

Detecting deception using comparable truth baselines

Glynis Bogaard, Ewout H. Meijer, Aldert Vrij & Galit Nahari

To cite this article: Glynis Bogaard, Ewout H. Meijer, Aldert Vrij & Galit Nahari (2023) Detecting deception using comparable truth baselines, *Psychology, Crime & Law*, 29:6, 567-583, DOI: [10.1080/1068316X.2022.2030334](https://doi.org/10.1080/1068316X.2022.2030334)

To link to this article: <https://doi.org/10.1080/1068316X.2022.2030334>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 22 Jan 2022.



Submit your article to this journal [↗](#)



Article views: 3310



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 3 View citing articles [↗](#)

Detecting deception using comparable truth baselines

Glynis Bogaard^a, Ewout H. Meijer ^a, Aldert Vrij ^b and Galit Nahari^c

^aDepartment of Clinical Psychological Science, Maastricht University, Maastricht, The Netherlands;

^bDepartment of Psychology, University of Portsmouth, Portsmouth, UK; ^cDepartment of Criminology, Bar-Ilan University, Ramat-Gan, Israel

ABSTRACT

Baselining – comparing the statements of interest to a known truthful statement by the same individual – has been suggested to improve lie detection accuracy. A potential downside of baselining is that it might influence the characteristics of a subsequent statement, as was shown in previous studies. In our first experiment we examined this claim but found no evidence that a truthful baseline influenced the characteristics of a subsequent statement. Next, we investigated whether using a truthful baseline statement as a within-subject comparison would improve lie detection performance by investigating verbal cues (Experiment 1) and intuitive judgements of human judges (Experiment 2). Our exploratory analyses showed that truth tellers included more auditory and temporal details in their target statement than in their baseline than liars. Observers did not identify this verbal pattern. Exposure to a truthful baseline statement resulted in a lower truth accuracy but no difference in lie accuracy.

ARTICLE HISTORY

Received 6 April 2021



Accepted 4 January 2022

KEYWORDS

Baselining; verbal credibility assessment; cues to deception; lie detection accuracy

Human lie detection performance is typically poor. On average, both lay people and professionals detect approximately 54% of the truths and lies, whereas 50% is expected by chance (Bond & DePaulo, 2008, 2006). Among the reasons for people's poor lie detection skills is that they pay too much attention to behavioral cues such as gaze aversion and body movements (Bogaard et al., 2016; Masip & Herrero, 2015; Vrij, 2008a), while the relationship between such cues and deception is faint and unreliable (DePaulo et al., 2003; Hartwig & Bond, 2011). Research has also shown that to improve truth/lie detection accuracy, observers should primarily focus on the content of people's statements (for reviews see Vrij, 2008b; Vrij, 2019). Especially looking at verbal characteristics of a statement, has been shown most promising (Amado et al., 2016; Luke, 2019; Masip et al., 2005b; Oberlader et al., 2016; Ormerod & Dando, 2015).

Yet, even verbal lie detection tools have a substantial error rate, with one potential source of this error rate being individual differences in liar's (verbal) behavior. Meta-analytic research has shown large individual differences in the transparency of people's lies.

CONTACT Glynis Bogaard  Glynis.bogaard@maastrichtuniversity.nl  Department of Clinical Psychological Science, Maastricht University, PO Box 616, Maastricht MD 6200, The Netherlands

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

That is, whether observers are able to detect a lie depends largely on the qualities of someone's lie-telling abilities (Bond & DePaulo, 2008). People who are fantasy prone, for example, are better at formulating believable lies (Merckelbach, 2004; Schelleman-Offermans & Merckelbach, 2010). And verbally skilled people get away with their lies more often as they tend to include more details in their deceptive stories (Kashy & DePaulo, 1996; Vrij et al., 2002). This, in turn, fits with the finding that self-reported good liars also report to highly rely on verbal strategies when lying (Verigin et al., 2019a). Research has also shown that females provide more details than males when telling the truth (Nahari & Pazuelo, 2015). Veracity assessment tools typically do not consider such individual differences.

One potential way to include liar's individual differences in a lie detection procedure – and thereby increase the predictive validity – is through baselining, i.e. using a known truthful statement or part of a statement for comparison (see e.g. Ekman & Friesen, 1974; Moston & Engelberg, 1993; Vrij, 2016). Two types of baselines have been introduced: small talk baselines and comparable truth baselines. Small talk baselines are popular among practitioners, who use an initial 'small talk' part of the interview as a comparison (Moston & Engelberg, 1993). The assumption is that people tell the truth during small talk, and any behavioral differences between this small talk and the part of the interview dealing with the crime under investigation is interpreted as a sign of deception. The problem here is that this comparison is confounded. Topics usually differ between small talk and the investigative phase and, depending on the topic and its personal relevance, people can respond differently (Davis & Hadiks, 1995). Furthermore, the stakes contrast substantially. Whereas small talk is not associated with negative consequences, the investigative phase is. Because of these confounds, interviewees' behavior differs between the baseline and the investigative part of the interview regardless of their veracity status (Ewens et al., 2014).

For a truthful verbal baseline comparison not to be confounded, it should be comparable in topic and stakes to the element under investigation, so-called *comparable truth baseline* (CTB; Vrij, 2016). Palena et al. (2018) compared a small talk baseline to a CTB. Participants performed two tasks, each consisting of several sub-tasks. The first task included receiving an envelope, logging on to a computer, instructions on how to locate a key to a safe-deposit box, and instructions about sending an email. The second task consisted of locating and switching two USB sticks. In a subsequent interview, half of the participants lied about this second task, whereas the other half told the truth. Participants in the CTB condition told the truth about the first task, and this statement served as the CTB. Participants in the small talk condition told something truthfully about their last year as a student and/or as a worker which served as a baseline. For the small talk baseline condition, the truthful and deceptive statements did not differ in detail richness between baseline and the investigative part of the interview. For the CTB condition, the truthful and deceptive statements differed for spatial details ($d = 1.20$). That is, there was less difference in the number of reported spatial details for truth tellers and their CTB than for liars and their CTB. However, no significant differences emerged for temporal, visual, audio or action details. However, as this study did not include a no-baseline condition, it is unclear to what extent a CTB could improve deception detection.

One potential explanation for the limited effect of the CTB (it only worked for spatial details) is as follows. Recent research suggests that truthful and deceptive information

interact and influence detail richness (Verigin et al., 2019b). That is, compared to liars who only told lies, liars whose lies were flanked by truths included more details in their lies. In other words, liars calibrated their deceptive responses based on the truthful information they provided. Additionally, research showed that narratives written after a first narrative contained fewer details than the first narrative, particularly when the second narrative is deceptive (Tomas et al., 2021). Thus, if liars indeed calibrate the number of details they report in their lie based on a prior truthful response, a CTB might actually decrease lie/truth discriminability. However, if liars are unable to calibrate their responses, the use of a CTB has the possibility to improve the lie/truth discriminability.

Therefore, in two experiments, we investigated 1) whether providing a truthful baseline influences the detailedness of a subsequent target statement provided by the same interviewee and 2) whether using a truthful baseline statement as a within-subject comparison would improve lie detection performance by investigating verbal cues and intuitive judgements of human judges.

Experiment 1

Method

Participants

Based on a G*power analysis (Fixed effects, special, main effects and interactions), with $f = .25$, power of .80 and $\alpha = .05$, 160 participants should be included. In total 171 participants (124 females) took part in this study with a mean age of $M = 22.50$ ($SD = 6.14$). Participants received a 5-euro voucher or course credits as a reward for participation. The study was performed at [redacted] University in accordance with the ethical standards of the institution and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon request in the repository DataverseNL [10.34894/Y2TAJ9].

Design

We manipulated the use of a truthful baseline statement (present vs. absent) and deception during a subsequent target statement (truth vs. lie). This resulted in a 2 (Baseline) x 2 (Veracity) between-subject design.

Procedure

Participant first read an information sheet and signed the consent form. Next, they were randomly assigned to one of four conditions. Participants assigned to the baseline present condition started with a pre-interview that served as the baseline statement. Interviewees were told they would be interviewed about a memorable negative event that had happened in their lives – such as having witnessed a crime or having been in a traffic accident and that it was their job to appear honest to the interviewer. We deliberately chose a negative event as this resembles the typical statements that are commonly provided in a legal setting. They were given time to prepare themselves. This baseline statement

was audiotaped while answering the free recall question *'Tell me about a memorable negative event you experienced in your life. It is important that you provide the experimenter with an honest account'*. After participants provided their statement, they were asked *'is there anything you would like to add'*.

Next, all participants were interviewed about the target event, which was also a negative autobiographical event. The interview was audiotaped and used a free recall format. Depending on whether participants were assigned to the truth or lie condition, participants received the following instruction. For truth tellers:

Please imagine this is a real interview setting and your honesty is being judged during this interview. Your statement should be about a real event that actually occurred in your life. Thus, convince the interviewer you are telling the truth by providing them with an explanation of the event that is as detailed as possible. You have some time to prepare your statement.

Liars received the following instruction *'Please imagine this is a real interview setting and your honesty is being judged during this interview. Your statement should be about a fabricated story and should consist of an event that never actually occurred in your life. Thus, convince the interviewer you are telling the truth by providing them with an explanation of the event that is as detailed as possible. You have some time to prepare your statement.'* Truth tellers and liars were then requested to provide their statement *'Tell me about a memorable negative event you experienced'* followed by *'is there anything you would like to add'*.

Lastly, participants were asked how motivated they were, how difficult the task was, how nervous they were and whether they thought they were good liars. All answers were provided on a 9-point Likert scale (1 = not at all, 9 = very much). We also asked them to what extent they told the truth in the pre-interview (for participants in baseline conditions) and the target interview (1 = completely fabricated, 9 = completely truthful).

Reality monitoring scoring

The statements provided covered several topics such as having been a victim of a crime (e.g. theft) harassment; a stressful event at school or at work; death of a close relative; relationship difficulties. To score the statements' richness in detail, we used the Reality Monitoring (RM) approach. Like Palena et al. (2018), the frequencies of the following RM subcategories were scored. The underlined parts illustrate the details coded: (1) Visual details (e.g. *'I remember one of my classmates got hurt on his head'* (3 details)), (2) auditory details (e.g. *'we heard really loud screams'* (1 detail)), (3) action details (e.g. *'I took my phone'* (1 detail)) (4) spatial details (e.g. *'my father came to pick me up at school'* (1 detail)), (5) temporal details (e.g. *'Our school was finished at 1:30 in the afternoon'* (1 detail)) and (6) Affect (e.g. *'I was really scared'* (1 detail)). We selected these criteria as they are among the most revealing veracity indicators (DePaulo et al., 2003; Hauch et al., 2017; Masip et al., 2005b). Details were only considered once, so repetitions were not counted. One rater coded all statements while a second rater coded 20% of the statements. All raters were blind to the veracity of the statements.

Both raters received RM coding training. Two sessions were supervised by the first author of this paper. In the first training session (1.5-2 h) the Reality Monitoring Criteria were described along with the examples. Both raters were then given five statements

that they had to code independently at home. During the second session, each transcript was discussed line by line to ensure raters were coding the same information units with the same criteria. Any inconsistencies and disagreements were discussed, and the raters had to arrive at an agreement on how to best score this information.

Inter-rater agreement

We used the Interclass Correlation Coefficient (ICC) two-way mixed model to calculate the coders' consistency. All consistencies were higher than .80, except for the criterion action details in the target statements (ICC = .58). Overall, the consistency ranged between ICC = .58 and ICC = .97 with an average interrater reliability of ICC = .90 for baseline statements and ICC = .85 for the target statements. Hence, there was a good to excellent agreement between raters.

Separate criteria scores were summed up to get an **RM total score** for the baseline statement, and one for the target statement.

Results

Inspection of our data revealed five outliers ($M \pm 2.5$ SD) who had high RM scores in either their baseline or target statement. They were removed for the subsequent analyses.¹ This resulted in the following cell count: No Baseline-Truth $n = 43$; No Baseline-Lie $n = 40$; Baseline-Truth $n = 45$; and Baseline-Lie $n = 38$. We report effect sizes Cohens f for the ANOVAs, with the following interpretation $f = 0.1$ is a small effect, $f = 0.25$ is a medium effect, and $f = 0.4$ is a large effect (Cohen, 1988).

Pre-analyses check

With two-way ANOVAs we examined whether there were any differences in motivation, difficulty, nervousness and self-reported lying ability between groups. Overall, participants were quite motivated and experienced some nerves. There were no main effects for motivation and nervousness of Baseline or Veracity and no interaction effect (all F s < 2.32 and all p 's $> .13$).

Table 1 shows that participants found the task relatively easy, but a significant interaction effect emerged for task difficulty, $F(1,162) 8.71, p = .004, f = .22$. Investigating the effect for truth tellers and liars separately showed that while no differences in task

Table 1. Overview of means and standard deviations for motivation, nervousness, difficulty, lying ability and truthfulness of target statement separated per factor.

Baseline	Veracity	Motivation		Nervousness		Difficulty		Lying ability		Truthfulness target	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Absent	TT	6.72	1.89	4.09	2.27	2.77	1.95	5.02	1.92	8.16	1.68
	L	6.70	1.80	4.75	2.29	4.38	2.05	4.85	2.35	3.17	2.14
Present	TT	6.68	1.39	4.09	2.47	3.84	2.26	5.51	2.40	8.42	1.39
	L	6.43	1.36	3.66	2.20	3.50	2.24	5.10	2.12	1.92	1.17

Note. All answers were provided on a 9-point Likert scale (1 = not at all, 9 = very much). TT indicates truth tellers and L indicates liars.

difficulty emerged for liars, $F(1,76) 3.25, p = .08, f = .17$, truth tellers found the baseline task more difficult than only providing a target statement, $F(1,86) 5.72, p = .02, f = .23$. Participants overall thought they were average liars ($M = 5.19, SD = 2.18$), and no interaction or main effects emerged for self-reported lie abilities (all $F_s < 1.17$ and all p 's $> .28$).

Manipulation check

A two-way ANOVA with Veracity and Baseline as between-subjects factors and the truthfulness in the target statement as dependent variable showed a significant Veracity main effect, $F(1,162) 507.42, p < .001, f = 1.75$, a marginally significant Baseline main effect, $F(1,162) 3.80, p = .05, f = .13$, and a significant interaction effect, $F(1,162) 8.80, p = .003, f = .22$. The interaction effect is the most informative effect of these three effects. Although truth tellers reported to have been more truthful than liars in both baseline conditions, the difference was larger in the baseline present group, $F(1,76) 10.10, p = .002, f = .34$ than in the baseline absent group $F(1,86) .62, p = .43, f < .001$. See [Table 1](#) for Means and SDs. These results show our manipulation was successful.

Word count

On average, the *target statements* were 214 words long ($SD = 127.17$). A two-way ANOVA with Veracity and Baseline as between-subject factors showed that truth tellers provided longer target statements ($M = 234.02, SD = 142.16$) than liars ($M = 192.95, SD = 104.49$) $F(1,162) 4.54, p = .03, f = .15$. Results showed no significant difference in the length of the target statements between participants who provided a baseline ($M = 205.03, SD = 109.22$) or not ($M = 224.41, SD = 142.90$) $F(1,162) 1.25, p = .27, f = .04$ and no significant interaction effect $F(1,162) 1.61, p = .21, f = .06$.

Investigation of *the baseline statements* showed that the provided baselines were on average 188 words long ($SD = 100.54$) and did not differ in the length between truth tellers ($M = 203.47, SD = 112.18$) and liars ($M = 169.57, SD = 82.44$) $F(1,81) 2.38, p = .13, f = .13$.

RM-scores: hypotheses testing

For our main analyses, we report the effect size Cohens f and accompanying Bayesian Factors (BF). Evidence for the interaction model is calculated as [interaction model]/[main factors] (see Wagenmakers et al., 2018). We use the following classification scheme: $BF_{10} > 100$ extreme evidence for H_1 , 30–100 very strong evidence, 10–30 strong evidence; 3–10 moderate evidence and 1–3 weak evidence for H_1 (Jarosz & Wiley, 2014; Lee & Wagenmakers, 2013). Similar interpretations are used for BF_{01} and its support for H_0 .

Do baselines influence subsequent statements?

To test whether providing a truthful baseline influences the richness of detail of a subsequent target statement, we ran a two-way (Bayesian) ANOVA with Veracity and Baseline as between-subject factors on the target's statement RM total score. The ANOVA showed no main effect for Veracity, $F(1,162) .43, p = .51, f < .001, BF_{01} = 5.02$. Results also showed

no significant findings for baseline, $F(1,162) 2.71, p = .10, f = .10, BF_{01} = 2.07$ or an interaction effect $F(1,162) 3.32, p = .07, f = .12, BF_{01} = 1.00$. However, the low Bayes Factors ($BF < 3$) show that there is only little more support for the null hypotheses than for the alternative hypotheses. At the very least, these results suggest that we found no clear evidence that truth tellers and liars match the detailedness of their subsequent statements using their truthful baselines.

Do truths and lies differ in their detail richness to the CTB?

For the subsequent analyses, we included only the participants in the baseline group. We first checked whether the RM baseline score did not differ between truth tellers ($M = 28.76, SD = 12.87$) and liars ($M = 26.21, SD = 10.64$). Results showed this was not the case, $F(1,81) .94, p = .34, f < .001, BF_{01} = 2.89$. A lack of pre-existing RM differences between truth tellers and liars in their baseline allows us to reliably investigate the RM difference score as an indicator of deception.

To investigate whether truths and lies differ in detail richness from their CTB, previous research (Palena et al., 2018) has investigated the *absolute RM difference score*. This measure does not examine whether the baseline was less or more detailed than the target statement. However, it is possible that this absolute score is not diagnostic because truth tellers give a more detailed target statement than their baseline, while liars give a less detailed statement than their baseline. Using the absolute score does not allow us to investigate this effect. As previous studies have used this measure, we reported the analysis in Appendix A. Nonetheless, we believe it is more relevant to investigate the *directional RM difference* as [Baseline RM score – Target RM score]. A negative score means that the target statement is more detailed than the baseline statement, while a positive score indicates the opposite.

To this end, we ran an ANOVA with the *directional RM difference score* as the dependent variable. Although the difference was in the expected direction, positive for liars ($M = 1.61, SD = 10.31$) and negative for truth tellers ($M = -1.51, SD = 13.46$), the directional RM difference score for truth tellers did not differ significantly from liars, $F(1,81) 1.36, p = .25, f = .07$. Yet, support for the null hypothesis was only weak ($BF_{01} = 2.41$).

Can baselines improve truth/lie detection?

To test whether including a CTB can increase truth/lie detection discrimination, we ran an ANOVA with the RM total target score as the dependent variable, and an ANCOVA with the RM total target scores as the dependent variable and the RM Baseline scores as a covariate. Without taking the RM baseline score into consideration, results showed truth tellers ($M = 30.26, SD = 12.71$) had a significantly higher RM total score than liars ($M = 24.60, SD = 9.24$) $F(1,81) 5.21, p = .03$. The ANCOVA showed similar results $F(1,80) 4.22, p = .04$. The covariate was significantly related to the RM target score $F(1,80) 20.71, p < .001, f = .50, BF_{10} > 100$. We are specifically interested in the effect size of the difference between truth tellers and liars with and without including the RM baseline score. Therefore, we calculated Cohens d . Results showed that the effect size for the ANCOVA ($d = .56, 95\% CI [0.12, 1.00]$) is comparable to the ANOVA ($d = 0.50, 95\% CI [0.06, 0.94]$). So, both analyses significantly differentiated between truth tellers and liars, yet Bayesian analyses only showed weak support for this effect.

RM Scores: Exploratory analyses

In the previous analyses, we used the total RM score. In the present set of analyses, we used the individual RM criteria to explore which criteria show most potential for baselining. We examined whether there were any differences for the *directional RM-difference scores* for the separate criteria. Results of the MANOVA showed no multivariate Veracity effect, $F(6,76) 1.73, p = .12, f = .23$. Yet, at the univariate level, results showed that truth tellers included more temporal and auditory details in their target statement than in their baseline statement, whereas liars included fewer temporal and auditory details in their target statement than in their baseline statement. These results are in line with the baseline predictions (see Table 2).

Discussion

We investigated whether truth tellers and liars try to match the detailedness of their subsequent target statement to their truthful baseline but expected liars to be poorer in doing so than truth tellers. We failed to find evidence that truth tellers and liars matched the detailedness of their subsequent statements using their truthful baselines. This contradicts Tomas et al. (2020), who found that second narratives included less details than first narratives. However, unlike Tomas et al. (2020) we did not counterbalance the veracity of both statements. The first statement was always a truthful baseline as this would best reflect how a baseline could be used in practice. Our findings also contrast the findings of Verigin et al. (2019b) who showed that lies that were flanked by truths included more details than lies that were flanked by lies. Our study only included two statements, instead of three, as Verigin et al. Perhaps the majority of a statement needs to be truthful before liars are able to match the level of detailedness of their lies to their truthful statements. Yet, our results showed no veracity differences between truth tellers and liars to begin with. Therefore, we should be careful with interpreting these null findings.

Lacking evidence of matching or calibration, we investigated whether truths and lies differ in their detail richness from the CTB. First, we examined whether the difference in detail richness (RM total score) was higher between CTBs and truthful target statements than between CTBs and fabricated target statements. The results showed no significant difference for truths and lies in detail richness to the CTB, which is in line with the findings of Palena et al. (2018). Yet, when they investigated the RM criteria separately,

Table 2. Ms, SDs, CIs and significance for the separate RM criteria for the directional* RM difference score.

	Criteria	Truth tellers		Liars		F	p	f
		M	SD	M	SD			
Directional RM difference	Visual	0.33	6.64	0.50	3.91	0.02	.892	.000
	Auditory	-1.29	3.55	0.39	2.63	5.85	.018	.242
	Action	0.47	2.95	0.07	2.97	0.35	.554	.000
	Spatial	0.84	5.55	0.39	4.75	0.15	.696	.000
	Temporal	-1.49	5.00	0.61	4.05	4.29	.042	.200
	Affect	-0.38	2.28	-0.37	1.48	0.00	.983	.000

Note. *directional RM difference score = [Baseline RM score - Target statement RM score]. Criteria indicated in bold show a significant difference between truth tellers and liars ($p < .05$).

they found more similarity in the number of used spatial details for truth tellers and their CTB than for liars, whereas none of the other investigated criteria significantly differed.

In contrast, our explorative analyses for the *absolute RM difference score* revealed that truth tellers show *less* rather than more differences in detail richness than liars. These findings might be the result of truth tellers giving a more detailed target statement than their baseline, while liars give a less detailed statement than their baseline, which was investigated by *the directional RM difference score*. Our exploratory analyses indeed revealed that truth tellers tended to give a target statement that was more detailed in temporal and auditory details than their baseline statement, while liars' results showed the opposite pattern. Truth tellers are probably more motivated to report a full detailed account for their target statement than baseline statement. Theoretically this can be explained by the finding that truth tellers tend to 'tell it as it happened'. That is, they are more forthcoming and tell the event including relevant details (Stromwall et al., 2006). Liars, on the other hand, often use strategies that include being intentionally vague or avoiding specific details and keeping the story clear and simple (Hartwig et al., 2007; Leins et al., 2013; Verigin et al., 2019a). So, when liars move from their truthful baseline to their deceptive target statement, they might provide less details.

Lastly, we examined whether using a CTB can increase truth/lie detection discrimination. Our findings showed that taking into consideration the RM baseline score did not increase the effect size for differentiating between truth tellers and liars. We found a medium effect size, which is in line with meta-analytic research investigating lie detection methods based on verbal characteristics (Amado et al., 2016; Masip et al., 2005b).

In sum, our results showed no evidence that truth tellers and liars match the detailedness of their target statement to their truthful baseline. There was no evidence that a baseline can improve truth/lie discrimination. Our exploratory analyses investigating the separate RM criteria showed that truth tellers gave a target statement that was more detailed in temporal and auditory details than their baseline, while liars' results showed the opposite pattern. These findings suggest that we should pay attention to the quality of the details, and not the quantity of details per se (Nahari & Nisin, 2019). However, would observers who make intuitive veracity judgements be able to decipher the different patterns emerged in providing details in baseline and target conditions? Experiment 2 examined this question.

Experiment 2

To date, three experiments investigated whether baseline interviews can aid intuitive truth/lie detection judgements. In one experiment, lay observers using CTB outperformed observers who used non-comparable baselining such as small talk when making lie detection judgements (56% vs. 47% accuracy; Caso et al., 2019b). The statements showed to observers were about a mission participants had completed. The first half of the mission was used as the CTB and the second half was used as the target statement. For the small talk baseline, statements included personal information about the participant (e.g. describe your last year as a worker). In another experiment, this time with Italian police officers as observers, using a similar CTB improved the detection of lies (55%) as compared to using no baseline (39%), but not of truths (52% vs. no baseline 60%) or total accuracy. In the most recent study, students were instructed to report

their alibi for a specific day (Verigin et al., 2020). Statements were either completely truthful or included a lie about what participants did between 1.00pm-3.00pm, but recollections about the remainder of the day were truthful. When judging the veracity of the statements, half of the observers were told that the surrounding truthful elements of the statement could be used as a baseline, while the other half received no such instruction. Instructing observers to use a baseline did not improve lie/truth discrimination (55% vs no instruction 51%).

The current Experiment 2 aimed to re-examine whether a comparable truth baseline improves truth/lie detection. The verbal difference between truth tellers and liars in Experiment 1 was only small and perhaps difficult for observers to detect. Nonetheless, it can still be that presenting a baseline improves observers' truth lie accuracy, as observers might pay attention to verbal cues that were not considered in Experiment 1. We used the statements from Experiment 1 – statements about negative autobiographical memories – as material. We asked participants to make credibility judgements about these statements and expected participants who received both the baseline and target statement to obtain higher accuracy ratings than participants who only received the target statement.

Method

Participants

Based on a G*power analysis (Repeated measures, within-between interaction), with $f = .20$, power of .80 and $\alpha = .05$, a minimum of 66 participants was needed. This experiment was set up as an online study, hence not all participants completed the survey. After deleting incomplete entries, 138 participants (119 females) took part in the experiment with a mean age of $M = 21.55$ ($SD = 2.98$). Participants received a 5-euro voucher or course credits as a reward for participation. The study was performed at [redacted] University in accordance with the ethical standards of the institution and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon request. After publication, the data will be available upon request in the repository DataverseNL [<https://dataverse.nl>; direct link to the data will be added later].

Design

We used a 3×2 mixed design with Type of Statement (control, target, baseline-target; see Procedure) as a between-subjects factor and Veracity (true vs. lie) as a within-subjects factor.

Procedure

This study was conducted online via Qualtrics. After signing the online consent form, participants were randomly assigned to one of three conditions. Participants had to answer each question before they could continue to the next one. In the **control** condition, participants received eight target statements. Four were randomly drawn from the pool of statements of participants in Experiment 1 who did not provide a baseline and gave a

truthful target statement, and the other four were randomly drawn from participants who did not provide a baseline and gave a fabricated target statement. In the **target** condition, participants received four target statements randomly drawn from the pool of participants in Experiment 1 who provided both a baseline and a truthful target statement, and four randomly drawn from the pool of participants who provided both a baseline and a fabricated target statement. Note that we only gave observers the target statement of these participants, not their baseline statement. In the **baseline-target** condition, participants received four target statements randomly drawn from the pool of participants in Experiment 1 who provided both a baseline and a truthful target statement, and four randomly drawn from the pool of participants who provided both a baseline and a fabricated target statement. These observers received both the **baseline** and **target** statement of the same person. Moreover, they were told that the baseline statement was a known truthful statement and that they should use it to make a credibility decision on the target statement. Because the study was offered online and to ensure participants carefully read the statements, we included a 60sec time delay when presenting the statements. That is, participants had to wait 60 s before they could skip to the next question.

Next, all participants were asked after every statement: 'How credible do you judge this statement?' Answers were given on 7-point rating scales (**credibility score**: 1 = very unbelievable to 7 = very believable). They were also asked the same question with a forced choice binary judgement (true vs. lie).

After completing the eight judgements, participants were asked to indicate their agreement with several statements on a 7-point scale (1 = strongly disagree to 7 = strongly agree). They were asked about their motivation, the difficulty of the task, how well they thought they performed and whether they usually get away with their lies.

Results

Pre-analyses check

Overall participants agreed with the statement 'I was motivated to participate in this study' ($M = 5.59$, $SD = 1.07$) and this motivation did not differ between conditions, $F(2,63) 2.84$, $p = .07$, $f = .26$. The statement 'It was easy for me to judge the veracity of the statements' was answered with a neutral response ($M = 3.39$, $SD = 1.36$), and again there were no differences between conditions, $F(2,63) .19$, $p = .83$, $f < .001$. When asked 'I did well on judging the veracity of the statements' participants tended to give a neutral answer ($M = 3.95$, $SD = 1.06$) and this did not differ between conditions, $F(2,63) .82$, $p = .44$, $f < .001$. Participants somewhat agreed with the statement that they usually get away with their lies ($M = 4.17$, $SD = 1.53$) and there were again no differences between conditions, $F(2,63) 1.29$, $p = .28$, $f = .09$.

Hypothesis testing

To investigate whether the baseline statement improved participants' accuracy, we first calculated the number of correct judgements for true and deceptive statements based on our participants' forced choice responses (max score = 4) and recoded this as percentage correct for truths and lies in the following way [percentage correct = (number correct

/ 4) * 100]. Next, we calculated the average credibility score for true and deceptive statements separately as follows: Participants' answers to the question 'How credible do you judge this statement?' (7-point scale) [average credibility truth = (sum credibility scores for truths / 4) and average credibility lie = (sum credibility scores for lies / 4). We carried out two mixed ANOVAs, with Type of Statement as a between-subjects variable (control, target, baseline-target) and Veracity (true vs. lie) as a within-subjects variable. First, we investigated credibility as the dependent variable, followed by accuracy.

Credibility judgement

For the 7-point judgement, we found no significant effect of Type of Statement, $F(2,135) 1.98, p = .14, f = .12, BF_{01} = 4.05$. Results showed weak support for the finding that truths were rated as more credible than lies, $F(1,135) 4.75, p = .03, f = .17, BF_{10} = 1.39$. Results also showed weak support for the interaction effect $F(2,135) 3.25, p = .04, f = .18, BF_{10} = 1.45$. We investigated the simple effects by examining the difference between the three Type of Statement conditions for truths and lies separately. For lies, there was no significant difference between conditions, $F(2,135) 1.34, p = .26, f = .07, BF_{01} = 4.60$. However, for truths, results showed weak support for a significant difference between conditions, $F(2,135) 3.87, p = .02, f = .20, BF_{10} = 1.81$ (see Table 2). Tukey post hoc tests showed that the baseline-target condition had a significantly lower credibility score than the target condition ($M_{diff} = .47, p = .03, 95\% \text{ CI } [0.05,0.88]$). See Table 3 for Means and SDs.

Accuracy judgement

For the binary judgement, we found a significant Veracity main effect, $F(1,135) 53.49, p < .001, f = .62, BF_{10} > 100$. Truths were judged more accurately than lies, see Table 3. We also found a significant effect for Type of Statement, $F(2,135) 4.69, p = .01, f = .23$, however, support for this finding was only weak ($BF_{10} = 1.20$). Tukey post hoc tests showed one effect: The baseline-target condition obtained a significantly lower accuracy score than the target condition ($M_{diff} = 11.23\%, p = .009, 95\% \text{ CI } [2.22,20.24]$).

The interaction effect was also significant, $F(2,63) 4.18, p = .02, f = .31, BF_{10} = 3.97$. Therefore, we investigated the simple effects by examining the difference between the three Type of Statement conditions for truths and lies separately. For lies, there was no significant difference between conditions, $F(2,135) 1.02, p = .36, f = .02, BF_{01} = 6.01$ (see Table 3). However, for truths, results showed a significant difference between conditions $F(2,135) 7.76, p = .001, f = .31$, which was strongly supported by the Bayes Factor ($BF_{10} = 43.21$). Tukey post hoc tests showed that the baseline-target condition had a significantly

Table 3. Overview of the percentage of correct judgements (binary judgement) and the average credibility score (7-point scale) for each condition.

Condition	Truths			Lies			Total
	% correct	Credibility		% correct	Credibility		
		<i>M</i>	<i>SD</i>		<i>M</i>	<i>SD</i>	
Control	66	4.76	0.90	38	4.47	0.81	52
Target	67	4.67	0.84	45	4.18	0.97	56
baseline target	49	4.29	0.83	40	4.43	0.94	45

Note. All answers were provided on a 7-point Likert scale (1 = not at all, 7 = very much).

lower accuracy than the target ($M_{diff} = 17.82\%$, $p = .002$, 95% CI [5.74,29.89]) and control conditions ($M_{diff} = 16.24\%$, $p = .003$, 95% CI [4.73,27.76]).

Discussion

Experiment 2 investigated to what extent verbal baselining could improve intuitive veracity judgements. Our main hypothesis was that participants in the baseline-target condition would outperform those in the control and target condition. This was not the case; the baseline-target condition obtained the lowest accuracy overall, caused by a low truth accuracy in this condition.

Our accuracy findings differ from Caso et al. (2019a, b) who reported an improved lie accuracy rate for observers using a comparable truth baselining. We can think of two possible reasons for these contrasting findings. First, we used a different type of baseline. Whereas Caso et al. (2019a, b) used the first part of a mission as a baseline, we asked participants to report a negative autobiographical memory. As a consequence, baseline and target statement were much more similar in Caso et al. than in our study, which may have facilitated successful lie detection in Caso et al. This explanation would fit with research showing that lie/truth discriminability was higher when people told the truth and lied about similar themes than different themes (Palena et al., 2019). A second explanation is that our participants were students and community members, whilst Caso et al. (2019a) tested police officers. The latter group has been shown to favor lie decisions (Masip et al., 2005a; Meissner & Kassin, 2002; Narchet et al., 2011), while the former group usually show a tendency to favor truth decisions (Levine, 2014). Verigin et al. (2020) found that observers – regardless of the baseline instruction – did not differ in truth/lie discrimination. Although more in line with our findings, there was no evidence in Verigin et al. (2020) that a baseline could decrease truth accuracy, as in our experiment. Again, this difference might be explained because we used separate statements for baseline and target about different topics, while Verigin et al. (2020) used parts of the same statement that were compared as baseline and target.

General discussion

Two experiments investigated to what extent baselining would influence lie detection accuracy as measured by Reality Monitoring criteria and intuitive judgements made by observers. Overall, we found no evidence that 1) participants calibrated their target statement to their baseline, 2) that similarity in verbal cues between CTB and target statement could distinguish between truthful and deceptive statements, and 3) that including a CTB would increase intuitive truth/lie discrimination. Yet, our exploratory analyses showed an interesting verbal pattern: Truth tellers' target statements included more temporal and auditory details than their baseline statement, while liars' results showed the opposite pattern. However, Experiment 2 showed that naïve observers were unable to spot this veracity difference when making veracity judgements using a truthful baseline.

Two methodological issues deserve attention. First, we used statements about negative autobiographical memories as a baseline. However, practitioners are often interested in peoples' whereabouts (e.g. alibis). Hence, the suitability of known truthful alibi statements as baselines might be interesting to investigate in future studies. Interviewees

could be asked about their whereabouts the day before or after the offence was committed and their verbal behavior can be compared to the statement of their whereabouts on the day of the offence.

Second, we used naïve observers in Experiment 2. This may not reflect real life when professionals make veracity judgements. Research has shown that informed observers, who are aware of the working of the verbal veracity assessment tool under investigation, performed much better than their naïve counterparts (Mac Giolla & Luke, 2021; Nahari, 2017). Someone could argue that when testing the efficacy of a verbal veracity assessment tool, the performance of informed observers is particularly relevant. This has not been investigated in the current experiment.

Verbal baselining can be introduced in a different way than examined in this article: By asking interviewees to discuss the same event but with different instructions or in different formats (Vrij, 2016). For example, people can be asked to provide their alibi twice, once before being provided with a detailed example statement (i.e. Model Statement) and once after listening to such a Model Statement. Research suggests that truth tellers and liars add different types of detail in response to a Model Statement (Vrij et al., 2018).

In conclusion, Experiment 1 showed no evidence that using a comparable truth baseline improves truth/lie accuracy. Experiment 2 showed that observers who used verbal baselining became worse at detecting truths, but equally accurate at detecting lies as compared to observers who did not use verbal baselining. Given support for most of our data was only weak, using differences in detail richness between the CTB and the target statement as a lie detection tool seems premature.

Note

1. Interpretations of the results of the statistical main analyses (hypotheses testing) were similar with and without the outliers, except for the ANCOVA. Without outliers, the results showed no veracity main effect $F(1,85) = 1.80, p = .18, f = .10$. The covariate, RM Baseline score, was still significantly related to the RM target score $F(1,85) = 34.86, p < .001, f = .62, R^2 = .29$.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This publication is part of the project "Outsmarting liars" with project number VI.Veni.201G.016 of the research programme Veni which is financed by the Dutch Research Council (NWO) and by the University Fund Limburg (UFL) SWOL [Grant Number CoBes18.038].

ORCID

Ewout H. Meijer  <http://orcid.org/0000-0001-9590-3699>

Aldert Vrij  <http://orcid.org/0000-0001-8647-7763>

References

- Amado, B. G., Arce, R., Farina, F., & Vilarino, M. (2016). Criteria-Based content analysis (CBCA) reality criteria in adults: A meta-analytic review. *International Journal of Clinical and Health Psychology, 16*(2), 201–210. <https://doi.org/10.1016/j.ijchp.2016.01.002>
- Bogaard, G., Meijer, E. H., Vrij, A., & Merckelbach, H. (2016). Strong, but wrong: Lay people's and police officers' beliefs about verbal and nonverbal cues to deception. *PLoS ONE, 11*(6), e0156615. <https://doi.org/10.1371/journal.pone.0156615>
- Bond, C. F., Jr., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review, 10*(3), 214–234. https://doi.org/10.1207/s15327957pspr1003_2
- Bond, C. F., & DePaulo, B. M. (2008). Individual differences in judging deception: Accuracy and bias. *Psychological Bulletin, 134*(4), 477–492. <https://doi.org/10.1037/0033-2909.134.4.477>
- Caso, L., Palena, N., Carlessi, E., & Vrij, A. (2019a). Police accuracy in truth/lie detection when judging baseline interviews. *Psychiatry, Psychology and Law, 26*(6), 841–850. <https://doi.org/10.1080/13218719.2019.1642258>
- Caso, L., Palena, N., Vrij, A., & Gnisci, A. (2019b). Observers' performance at evaluating truthfulness when provided with comparable truth or small talk baselines. *Psychiatry, Psychology and Law, 26*(4), 571–579. <https://doi.org/10.1080/13218719.2018.1553471>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Erlbaum Associates, Hillsdale.
- Davis, M., & Hadiks, D. (1995). Demeanor and credibility. *Semiotica, 106*(1-2), 5–54. <https://doi.org/10.1515/semi.1995.106.1-2.5>
- 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. <https://doi.org/10.1037/0033-2909.129.1.74>
- Ekman, P., & Friesen, W. V. (1974). Detecting deception from the body or face. *Journal of Personality and Social Psychology, 29*(3), 288–298. <https://doi.org/10.1037/h0036006>
- 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. <https://doi.org/10.1002/jip.1414>
- Hartwig, M., & Bond Jr, C. F. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin, 137*(4), 643–659. <https://doi.org/10.1037/a0023589>
- Hartwig, M., Granhag, P. A., & Strömwall, L. A. (2007). Guilty and innocent suspects' strategies during police interrogations. *Psychology, Crime & Law, 13*(2), 213–227. <https://doi.org/10.1080/10683160600750264>
- Hauch, V., Sporer, S. L., Masip, J., & Blandón-Gitlin, I. (2017). Can credibility criteria be assessed reliably? A meta-analysis of criteria-based content analysis. *Psychological Assessment, 29*(6), 819–834. <https://doi.org/10.1037/pas0000426>
- Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving, 7*(1), 2–9. <https://doi.org/10.7771/1932-6246.1167>
- Kashy, D. A., & DePaulo, B. M. (1996). Who lies? *Journal of Personality and Social Psychology, 70*(5), 1037–1051. <https://doi.org/10.1037/0022-3514.70.5.1037>
- Lee, M. D., & Wagenmakers, E. J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.
- Leins, D. A., Fisher, R. P., & Ross, S. J. (2013). Exploring liars' strategies for creating deceptive reports. *Legal and Criminological Psychology, 18*(1), 141–151. <https://doi.org/10.1111/j.2044-8333.2011.02041.x>
- Levine, T. R. (2014). Truth-Default theory (TDT). *Journal of Language and Social Psychology, 33*(4), 378–392. doi:10.1177/0261927X14535916
- Luke, T. J. (2019). Lessons from pinocchio: Cues to deception may be highly exaggerated. *Perspectives on Psychological Science, 14*(4), 646–671. <https://doi.org/10.1177/1745691619838258>
- Mac Giolla, E., & Luke, T. J. (2021). Does the cognitive approach to lie detection improve the accuracy of human observers? *Applied Cognitive Psychology, 35*(2), 385–392. <https://doi.org/10.1002/acp3777>

- Masip, J., Alonso, H., Garrido, E., & Antón, C. (2005a). Generalized communicative suspicion (GCS) among police officers: Accounting for the investigator bias Effect1. *Journal of Applied Social Psychology, 35*(5), 1046–1066. <https://doi.org/10.1111/j.1559-1816.2005.tb02159.x>
- Masip, J., & Herrero, C. (2015). Police detection of deception: Beliefs about behavioral cues to deception are strong even though contextual evidence is more useful. *Journal of Communication, 65*(1), 125–145. <https://doi.org/10.1111/jcom.12135>
- Masip, J., Sporer, S. L., Garrido, E., & Herrero, C. (2005b). The detection of deception with the reality monitoring approach: A review of the empirical evidence. *Psychology, Crime & Law, 11*(1), 99–122. <https://doi.org/10.1080/10683160410001726356>
- 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
- Merckelbach, H. (2004). Telling a good story: Fantasy proneness and the quality of fabricated memories. *Personality and Individual Differences, 37*(7), 1371–1382. <https://doi.org/10.1016/j.paid.2004.01.007>
- Moston, S., & Engelberg, T. (1993). Police questioning techniques in tape recorded interviews with criminal suspects. *Policing and Society, 3*(3), 223–237. <https://doi.org/10.1080/10439463.1993.9964670>
- Nahari, G. (2017). Top-down processes in interpersonal reality monitoring assessments. *Psychology, Public Policy, and Law, 23*(2), 232–242. <https://doi.org/10.1037/law0000110>
- Nahari, G., & Nisin, Z. (2019). Digging further into the speech of liars: Future research prospects in verbal lie detection. *Frontiers in Psychiatry, 10*, 56. <https://doi.org/10.3389/fpsy.2019.00056>
- Nahari, G., & Pazuelo, M. (2015). Telling a convincing story: Richness in detail as a function of gender and information. *Journal of Applied Research in Memory and Cognition, 4*(4), 363–367. <https://doi.org/10.1016/j.jarmac.2015.08.005>
- Narchet, F. M., Meissner, C. A., & Russano, M. B. (2011). Modeling the influence of investigator bias on the elicitation of true and false confessions. *Law and Human Behavior, 35*(6), 452–465. <https://doi.org/10.1007/s10979-010-9257-x>
- Oberlader, V. A., Naefgen, C., Koppehele-Gossel, J., Quinten, L., Banse, R., & Schmidt, A. F. (2016). Validity of content-based techniques to distinguish true and fabricated statements: A meta-analysis. *Law and Human Behavior, 40*(4), 440–457. <https://doi.org/10.1037/lhb0000193>
- Ormerod, T. C., & Dando, C. J. (2015). Finding a needle in a haystack: Toward a psychologically informed method for aviation security screening. *Journal of Experimental Psychology: General, 144*(1), 76–84. <https://doi.org/10.1037/xge0000030>
- Palena, N., Caso, L., & Vrij, A. (2019). Detecting lies via a theme-selection strategy. *Frontiers in Psychology, 9*, 2775. <https://doi.org/10.3389/fpsyg.2018.02775>
- 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. <https://doi.org/10.1002/jip.1495>
- Schelleman-Offermans, K., & Merckelbach, H. (2010). Fantasy proneness as a confounder of verbal Lie detection tools. *Journal of Investigative Psychology and Offender Profiling, 7*(3), 247–260. <https://doi.org/10.1002/jip.121>
- Stromwall, L. A., Hartwig, M., & Granhag, P. A. (2006). To act truthfully: Nonverbal behaviour and strategies during a police interrogation. *Psychology, Crime & Law, 12*(2), 207–219. doi:10.1080/10683160512331331328
- Tomas, F., Dodier, O., & Demarchi, S. (2021). Baseline affects the production of deceptive narratives. *Applied Cognitive Psychology, 35*(1), 300–307. <https://doi.org/10.1002/acp.3768>
- Verigin, B. L., Meijer, E. H., Bogaard, G., & Vrij, A. (2019a). Lie prevalence, lie characteristics and strategies of self-reported good liars. *PLoS ONE, 14*(12), e0225566. <https://doi.org/10.1371/journal.pone.0225566>
- Verigin, B. L., Meijer, E. H., & Vrij, A. (2021). A within-statement baseline comparison for detecting lies. *Psychiatry, Psychology and Law, 28*(1), 94–103. <https://doi.org/10.1080/13218719.2020.1767712>
- Verigin, B. L., Meijer, E. H., Vrij, A., & Zauzig, L. (2019b). The interaction of truthful and deceptive information. *Psychology, Crime & Law, 26*(4), 367–383. <https://doi.org/10.1080/1068316X.2019.1669596>

- Vrij, A. (2008a). Beliefs about nonverbal and verbal cues to deception. In A. Vrij (Ed.), *Detecting lies and deceit* (pp. 115–140). Wiley.
- Vrij, A. (2008b). Nonverbal dominance versus verbal accuracy in Lie detection. *Criminal Justice and Behavior*, 35(10), 1323–1336. <https://doi.org/10.1177/0093854808321530>
- Vrij, A. (2016). Baseline as a Lie detection method. *Applied Cognitive Psychology*, 30(6), 1112–1119. <https://doi.org/10.1002/acp.3288>
- Vrij, A. (2019). Deception and truth detection when analyzing nonverbal and verbal cues. *Applied Cognitive Psychology*, 33(2), 160–167. <https://doi.org/10.1002/acp.3457>
- Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2002). Will the truth come out? The effect of deception, age, status, coaching, and social skills on CBCA scores. *Law and Human Behavior*, 26(3), 261–283. doi:10.1023/A:1015313120905
- Vrij, A., Leal, S., & Fisher, R. P. (2018). Verbal deception and the model statement as a lie detection tool. *Frontiers in Psychiatry*, 9, 492. <https://doi.org/10.3389/fpsy.2018.00492>
- Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E. J., van Doorn, J., Smira, M., Epskamp, S., Etz, A., Matzke, D., de Jong, T., van den Bergh, D., Sarafoglou, A., Steingroever, H., Derks, K., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, 25(1), 58–76. <https://doi.org/10.3758/s13423-017-1323-7>

Appendix A

Absolute RM difference scores

To investigate whether truths and lies differ in detail richness from their CTB, we followed Palena et al. (2018b) procedure and conducted an ANOVA comparing the absolute RM-difference score between truth tellers and liars. The *absolute RM difference score* was calculated as follows: $\text{abs}[\text{Baseline RM score} - \text{Target statement RM score}]$. Results of the ANOVA showed that the *absolute RM difference score* for truth tellers ($M = 10.44$, $SD = 8.48$) did not differ significantly from liars ($M = 7.39$, $SD = 7.26$), $F(1,81) = 3.031$, $p = .09$, $f = .16$. Yet, support for the null hypothesis is only weak ($BF_{01} = 1.17$). This means that we found no evidence for our hypothesis that there is more similarity in detail richness in truth tellers' statements and their CTB than in liars' statements and their CTB.

In the previous analysis, we used the total RM score. Here we used the individual RM criteria to explore which criteria show most potential for baselining. Therefore, a MANOVA was conducted, with Veracity as the between-subjects factor and the *absolute RM difference score* for each criterion separately as the dependent variable. Results showed a multivariate Veracity effect $F(6,76) = 2.87$, $p = .01$, $f = .37$. At the univariate level, truth tellers showed more differences in detail richness in their reported visual, auditory and affect details than liars, which is contrary to what we expected (see Table 4).

Table 4. Ms, SDs and significance for the separate RM criteria for the absolute* RM difference score.

	Criteria	Truth tellers		Liars		F	p	f
		M	SD	M	SD			
Absolute RM difference	Visual	5.22	4.04	2.82	2.72	9.73	.003	.320
	Auditory	2.98	2.28	2.03	1.68	4.52	.036	.203
	Action	2.38	1.77	2.29	1.86	0.05	.826	.000
	Spatial	4.22	3.64	3.34	3.36	1.29	.259	.060
	Temporal	4.16	3.10	2.87	2.89	3.79	.055	.181
	Affect	1.67	1.58	1.00	1.14	4.69	.033	.201

Note. *absolute RM difference score = $\text{abs}[\text{Baseline RM score} - \text{Target statement RM score}]$. Criteria indicated in bold show a significant difference between truth tellers and liars ($p < .05$).