised in the police literature, the truthful baseline behaviour was established prior to the interview through non-threatening questions. The investigative part of the interview then followed in which the interviewee was aware they would be assessed on whether they were lying. During the investigative part, interviewees either discussed the job they had (truth tellers, $n = 128$) or pretended to have (liars, $n = 115$). Findings revealed that both liars and truth tellers’ behavioural patterns differed between the baseline behaviour and the investigative part of the interview. The findings suggest small-talk should not be used as a baseline comparison with the investigative part of the interview when determining if the interviewee is being deceitful. An alternative way of using a baseline lie detection method, the comparable truth method, is discussed.

*Keywords:* baseline behaviour, comparable truths, deception
Drop the Small Talk when Establishing Baseline Behaviour in Interviews

Police officers 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). For example, in the Improving Interpersonal Evaluations for Law Enforcement and National Security technique (IIE), developed by Ekman (2001), investigators are instructed to observe the subjects 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 subjects 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 behavioural patterns observed during the investigative interview that differ from the truthful, baseline, response could be interpreted as a sign of deceit. This baseline lie detection technique is commonly employed by those trained in such interrogation techniques (Moston & Engelberg, 1993).

The questioning which establishes these baseline behaviours 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 which 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 to the actual interview (Vrij, 1995) and this ‘apple-orange’ comparison will be prone to incorrect judgements (Moston & Engelberg, 1993).
When making baseline comparisons in the way investigators are advised to do, they fail to take situational factors into account. That is, 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 that affect someone’s behaviour. People react differently in (i) 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 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 Hirsch and Wolf’s (2001) real life high-stakes deception study in which they observed 23 verbal and nonverbal cues displayed by former US President Bill Clinton during his Grand Jury Testimony about his alleged relationship with Monica Lewinsky. Hirsch and Wolf (2001) examined a 23-minute segment of the videotape in which Clinton denied having committed various acts of a sexual nature with Monica Lewinsky and compared this 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 minutes 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, it has received little attention from researchers to date. We filled the gap by carrying out a baseline study. Similar to what has been advised in the police literature, the truthful baseline behaviour was established prior to the interview through non-threatening questions. We then continued with the actual interview in which the interviewee was aware they would be assessed on whether they were lying. We used two target periods in the actual interview. 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. Since people are likely to display different behaviours during small talk and interviews, we predicted that both liars and truth tellers’ behavioural patterns will differ between small-talk and the investigative part of the interview (Hypothesis 1). Since people’s behaviour depends on the topic of investigation and changes over time, liars and truth tellers’ behavioural patterns may also differ when target periods 1 and 2 are compared (Hypothesis 2). These two hypotheses go against the baseline lie detection method. In this method, a Veracity X Period interaction effect is expected with liars, but not truth tellers, displaying different behaviours between baseline and target periods (Baseline Hypothesis, Hypothesis 3). To formulate uni-directed hypotheses (Hypotheses 1 and 2) rather than directed hypotheses is in alignment with the baseline technique where differences in behaviours are expected in liars when comparing small talk with the actual interrogation, without stating what these differences actually will be.
In the present study we rated the interviewees’ behaviour in terms of nervousness, thinking hard, and controlling self. These are three factors commonly associated with deception (Zuckerman, DePaulo, & Rosenthal, 1981). Emotions, including guilt and fear, can influence a liar’s behaviour and lead to an increase in cues such as self adapters (touching own clothes, hair etc), speech hesitations (uhm’s and er’s), speech errors (stutters, repetitions), and fundamental frequency of pitch (e.g., higher pitched voice) (Vrij, 2008). Furthermore, regarding thinking hard, liars have many aspects to think about telling when fabricating their story. Not only do they need to make up their story, they also need to monitor their fabrications to make sure they sound plausible and remember their earlier statement to keep it consistent. Thinking hard leads to an increase in pauses, speech hesitations and speech errors (Goldman-Eiser, 1968), and also leads to fewer hand and arm movements (Ekman, 1997; Ekman & Friesen, 1972). Finally, liars may attempt to control their behaviour because they realise observers pay attention to their reactions (Buller & Burgoon, 1996; Burgoon & Buller 1994; Burgoon, Buller, Floyd, & Grandpre, 1996). However, even if people try to appear credible they may not be successful. Since the common view is that liars will increase their movements, liars will probably attempt to move more deliberately and avoid movements that are not essential. This will result in behaviours appearing unusually rigid, more rehearsed and lacking spontaneity, because people usually make movements with their body, hands, feet, etc, that are not essential (Burgoon & Buller, 1994; DePaulo & Kirkendol, 1989). Thus, we predict that liars will appear to be more nervous, thinking harder and controlling themselves more than truth tellers (Hypothesis 4).

Method

Participants

A total of 243 participants (145 Females and 98 Males) took part in the study. Ages ranged from 16-75 years with an average age of $M = 26.44$ years ($SD = 10.91$).
Design

We employed a 2 (Veracity: Truth vs Lie) x 3 (Period: Baseline vs target 1 vs target 2) mixed subjects design. An ‘occupation scenario’ similar to Mann et al. (2012) and Vrij, Mann, Leal, and Fisher (2012) was used whereby participants either told the truth or lied about their job in an interview. Participants were allocated randomly to the truth telling (n = 128) or lying (n = 115) condition. The dependent variables were ratings of nervousness, thinking hard and controlling self.

Procedure

Participants were recruited via intranet announcements and advertising posters which asked for volunteers to take part in a study called ‘Improving Communications in Interviews’. They were informed the study would last 25-45 minutes, that they would be interviewed about their real or pretend job and that they would receive a £5 cash reward.

The first participants who responded via email acted as participants (consisting of university administrative staff, maintenance staff, students and the general public). Initially they were sent via email a ‘selection briefing form’ which consisted of an informed consent form and a list of 17 different jobs for which they were asked to indicate how much they knew about each of them (on a rating scale ranging from [1] very little to [7] a lot). They were further asked which job, if any, they currently had.

The truth tellers were informed that they would be interviewed about their actual job. The liars were told that their task would be to pretend during the interview to have a job that they did not have. We emailed liars the job we wanted them to pretend to have and this decision was based on their responses given in the selection briefing form. The lying job chosen was always one that participants had indicated they did not know too much about (a score of 2 or 3 on the selection form). This score would guarantee that the participants were in fact lying but it was not a completely impossible lie and thus reflects a real life situation.
An appointment for when the interview would take place was then made with each participant. To make sure that they had sufficient time to prepare themselves the interview was scheduled for at least three days after the veracity allocation. In order to motivate participants to being convincing, we informed them that they would receive a £5 reward if the interviewer believed them to be telling the truth. Participants were further told that if they were not believed they would have to write a statement detailing why they thought they were not believed. For ethical reasons, at the end of the study all participants were told that the interviewer believed them and received their £5 reward.

Truth tellers had 56 different professions, with cleaner ($n = 13$) and waiter/waitress ($n = 12$) being the most frequently mentioned. Other professions occupied by more than one truth telling participant included teacher, bar tender, sales assistants and administrators. The jobs that liars pretended to have during the interview matched those of truth tellers.

On arrival, participants were greeted by the experimenter. Truth tellers were reminded to answer the questions truthfully and liars were reminded which occupation they needed to pretend they had during the interview. All participants said that they had understood the instructions and all liars said they had remembered the jobs they had been allocated previously. A check of the transcripts revealed that all truth tellers discussed their actual job and all liars discussed their allocated pretend job.

Participants were then brought to the interview room and introduced to the interviewer. The same, two female, interviewers were used for all of the 243 interviews and they were both blind to the veracity of the participant.

The interview consisted of ten questions, however only three were utilised in this study. The interview commenced with an introductory baseline question: ‘You just read and signed an informed consent form, could you please tell me what you remember about it and what it said’. At this point all participants were telling the truth. The remaining part of the
interview consisted of three short answer questions (questions 2, 3 and 4) and five detail eliciting questions (questions 5 to 9). For target period 1 we used question 5, the first detail eliciting question (target period 1): ‘Please describe your place of work in as much detail as you can.’, and for target period 2 we used question 9, the last detail eliciting question: ‘If you were training me to do your job for a day, what things would I need to know about it?’ When answering questions 5 and 9 participants were either in the lying or truth telling condition.

The interviews were video and audio recorded. The truthful \(M = 933.46\) seconds, \(SD = 330.22\) and deceptive \(M = 893.44, SD = 373.49\) interviews were of a similar length, \(F(1, 235) = .32, p = .58, ns, d = .11, \text{eta}^2 = .00.\) Following the interview participants were then thanked, debriefed, and received the £5.

Coding

Nervousness. A coder blind to the hypotheses and experimental conditions watched the interviews and rated nervousness on a 7-point Likert scale rating from [1] not nervous to [7] very nervous. Participants who were not nervous showed a relaxed demeanour. Those who were nervous fidgeted more e.g., leg and finger tapping. They also spoke in a faster pace, with a nervous tone in their voice and showed more self manipulations. A second coder, blind to the hypotheses and experimental conditions, rated nervousness in a sample of 61 transcripts (25%). The inter-rater reliability (Intra-class Correlation Coefficient (ICC) for nervousness was .31. Due to low reliability the nervousness variable was removed from the study.

Thinking hard. The coder further rated thinking hard on a 7-point Likert scale rating from [1] ‘not thinking hard’ to [7] ‘thinking very hard’ for each of the three questions. Participants who were not thinking hard answered the questions instantly without hesitation. Those who were thinking hard needed to think about their answers. They also thought more during their
answers about what they wanted to say next and hesitated with uhms and erms. The inter-rater reliability (ICC) for thinking hard was .61.

**Controlling self.** The coder finally rated controlling self on a 7-point Likert scale rating from [1] not controlling self to [7] controlling self a lot. Participants who were not controlling themselves were very expressive with their hands and body movements with a more relaxed and natural posture. Those who controlled themselves a lot sat more rigid with no movement and fixed their gaze. The inter-rater reliability (ICC) for controlling self was .70.

**Results**

A mixed 2 (Veracity) x 3 (Period) MANOVA design was conducted with Veracity (truth versus lie) as a between-subjects factor and Period (baseline, target 1, target 2) as a within-subjects factor. The ratings of thinking hard, and controlling self were the dependent variables. At a multivariate level the Veracity main effect, $F(2, 239) = 4.71, p = .001, \eta^2 = .04$ and the Period main effect, $F(4, 237) = 31.23, p < .001, \eta^2 = .35$ were both significant.

At a univariate level two Veracity main effects were significant. Liars ($M = 3.40, SD = 1.49$) had to think harder than truth tellers ($M = 2.91, SD = 1.36$), $F(1, 240) = 7.06, p = .008, d = .34, \eta^2 = .029$, and liars ($M = 2.30, SD = 1.17$) controlled themselves more than truth tellers ($M = 1.97, SD = 1.06$), $F(1, 240) = 5.26, p = .023, d = .29, \eta^2 = .021$. This supports Hypothesis 4.

At a univariate level all three Period main effects were significant and the results are depicted in Table 1.

Table 1 about here

Table 1 shows that participants made the impression of having to think harder and controlling themselves more in the baseline period than in the two target periods, supporting Hypothesis 1. Participants also gave the impression of controlling themselves more in target period 2 compared to target period 1. This supports Hypothesis 2.
For baseline lie detection the Veracity X Period interaction is relevant, as liars, but not truth tellers, are supposed to display different behaviours during baseline and target periods. The interaction effect was not significant on a multivariate level, $F(4, 237) = 0.23, p = .92, \text{eta}^2 = .004$, neither were any of the univariate effects significant, all $F’s < .22$, all $p’s > .638$. The Baseline Hypothesis was therefore rejected.

**Discussion**

In the present experiment we examined whether liars and truth tellers’ baseline behaviour, established before the interview begun, differed from their behaviour displayed during the investigative part of the interview. At the baseline period all participants were telling the truth when answering a non-threatening question. This behaviour was compared with their behaviour during the actual interview when participants were aware that their veracity was assessed. According to the baseline lie detection technique, differences in behaviour between the baseline part of the interview and the investigative part of the interview may indicate deception (Inbau, Reid, Buckley, & Jayne, 2013). However, we found that all participants, regardless of veracity, displayed different behaviours between the baseline and both target periods. In other words, liars’ behaviour did change, but so did truth tellers’ behaviour, and the Baseline Hypothesis was therefore rejected. Since most suspects (liars and truth tellers) appear to change their behaviour between baseline and target periods, it means that most suspects will make a suspicious impression when the baseline lie detection technique is applied. This suits those police officers who believe that the suspect they interview is guilty anyway, but it does not make them good lie detectors. Participants also showed different behaviours in target period 1 compared to target period 2, which supports the suggestion that behaviour can change over time and in fact even in the same interview.

Liars gave the impression they were thinking harder and controlling themselves more than truth tellers, which supports two of the three theoretical premises regarding deception.
When telling a story liars have many aspects to think about. Not only do they need to recall the lie, which is in itself cognitively more demanding than telling the truth, they need to remember what they have previously said and monitor their responses so that they sound plausible and consistent. Attempting to control behaviours leaves liars looking more rigid and lacking the spontaneity that truth tellers have when telling their story.

Nervousness was disregarded in the analyses due to low inter-rater reliability. However, this lack of reliability is informative as it shows the difficulty of assessing ‘nervousness’ in an interviewee, a cue many people, including practitioners, seem to rely upon when making veracity judgements (Strömwall, Granhag, & Hartwig, 2004; Vrij & Granhag, 2007).

Despite rejection of the Baseline Hypothesis, we are not opposed to the principles of the baseline lie detection technique per se, but the way in which it is currently employed. For such a technique to work it is essential that the truthful baseline responses are made under similar conditions to the responses under investigation (comparable truth). Only truthful and deceptive responses which are truly comparable in every way apart from veracity should be analysed (Vrij, 2008). It is only then that differences in behaviour between truthful and deceptive responses can be attributed to the fact that someone is lying.

An example of a comparable truth can be seen in the work of Vrij and Mann (2001) who examined video footage of a murder suspect in their police interview. When asked what he had done on the day of the murder the suspect gave a detailed account of his activities during the morning (went to work), afternoon (visited a market) and evening (visited a neighbour). The police checked every detail the man gave and were able to confirm his whereabouts in the morning with several independent witnesses. The suspect’s whereabouts in the afternoon and evening could not be verified. Subsequently, conclusive evidence finally revealed that he did in fact meet his victim in the afternoon and killed her later that day.
Detailed analyses of the videotapes revealed a change in the suspect’s behaviour when he started to describe his activities in the afternoon and evening. The suspect’s truthful statements (about his morning activities) and deceptive statements (about his afternoon and evening activities) are comparable as they were obtained in the same interview under the same circumstances. This change in behaviour may not necessarily mean that he was lying, but it makes him suspicious as there is no real reason why different behaviours would emerge when describing activities at different times in a day.

Unlike the work of Mann and colleagues (Mann, Vrij, & Bull, 2004, 2006; Mann, Vrij, Fisher, & Robinson, 2008; Vrij & Mann, 2001), the scenario in the current study was relatively low stakes. We doubt that the baseline technique will become any more effective in high stakes situations. Indeed, in high stakes situations, the repercussions of not being believed makes it likely that liars will experience emotions such as fear, which may result in different behaviours between baseline and interrogation. However, truth tellers are also likely to fear the repercussions of not being believed in high stakes situations (Bond & Fahey, 1987), which makes it likely that they also will show different behaviours during the interrogation compared to baseline.

Conclusion

The baseline lie detection technique as it is currently used in interrogation techniques does not distinguish liars and truth tellers. Behavioural differences occur in both liars and truth tellers when comparing baseline behaviour to behaviour in the investigative part of the interview. Those wishing to compare someone’s truthful and deceptive nonverbal responses should make sure they are truly comparable. That is, these responses should be taken from the same interview setting; the person should be speaking about the same/similar topics; and the responses should be discussed within a short period of time from each other.
This work is funded by the High-Value Detainee Interrogation Group, J-FBI-12-194 awarded to the University of Portsmouth (UK). Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the U.S. Government.
References


Mann, S., Vrij, A., Fisher, R. & Robinson, M. (2008). See no lies, hear no lies: Differences in discrimination accuracy and response bias when watching or listening to police
doi: 10.1002/acp.1406

behavior under conditions of endogenously-present and exogenously-induced
motivational states. *Journal of Clinical Psychology*, 54, 15-26. doi:
10.1037/h0028651

interviews with criminal suspects. *Policing and Society: An International Journal of

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the
attribution process. In L. Berkowitz (Eds.), *Advances in experimental psychology* (pp.

Stiff, J., Corman, S., Krizek, B., & Snider, E. (1994). Individual differences and changes in
nonverbal behavior unmasking the changing faces of deception. *Communication

In P. A. Granhag & L. A. Strömwall (Eds.), *Deception detection in forensic contexts* (pp.

of Psychology: Interdisciplinary and Applied*, 11, 47-53. doi:
10.1080/00223980.1995.9914944

detection. *Psychology, Crime & Law*, 12, 193-206. doi:
10.1080/10683160512331331319

and Sons.


Table 1

*Impression Ratings as a Function of Period*

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Target 1</th>
<th>Target 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Thinking hard</td>
<td>3.60</td>
<td>1.7</td>
<td>2.90</td>
</tr>
<tr>
<td>Controlling Self</td>
<td>2.55</td>
<td>1.4</td>
<td>1.83</td>
</tr>
</tbody>
</table>

Note: Only mean scores with a different superscript differ significantly from each other (p < .05)