



Detecting Deception through Telephone Sketches in an Insurance Setting

Aldert Vrij* ; Haneen Deeb; Sharon Leal

Department of Psychology, University of Portsmouth, UK.

*Corresponding Author(s): Aldert Vrij

Department of Psychology, University of Portsmouth,
King Henry Building, King Henry 1 Street, PO1 2DY,
Hants, United Kingdom.
Email: aldert.vrij@port.ac.uk

Abstract

In the present experiment, carried out via Zoom, we examined the effect of (i) sketching and (ii) having the video turned on or off on the verbal accounts of truth tellers and lie tellers. Truth tellers reported an incident of loss, theft or accidental damage they experienced in the last 24 months, whereas lie tellers made up an incident. Half of the participants were asked to sketch the event whilst reporting it, whereas the other half did not sketch. Half of the participants were interviewed with the video turned on and the other half with the video turned off. We measured, total details, complications, plausibility and 'it could have been worse' (playing down the incident). We expected truth tellers to provide more total details, more complications, more 'It could have been worse' comments and more plausible statements than lie tellers, particularly when they sketch with the video turned off. Truth tellers provided more complications, more plausible statements and, particularly, more 'It could have been worse' comments than lie tellers. The sketching and video factors yielded no effects.

Received: Jun 28, 2022

Accepted: Jul 20, 2022

Published Online: Jul 22, 2022

Journal: Annals of Forensic Science and Research

Publisher: MedDocs Publishers LLC

Online edition: <http://meddocsonline.org/>

Copyright: © Vrij A (2022). *This Article is distributed under the terms of Creative Commons Attribution 4.0 International License*

Keywords: Deception; Online interviews; Sketching; Video turned off or on; Insurance claims; Complications; Plausibility; 'It could have been worse' comments.

Detecting deception through telephone sketches in an insurance setting

Current technology enables interviews to more frequently be held remotely. During the COVID-19 related lockdown measures, this facility has probably proved to be lifesaving. Moving forwards, many organisations and individuals are likely to choose to continue using such remote modes of communication to a greater degree, lessening the need for travel to a greater extent than before, even when restrictions are removed (<https://www.theguardian.com/technology/2020/may/12/twitter-coronavirus-covid19-work-from-home>). In this experiment we will manipulate the use of sketching in online inter-

views to elicit cues to deception and examine the effect on such cues of having the video turned on or off.

Sketching while narrating (that is, reporting an event and sketching the reported event at the same time) facilitates verbal recall in truth tellers [1,2,3]. The literature provides five reasons for this, summarised by [4]: (i) sketching mentally reinstates the context of the interviewee's experience; (ii) sketching one aspect of an event may cue retrieval of other aspects of that event; (iii) sketching is a visual output compatible with visually experienced events; (iv) sketching is a time-consuming activity that slows down the thinking process and thus gives interviewees good opportunity to search their memory; and (v) sketch-

Cite this article: Vrij A, Deeb H, Leal S. Detecting Deception through Telephone Sketches in an Insurance Setting. *Ann Anat Res.* 2022; 1(1): 1003.



ing automatically leads to providing spatial information because someone must situate each person or object in a specific location in the sketch.

Research has shown that sketching while narrating results in more pronounced verbal differences between truth tellers and lie tellers than narrating without sketching [4,5]. A truth teller's memory of a truly experienced event is likely to be richer in detail than a lie teller's memory of a fabricated event and a richer memory enhances the sketching effect. In most experiments to date the sketching instruction was introduced in face-to-face interviews [6], but in the present experiment it was introduced in an online setting.

The interviews in the current online experiment were held with the video being turned on or off. When the video is turned on, the interviewee and interviewer can see each other, similar to a face-to-face interview. We therefore expected to replicate the findings found in face-to-face interviews. When the video is turned off, the interviewer and interviewee cannot see each other and the interviewer cannot check whether the interviewee complies with the instruction to sketch. This inability to check the interviewee may have a differential effect on truth tellers and lie tellers. Lie tellers are inclined to be less cooperative than truth tellers [7,8], an inclination that may become particularly activated when the interviewer cannot check an interviewee's level of cooperation. This may result in lie tellers putting less effort in their sketching than truth tellers when the interviewer cannot see them. Since sketching facilitates reporting information, the difference between truth tellers and lie tellers in reporting information should become most pronounced in the sketch condition with the video function turned off.

Four verbal cues

We examined four verbal cues. Two of them, total details and complications, have been examined extensively (details) and rather extensively (complications) before. Researchers seem reluctant to examine the third cue, plausibility [9], but it has shown to have good potential as a veracity indicator. We are not aware of research examining the fourth cue, 'it could have been worse', but we believe this cue to have great potential in insurance claim settings, the deception scenario we used in the current experiment.

A detail refers to a unit of perceptual information (what the interviewee saw, heard, smelled etc.). Truth tellers typically report more details than lie tellers [10,7,11]. Two reasons are thought to cause this effect. First, truth tellers are more willing to provide details than lie tellers. This is reflected in the strategies truth tellers and lie tellers report to have used in interviews to sound credible and to avoid detection. A prominent strategy amongst truth tellers is to 'tell it all', whereas a prominent strategy amongst lie tellers is to 'keep it simple' [12,13]. Lie tellers may be afraid that the details they provide will incriminate them if checked by the investigator [14] or that they will forget those details when interviewed again about the event later. Second, truth tellers are more able than lie tellers to provide details because lie tellers lack the imagination to fabricate enough information that sounds plausible [15].

Complications are clusters of details that make the story more complex (e.g., "Initially we did not see our friend, as he was waiting at a different entrance"). Truth tellers typically report more complications than lie tellers [16]. Complications are typically not key elements of an experience, and an experience

is typically well understood if the complications are left out. Not reporting complications therefore aligns with the 'keep it simple' strategy and reporting complications aligns with the 'tell it all' strategy. Lie tellers may also leave out reporting complications because they think that such complications make a story sound less believable [17].

Plausibility addresses the question *how likely is it that the activities happened in the way described* [18]. For plausibility, context is important [19]. Events typically happen in certain ways and deviation from these norms will make a statement implausible. For example, claiming to have visited many landmarks in London in one day (Windsor Castle, Buckingham Palace, Madame Tussauds and Tate Gallery) sounds implausible because the interviewee will lack the time to have done this. Apart from context, plausibility is positively correlated with reporting details and complications: The more details and complications someone reports, the more plausible a story sounds [9].

It could have been worse means that the interviewee plays down the experience. In an insurance context it refers to playing down the effects of what is lost, stolen or damaged. For example, one interviewee who reported that her disabled son tried to climb on top of the TV after which both son and TV fell on the floor, said that she was relieved that her son did not get hurt in the incident. We expect truth tellers more than lie tellers to make more of these 'it could have been worse' comments. Truth tellers may feel embarrassed or annoyed with themselves by their experience of loss, theft or accidental damage because they may feel some responsibility for the event. Playing down the seriousness of the event would then be a good way to cope with it. In contrast, lie tellers who make a false claim may wish to get the most out of it. Making playing down comments may not work in their advantage to achieve this goal.

Hypotheses

We tested the following four pre-registered hypotheses (<https://osf.io/vs2hx/>), albeit that the pre-registered hypotheses referred to details and complications only.

Hypothesis 1: Truth tellers will provide more total details, more complications, more 'It could have been worse' comments and more plausible statements than lie tellers (Veracity main effect).

Hypothesis 2: Participants in the sketch condition will provide more total details, more complications, and more plausible statements than participants in the no sketch condition (Sketch main effect).

Hypothesis 3: Participants in the video enabled condition will provide more total details, more complications, and more plausible statements than participants in the video disabled condition (Video main effect).

Hypothesis 4: The most profound differences between truth tellers and lie tellers will be evident when they sketch in the video turned off condition (Veracity X Sketch X Video interaction effect).

Method

Ethics

A favourable ethical review decision was given by the relevant ethics committee of the university.

Participants

A G*Power analysis revealed that at least 197 participants are required for the experiment to have high statistical power (.97), alpha level of 0.05, and a medium effect size ($f^2 = .06$). A total of 199 participants were initially recruited but the data of two participants were deleted due to failing to follow instructions or due to an interviewer error. A total of 197 participants took part, 48 were males and 149 were females. Their average age was $M = 24.81$ ($SD = 8.48$). Most participants ($n = 44$) identified themselves as white British, followed by Asian ($n = 39$), white ($n = 37$), European ($n = 30$), British without mentioning black or white ($n = 14$), Black British ($n = 11$), African ($n = 11$), mixed ($n = 5$), Arab ($n = 3$) and South American ($n = 3$).

Allocation to the Veracity factor was determined by having experienced loss, theft or damage in the last 24 months (truth tellers, $n = 97$) or not (lie tellers, $n = 100$). Allocation to the sketch and video conditions occurred randomly. A total of 99 participants were allocated to the sketch-present condition and 98 participants to the sketch-absent condition. A total of 99 participants were allocated to the video turned on condition and the remaining 98 participants to the video turned off condition. Cell sizes varied from 23 to 25.

Design

Data were analysed utilising a Veracity (truth vs lie) X Sketch (present vs absent) X Video (turned on vs turned off) design. Dependent variables were total details, complications, plausibility and 'it could have been worse'.

Procedure

The experiment was advertised as an insurance deception study and that we were looking for people who have or have not experienced loss, theft or accidental damage in the last 24 months. Those who have experienced such an event were allocated to the truth teller condition and the others to the lie teller condition. Allocation to conditions took place at least 24 hours before the experiment started to give participants the opportunity to prepare themselves for the interview. The type of claim lie tellers were asked to fabricate matched the truth tellers' claims, but we left it up to the lie teller how to phrase their claim. For example, if a truth teller reported that a telephone had been stolen on a train, a lie teller was told to discuss a stolen phone. The participant information sheet and consent form were emailed to the participants at least 24 hours prior to their appointment and they had the opportunity to ask any questions about the experiment prior to and during their appointment.

The experiment was conducted via Zoom. At the beginning of the appointment participants were reminded of their veracity condition and of the event they were asked to report about. To motivate them to appear convincing, the experimenter told them that their name would be added to a draw with three prizes (Amazon vouchers worth £150, £100 and £50) if the interviewer believes their account. If the interviewer does not believe their account, participants were told that they must write a report of their claim at the end of the experiment.

Pre-Interview Questionnaire

When participants told the experimenter to be ready for the interview, they completed a pre-interview questionnaire. It measured background characteristics (age, gender, ethnicity), motivation, preparation thoroughness and preparation time. Participants were asked how motivated they were to perform

well during the interview on a 5-point scale ranging from 1 (*not at all motivated*) to 5 (*very motivated*). Participants then indicated their preparation thoroughness via three items (1 = *shallow* to 7 = *thorough*; 1 = *insufficient* to 7 = *sufficient*; and 1 = *poor* to 7 = *good*). The answers to the three questions were averaged (Cronbach's alpha = .92). Finally, one question was asked for preparation time: 'Do you think the amount of time you were given to prepare was' (1 = *insufficient* to 7 = *sufficient*).

The Interview

Participants were then interviewed in a breakout room by an interviewer (research assistant) who was blind to their veracity condition. The interviewer disabled the video feature for half of the participants and enabled it for the other half. Participants in the video turned on condition were told to treat the interview as a face-to-face interview, and those in the video turned off condition were told to treat it as a telephone interview.

The interviewer started the interview as follows: "OK, just so I can understand, I am going to need you to take me back to the day of the incident when you [Event], and tell me in as much detail as possible everything that happened from just before the incident occurred until after the incident had finished, and to describe your reactions to it"? We label this part of the interview Phase 1. After responding to the first question, all participants received the following instructions: "I would now like you to go back in your memory to the day of the incident. Please think about the incident and recall what you could see, what you could hear, what you could feel and what you could smell and let me know when you have done that"? Once ready, participants in the no sketch condition were instructed as follows: "Now please think about what happened from just before the incident occurred until after the incident had finished and whilst doing so, talk me through everything you experienced". Participants in the sketch condition were asked to prepare a piece of paper and pencil/pen and were instructed as follows: "Now please draw for me what happened from just before the incident occurred until after the incident had finished and whilst doing so, talk me through everything you experienced". We label this part of the interview Phase 2. After the interview, participants in the sketch condition were asked to take a photo of their sketch and to email it to the interviewer.

Post-Interview Questionnaire and Debrief

All participants completed a post-interview questionnaire on Qualtrics measuring rapport with the interviewer, percentage of lie telling/truth telling and (in the sketch present condition only) their expectation that they would be asked to send their sketch, and effort spent on sketching.

We measured rapport with the interviewer, because it is an important motivator for a productive interview [20]. It was measured via the nine-item Interaction Questionnaire [21]. It contains items such as smooth, bored, engrossed and involved, using 7-point scales ranging from 1 (*not at all*) to 7 (*extremely*), Cronbach's alpha = .84. Participants also rated the extent to which they told the truth in the interview (on an 11-point Likert scale ranging from 0% to 100%). Participants in the sketch condition were finally asked the extent to which they expected to be asked to send their sketch (on a 7-point Likert scale from 1 = *not at all* to 7 = *very much so*), and the effort spent on sketching (on a 7-point Likert scale from 1 = *not at all* to 7 = *very much so*).

After completing the questionnaire, participants were sent the debrief form. They were told by the experimenter that they

were believed by the interviewer and were therefore eligible for a £10 payment or one course credit and entry into a draw for Amazon vouchers as thanks for their contribution.

Coding

The interview recordings were transcribed and the transcripts were used for coding. One rater, blind to the Veracity condition and hypotheses, was taught the coding scheme by the first author who has more than 25 years of experience in coding verbal indicators to deception.

A *detail* is defined as a non-redundant unit of information that describes the interviewee’s experiences. For example, the following sentence contains five details: “So erm I decided to obviously take my umbrella with me to the bus stop, it’s erm about 15 minutes-walk away so I didn’t want to get wet”. In Phase 2 only new details (details not reported in Phase 1) were coded.

A *complication* is an occurrence that affects the storyteller and makes a situation more complex [9]. Again, in Phase 2 only new complications were coded. The following sentence has two complications: “When we returned to the restaurant it was already closed and there was a security guard there, but he said that he didn’t have a key to the restaurant”.

Plausibility was defined as *How likely is it that the activities happened in the way described* [18]. Plausibility was rated on a 7-point scale from 1 = *not at all plausible* to 7 = *very plausible*. For example, the following statement was considered plausible (score of 6). The statement includes contextual information (when it happened) and details about the accident, its consequences, and the aftermath.

From what I remember it was a weekday in October of 2019. I just came back from my squash practice in the evening and I got the stuff out of my bag and I left my water bottle on my desk. I went to get my computer to start my assignment so got my water bottle, I had a sip and I left it without realising I left it open on the side. I picked up my chair, tucked it in and desk tilted and the water bottle fell over, covering my keyboard in water. After tapping on the keyboard the computer completely switched off. The screen went black it wouldn’t charge when turned on. It was very upsetting to me at the time because I’d worked the previous summer over 5 or 6 months I saved up to buy this computer and now it was completely ruined, so it was very frustrating for me as well, because I couldn’t get any of my university work done. I had to borrow a laptop so it was quite an ordeal for myself.

The following statement was considered implausible (score of 2). It lacks detail and the action (getting out of the car to see her mum seems unlikely in this situation).

It was last year. I went with my mum to the supermarket. I decided to stay in the car when my mum went to the supermarket so I was alone in the car. Then a thief came to the car and hit the window. I then went out and I started to run to the supermarket and went to my mum. I said to her “Someone tried to steal the car”. Then we went back to the car and mum’s bag wasn’t there.

It could have been worse was defined as a comment from the interviewee that played down the event. Examples are: (i) I think my girlfriend was at work but she was coming back in about an hour or two, so I was only going to be locked out for a couple of hours; (ii) “I fortunately saved all of my important files on Google Drive” and (iii) “Luckily I took out my purse, and my phone before my bag was stolen”.

A second rater coded a random sample of 73 transcripts for total details, complications and plausibility and, since the variable is new, all transcripts for the ‘it could have been worse’ variable. Inter-rater reliability between the two coders, using the two-way random effects model measuring consistency, was good for total details (Single Measures ICC = .72), complications (Single Measures ICC = .94), plausibility (Single Measures ICC = .64) and ‘it could have been worse’ (Average Measures ICC = .91). For ‘it could have been worse’, the first rater looked at the disagreements and decided whether or not the comment qualified as an ‘it could have been worse’ comment.

Results

Questionnaire Variables

A 2 (Veracity: Truth vs lie) X 2 (Sketch: Present vs absent) X 2 Video (turned on vs turned off) MANOVA was carried out with the following five questionnaire variables as dependent variables: Motivation, preparation thoroughness, preparation time, rapport and percentage truth telling. At a multivariate level the analysis revealed a significant main effect for Veracity, $F(5, 185) = 76.53, p < .001, \eta_p^2 = .67$. All other effects were not significant, all $F_s < 1.63$, all $p_s > .154$.

The univariate Veracity results are presented in Table 1. Truth tellers reported their rapport with the interviewer to be better than lie tellers and truth tellers also reported to have told the truth more than lie tellers. Although lie tellers rated the preparation time given to them to be better than truth tellers, the Bayes analyses showed no support for this significant effect.

The mean scores for the total sample showed that motivation to perform well in the experiment (measured on a 5-point Likert scale) was very high amongst participants ($M = 4.21, SD = 0.82$). Participants were satisfied with their preparation thoroughness ($M = 5.08, SD = 1.31$) and preparation time ($M = 6.12, SD = 1.25$) and that they had good rapport with the interviewer ($M = 5.47, SD = 0.99$). The latter three variables were all measured on 7-point Likert scales.

Table 1: Questionnaire Variables and Verbal Cues as a Function of Veracity.

	Truth			Lie			NHST			BF ₁₀
	M	(SD)	95% CI	M	(SD)	95% CI	F	p	d (95% CI)	
Questionnaire variables										
Motivation	4.30	0.86	4.14,4.46	4.12	0.79	3.96,4.28	2.45	.119	0.22 (-0.07,0.49)	0.46
Preparation thoroughness	5.10	1.30	4.83,5.36	5.06	1.33	4.80,5.32	0.04	.837	0.03 (-0.25,0.31)	0.16
Preparation time	5.93	1.43	5.68,6.18	6.31	1.02	6.06,6.56	4.56	.034	0.31 (0.02,0.58)	1.36

Rapport	5.68	0.97	5.48,5.87	5.27	0.96	5.08,5.46	8.72	.004	0.42 (0.14,0.70)	9.47
Percentage truth telling	94.75	17.55	89.96,99.69	29.37	29.06	24.58,34.16	357.63	< .001	2.71 (2.29,3.06)	5.453 × 10 ⁴²
Verbal cues Phase 1										
Total details	21.58	(10.22)	19.69,23.50	21.54	(8.66)	19.67,23.41	0.002	.968	0.00 (-0.28,0.28)	0.16
Complications	4.44	(3.01)	3.83,5.04	3.39	(3.03)	2.80,3.98	6.00	.015	0.35 (0.06,0.62)	2.48
Plausibility	4.44	(0.97)	4.26,4.62	4.09	(0.85)	3.91,4.27	7.35	.007	0.38 (0.10,0.66)	4.72
It could have been worse	0.55	(0.61)	0.44,0.65	0.27	(0.47)	0.16,0.38	12.52	<.001	0.52 (0.22,0.79)	51.76
New Verbal cues Phase 2										
Total details	6.05	(5.32)	5.03,7.12	5.85	(5.20)	4.85,6.92	0.09	0.77	0.04 (-0.24,0.32)	0.16
Complications	1.26	(1.53)	0.98,1.53	0.78	(1.25)	0.52,1.07	5.80	.017	0.34 (0.06,0.62)	2.25
Plausibility	4.60	(1.29)	4.36,4.84	4.26	(1.09)	4.04,4.51	3.98	.047	0.29 (0.00,0.56)	0.98
It could have been worse	0.04	(0.20)	0.01,0.07	0.01	(0.10)	-0.02,0.04	1.93	.166	0.19 (-0.09,0.47)	0.38

Note: NHST = Null-Hypothesis Significance Testing.

Participants in the sketch present condition were asked the extent to which they expected a screenshot of their sketch to be taken and the effort they put in their sketching. A 2 (Veracity) X 2 (Video) MANOVA with those two variables as dependent variables did not reveal a multivariate significant effect, all *F*s < 2.99, all *p*s > .054. The average means showed that the screenshot was to some extent expected (*M* = 4.19, *SD* = 2.46) and that the participants put a limited amount of effort in their sketches (*M* = 3.64, *SD* = 1.77).

Hypotheses-Testing

To test our hypotheses, we carried out frequentist analyses and Bayesian analyses [22]. Bayesian analyses test the likelihood of the data under both the Null Hypothesis (H0) and the Alternative Hypothesis (H1). Bayes Factors (*BF*₁₀) between 1 and 3 indicate weak evidence for the Alternative Hypothesis (H1), between 3 and 20 indicate positive evidence, between 20 and 150 indicate strong evidence, and above 150 indicate very strong evidence [22]. A Bayes Factor close to 1 means no evidence can be derived from the data for either the null or the Alternative Hypothesis. The inverse of *BF*₁₀ is *BF*₀₁ (1/*BF*₁₀) which

is the likelihood of supporting evidence for the Null Hypothesis (H0) compared to the Alternative Hypothesis (H1). We report only *BF*₁₀ statistics in Tables 1 and 2 because *BF*₀₁ can be inferred by inverting *BF*₁₀.

Hypotheses-Testing: Phase 1

A 2 (Veracity: Truth vs lie) X 2 (Sketch: Present vs absent) X 2 (Video: Turned on vs turned off) MANOVA was carried out on the Phase 1 data with total details, complications, plausibility and ‘it could have been worse’ as dependent variables. At a multivariate level only the Veracity main effect was significant, *F*(4, 185) = 5.14, *p* < .001, *η*_{*p*}² = .10. All other *F*s < 2.11, all *p*s > .080. The univariate effects are presented in Table 1. Truth tellers provided significantly more complications, more ‘it could have been worse’ comments and more plausible statements than lie tellers. The Bayesian analysis showed strong evidence for ‘it could have been worse’, positive evidence for plausibility and weak evidence for complications. This supports Hypothesis 1 except for total details. The absence of a Video effect means that Hypothesis 2 was not supported. The Bayes Factors, many of them close to zero, show that there was no difference between the two video conditions.

Table 2: Verbal Cues as a Function of Video.

	Video turned off			Video turned on			NHST			<i>BF</i> ₁₀
	<i>M</i>	(<i>SD</i>)	95% CI	<i>M</i>	(<i>SD</i>)	95% CI	<i>F</i>	<i>p</i>	<i>d</i> (95% CI)	
Verbal cues Phase 1										
Total details	22.33	(10.14)	20.45,24.24	20.80	(8.67)	18.91,22.68	1.31	.254	0.16 (-0.12,0.44)	0.28
Complications	4.35	(3.60)	3.75,4.94	3.47	(2.35)	2.88,4.07	4.15	.043	0.29 (0.00,0.57)	1.03
Plausibility	4.37	(0.92)	4.19,4.55	4.16	(0.92)	3.98,4.34	2.50	.115	0.23 (-0.06,0.50)	0.49
It could have been worse	0.39	(0.55)	0.28,0.50	0.42	(0.57)	0.32,0.53	0.23	.629	0.05 (-0.23,0.33)	0.17
New Verbal cues Phase 2										
Total details	6.50	(5.11)	5.49,7.57	5.40	(5.35)	4.36,6.43	2.34	.128	0.21 (-0.07,0.49)	0.43
Complications	1.17	(1.48)	0.90,1.45	0.86	(1.33)	0.58,1.13	2.58	.110	0.22 (-0.06,0.50)	0.49
Plausibility	4.61	(1.17)	4.38,4.85	4.24	(1.20)	4.01,4.48	4.76	.030	0.31 (0.03,0.59)	1.42
It could have been worse	0.03	(0.17)	-0.01,0.06	0.03	(0.16)	-0.01,0.05	0.22	0.643	0.00 (-0.28,0.28)	0.17

Note: NHST = Null-Hypothesis Significance Testing.

Hypotheses-Testing: Phase 2

A 2 (Veracity) X 2 (Sketch) X 2 (Video) MANOVA was carried out on the Phase 2 data with new total details, new complications, plausibility, and new 'it could have been worse' as dependent variables. At a multivariate level, none of the effects were significant, all $F_s < 2.37$, all $p_s > .053$. The average mean of reported details was small ($M = 5.95$, $SD = 5.25$). This means that Hypotheses 3 and 4 were not supported.

The Relationship between Plausibility and the Other Verbal Output Variables

Through a regression analysis we examined which verbal output variables contributed to the plausibility ratings. A forced entry method regression analysis was conducted with total details, complications and 'it could have been worse' as predictors and plausibility as the outcome variable. Complications contributed to the model ($\beta = .47$, $p < .001$) but total details ($\beta = -.03$, $p = 0.693$) and 'it could have been worse' ($\beta = .10$, $p = .136$) did not.

Discussion

The statements of truth tellers reporting a genuine account of a loss, theft or accidental damage differed from the statements of lie tellers reporting a false incident. Truth tellers provided more complications, more 'it could have been worse' comments and a more plausible account than lie tellers. This is the first time that complications have been examined in an insurance claim setting and the results showed some evidence that it can be used in such a setting to distinguish between truth and deceit. This is good news for practitioners dealing with insurance claims. Interviewers can count complications in real time [23], which means that the interviews do not have to be transcribed first. However, given that the evidence was weak and the effect size small, our results suggest that practitioners may struggle to classify claimants as truth tellers and lie tellers based on the number of reported complications.

The 'it could have been worse' variable was examined for the first time and showed potential as a veracity indicator. As with complications, interviewers can count such comments in real time. Research attempting to replicate this finding is required to show the robustness of the effect. Such research could also examine the underlying mechanism as to why truth tellers and lie tellers differ in reporting 'it could have been worse' comments.

Truth tellers also sounded more plausible than lie tellers, providing further support that plausibility is a diagnostic veracity indicator [9]. Researchers are reluctant to examine plausibility or to recommend practitioners to use it as a veracity indicator, probably due to the subjective nature of the variable: What exactly makes a statement plausible or implausible?. Apart from the context, it has been suggested that reporting complications and reporting details contribute to plausibility [9]. The current dataset supported the findings for complications, but not for details. A similar pattern of results (plausibility is positively correlated with complications but not with details) emerged elsewhere [24]. This suggests that out of these two variables -complications and details- complications is the strongest predictor of plausibility.

Total details did not emerge as a veracity indicator. This is an atypical finding because total details is often a strong veracity indicator [25]. Perhaps a different finding would have emerged if we would have considered the verifiability of the details. That

is, do the details provide leads that the investigators can check (receipts, named witnesses, CCTV footage)?. Truth tellers typically report more verifiable details than lie tellers [26]. One insurance claim experiment revealed that truth tellers and lie tellers did not differ in the number of details they reported but that differences emerged when a distinction was made between details that could be verified and details that could not. Truth tellers reported more verifiable details than lie tellers whereas lie tellers reported more unverifiable details than truth tellers [27]. For differences in verifiable details to emerge between truth tellers and lie tellers in insurance claim settings, interviewees need to be encouraged at the beginning of the interview to try to include details the investigator can check [26]. We did not do this in the present experiment because the context reinstatement and sketching instructions were extensive. We were afraid to overwhelm interviewees if we would add another instruction to the list.

Whether during the zoom meeting the video was turned on or off did not affect the results. The Bayes analyses results were close to zero which suggests that the video turned on or off really did not matter. In many zoom meetings participants are reluctant to turn on their video and the results thus suggests that this is an irrelevant factor for lie detection purposes when paying attention to the verbal cues we examined in the experiment.

Sketching had no effect on the results. The non-significant effect in Phase 1 is unsurprising because the sketching factor was introduced after Phase 1. However, the results were also not significant in Phase 2. The likely reason for this null effect is the little amount of new detail added in Phase 2 (on average only six details). If interviewees do not report much new information, any manipulation will have no effect due to a floor effect of the results. Two factors may have contributed to the little amount of new information provided in Phase 2. First, participants reported not to have spent much effort in their sketching. For a sketching instruction to work participants should fully engage with the sketching task. This could be the result of conducting the interviews online rather than face-to-face. Perhaps interviewees put less effort in their sketching exercise in online interviews than in face-to-face interviews. Future research could examine this. Second, the context reinstatement instruction we gave ("Please think about the incident and recall what you could see, what you could hear, what you could feel and what you could smell and let me know when you have done that?") may have focussed the interviewees' minds in an unexpected manner. Reading the transcripts showed that many interviewees focussed on the latter two senses (feel and smell) and had little to no information to report about these senses. Since a sketch is a visual output, it is most likely to facilitate recall of visual information (what people could see), but the context reinstatement instruction did not focus enough on that aspect of the experience.

Author contributions

The first author designed the experiment (together with the third author), analysed the data (together with the second author) and wrote the initial draft (together with the second author). The second author also set up the experiment. The third author commented on the initial draft.

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflict of interest

The authors have no conflict of interest to declare.

References

1. Dando C, Wilcock R, Milne R. The Cognitive Interview: The efficacy of a modified mental reinstatement of context procedure for frontline police investigators. *Applied Cognitive Psychology*. 2009; 23: 138-147.
2. Leins D, Fisher RP, Pludwinsky L, Robertson B, Mueller DH. Interview protocols to facilitate human intelligence sources' recollections of meetings. *Applied Cognitive Psychology*. 2014; 28: 926-935.
3. Mattison MCL, Dando CJ, Ormerod TC. Sketching to remember: Episodic free recall task support for child witnesses and victims with autism spectrum disorder. *Journal of Autism and Developmental Disorders*. 2015; 45: 1751-1765.
4. Vrij A, Mann S, Leal S, Fisher RP, Deeb H. Sketching while narrating as a tool to detect deceit. *Applied Cognitive Psychology*. 2020; 34: 628-642.
5. Vrij A, Leal S, Fisher RP, Mann S, Dalton G, et al. Sketching as a technique to elicit information and cues to deceit in interpreter-based interviews. *Journal of Applied Research in Memory and Cognition*. 2018; 7: 303-313.
6. Mac Giolla E, Granhag PA, Vernham Z. Drawing-based deception detection techniques: A state-of-the-art review. *Crime Psychology Review*. 2017; 3: 23-38.
7. DePaulo BM, Lindsay JL, Malone BE, Muhlenbruck L, Charlton K, et al. Cues to deception. *Psychological Bulletin*. 2003; 129: 74-118.
8. Vrij A. *Detecting lies and deceit: Pitfalls and opportunities*, second edition. Chichester: John Wiley and Sons. 2008.
9. Vrij A, Deeb H, Leal S, Granhag PA, Fisher RP. Plausibility: A verbal cue to veracity worth examining?. *The European Journal of Psychology Applied to Legal Context*. 2021; 13: 47-53.
10. Amado BG, Arce R, Fariña F, Vilarino M. Criteria-Based Content Analysis (CBCA) reality criteria in adults: A meta-analytic review. *International Journal of Clinical and Health Psychology*. 2016; 16: 201-210.
11. Gancedo Y, Fariña F, Seijo D, Vilarino M, Arce R. Reality monitoring: A meta-analytical review for forensic practice. *The European Journal of Psychology Applied to Legal Context*. 2021; 13: 99-110.
12. Colwell K, Hiscock-Anisman C, Memon A, Woods D, Michlik P M. Strategies of impression management among deceivers and truth tellers: How liars attempt to convince. *American Journal of Forensic Psychology*. 2006; 24: 31-38.
13. Hartwig M, Granhag PA, Strömwall L. Guilty and innocent suspects' strategies during police interrogations. *Psychology, Crime, & Law*. 2007; 13: 213-227.
14. Nahari G. Verifiability approach: Applications in different judgmental settings. In T. Docan-Morgan, the *Palgrave Handbook of Deceptive Communication* New York, NY: United States: Palgrave Macmillan. 2019; 213-225.
15. Köhnken G. Statement Validity Analysis and the 'detection of the truth'. In PA Granhag & LA Strömwall. *Deception detection in forensic contexts*. Cambridge, England: Cambridge University Press. 2004: 41-63.
16. Vrij A, Palena N, Leal S, Caso L. The relationship between complications, common knowledge details and self-handicapping strategies and veracity: A meta-analysis. *The European Journal of Psychology Applied to Legal Context*. 2021; 13: 55-77.
17. Maier BG, Niehaus S, Wachholz S, Volbert R. The strategic meaning of CBCA criteria from the perspective of deceivers. *Frontiers in Psychology*. 2018; 9: 855.
18. Leal S, Vrij A, Deeb H, Kamermans K. Encouraging interviewees to say more and deception: The Ghostwriter method. *Legal and Criminological Psychology*. 2019; 24: 273-287.
19. Blair J P, Levine T, Shaw A. Content in context improves deception detection accuracy. *Human Communication Research*. 2010; 36: 423-442.
20. Brimbal L, Dianiska RE, Swanner JK, Meissner CA. Enhancing cooperation and disclosure by manipulating affiliation and developing rapport in investigative interviews. *Psychology, Public Policy, and Law*. 2019; 25: 107-115.
21. Vallano JP, Schreiber Compo N. A comfortable witness is a good witness: Rapport-building and susceptibility to mis-information in an investigative mock-crime interview. *Applied Cognitive Psychology*. 2011; 25: 960-970.
22. Jarosz A F, Wiley J. What are the odds?. A practical guide to computing and reporting Bayes Factors. *The Journal of Problem Solving*. 2014; 7.
23. Vrij A, Fisher R P, Leal S. How researchers can make verbal lie detection more attractive for practitioners. *Psychiatry, Psychology, Law*. 2022.
24. Vrij A, Leal S, Deeb H, Castro-Campos C, Fisher RP, et al. The effect of using countermeasures in interpreter-absent and interpreter-present interviews. *The European Journal of Psychology Applied to Legal Context*. 2022.
25. Verschuere B, Bogaard G, Meijer EH. Discriminating deceptive from truthful statements using the verifiability approach: A meta-analysis. *Applied Cognitive Psychology*. 2021; 35: 374-384.
26. Palena N, Caso L, Vrij A, Nahari G. The Verifiability Approach: A meta-analysis. *Journal of Applied Research in Memory and Cognition*. 2021; 10: 155-166.
27. Vrij A, Nahari G, Isitt R, Leal S. Using the verifiability lie detection approach in an insurance claim setting. *Journal of Investigative Psychology and Offender Profiling*. 2016; 13: 183-197.