



The effects of sketching while narrating on information elicitation and deception detection in multiple interviews[☆]

Haneen Deeb^{*}, Aldert Vrij, Sharon Leal, Jennifer Burkhardt

University of Portsmouth, Department of Psychology, United Kingdom

ARTICLE INFO

Keywords:

Deception
Lie detection
Sketch
Multiple interviews

ABSTRACT

Sketching while narrating is effective for eliciting information and veracity cues in single interviews. In the current research, we tested this technique in multiple interviews. Participants were interviewed three times over three weeks about a genuine (truth tellers) or a fabricated (lie tellers) memorable event. They sketched while narrating in Week 1, Week 2, Weeks 1 and 2, or not at all (verbal statement only). Statements were coded for total, core, peripheral, and common knowledge details, self-handicapping strategies, complications, plausibility, and proportions of complications and core details. In the third interview and across interviews, the Sketch instruction resulted in a higher proportion of core details. Truth tellers reported more total and core details and complications and fewer common knowledge details and exhibited a higher proportion of complications than lie tellers. Truth tellers' stories also sounded more plausible than lie tellers' stories. The interaction effects were not significant. Thus, sketching while narrating seemed to have a similar effect on truth tellers and lie tellers in the current study.

1. Introduction

In the last decade, interview techniques have been developed to enhance or elicit verbal cues to deceit (Granhag & Hartwig, 2015; Nahari, 2019; Vrij et al., 2017). Such techniques were typically tested in single interview settings. This contrasts with real life interviews where suspects are often interviewed on multiple occasions over a period of time (Alison et al., 2014; Home Office, 2020). In the current experiment, we examined the efficiency of the sketching while narrating interview technique in multiple interviews.

1.1. Sketching while narrating as an effective interview technique to elicit information

Sketching while narrating, that is sketching while talking at the same time, leads to more information from truth tellers than just narrating (Dando et al., 2009; Eastwood et al., 2018; Leins et al., 2014). At least five reasons contribute to this effect (Vrij, Mann, et al., 2020). First, sketching mentally reinstates the context of the interviewees'

experiences which in turn enhances recall. Second, unlike verbal statements, sketches are visual outputs and therefore a more compatible output format when recalling visually experienced events. Third, sketching is a time-consuming activity which gives (truthful) interviewees good opportunity to search their memory, thus facilitating recall. Fourth, sketching one aspect of the event cues retrieval of other aspects of the same event. Fifth, sketching automatically leads to the provision of spatial information as interviewees must situate each described person or object somewhere in the sketched location. This contrasts with verbal statements in which interviewees do not always spontaneously report where persons and objects are exactly.

1.2. Verbal cues to deception

Lie tellers and truth tellers differ in the amount and type of detail they provide (Amado et al., 2016; Vrij, Leal, Jupe, & Harvey, 2018). The deception research to date that tested the effectiveness of sketching while narrating (Vrij, Leal, Jupe, & Harvey, 2018; Vrij, Mann, et al., 2020) focused on the total amount of detail and on the following three

[☆] This work was supported by the High-Value Detainee Interrogation Group [Grant number 15F06718C0002515]. The funding body had no other involvement in the research than providing financial support. Any opinions, findings, conclusions, or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the U.S. Government. The data for this study will be made available on the public repository of the University of Portsmouth (<https://researchportal.port.ac.uk/portal/en/datasets/search.html>) following acceptance for publication.

^{*} Corresponding author at: Department of Psychology, University of Portsmouth, King Henry I Street, PO1 2DY Portsmouth, United Kingdom.

E-mail address: haneen.deeb@port.ac.uk (H. Deeb).

detail types: (a) common knowledge details which comprise scripted information that is commonly known by people (“The event had an Oscars theme so everybody was dressed up”); (b) self-handicapping strategies which are justifications presented by interviewees for why they cannot provide certain information (“There isn’t much to say about the actual bungee jump as it took only a few moments”); and (c) complications which are details that make the story more complex than necessary (“I started going down the mountain and I realised I haven’t clicked my boot into the ski correctly”). We therefore focus on these cues in the current research.

Generally, truth tellers report more total details and complications and fewer common knowledge details and self-handicapping strategies than lie tellers (Vrij et al., 2019; Vrij, Leal, Mann, Fisher, et al., 2018). As memory of a genuine event is often richer in detail than memory of an imagined event (Johnson & Raye, 1981; Sporer & Sharman, 2006), truth tellers are able to produce many details and complications (Amado et al., 2016). In contrast, lie tellers tend to keep their stories simple (DePaulo et al., 2003; Hartwig et al., 2007), so they may not be willing to provide many details or complications, because this will make a story complex (Vrij, 2005; Vrij & Vrij, 2020). The absence of a genuine experience of an event may prompt lie tellers to rely on common knowledge details when providing information or to avoid including certain information and providing a justification for that (Vrij, Leal, Mann, Fisher, et al., 2018).

In addition to the abovementioned four cues, we also examined core details and peripheral details. Core details are details that are central to the event under consideration, and, if altered, change the event, whereas peripheral details are details that do not change the main event (Herlihy et al., 2002; Heuer & Reisberg, 1990). As an example, an interviewee reporting about being stung by a jellyfish may state details concerning the core event (“I suddenly felt this really painful stinging in the bottom of my leg”), but then talk about peripheral details (“The hotel had sea turtles that laid their eggs on the beach”). Researchers generally agree that truth tellers are more likely to include core details in their statements than lie tellers, but there is less agreement concerning peripheral details (Deeb et al., 2018; Leal, Vrij, Vernham, et al., 2019). Where differences emerged, lie tellers reported more peripheral details than truth tellers (Deeb et al., 2018; Leal et al., 2018). This can be explained by lie tellers’ unwillingness to provide false core details as they want to keep their stories simple (Granhag & Hartwig, 2008), so they compensate for that by reporting peripheral details, possibly from past experience or from commonly known events (Deeb, Vrij, Leal, Verigin, & Kleinman, 2020).

Examining a mixture of cues to truthfulness (complications or core details) and cues to deceit (common knowledge details, self-handicapping strategies, and peripheral details) allows researchers to introduce within-subjects measures by comparing aspects of the same statement and using the interviewee as his or her own control (Vrij, Leal, & Fisher, 2018). Within-subjects measures are generally preferred by practitioners (Vrij, 2016, 2019) and scholars (Nahari et al., 2019; Nahari & Vrij, 2015). The proportion of complications is an example of a recently introduced within-subjects measure. This proportion score compares the number of complications in a single statement to the total number of complications, common knowledge details, and self-handicapping strategies in that statement (complications/[complications + common knowledge details + self-handicapping strategies]), and it is often found to be higher among truth tellers than lie tellers (Leal, Vrij, Deeb, & Kamermans, 2019; Vrij, Leal, Jupe, & Harvey, 2018; Vrij & Vrij, 2020). Another within-subjects measure that was examined in only one previous study (Deeb, Vrij, & Leal, 2020) and that was also found to be higher among truth tellers than lie tellers is the proportion of core details (core details/[core details + peripheral details]).

A cue that has been shown to be one of the most robust veracity cues with the largest effect sizes to date is plausibility, or the extent to which the story looks believable (DePaulo et al., 2003; Sporer et al., 2020; Vrij, Deeb, et al., 2020). Truth tellers’ stories often sound more plausible than lie tellers’ stories. Plausibility is typically measured on a rating scale and

can be assessed in real time during the interview. This makes it potentially an appealing cue for practitioners.

1.3. Sketching while narrating may affect truth tellers and lie tellers differently

Sketching while narrating may have a different effect on truth tellers and lie tellers. Since, truth tellers’ memory of an event is typically richer than lie tellers’ memory, sketching should facilitate memory recall more among truth tellers than among lie tellers. It may therefore strengthen the differences in total details and complications between truth tellers and lie tellers (Vrij, Leal, Fisher, Mann, et al., 2018). If the request to sketch while narrating makes lie tellers think that they have to provide more information, it may strengthen veracity differences for common knowledge details and self-handicapping strategies. Also, if truth tellers focus more on core details whereas lie tellers focus more on peripheral details, veracity differences may be strengthened further as a result of the sketching instruction. Derived from this, strengthening these veracity effects would automatically make the proportion of complications and core details scores more diagnostic. In a similar vein, the richer additional details truth tellers report as a function of sketching may sound more plausible than the additional details lie tellers report. If so, the sketch instruction will also strengthen the plausibility veracity effect.

1.4. Sketch and veracity effects in multiple interviews

In multiple interview settings, an earlier interview may have an effect on the performance in a following interview. An early recall attempt by truth tellers may strengthen episodic memory and cue retrieval of various aspects of an event (Gabbert et al., 2009). This in turn reduces the extent of forgetting and facilitates later recall by truth tellers who potentially add new information in subsequent interviews (Vrij, Leal, Fisher, Mann, et al., 2018). This may be particularly the case if memory-enhancement techniques such as sketching instructions are introduced in these earlier interviews (Izovotas et al., 2018). In contrast, lie tellers would be expected to report the same information over time, because they are typically concerned with appearing consistent during interviews (Granhag & Strömwall, 1999). To achieve this, lie tellers rehearse a statement prior to the first interview and make sure they repeat the same information in subsequent interviews (Deeb et al., 2017; Vredeveltdt et al., 2014).

2. The current study

We examined the effects of sketching while narrating in three interviews (Time 1, Time 2, and Time 3)—each interview one week apart. The instruction to sketch and narrate was presented in Week 1 only, Week 2 only, Weeks 1 and 2, or not at all (Sketch control condition). We tested the Sketch, Veracity, and Sketch × Veracity effects at Time 3 and across the three interviews (unique details). We expected the same results to emerge in both analyses; Hypotheses 1 to 3 reported below therefore apply to both analyses.

Hypothesis 1. Participants in the Sketch Weeks 1 and 2 condition will report the most total details, core details, peripheral details, common knowledge details, self-handicapping strategies, and complications, and exhibit the highest proportions of complications and core details, whereas those who are not asked to sketch and narrate (Sketch control condition) will report the least details (*Sketch main effect*).

Hypothesis 2. Truth tellers will provide more total details, core details, and complications and fewer peripheral details, common knowledge details, and self-handicapping strategies, exhibit higher proportions of complications and core details, and sound more plausible than lie tellers (*Veracity main effect*).

Hypothesis 3. The Veracity effects will be most profound in the Sketch Weeks 1 and 2 condition and least profound in the Sketch control condition (*Sketch × Veracity interaction effect*).

In addition to testing the above hypotheses, we also wanted to get insight into different strategies that truth tellers and lie tellers prepare and use in multiple interviews, so we asked participants about their preparation strategies and their convincing strategies during the interviews. Also, to examine differences in perceptions between truth tellers and lie tellers concerning the sketching while narrating task, we asked those who sketched about the extent to which they thought this task was difficult.

3. Method

3.1. Participants and design

The required sample size was determined using G*Power software. The analysis revealed that to achieve high power of 0.99, a moderate to large effect size of $f^2 = 0.11$, and an error probability of 1% ($\alpha = 0.01$), at least 217 participants were needed.

We recruited a total of 248 student and staff members at the University of Portsmouth between October 2019 and February 2020. We did not set any inclusion criteria other than that participants should be 18 years or older. One participant withdrew during the first session, and four others did not attend the following sessions. Therefore, the final sample included a total of 243 participants (70% female, $M_{age} = 24.05$, $SD_{age} = 9.87$).

Approximately half (57%) of the participants were British, and the remaining participants were European (17%), Asian (13%), African (2%), Arab (1%), of mixed ethnicity (7%), or of other background (2%). Due to a technical error, we could not know the ethnicity of four participants. Although 43% of the participants were non-native English speakers, we do not anticipate that this would have affected the results as it has been shown that non-native but highly proficient speakers provide statements that are similar to those of native speakers (Evans et al., 2017). Also, English is a requirement at the corresponding university, so participants are expected to be highly proficient English speakers. To corroborate this conclusion, we examined potential differences between the British and non-British participants in our sample. We ran two sets of one-way multivariate analyses of variance (MANOVAs) with Ethnicity as the independent variable. In one set—for each of the Time 3 details and the unique details—we included core details, peripheral details, common knowledge details, self-handicapping strategies, complications, and plausibility as dependent variables, and in another set—also for each of the Time 3 details and the unique details—we included the proportion of complications, the proportion of core details, and total details as dependent variables. Missing cases were removed from the analyses. No significant differences emerged (all Pillai's Trace < 0.21, all $ps > .390$).

Participants either received two course credits or £30 for taking part in the study. They were also included in a draw to win one of three prizes (£50, £75, or £150). The experiment was approved by the standing ethics committee, and it complied with recognised ethics standards including the Declaration of Helsinki and the UK Integrity Office Code of Practice.

Participants were randomly allocated to the Sketch and Veracity conditions. The sample included 122 truth tellers and 121 lie tellers. Among truth tellers, 30 participants were allocated to the Sketch Week 1 or the Sketch Weeks 1 and 2 conditions, and 31 participants to the Sketch Week 2 or control conditions. Among lie tellers, 30 participants were allocated to the Sketch Week 1, Sketch Week 2, or Sketch Weeks 1 and 2 conditions, and 31 participants to the control condition.

To test the hypotheses, data were analysed using a 4 (Sketch: Week 1, Week 2, Weeks 1 and 2, control) × 2 (Veracity: truth teller, lie teller) between-subjects design with core details, peripheral details, common knowledge details, self-handicapping strategies, complications,

plausibility, proportion of complications, proportion of core details, and total details as dependent variables.

3.2. Materials

Before each interview at Times 1, 2, and 3, participants completed an automated pre-interview questionnaire that was created via Qualtrics software. In each pre-interview questionnaire, they rated on 7-point Likert scales ranging from 1 = *not at all* to 7 = *completely* (a) their motivation to appear convincing during the interview, and (b) the extent to which their preparation for the interview was (i) sufficient, (ii) good, and (iii) thorough. They were also asked whether or not they prepared any strategies for the interview and to write down the preparation strategy or the reason for not preparing one. In addition, at Time 1, participants were asked about their background characteristics (age, sex, and ethnicity), and at Times 2 and 3, they were asked about any preparations they made during the past week outside the experiment setting.

Participants also completed an automated post-interview questionnaire that was created via Qualtrics software after the interview at Time 3. In this questionnaire, participants rated on 7-point scales (1 = *not at all* to 7 = *completely*) their motivation, perceived believability by the interviewer, likelihood of writing a statement, and probability of winning the prize. Also, those who sketched rated the extent to which they thought the sketching task was difficult. As a manipulation check, participants rated on percentage scales (0–100%) the extent to which they were truthful in the three interviews. In addition, participants were asked open questions about their convincing strategy.

For the analyses, we calculated an average score for the three pre-interview motivation ratings and the post-interview motivation rating combined ($\alpha = 0.87$). We also calculated an average preparation score (preparation efficiency) for the three pre-interview ratings combined about whether participants' preparation was thorough, sufficient, and good ($\alpha = 0.93$).

3.3. Procedure

The procedure was adapted from Leal et al. (2018). After participants arrived for their appointment at the Department of Psychology at Time 1, they signed a consent form and were then randomly allocated to the truth teller or lie teller condition. Truth tellers were asked to think of a memorable event that happened to them in the past two years. The event should have been out of the ordinary that they (or others) do not experience daily. Truth tellers were given examples of such events (e.g., seeing a famous person when they were out for dinner, or going on a pheasant shoot). Lie tellers and truth tellers were matched for memorable events. That is, lie tellers were asked to lie about an event previously chosen by a truth teller and to pretend they experienced that event. Lie tellers were given information only about the main theme of the event, so they were not provided with specific details about the event. For example, if a truth teller reported about how she celebrated her 21st birthday by going for a bungee jump, the lie teller was instructed to lie about how she celebrated her 21st birthday. Thus, lie tellers were free to fabricate the details they wanted about the story. The Sketch condition was randomly allocated to truth tellers, but lie tellers were allocated to the same Sketch condition as the truth tellers with whom they were matched.

All participants were informed that they will be interviewed about the event in three interviews—each one week apart—and that they will need to convince the interviewer they are honest. To motivate participants, they were told that if they are convincing, their names will be entered in a draw to win £50, £75, or £150. However, if they are not convincing, they will be asked to write a statement about the event and their names will not be entered in the draw. At Times 2 and 3, participants were given the same motivating instructions again and told that their name will be entered in the draw for a second (or third) time if they

are convincing in the corresponding interview. If they are not convincing, they will have to write a statement. In reality, all participants were entered in the draw and none of them was asked to write a statement. All participants were given as much time as they need to prepare for the interview.

Before each interview, participants completed an automated pre-interview questionnaire. After that, they were taken to the interview room where they were interviewed by one of four research assistants (interviewers), blind to the participants' Veracity condition and to the study hypotheses. The same interviewer interviewed each participant at Times 1, 2, and 3. However, for reasons beyond our control, the same interviewer could not be available for eight participants, so a different interviewer took this role. Four one-way MANOVAs were conducted on the Time 3 details and the unique details to examine the effects of having the same or a different interviewer across interviews. For each of the Time 3 details and the unique details, one analysis included core details, peripheral details, common knowledge details, self-handicapping strategies, complications, and plausibility as dependent variables, and another analysis included proportion of complications, proportion of core details, and total details as dependent variables. No significant differences emerged (all Pillai's Trace < 0.02, all $ps > .682$). Thus, changing interviewers did not have any effect on the results.

The experimenter gave the interviewer a slip of paper on which the event, the event date, and the participant's Sketch condition were written. The theme of the event was kept general (e.g., tell me about celebrating your 21st birthday in May 2019). The interviewer's knowledge of the participant's Sketch condition was not expected to affect their demeanor, because the interviewer did not know the study hypotheses or the participant's Veracity condition. There is also no empirical evidence demonstrating an interviewer's effect when the interviewee is verbally reporting versus sketching an event.

For all participants, the interview started as follows:

Could you please tell me in as much detail as possible everything about that event, but before doing so I would like you to take a few moments to picture in your mind that memorable event and think about where you were and what you saw, heard, felt and smelled during that time. Take a moment to think about all your senses during the event and then please let me know when you have done that?

After indicating that they were ready, participants who were not assigned to sketch were asked, "Ok, now tell me in as much detail as possible everything you remember". Participants who were assigned to a sketch condition were asked, "Ok, now could you please draw for me exactly what you could see at that time and whilst doing so, please describe in as much detail as possible everything that you remember about that event".

Participants sketched while narrating at Time 1 if they were allocated to the Sketch Week 1 condition, at Time 2 if they were allocated to the Sketch Week 2 condition, and at Times 1 and 2 if they were allocated to the Sketch Weeks 1 and 2 condition. Participants in the Sketch control condition did not provide a sketch at all.

After the interviews at Times 1 and 2, participants were scheduled for the next session. After the Time 3 interview, they were asked to complete an automated post-interview questionnaire, and then they were fully debriefed, rewarded, and thanked.

3.4. Coding

All interviews were audiotaped, transcribed and coded for details. Every noun, verb, adjective, and adverb was considered a single detail, but other details such as conjunctions, prepositions, and pronouns were not counted because they are not precise and thus not informative. For example, the statement "The festival was in a field in Glastonbury" includes three details. A distinction was made between core and

peripheral details. A participant talking about going to a concert reported, "I went with my friends to the concert where it was pretty crowded" and "I sat in the terminal and waited for ages for the ferry". The first statement includes four core details central to the concert event, whereas the second statement includes five peripheral details happening after the concert and thus irrelevant to the concert event. Details that were repeated in a single interview were coded only once. Details mentioned in the second or third interview but not in a previous interview were coded as new details and classified into core or peripheral details.

Two coders—both blind to participants' veracity conditions—coded the transcripts independently for core and peripheral details. One coder coded all the transcripts, and the other coder coded 40 transcripts (16%). Inter-rater reliability analyses were computed using the Intra-Class Correlation (ICC) coefficient (single measures scores). ICC coefficients less than 0.40 are poor, between 0.40 and 0.59 are fair, between 0.60 and 0.74 are good, and between 0.75 and 1.0 are excellent (Hallgren, 2012). ICC was good for core details (ICC = 0.72) and peripheral details (ICC = 0.68).

Two other coders—blind to participants' veracity conditions—coded the transcripts independently for common knowledge details which are strongly invoked stereotypical knowledge about events ("The concert smelled of cigarettes and alcohol"), self-handicapping strategies which are explicit or implicit justifications as to why someone is not able to provide information ("We were lazy and we didn't want to go to touristy places"), complications which are occurrences that make a situation more complex than necessary ("I remember coming into LAX actually having a cross wind landing, it was some quite strong winds trying to come in the landing, that was fun"), and plausibility which was defined as how likely is it that the activities happened in the way described and it was measured on a 7-point scale (1 = implausible to 7 = plausible). One coder coded 70 transcripts (29%) and the other coder coded all the transcripts. Inter-rater reliability was excellent for common knowledge details (ICC = 0.91), self-handicapping strategies (ICC = 1.00), complications (ICC = 0.95), and plausibility (ICC = 0.81).

Based on the coded details, the total number of details and the proportion scores of (i) complications and (ii) core details were calculated for the Time 3 details and the unique details. The total number of details at Time 3 included the total number of core and peripheral details at Time 3. The proportion of complications at Time 3 was calculated as: complications at Time 3 / (complications + common knowledge details + self-handicapping strategies at Time 3). The proportion of core details at Time 3 was calculated as: core details at Time 3 / (core details + peripheral details at Time 3).

Unique total details were new core and peripheral details mentioned at Time 2 or Time 3 added to the core and peripheral details mentioned at Time 1. The proportion of unique complications was the total number of unique complications across the three interviews divided by the total number of unique complications, common knowledge details, and self-handicapping strategies across interviews. The proportion of unique core details was calculated by dividing the total number of unique core details across the three interviews by the total number of unique core and peripheral details across interviews. We also calculated unique plausibility by averaging plausibility scores at Times 1, 2, and 3.

Participants' open responses in the pre-interview and post-interview questionnaires regarding strategies used and reasons for (not) using a strategy were coded independently by the two coders who also coded core and peripheral details, both blind to the veracity conditions. One coder formulated categories based on participants' responses to the questions and coded all the responses. Similar responses were grouped together in a single category, and each category was labelled to describe the theme of responses. When the same response could fit in more than one category, it was allocated to those corresponding categories. To assess inter-rater reliability, a second coder also coded all participants' responses based on the corresponding categories. Disagreements were discussed and resolved. Inter-rater agreement was excellent, Cohen's κ

= 0.81. The strategies are discussed in the Results.

4. Results

4.1. Pre-interview and post-interview questionnaires

The results of responses to the questionnaires are included in the supplementary material. Compared to lie tellers, truth tellers were significantly more truthful, more likely to think they were believed by the interviewer, and less likely to believe they will have to write a statement. Truth tellers and lie tellers were equally and highly motivated, and they did not significantly differ in their perceptions on the likelihood of winning a prize, the difficulty of the sketching while narrating task, and the efficiency of their preparations. Significantly more lie tellers than truth tellers reported preparing (or using) a convincing strategy during the experimental sessions and prior to the third session. Truth tellers predominantly reported that they were truthful and detailed, whereas lie tellers' most frequently reported strategies were to use an embedded lie (a truthful account that includes fabricated details), remain consistent across interviews (e.g., stick to the same story), control nonverbal behaviour (e.g., not hesitate, stay calm, be confident), rehearse using external resources (e.g., google relevant details, write down notes), and manage information (e.g., include specific core, perceptual, verifiable or other details, add new details in subsequent interviews, justify/avoid/make mistakes, pretend to forget). Lie tellers who did not prepare preferred to provide spontaneous responses or felt confident enough to lie.

4.2. Hypotheses testing

We carried out multivariate analyses of variance (MANOVAs) on the Time 3 details and the unique details. To test the multicollinearity assumption of MANOVA we examined correlations between variables that are higher than 0.80 (Grewal et al., 2004). The correlations are tabulated in Tables 1 and 2. High correlations emerged between core details and total details for the Time 3 details and for unique details ($r = 0.98$), and between peripheral details and proportion of core details for the Time 3 details ($r = -0.82$) and for unique details ($r = -0.80$). To avoid including highly correlated variables in a single analysis, we ran two sets of analyses on each of the Time 3 details and the unique details, one with the proportion scores (proportion of complications and proportion of core details) and total details as dependent variables, and another one with the remaining cues as dependent variables (see Deeb, Vrij, & Leal, 2020 for similar analyses).

Table 1
Correlations between Time 3 details.

	Core details	Peripheral details	Common knowledge details	Self-handicapping strategies	Complications	Plausibility	Proportion of complications	Proportion of core details	Total details
Core details		0.18**	-0.10	-0.09	0.71***	0.43***	0.24***	0.17**	0.98***
Peripheral details	0.18**		-0.003	-0.04	0.17**	-0.04	0.16*	-0.82***	0.37***
Common knowledge details	-0.10	-0.003		0.34***	-0.14*	-0.33***	-0.35***	-0.03	-0.09
Self-handicapping strategies	-0.09	-0.04	0.34***		-0.11	-0.26***	-0.24***	-0.02	-0.10
Complications	0.71***	0.17**	-0.14*	-0.11		0.48***	0.45***	0.06	0.71***
Plausibility	0.43***	-0.04	-0.33***	-0.26***	0.48***		0.42***	0.22***	0.40***
Proportion of complications	0.24***	0.16*	-0.35***	-0.24***	0.45***	0.42***		-0.05	0.26***
Proportion of core details	0.17**	-0.82***	-0.03	-0.02	0.06	0.22***	-0.05		0.003
Total details	0.98***	0.37***	-0.09	-0.10	0.71***	0.40***	0.26***	0.003	

* $p < .05$.
 ** $p < .01$.
 *** $p < .001$.

We report below the results of the null hypotheses significance testing (NHST) along with Hedge's g which is the unbiased correction of Cohen's d effect size. An effect size of 0.2 is considered small, of 0.5 is considered medium, and of 0.8 is considered large. We also report Cohen's U_3 and the Probability of Superiority (PS). Cohen's U_3 (Cohen, 1988) represents the percentage of non-overlap between means (i.e., the percentage of a population A which the upper half of the cases of a population B exceeds). For example, if Cohen's $U_3 = 64\%$ then the upper half of the experimental group (population B) exceeds 64% of the control group (population A). The PS demonstrates the probability that a randomly selected person from one group has a higher observed score than a randomly selected person from the other group (Fritz et al., 2012; Lakens, 2013; Magnusson, 2020). To further corroborate our results, we carried out Bayesian analyses of variance that test the likelihood of the data under both the null hypothesis (H_0) and the alternative hypothesis (H_1). Bayes factors (BF_{10}) between 1 and 3 indicate weak evidence for the alternative hypothesis (H_1), between 3 and 20 indicate positive evidence, between 20 and 150 indicate strong evidence, and above 150 indicate very strong evidence (Jarosz & Wiley, 2014). The inverse of BF_{10} is BF_{01} ($1/BF_{10}$) which is the likelihood of supporting evidence for the null hypothesis (H_0) compared to the alternative hypothesis (H_1). We report below BF_{10} statistics only as BF_{01} can be inferred by inverting BF_{10} . Note that the prior probability distribution for each model was 0.20, a default uniform prior that is distributed equally across models (see Wagenmakers et al., 2018). We also report inclusion probabilities for matched models ($BF_{Inclusion}$) that compare all models with interaction effects to all models that include the same predictors but without the interaction effects (van den Bergh et al., 2020). We conducted the analyses for the NHST using SPSS statistical package and for the Bayesian testing using JASP software.

4.2.1. Time 3 details

A 4 (Sketch: Week 1, Week 2, Weeks 1 and 2, control) \times 2 (Veracity: lie teller, truth teller) MANOVA with core details, peripheral details, common knowledge details, self-handicapping strategies, complications, and plausibility at Time 3 as dependent variables, revealed a significant multivariate effect of Veracity, Pillai's Trace = 0.32, $F(6, 230) = 18.25, p < .001, \eta^2 = 0.32$. No significant effects emerged for Sketch, Pillai's Trace = 0.10, $F(18, 696) = 1.28, p = .191, \eta^2 = 0.03$, or the Sketch \times Veracity interaction effect, Pillai's Trace = 0.06, $F(18, 696) = 0.83, p = .668, \eta^2 = 0.02$. The Veracity effects revealed that at Time 3 (see Table 3), truth tellers reported more core details and complications and fewer common knowledge details and self-handicapping strategies than lie tellers. Also, truth tellers' statements sounded more

Table 2
Correlations between unique details.

	Core details	Peripheral details	Common knowledge details	Self-handicapping strategies	Complications	Plausibility	Proportion of complications	Proportion of core details	Total details
Core details		0.21**	0.04	-0.02	0.75***	0.43***	0.21**	0.19**	0.98***
Peripheral details	0.21**		0.12	0.05	0.18**	-0.04	0.11	-0.80***	0.41***
Common knowledge details	0.04	0.12		0.43***	-0.09	-0.38***	-0.47***	-0.06	0.06
Self-handicapping strategies	-0.02	0.05	0.43***		-0.07	-0.31***	-0.27***	-0.05	-0.01
Complications	0.75***	0.18**	-0.09	-0.07		0.51***	0.33***	0.06	0.74***
Plausibility	0.43***	-0.04	-0.38***	-0.31***	0.51***		0.48***	0.23***	0.39***
Proportion of complications	0.21**	0.11	-0.47***	-0.27***	0.33***	0.48***		-0.05	0.22**
Proportion of core details	0.19**	-0.80***	-0.06	-0.05	0.06	0.23***	-0.05		0.001
Total details	0.98***	0.41***	0.06	-0.01	0.74***	0.39***	0.22**	0.001	

** $p < .01$.
*** $p < .001$.

Table 3
Descriptive and inferential statistics for Time 3 details as a function of veracity.

Detail type	Truth tellers		Lie tellers		<i>F</i>	<i>p</i>	Hedge's <i>g</i> [95% CI]	Cohen's <i>U</i> ₃	<i>PS</i>	<i>BF</i> ₁₀	<i>BF</i> _{Inclusion} Veracity effect	<i>BF</i> _{Inclusion} Interaction effect
	<i>n</i> = 122 <i>M</i> (<i>SD</i>)	95% CI	<i>n</i> = 121 <i>M</i> (<i>SD</i>)	95% CI								
Core details	87.80 (77.56)	73.90, 101.71	53.52 (30.73)	47.99, 59.05	20.39	<.001	0.58 [0.32, 0.84]	71.90	65.90	1669.05	1681.35	0.08
Peripheral details	08.49 (12.84)	06.19, 10.79	08.69 (12.73)	06.40, 10.98	00.02	.886	0.02 [-0.24, 0.27]	50.80	50.60	00.14	00.14	0.14
Common knowledge details	00.16 (00.58)	00.06, 00.27	00.55 (01.11)	00.35, 00.75	11.49	.001	0.44 [0.18, 0.70]	67.00	62.20	33.80	33.79	0.08
Self-handicapping strategies	00.02 (00.13)	-0.01, 00.04	00.07 (00.29)	00.02, 00.13	04.05	.045	0.22 [-0.03, 0.48]	58.70	56.20	00.92	00.92	0.17
Complications	04.64 (05.96)	03.57, 05.71	01.55 (01.92)	01.21, 01.90	28.95	<.001	0.70 [0.44, 0.96]	75.80	69.00	83,379.75	83,675.94	0.05
Plausibility	05.20 (01.17)	04.99, 05.41	03.58 (01.23)	03.36, 03.80	108.62	<.001	1.34 [1.07, 1.63]	91.00	82.8	2.20 × 10 ¹⁹	2.211 × 10 ¹⁸	0.06
Proportion of complications	00.78 (00.39)	00.71, 00.85	00.54 (00.47)	00.46, 00.63	18.43	<.001	0.55 [0.30, 0.81]	70.90	65.10	723.78	723.59	0.11
Proportion of core details	00.91 (00.12)	00.89, 00.93	00.87 (00.16)	00.84, 00.90	05.05	.026	0.28 [0.03, 0.54]	61.00	57.80	01.24	01.55	0.23
Total details	96.30 (81.71)	81.65, 110.94	62.21 (34.43)	56.01, 68.40	17.66	<.001	0.54 [0.29, 0.80]	70.50	64.90	535.65	536.75	0.07

plausible than lie tellers' statements. Bayesian analyses revealed strong to very strong evidence for all these variables except for self-handicapping strategies which received weak evidence ($BF_{10} = 0.92$).

A 4 (Sketch: Week 1, Week 2, Weeks 1 and 2, control) × 2 (Veracity: lie teller, truth teller) MANOVA with proportion of complications, proportion of core details, and total details at Time 3 as dependent variables, revealed significant multivariate effects of Sketch, Pillai's Trace = 0.10, $F(9, 705) = 2.72, p = .004, \eta^2 = 0.03$, and Veracity, Pillai's Trace = 0.14, $F(3, 233) = 12.43, p < .001, \eta^2 = 0.14$, but the Sketch × Veracity interaction effect was not significant, Pillai's Trace = 0.04, $F(9, 705) = 0.99, p = .448, \eta^2 = 0.01$. There was a significant Sketch main effect for the proportion of core details, $F(3, 235) = 7.56, p < .001, \eta^2 = 0.09, BF_{10} = 209.14$. At Time 3, the Sketch control condition ($M = 0.83, SD = 0.19, 95\% CI [0.79, 0.88]$) showed a lower proportion of core details than the Sketch Week 1 condition ($M = 0.93, SD = 0.11, 95\% CI [0.90, 0.96]$), $p = .001$, Hedge's $g = 0.64, 95\% CI [0.28, 1.01]$, Cohen's $U_3 = 73.90, PS = 67.50$, and the Sketch Weeks 1 and 2 condition ($M = 0.94, SD = 0.11, 95\% CI [0.91, 0.96]$), $p < .001$, Hedge's $g = 0.70, 95\% CI [0.34, 1.08]$, Cohen's $U_3 = 75.80, PS = 69.00$. As for the Veracity effects at Time 3, Table 3 shows that the proportion scores of complications and core

details and total details were higher among truth tellers than lie tellers, but the evidence for the proportion of core details was weak ($BF_{10} = 1.24$).

4.2.2. Unique details

A 4 (Sketch: Week 1, Week 2, Weeks 1 and 2, control) × 2 (Veracity: lie teller, truth teller) MANOVA with unique core, peripheral, and common knowledge details, self-handicapping strategies, complications, and plausibility as dependent variables, revealed significant multivariate effects of Sketch, Pillai's Trace = 0.15, $F(18, 696) = 1.96, p = .010, \eta^2 = 0.05$, and Veracity, Pillai's Trace = 0.31, $F(6, 230) = 17.37, p < .001, \eta^2 = 0.31$. The Sketch × Veracity interaction effect was not significant, Pillai's Trace = 0.05, $F(18, 696) = 0.69, p = .822, \eta^2 = 0.02$. A significant Sketch main effect emerged for unique peripheral details, $F(3, 235) = 4.38, p = .005, \eta^2 = 0.05, BF_{10} = 5.14$. The Sketch Week 2 condition ($M = 21.61, SD = 26.14, 95\% CI [14.91, 28.30]$) elicited more peripheral details than the Sketch Weeks 1 and 2 condition ($M = 10.33, SD = 19.12, 95\% CI [5.39, 15.27]$), $p = .033$, Hedge's $g = 0.49, 95\% CI [0.13, 0.86]$, Cohen's $U_3 = 68.80, PS = 63.60$. As for veracity effects (see Table 4), positive to very strong evidence emerged that truth tellers'

Table 4
Descriptive and inferential statistics for unique details as a function of veracity.

Detail type	Truth tellers		Lie tellers		<i>F</i>	<i>p</i>	Hedge's <i>g</i> [95% CI]	Cohen's <i>U</i> ₃	<i>PS</i>	<i>BF</i> ₁₀	<i>BF</i> _{inclusion} Veracity effect	<i>BF</i> _{inclusion} Interaction effect
	<i>n</i> = 122 <i>M</i> (<i>SD</i>)	95% CI	<i>n</i> = 121 <i>M</i> (<i>SD</i>)	95% CI								
Unique core details	143.32 (122.38)	121.38, 165.25	91.01 (51.80)	81.68, 100.33	18.83	<.001	0.55 [0.30, 0.81]	70.90	65.10	789.18	815.27	0.07
Unique peripheral details	15.46 (22.98)	11.34, 19.58	16.23 (22.02)	12.27, 20.19	00.09	.771	0.03 [-0.22, 0.29]	51.20	50.80	00.15	00.15	0.09
Unique common knowledge details	00.42 (00.92)	00.25, 00.58	01.18 (01.82)	00.85, 01.51	17.23	<.001	0.53 [0.27, 0.78]	70.20	64.60	383.76	394.58	0.06
Unique self-handicapping strategies	00.02 (00.13)	-0.01, 00.04	00.17 (00.59)	00.07, 00.28	08.50	.004	0.35 [0.10, 0.61]	63.70	59.80	06.95	06.96	0.26
Unique complications	08.37 (11.43)	06.32, 10.42	02.81 (03.14)	02.24, 03.38	26.31	<.001	0.66 [0.40, 0.92]	74.50	68.00	25,104.27	25,148.55	0.06
Unique plausibility	04.95 (01.00)	04.77, 05.13	03.66 (00.97)	03.48, 03.83	103.92	<.001	1.30 [1.03, 1.59]	90.30	82.10	3.35 × 10 ¹⁸	3.377 × 10 ¹⁸	0.07
Proportion of unique complications	00.86 (00.28)	00.81, 00.91	00.62 (00.41)	00.54, 00.69	28.42	<.001	0.68 [0.42, 0.94]	75.20	68.50	63,602.36	63,624.37	0.06
Proportion of unique core details	00.90 (00.13)	00.88, 00.92	00.86 (00.16)	00.83, 00.89	05.49	.020	0.27 [0.02, 0.53]	60.60	57.60	01.39	01.87	0.21
Total unique details	158.78 (130.71)	135.35, 182.21	107.24 (58.95)	96.63, 117.85	15.48	<.001	0.51 [0.25, 0.76]	69.50	64.10	194.74	195.81	0.06

statements included more unique core details and complications and fewer unique common knowledge details and self-handicapping strategies than lie tellers' statements. Also, very strong evidence emerged that truth tellers' statements sounded more plausible than lie tellers' statements.

Another 4 (Sketch: Week 1, Week 2, Weeks 1 and 2, control) × 2 (Veracity: lie teller, truth teller) MANOVA was conducted with the proportion of unique complications, proportion of unique core details, and total unique details as dependent variables. The analysis revealed a significant main effect of Sketch, Pillai's Trace = 0.15, $F(9, 705) = 4.00$, $p < .001$, $\eta^2 = 0.05$, and Veracity, Pillai's Trace = 0.17, $F(3, 233) = 16.06$, $p < .001$, $\eta^2 = 0.17$. The Sketch × Veracity interaction effect was not significant, Pillai's Trace = 0.03, $F(9, 705) = 0.69$, $p = .716$, $\eta^2 = 0.01$. A significant Sketch main effect emerged for the proportion of unique core details, $F(3, 235) = 10.86$, $p < .001$, $\eta^2 = 0.12$, $BF_{10} = 10,893.99$. This proportion score was lower in the Sketch control condition ($M = 0.81$, $SD = 0.20$, 95% CI [0.76, 0.86]) than in the Sketch Week 1 condition ($M = 0.92$, $SD = 0.12$, 95% CI [0.89, 0.95]), $p < .001$, Hedge's $g = 0.66$, 95% CI [0.30, 1.03], Cohen's $U_3 = 74.50$, $PS = 68.00$, and in the Sketch Weeks 1 and 2 condition ($M = 0.93$, $SD = 0.10$, 95% CI [0.91, 0.96]), $p < .001$, Hedge's $g = 0.75$, 95% CI [0.39, 1.13], Cohen's $U_3 = 77.30$, $PS = 70.20$. The Veracity effects (Table 4) demonstrated that the proportions of unique complications and unique core details and total unique details were higher among truth tellers than lie tellers, but the proportion of unique core details received weak evidence ($BF_{10} = 1.39$).

Overall, the results for the Time 3 details and unique details were similar. In both analyses, Hypothesis 1 that predicted a Sketch main effect at Time 3 and for unique details was supported only for the proportion of core details, with peripheral details showing positive evidence for unique details only. Hypothesis 2 that predicted Veracity main effects at Time 3 and for unique details was supported for all details except for peripheral details (not significant and received weak evidence in both analyses) and for proportion of core details (significant in both analyses but received weak evidence). Self-handicapping strategies was a diagnostic Veracity cue in the unique details analysis but not in the Time 3 analysis (significant but weak evidence). Hypothesis 3 that predicted interaction effects was not supported.

4.3. Exploratory analyses

4.3.1. Time 1 details

The absence of Sketch related effects goes against the general trend found in the interviewing literature where sketching while narrating has been found to be beneficial for eliciting information and veracity cues. A possible explanation is that we analysed details provided at Time 3 and unique details provided throughout the three interviews rather than details provided at Time 1. Since to date all sketching deception research measured the effect of sketching at Time 1, we carried out exploratory analyses on the Time 1 details. We dichotomised the Sketch factor into Sketch absent versus Sketch present at Time 1.

A 2 (Sketch: absent vs present) × 2 (Veracity: truth teller, lie teller) MANOVA was carried out with core details, peripheral details, common knowledge details, self-handicapping strategies, complications, and plausibility at Time 1 as dependent variables. Significant multivariate effects emerged for Sketch, Pillai's Trace = 0.12, $F(6, 234) = 5.26$, $p < .001$, $\eta^2 = 0.12$, and Veracity, Pillai's Trace = 0.25, $F(6, 234) = 13.26$, $p < .001$, $\eta^2 = 0.25$. The Sketch × Veracity interaction effect was not significant, Pillai's Trace = 0.02, $F(6, 234) = 0.82$, $p = .555$, $\eta^2 = 0.02$. A significant Sketch main effect emerged for core details, $F(1, 239) = 13.70$, $p < .001$, $\eta^2 = 0.05$, $BF_{10} = 54.34$, and peripheral details, $F(1, 239) = 11.16$, $p = .001$, $\eta^2 = 0.05$, $BF_{10} = 26.23$. Participants who sketched at Time 1 provided more core details ($M = 89.27$, $SD = 61.28$, 95% CI [78.19, 100.34]) than those who did not sketch ($M = 64.11$, $SD = 47.49$, 95% CI [55.63, 72.58]), $p < .001$, Hedge's $g = 0.46$, 95% CI [0.20, 0.71], Cohen's $U_3 = 67.70$, $PS = 62.80$. Also, those who sketched at Time 1 provided fewer peripheral details ($M = 6.08$, $SD = 12.48$, 95% CI [3.82, 8.33]) than those who did not sketch ($M = 12.75$, $SD = 17.99$, 95% CI [9.54, 15.96]), $p = .001$, Hedge's $g = 0.43$, 95% CI [0.18, 0.69], Cohen's $U_3 = 66.60$, $PS = 61.90$. The Veracity effects (Table 5) demonstrated that at Time 1, truth tellers' statements included more core details and complications and fewer common knowledge details than lie tellers' statements. Truth tellers' statements also sounded more plausible than lie tellers' statements.

Another 2 (Sketch: absent vs present) × 2 (Veracity: truth teller, lie teller) MANOVA was carried out with proportion of complications, proportion of core details, and total details at Time 1 as dependent variables. Significant multivariate effects emerged for Sketch, Pillai's

Table 5
Descriptive and Inferential Statistics for Time 1 Details as a Function of Veracity.

Detail type	Truth tellers		Lie tellers		F	p	Hedge's g [95% CI]	Cohen's U ₃	PS	BF ₁₀	BF _{Inclusion} Veracity effect	BF _{Inclusion} Interaction effect
	n = 122 M (SD)	95% CI	n = 121 M (SD)	95% CI								
Core details	89.93 (67.26)	77.88, 101.99	63.02 (37.55)	56.26, 69.77	15.79	<.001	0.49 [0.24, 0.75]	68.80	63.60	132.83	202.12	0.29
Peripheral details	09.01 (16.85)	05.99, 12.03	09.90 (14.81)	07.24, 12.57	00.22	.642	0.06 [-0.20, 0.31]	52.40	51.70	00.15	00.15	0.21
Common knowledge details	00.25 (00.68)	00.12, 00.37	00.75 (01.19)	00.54, 00.97	16.55	<.001	0.51 [0.26, 0.77]	69.50	64.10	306.02	306.24	0.18
Self-handicapping strategies	00.02 (00.13)	-00.01, 00.04	00.09 (00.48)	00.01, 00.17	03.14	.078	0.21 [-0.04, 0.46]	58.30	55.90	00.61	00.61	0.22
Complications	04.48 (05.33)	03.52, 05.43	01.86 (02.24)	01.46, 02.26	25.45	<.001	0.64 [0.38, 0.90]	73.90	67.50	11,470.63	12,051.18	0.64
Plausibility	04.66 (00.89)	04.50, 04.82	03.73 (00.77)	03.59, 03.87	76.67	<.001	1.11 [0.85, 1.39]	86.70	78.40	8.320 × 10 ¹²	8.574 × 10 ¹²	0.97
Proportion of complications	00.79 (00.37)	00.73, 00.86	00.57 (00.44)	00.49, 00.65	18.82	<.001	0.54 [0.28, 0.80]	70.50	64.90	815.96	826.57	0.21
Proportion of core details	00.91 (00.14)	00.89, 00.94	00.88 (00.15)	00.85, 00.91	03.07	.081	0.21 [-0.05, 0.46]	58.30	55.90	00.50	00.60	0.18
Total details	98.94 (70.76)	86.26, 111.63	72.92 (42.66)	65.24, 80.60	12.41	.001	0.44 [0.19, 0.70]	67.00	62.20	37.69	42.07	0.25

Trace = 0.15, $F(3, 237) = 13.69, p < .001, \eta^2 = 0.15$, and Veracity, Pillai's Trace = 0.12, $F(3, 237) = 10.53, p < .001, \eta^2 = 0.12$. The Sketch × Veracity interaction effect was not significant,

Pillai's Trace = 0.004, $F(3, 237) = 0.34, p = .799, \eta^2 = 0.004$. A significant Sketch main effect emerged for total details, $F(1, 239) = 6.25, p = .013, \eta^2 = 0.03, BF_{10} = 2.26$, and proportion of core details, $F(1, 239) = 29.61, p < .001, \eta^2 = 0.11, BF_{10} = 83,721.79$. Bayesian analyses revealed weak evidence for total details ($BF_{10} = 2.26$). The proportion of core details at Time 1 was higher among those who sketched ($M = 0.94, SD = 0.10, 95\% CI [0.93, 0.96]$) than among those who did not sketch ($M = 0.85, SD = 0.17, 95\% CI [0.82, 0.88]$), $p < .001$, Hedge's $g = 0.64, 95\% CI [0.39, 0.90]$, Cohen's $U_3 = 73.90, PS = 67.50$. As for the Veracity main effect, truth tellers reported significantly more total details and a higher proportion of complications than lie tellers (see Table 5).

The results thus demonstrated that core details and peripheral details differed between the Sketching conditions and the Sketch control condition at Time 1 but no longer at Time 3. To understand why core and peripheral details stopped having an effect over time as a function of Sketch instruction, we ran a mixed MANOVA with Sketch (Week 1, Week 2, Weeks 1 and 2, control) as between-subjects factor, Time as within-subjects factor, and the number of core details and peripheral details in the three interviews as the dependent variable. Significant effects emerged for Sketch, Pillai's Trace = 0.09, $F(6, 478) = 3.79, p = .001, \eta^2 = 0.05$, for Time, Pillai's Trace = 0.11, $F(4, 236) = 7.38, p < .001, \eta^2 = 0.11$, and for Time × Sketch, Pillai's Trace = 0.15, $F(12, 714) = 3.05, p < .001, \eta^2 = 0.05$.

As we were interested in the interaction effect, we only report the results for the Time × Sketch effect. At the univariate level, there was an effect for core details, $F(4.99, 397.77) = 5.03, p < .001, \eta^2 = 0.06$, but not for peripheral details, $F(4.65, 370.72) = 1.00, p = .418, \eta^2 = 0.01$. Simple effects revealed that for the Sketch Week 1 condition, more core details were elicited at Time 1 ($M = 89.82, SD = 60.75, 95\% CI [74.12, 105.51]$) than at Time 2 ($M = 74.80, SD = 55.61, 95\% CI [60.43, 89.17]$), and at Time 3 ($M = 73.18, SD = 63.77, 95\% CI [56.71, 89.66]$), $p < .001$. For the Sketch Week 2 condition, more core details were elicited at Time 2 ($M = 79.00, SD = 48.09, 95\% CI [66.68, 91.32]$) than at Time 3 ($M = 71.13, SD = 44.83, 95\% CI [59.65, 82.61]$), $p = .005$. For the Sketch Weeks 1 and 2 condition, more core details were elicited at Time 1 ($M = 88.72, SD = 62.32, 95\% CI [72.62, 104.82]$), $p = .019$ and at Time 2 ($M = 89.15, SD = 71.10, 95\% CI [70.78, 107.52]$), $p = .001$

than at Time 3 ($M = 79.83, SD = 72.04, 95\% CI [61.22, 98.44]$). No differences emerged for the Sketch control condition. These results demonstrate that participants reported more core details when they were asked to sketch and narrate than when they were asked to provide a verbal statement, but this effect did not last beyond the interview in which participants were asked to sketch.

4.3.2. Diagnostic efficiency of the dependent variables

We ran diagnostic tests by calculating the area under the receiver operating characteristic curve (AUC) for the Time 1, Time 3, and unique data to understand the diagnostic efficiency of the verbal cues assessed in the current study. As peripheral, details, common knowledge details, and self-handicapping strategies were higher among lie tellers than truth tellers, we reversed the values of those cues by subtracting them from one point higher than the highest value of the corresponding cue. The results are summarised in Table 6. Values above 0.5 indicate that the cue is significantly diagnostic; and the higher the value, the more diagnostic the cue. All cues except for peripheral details and self-handicapping strategies, and to some extent proportion of core details, differed

Table 6
Area under the curve (AUC) for Time 1 details, Time 3 details, and unique details.

Detail type	Time 1 details [95% CI]	Time 3 details [95% CI]	Unique details [95% CI]
Core details	0.64 [0.57, 0.70]	0.67 [0.60, 0.74]	0.66 [0.59, 0.73]
Peripheral details	0.54 [0.47, 0.62]	0.51 [0.43, 0.58]	0.52 [0.45, 0.59]
Common knowledge details	0.62 [0.55, 0.69]	0.59 [0.52, 0.66]	0.62 [0.55, 0.69]
Self-handicapping strategies	0.52 [0.45, 0.59]	0.53 [0.45, 0.60]	0.55 [0.48, 0.63]
Complications	0.68 [0.62, 0.75]	0.70 [0.63, 0.76]	0.72 [0.66, 0.78]
Plausibility	0.77 [0.71, 0.83]	0.83 [0.78, 0.88]	0.83 [0.77, 0.88]
Proportion of complications	0.64 [0.58, 0.71]	0.64 [0.57, 0.71]	0.67 [0.60, 0.73]
Proportion of core details	0.57 [0.50, 0.64]	0.56 [0.49, 0.64]	0.58 [0.50, 0.65]
Total details	0.62 [0.55, 0.69]	0.65 [0.58, 0.72]	0.64 [0.57, 0.71]

significantly from chance (0.5). Plausibility showed the highest diagnosticity (average diagnosticity = 0.810) followed by complications, proportion of complications, core details, and total details (average diagnosticity = 0.70, 0.649, 0.653, and 0.635 respectively).

We further explored if looking at all cues combined would enhance diagnosticity by averaging all the cues for Time 1 details, Time 3 details, and unique details. The results of this analysis revealed that the AUC for the Time 1 details was 0.64, 95% CI [0.58, 0.71], for the Time 3 details was 0.68, 95% CI [0.61, 0.75], and for unique details was 0.67, 95% CI [0.60, 0.74].

5. Discussion

5.1. Veracity effects: robust verbal differences

We compared the verbal responses of truth tellers and lie tellers at three occasions: (a) Time 1, (b) Time 3 and (c) across the three interviews (unique details). The three comparisons showed strong overlap. In each comparison, truth tellers reported more core details, complications, and total details, fewer common knowledge details, and a higher proportion of complications than lie tellers. In addition, in each comparison truth tellers' stories sounded more plausible than lie tellers' stories. The ROC curve analyses provided further support for the overlap of the results. This overlap means that if, for whatever reason, investigators only have access to the third interview, they would not be in a disadvantageous position for assessing veracity. It also means that lie tellers did not become better at the task during the course of the multiple interviews. This can be explained by the strategies they reported to have used, as their two favourite strategies "remain consistent throughout the interviews" and "control nonverbal behaviour" do not address the veracity cues examined here. The results thus add to the robustness of the examined veracity cues. Previous research has shown them to be diagnostic in single interview settings (Vrij, Deeb, et al., 2020; Vrij, Leal, Mann, Fisher, et al., 2018), but they also appear to be diagnostic in multiple interview settings (also see also Deeb, Vrij, & Leal, 2020).

As in previous research (Deeb, Vrij, & Leal, 2020; Vrij, Leal, Jupe, & Harvey, 2018; Vrij & Vrij, 2020), complications and the proportion of complications received stronger evidence than the "total details" variable. The proportion variable takes the differences in speech content between truth tellers and lie tellers better into account than the total details variable. That is, lie tellers often report fewer details than truth tellers, which makes the total details variable a diagnostic veracity cue. However, truth tellers and lie tellers also differ in the types of detail they report: Complications (truth tellers) versus common knowledge details and self-handicapping strategies (lie tellers). These differences between truth tellers and lie tellers in the types of detail they report are taken into consideration in the proportion of complications variable but not in the total details variable, making the proportion of complications variable more diagnostic.

Out of the three variables—complications, common knowledge details and self-handicapping strategies—complications was the most diagnostic veracity cue. This reflects the verbal veracity literature (e.g., Masip et al., 2005) where it is generally found that cues to truthfulness (cues more frequently reported by truth tellers than lie tellers) are more diagnostic than cues to deceit (cues more frequently reported by lie tellers than truth tellers).

Nonetheless, it is beneficial to consider a mixture of cues to truthfulness and deceit than just examining cues to truthfulness—see the proportion of complications versus total details discussion above. The lack of cues to deceit was recently addressed as one of the important shortcomings in the verbal deception literature (Nahari et al., 2019) and researchers should continue their search for them.

Plausibility was the most diagnostic veracity cue examined. Similar findings emerged in previous research (Sporer et al., 2020; Vrij, Deeb, et al., 2020; Vrij, Leal, et al., 2020). Despite the strong potential of plausibility as a veracity cue, it is rarely examined by verbal deception

researchers. One problem is that plausibility is a subjective, qualitative cue that cannot be quantified in frequency of occurrence. Subjectivity indicates a lack of standardisation which is problematic when used in real life settings. However, as plausibility seems to discriminate truth tellers and lie tellers relatively well, we think that it deserves attention. It is reasonable to suggest that plausibility consists of a cluster of verbal cues and it is worthwhile to examine which cues these are. One obvious candidate is the context in which the information is presented, whereby abnormal or impossible activities are considered implausible (Blair et al., 2010). However, there are probably more cues to consider.

The proportion of core details was not highly diagnostic although it differed significantly from chance levels (see ROC analysis). We thus partially replicated the findings of the only other deception study that examined the proportion of core details in multiple interviews (Deeb, Vrij, & Leal, 2020). In that study strong evidence was found that truth tellers provided a higher proportion of core details than lie tellers. Our data do not allow us to explain the discrepancy in findings, even more so because Deeb, Vrij, and Leal (2020) used the same "describe a memorable event" scenario used in the current experiment. More research is needed to examine the potential of the proportion of core details as a veracity cue.

5.2. Sketch effects: the importance of details surrounding the centrality of the event

The sketching while narrating interview technique elicited a high proportion of core details at Times 1 and 3, and for unique details. This means that the Sketch instruction given at Time 1 made participants more focused on reporting core elements even at Time 3 when a sketch instruction was no longer present. Since investigators are typically mostly interested in core details, this can be considered a positive effect. To our knowledge this is the first experiment where a distinction was made between core and peripheral details in sketching while narrating. We think this distinction is worthwhile to examine in future studies.

More core details were elicited in interviews involving a sketch than in interviews involving a verbal statement, but this effect did not last beyond the interview in which the sketch was introduced. In other words, in terms of the number of core details reported, sketching did not have a spill-over effect to later interviews; the Sketch effect was only visible in the proportion of core details as discussed in the previous paragraph. Nonetheless, core details were not examined in previous sketching research, and this cue seems promising for the purpose of eliciting valuable information in single interviews.

Contrary to our expectations, we did not find differences between the Sketching and control conditions for complications, common knowledge details, and self-handicapping strategies. Previous studies on sketching while narrating in single interviews also failed to elicit some of these cues, particularly when they occurred infrequently (Vrij et al., 2019; Vrij, Mann, et al., 2020). As Tables 3 through 5 show, common knowledge details and self-handicapping strategies occurred infrequently. This may be caused by the type of deception scenario we used: To describe a unique event that was memorable. Lie tellers may find it difficult to rely on common knowledge when describing a unique event and may think that self-handicapping strategies are not suitable when they are supposed to describe a clearly memorable event.

5.3. Sketch × Veracity interaction effects: potential impact of the event and the interview technique

We did not find a Sketch × Veracity interaction effect, which means that, unlike what was hypothesised, sketching did not facilitate lie detection. This could be the result of a ceiling effect: Core details at Time 1 (Table 5) and Time 3 (Table 3) did not differ much for truth tellers, meaning that they did not have enough additional details to elaborate on their stories at Time 3 compared to Time 1. This will decrease the chance of an interaction effect to occur.

Alternatively, perhaps a selection of participants described an event for which sketching may be less suitable. Sketching is a visual output, so it may not be very beneficial in situations when interviewees do not describe visual experiences. Also, a sketch represents a specific moment in time, and it will be particularly useful for someone specifying a single event happening at a certain time (e.g., being in a restaurant) rather than for someone describing a sequence of events (e.g., a day long hiking trip). Most participants reported a visual experience in the interviews, but these were mostly sequences of events.

5.4. Methodological considerations

Participants reported positive events only as we did not want to induce stress in the experiment. Future research could examine if the same cues are elicited when negative events are reported in an ethically approved experiment. For instance, memory research has shown that negative events, compared to positive and neutral events, have differential effects on core and peripheral details (Herlihy et al., 2002; Herlihy et al., 2012). Truth tellers are less able to recall peripheral details concerning negative events than concerning positive events. Negative or positive valence should have no effect on lie tellers, and therefore veracity effects may be most diagnostic if negative events are reported due to lie tellers focusing more than truth tellers on peripheral details.

5.5. Conclusions

Although sketching while narrating did not facilitate lie detection, it was effective for making interviewees focus on core aspects of the reported event; an effect that carried over to later interviews where a sketch was no longer introduced. Veracity effects remained constant across multiple interviews demonstrating that access to an initial or a third interview is equally effective for detecting deception.

CRedit authorship contribution statement

Haneen Deeb: Conceptualisation, Methodology, Project administration, Software, Investigation, Supervision, Formal analysis, Data curation, Writing – Original Draft.

Aldert Vrij: Conceptualisation, Methodology, Formal analysis, Writing – Review & Editing, Supervision, Funding acquisition.

Sharon Leal: Conceptualisation, Methodology, Writing – Review & Editing.

Jennifer Burkhardt: Investigation, Formal analysis.

Declaration of competing interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.actpsy.2020.103236>.

References

- Alison, L., Alison, E., Noone, G., Elntib, S., Waring, S., & Christiansen, P. (2014). Whatever you say, say nothing: Individual differences in counter interrogation tactics amongst a field sample of right wing, AQ inspired and paramilitary terrorists. *Personality and Individual Differences*, 68, 170–175. <https://doi.org/10.1016/j.paid.2014.04.031>
- Amado, B. G., Arce, R., Fariña, 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>
- van den Bergh, D., van Doorn, J., Marsman, M., Draws, T., van Kesteren, E.-J., Derks, K., ... Wagenmakers, E.-J. (2020). A tutorial on conducting and interpreting a Bayesian ANOVA in JASP. *L'Année Psychologique*, 120(1), 73–96. <https://doi.org/10.3917/anpsy1.201.0073>
- Blair, J. P., Levine, T., & Shaw, A. (2010). Content in context improves deception detection accuracy. *Human Communication Research*, 36(3), 423–442. <https://doi.org/10.1111/j.1468-2958.2010.01382.x>
- Cohen, J. (Ed.). (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates.
- Dando, C., Wilcock, R., Milne, R., & Henry, L. (2009). A modified cognitive interview procedure for frontline police investigators. *Applied Cognitive Psychology*, 23(5), 698–716. <https://doi.org/10.1002/acp.1501>
- Deeb, H., Granhag, P. A., Vrij, A., Strömwall, L. A., Hope, L., & Mann, S. (2018). Visuospatial counter-interrogation strategies by liars familiar with the alibi setting. *Applied Cognitive Psychology*, 32(1), 105–116. <https://doi.org/10.1002/acp.3383>
- Deeb, H., Vrij, A., Hope, L., Mann, S., Granhag, P. A., & Lancaster, G. L. (2017). Suspects' consistency in statements concerning two events when different question formats are used. *Journal of Investigative Psychology and Offender Profiling*, 14(1), 74–87. <https://doi.org/10.1002/jip.1464>
- Deeb, H., Vrij, A., & Leal, S. (2020). The effects of a model statement on information elicitation and deception detection in multiple interviews. *Acta Psychologica*, 207, Article 103080. <https://doi.org/10.1016/j.actpsy.2020.103080>
- Deeb, H., Vrij, A., Leal, S., Verigin, B. L., & Kleinman, S. M. (2020). When and how are lies told? And the role of culture and intentions in intelligence-gathering interviews. *Legal and Criminological Psychology*. <https://doi.org/10.1111/lcrp.12171>. Advance online publication.
- DePaulo, B. M., Lindsay, J. L., 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>
- Eastwood, J., Snook, B., & Luther, K. (2018). Measuring the effectiveness of the sketch procedure for recalling details of a live interactive event. *Applied Cognitive Psychology*, 32(6), 747–754. <https://doi.org/10.1002/acp.3454>
- Evans, J. R., Pimentel, P. S., Pena, M. M., & Michael, S. W. (2017). The ability to detect false statements as a function of the type of statement and the language proficiency of the statement provider. *Psychology, Public Policy, and Law*, 23(3), 290–300. <https://doi.org/10.1037/law0000127>
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141(1), 2–18. <https://doi.org/10.1037/a0024338>
- Gabbert, F., Hope, L., & Fisher, R. P. (2009). Protecting eyewitness evidence: Examining the efficacy of a self-administered interview tool. *Law and Human Behavior*, 33, 298–307. <https://doi.org/10.1007/s10979-008-9146-8>
- 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. <https://doi.org/10.1080/10683160701645181>
- Granhag, P. A., & Hartwig, M. (2015). The Strategic Use of Evidence (SUE) technique: A conceptual overview. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Deception detection: Current challenges and new approaches* (pp. 231–251). Wiley.
- Granhag, P. A., & Strömwall, L. A. (1999). Repeated interrogations—Stretching the deception detection paradigm. *Expert Evidence*, 7, 163–174. <https://doi.org/10.1023/A:1008993326434>
- Grewal, R., Cote, J. A., & Baumgartner, H. (2004). Multicollinearity and measurement error in structural equation models: Implications for theory testing. *Marketing Science*, 23(4), 519–529. <https://doi.org/10.1287/mksc.1040.0070>
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23–34. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3402032/>.
- Hartwig, M., Granhag, P. A., & Strömwall, L. (2007). Guilty and innocent suspects' strategies during police interrogations. *Psychology, Crime, & Law*, 13(2), 213–227. <https://doi.org/10.1080/10683160600750264>
- Herlihy, J., Jobson, L., & Turner, S. (2012). Just tell us what happened to you: Autobiographical memory and seeking asylum. *Applied Cognitive Psychology*, 26(5), 661–676. <https://doi.org/10.1002/acp.2852>
- Herlihy, J., Scragg, P., & Turner, S. (2002). Discrepancies in autobiographical memories – Implications for the assessment of asylum seekers: Repeated interviews study. *British Medical Journal*, 324, 324–327. <https://doi.org/10.1136/bmj.324.7333.324>
- Heuer, F., & Reisberg, D. (1990). Vivid memories of emotional events: The accuracy of remembered minutiae. *Memory and Cognition*, 18, 496–506. <https://doi.org/10.3758/BF03198482>
- Home Office. (2020). Interviewing suspects (version 7). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/864940/interviewing-suspects-v7.0.pdf.
- Izovotas, A., Vrij, A., Hope, L., Mann, S., Granhag, P. A., & Strömwall, L. A. (2018). Facilitating memory-based lie detection in immediate and delayed interviewing: The role of mnemonics. *Applied Cognitive Psychology*, 32(5), 561–574. <https://doi.org/10.1002/acp.3435>
- 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>
- Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, 88(1), 67–85. <https://doi.org/10.1037/0033-295X.88.1.67>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, Article 863. <https://doi.org/10.3389/fpsyg.2013.00863>
- Leal, S., Vrij, A., Deeb, H., & Jupe, L. (2018). Using the model statement to elicit verbal differences between truth tellers and liars: The benefit of examining core and peripheral details. *Journal of Applied Research in Memory and Cognition*, 7(4), 610–617. <https://doi.org/10.1016/j.jarmac.2018.07.001>

- Leal, S., Vrij, A., Deeb, H., & Kamermans, K. (2019). Encouraging interviewees to say more and deception: The ghostwriter method. *Legal and Criminological Psychology, 24*(2), 273–287. <https://doi.org/10.1111/lcrp.12152>
- Leal, S., Vrij, A., Vernham, Z., Dalton, G., Jupe, L. M., Nahari, G., & Rozmann, N. (2019). Using the model statement to elicit verbal differences between truth tellers and liars amongst Arab interviewees: A partial replication of Leal, Vrij, Deeb, and Jupe (2018). *Applied Cognitive Psychology, 33*(6), 1008–1017. <https://doi.org/10.1002/acp.3536>
- Leins, D., Fisher, R. P., Pludwinsky, L., Robertson, B., & Mueller, D. H. (2014). Interview protocols to facilitate human intelligence sources' recollections of meetings. *Applied Cognitive Psychology, 28*(6), 926–935. <https://doi.org/10.1002/acp.3041>
- Magnusson, K. (2020). Interpreting Cohen's d effect size: An interactive visualization (version 2.1.1) [web app]. R Psychologist. <https://tpsychologist.com/d3/cohend/>
- Masip, J., Sporer, S. L., Garrido, E., & Herrero, C. (2005). The detection of deception with the reality monitoring approach: A review of the empirical evidence. *Psychology, Crime and Law, 11*(1), 99–122. <https://doi.org/10.1080/10683160410001726356>
- Nahari, G. (2019). Verifiability approach: Applications in different judgmental settings. In T. Docan-Morgan (Ed.), *The Palgrave handbook of deceptive communication* (pp. 213–225). Palgrave Macmillan.
- Nahari, G., Ashkenazi, T., Fisher, R. P., Granhag, P. A., Hershkowitz, L., Masip, J., ... Vrij, A. (2019). "Language of lies": Urgent issues and prospects in verbal lie detection research. *Legal and Criminological Psychology, 24*(1), 1–23. <https://doi.org/10.1111/lcrp.12148>
- Nahari, G., & Vrij, A. (2015). Systematic errors (biases) in applying verbal lie detection tools: Richness in detail as a test case. *Crime Psychology Review, 1*(1), 98–107. <https://doi.org/10.1080/23744006.2016.1158509>
- Sporer, S., Manzanero, A. L., & Masip, J. (2020). Optimizing CBCA and RM research: Recommendations for analyzing and reporting data on content cues to deception. *Psychology, Crime, & Law*. <https://doi.org/10.1080/1068316X.2020.1757097>. Advance online publication.
- Sporer, S. L., & Sharman, S. J. (2006). Should I believe this? Reality monitoring of accounts of self-experienced and invented recent and distant autobiographical events. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 20*(6), 837–854. <https://doi.org/10.1002/acp.1234>
- Vredeveltdt, A., van Koppen, P. J., & Granhag, P. A. (2014). The inconsistent suspect: A systematic review of different types of consistency in truth tellers and liars. In R. Bull (Ed.), *Investigative interviewing* (pp. 183–207). Springer.
- Vrij, A. (2005). Criteria-based content analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law, 11*(1), 3–41. <https://doi.org/10.1037/1076-8971.11.1.3>
- 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., Deeb, H., Leal, S., Granhag, P. A., & Fisher, R. P. (2020). *Plausibility: A verbal cue to veracity worth examining?*. Manuscript submitted for publication. Department of Psychology, University of Portsmouth.
- 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. <https://doi.org/10.1111/lcrp.12088>
- Vrij, A., Leal, S., & Fisher, R. P. (2018). Verbal deception and the model statement as a lie detection tool. *Frontiers in Psychiatry, 9*, Article 492. <https://doi.org/10.3389/fpsy.2018.00492>
- Vrij, A., Leal, S., Fisher, R. P., Mann, S., Dalton, G., Jo, E., ... Houston, K. (2018). Sketching as a technique to eliciting information and cues to deceit in interpreter-based interviews. *Journal of Applied Research in Memory and Cognition, 7*(2), 303–313. <https://doi.org/10.1016/j.jarmac.2017.11.001>
- Vrij, A., Leal, S., Fisher, R. P., Mann, S., Deeb, H., Jo, E., ... Hamzeh, S. (2020). The efficacy of using countermeasures in a model statement interview. *The European Journal of Psychology Applied to Legal Context, 12*(1), 23–34. <https://doi.org/10.5093/ejpalc2020a3>
- Vrij, A., Leal, S., Fisher, R. P., Mann, S., Jo, E., Shaboltas, A., ... Houston, K. (2019). Eliciting information and cues to deceit through sketching in interpreter-based interviews. *Applied Cognitive Psychology, 33*(6), 1197–1211. <https://doi.org/10.1002/acp.3566>
- 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. <https://doi.org/10.1111/lcrp.12126>
- Vrij, A., Leal, S., Mann, S., Fisher, R. P., Dalton, G., Jo, E., ... Houston, K. (2018). Using unexpected questions to elicit information and cues to deceit in interpreter-based interviews. *Applied Cognitive Psychology, 32*(1), 94–104. <https://doi.org/10.1002/acp.3382>
- Vrij, A., Mann, S., Leal, S., Fisher, R. P., & Deeb, H. (2020). Sketching while narrating as a tool to detect deceit. *Applied Cognitive Psychology, 34*(3), 628–642. <https://doi.org/10.1002/acp.3646>
- Vrij, A., & Vrij, S. (2020). Complications travel: A cross-cultural comparison of the proportion of complication as a verbal cue to deceit. *Journal of Investigative Psychology and Offender Profiling, 17*(1), 3–16. <https://doi.org/10.1002/jip.1538>
- 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. <https://doi.org/10.3758/s13423-017-1323-7>