Go To

Go to old article view

Legal and Criminological Psychology *Explore this journal >*

View issue TOC
Volume 20, Issue 1
February 2015
Pages 129–146

Original Article

You cannot hide your telephone lies: Providing a model statement as an aid to detect deception in insurance telephone calls

Sharon Leal , Aldert Vrij, Lara Warmelink, Zarah Vernham, Ronald P. Fisher

First published:
16 May 2013 Full publication history

DOI:
10.1111/lcrp.12017 View/save citation

Cited by (CrossRef):
28 articles Check for updates | Citation tools

Am 2

Abstract

Deception research regarding insurance claims is rare but relevant given the financial loss in terms of fraud. In Study 1, a field study in a large multinational insurance fraud detection company, truth telling mock claimants (N = 19) and lying mock claimants (N = 21) were interviewed by insurance company telephone operators. These operators classified correctly only 50% of these truthful and lying claimants, but their task was particularly challenging: Claimants said little, and truthful and deceptive statements did not differ in quality (measured with Criteria-Based Content Analysis [CBCA]) or plausibility. In Study 2, a laboratory experiment, participants in the experimental condition (N = 43) were exposed to an audiotaped truthful and detailed account of an event that was unrelated to insurance claims (a day at the motor races). The number of words, quality of the statement (measured with CBCA), and plausibility of the participants' accounts were compared with

participants who were not given a model statement (N = 40). The participants who had listened to the model statement provided longer statements than control participants, truth tellers obtained higher CBCA scores than liars, and only in the model statement condition did truth tellers sound more plausible than liars. Providing participants with a model statement is thus an innovative and successful tool to elicit cues to deception. Providing such a model has the potential to enhance performance in insurance call interviews, and, as we argue, in many other interview settings.

Background

Deception researchers active in forensic contexts mainly examine lying in police interviews (Vrij & Granhag, **2012**). However, there is growing understanding that other areas are equally relevant and deserve attention. Only recently a plea was made to expand deception research to security and intelligence settings (Brandon, **2011**; Loftus, **2011**), and, indeed, research publications have begun to emerge about deception related to intentions (Clemens, Granhag, & Strömwall, **2011**; Granhag & Knieps, **2011**; Vrij, Granhag, Mann, & Leal, **2011**; Warmelink *et al*., **2011**), security screen interviewing (Leal, Vrij, Mann, & Fisher, **2010**), undercover interviewing (Vrij, Mann, Jundi, Hope, & Leal, **2012**), and collective interviewing (Driskell, Salas, & Driskell, **2012**; Vrij *et al*., **2012**). In this article, we address another relevant setting: Lying to insurance companies in telephone interviews. One reason for insurance companies to call customers who claim theft, loss, or accidental damage is to determine the truthfulness of these claims. It is a sensible activity for insurance companies to actively attempt to mitigate the effect of fraud, as many insurance holders make false claims to them. The Association of British Insurers (Association of British Insurers, **2009**) reported that 20% of insurance holders say that they would consider making an exaggerated or completely made up insurance claim in the future. The ABI further estimated that undetected general insurance claims fraud totals £1.9 billion ($2.95 billion) a year in the United Kingdom alone.

In overview, the goal of the research was twofold. The first goal was to determine base-level performance in an ecologically valid setting: How well do insurance agencies currently discriminate between truthful and fraudulent claims? (Study 1, a field study). The second goal was to assess whether an experimental manipulation can enhance insurance companies' abilities to discriminate between truthful and fraudulent claims (Study 2, a laboratory experiment). In Study 2, the claimants were asked (or not) first to listen to a model account of a truthful, unrelated event before speaking to the mock insurance agent. We expected that listening to the model account would elicit more verbal cues to deception, thereby increasing the differences – and ability to discriminate- between the accounts of truthful and deceptive claimants.

Practitioners often mention an apparent lack of ecological validity as a limitation of police deception research (Inbau, Reid, Buckley, & Jayne, **2013**; Mann, Vrij, Fisher, & Robinson, **2008**). Field deception research, in which real life cases are examined, has the highest ecological validity, but that research is rare in a police context (but see Mann, Vrij, & Bull, **2002**; Ten Brinke & Porter, **2012**; Ten Brinke, Porter, & Baker, **2012**; and Vrij & Mann, **2001**, for exceptions). One major problem is establishing ground truth, that is, determining which suspects are actually lying and which are not. Researchers therefore tend to carry out research with mock suspects in simulated police interviews. In an insurance context, both anecdotal accounts from fraud investigators and ethnographic insurance research (Morley, Ball, & Ormerod, **2006**) show that claim outcome is typically used as a measurement of claim veracity. Such a

presumption has ground truth problems, as a withdrawn claim does not necessarily mean it was false. Likewise, a continued claim does not necessarily mean it was genuine. We considered it to be possible to conduct insurance deception research that has high ecological validity. Although establishing the ground truth in insurance claims is as difficult as establishing the ground truth in police interviews, it is possible for researchers to recruit participants where the incident is verifiable, and to feed these mock claimants into the natural, existing insurance claim system. We followed this procedure in Study 1. In the large multinational insurance fraud detection company we were collaborating with, claimants who have filed their claims with the insurance company are called by an insurance company telephone operator to discuss these claims. With permission of the insurance company we have for the purpose of the study filed claims of genuine and fraudulent mock claimants. The genuine mock claimants were participants, who had genuinely experienced theft, loss or accidental damage, but the items were never covered by insurance and hence they never submitted the claim by our or any other insurance company. Fraudulent mock claimants were participants who were instructed by us to pretend theft, loss or accidental damage. After the mock claims were filed, they were treated as normal claims. Thus, the mock claimants were called by the insurance company operators at a predetermined date and time to discuss these mock claims, with the operators believing that they were dealing with real claimants rather than mock claimants.

The insurance company that took part in the study has a standardized telephone interview protocol. An insurance company telephone operator (called 'insurance caller' hereafter) always starts the interview with an open-ended question inviting the claimant to describe in as much detail as possible what happened. We expect claimants' answers to this question to be short, because their beliefs about how much detail is expected from them are likely to be inadequate. If an interviewee does not know his/her conversation partner well, which is the case in insurance call interviews and, in fact, most formal interview settings, they tend to give short answers (Fisher, **2010**; Fisher, Milne, & Bull, **2011**). This may occur particularly in telephone interviews because such interviews are more distant and less intimate than, for example, face-to-face interviews.

Short interviews are problematic for lie detection purposes. Verbal cues to deceit are more likely to occur in longer than shorter statements (Vrij, Mann, Kristen, & Fisher, **2007**), as words are the carriers of verbal cues to deceit. Also, when truth tellers do not say much, they make the task easier for liars, who do not need to become talkative themselves to match truth tellers. We therefore envisaged that (a) the participants (truth tellers and liars alike) would not say much and (b) the statements of truth tellers and liars would be similar in detail and would sound equally plausible.

Insurance callers use their own coding system, and that coding system is confidential for commercial reasons. However, they do not use recognized verbal coding systems such as Criteria-Based Content Analysis (CBCA; Köhnken, **2004**; Köhnken & Steller, **1988**) or RM (Masip, Sporer, Garrido, & Herrero, **2005**; Sporer, **2004**). In Study 1, we examined the insurance callers' performance in terms of truth and lie detection. Research has overwhelmingly demonstrated that lie detection is a difficult task, with people averaging 53% accuracy rate when listening to audiotapes (Bond & DePaulo, **2006**). Vrij (**2008**) found in his review no difference in accuracy between lay persons and professionals. We have no reason to believe that the insurance callers' performance would be better than the performance typically achieved by observers and therefore expect an accuracy rate at or just above the level of chance.

Differences between truth tellers and liars may emerge if truth tellers provide longer statements. Talkative truth tellers raise the standard for liars, who also need to become more talkative to match truth tellers. In becoming more talkative, liars potentially increase exposing their deception. A possible way to make truth tellers more talkative, which we examine in Study 2, is to provide participants with a detailed, model statement – albeit about an unrelated topic. The underlying assumption is that if truth tellers hear a detailed model statement, their views on what is expected from them may change and, as a result, they may provide a more detailed answer themselves. In alignment with these theoretical assumptions, an experiment with cooperative participants (truth tellers) has shown that providing people with a detailed model statement resulted in their providing more detailed statements (Hirn, Fisher, & Carol, **2012**).

Truth tellers' inclination to provide more detail after being exposed to a detailed model statement may not be replicated by liars. First, liars face the problem that they should not say too much, as the information they give may indicate that they are lying. For example, they may say something that the interviewer knows to be false or easily can find out to be false. Second, liars typically prepare themselves for interviews (Granhag & Hartwig, **2008**), by preparing answers to expected questions (Colwell, Hiscock-Anisman, Memon, Woods, & Michlik, **2006**; Hartwig, Granhag, & Strömwall, **2007**; Hartwig, Granhag, Strömwall, & Doering, **2010**; Strömwall, Hartwig, & Granhag, **2006**). This means that liars in all likelihood have planned what they are going to say when asked to describe their claims. However, it is unlikely that they have prepared as much detail as the detailed model statement implies they should provide. A model statement therefore puts pressure on liars to include more detail than they have initially prepared. Perhaps liars lack the imagination and skills to generate the same amount of extra detail as truth tellers do. If so, then truth tellers will give longer answers that contain more detail than liars, particularly after being exposed to a detailed model statement (veracity × model statement interaction). This finding would be in alignment with the deception literature where spontaneous lies often reveal more cues to deceit than prepared lies (DePaulo *et al*., **2003**; Sporer & Schwandt, **2006**).

An alternative outcome is possible. After listening to a detailed, model statement liars may manage to lengthen their answers and provide additional detail. However, this additional information may not sound as plausible as the additional information truth tellers provide. If this is the case, then number of words and amount of detail will not differ between truth tellers and liars, but plausibility would, with truth tellers' answers sounding more plausible, particularly after being exposed to a detailed model statement.

We measured detail with CBCA, a list of 19 criteria thought to be more present in truthful than deceptive statements. CBCA-trained evaluators judge the presence or absence of each of these criteria in a statement, and the presence of each criterion strengthens the hypothesis that the account is based on genuine personal experience (Köhnken & Steller, **1988**; Steller & Köhnken, **1989**). In other words, truth tellers are expected to obtain higher CBCA scores than liars. Although CBCA was developed to analyse children's statements in sexual abuse cases, it has been argued that CBCA analyses are also suitable for statements from adults who talk about issues other than sexual abuse (Köhnken, Schimossek, Aschermann, & Höfer, **1995**). Research has convincingly supported this view (Vrij, **2005**, **2008**). We had several reasons to carry out CBCA analyses. First, it is the most widely researched verbal veracity assessment tool, and it has been used in over 50 published studies, mostly with adult participants (Vrij, **2008**). Second, research has revealed that CBCA can differentiate between truth tellers and liars. Vrij's (**2008**) review of CBCA research included 20 samples in which total CBCA

scores were reported. In 16 of those 20 samples (80%), truth tellers' total CBCA scores were significantly higher than liars' total CBCA scores. This is impressive, particularly considering that the vast majority of cues to deceit are faint and unreliable (DePaulo *et al*., **2003**). In terms of accuracy, Vrij's (**2008**) review included 24 samples in which accuracy rates were reported. On average, 70.81% of truthful statements and 71.12% of deceptive statements were classified correctly based on CBCA assessments. Third, although there is some controversy over CBCA, for example about its theoretical foundation (Sporer, **2004**), the idea that CBCA measures the quality of a statement is undisputed. Vrij (**2005**, **2008**) provides a detailed description of the CBCA criteria and definitions of the criteria are given in Table 1.

Go To

Table 1. Criteria-Based Content Analysis criteria used in the studies

General characteristics	
1. Logical structure	Coherency of the statement in terms of not containing logical inconsistencies or contradictions
2. Unstructured production	The presentation of the information in a (non) chronological order
3. Quantity of details	The inclusion of specific descriptions of place, time, persons, objects and events
Specific contents	
4. Contextual embedding	Events being placed in time and location, and actions being connected with other daily activities and/or customs
5. Descriptions of interactions	Information that interlinks at least the alleged perpetrator and witness
6. Reproduction of conversation	Parts of the conversation are reported in original form or if the different speakers are recognizable in the reproduced dialogues
7. Unexpected complications during the incident	Elements incorporated in the statement that are somewhat unexpected
8. Unusual details	Details of people, objects, or events that are unique, unexpected or surprising but meaningful in the context
9. Superfluous details	Details in connection with the allegations that are not essential for the accusation
11. Related external associations	Events are reported that are not actually part of the alleged offence but are merely related to the offence
12. Accounts of subjective mental state	Development and change in feelings experienced at the time of the incident. This criterion also includes reports of thoughts
Motivation-related contents	

14. Spontaneous corrections	Corrections that are made or information that is added to material previously provided in the statement without having been prompted by the interviewer
15. Admitting lack of memory	An unprompted interviewee admitting lack of memory by either saying 'I don't know' or 'I don't remember'
16. Raising doubts about one's own testimony	Interviewee indicating that part of his or her description sounds odd, implausible, unlikely, etc.
17. Self-deprecation	Inclusion by the interviewee of personally unfavourable, self-incriminating details

STUDY 1

Method

Participants

A total of 40 participants (14 women and 26 men) took part as mock claimants. Their average age was M = 35.61 years (SD = 11.63). A total of 22 participants reported accidental damage, 13 participants reported theft, and the remaining five reported loss. Those 40 claims were processed by a random sample of nine insurance callers (six women and three men) who happened to be on duty at the time the telephone call with the mock claimant was placed.

Procedure

To guarantee ground truth, all truth telling mock claimants (N = 19) were known personally to the experimenter. Ground truth was either established by the experimenter having personally witnessed the event or by the known person spontaneously mentioning an accident to the experimenter. For example, a friend had just dropped her phone down the toilet, but had no mobile phone insurance. The majority of lying mock claimants (14 out of the 21 liars that took part) were also known to the experimenter and were asked verbally to take part. The remaining lying mock claimants were friends of the recruited truthful mock claimants. The liars were asked to generate an incident (accidental damage, theft or loss) that they would like to lie about.

In both lie and truth telling conditions mock claims were related to accidental damage, theft, and loss. Accidental damage included events such as spilling paint on a carpet or dropping a mobile phone. This category did not include car accidents because these often involve medical reports and are dealt with by a different investigation team. Theft was any item or items usually covered by insurance policies and ranged from a single mobile phone being stolen to house burglary with several valuable items stolen. Loss would include any item that would be potentially covered by insurance that was lost by the mock policy holder. Examples included cameras and phones.

Mock claimant names were entered into the incident sheet provided by the insurance company and were allocated individual claim numbers by the insurance company line manager. The line manager

also checked whether the claim would be covered by the insurance policy (all claims were). These mock claims were processed together with real claims (claims processed by the insurance company that were not part of the study). The line managers knew which were mock claimants and which were real claimants, but they did not know whether these mock claimants were truth tellers or liars. Although insurance callers were aware that research was being carried out in their company, they did not know on which days the mock claim interviews would take place. Not a single insurance caller (10 were involved) raised suspicion about any of the mock claim interviews, a strong indication that these mock claimants appeared real to them.

The experimenter phoned the mock claimants and indicated when they could expect to receive the phone call from the insurance company to make sure that they would be available. Truthful mock claimants were instructed to answer the questions truthfully whereas deceptive mock claimants were instructed to answer the questions in such a way that the interviewer would believe that indeed something had been accidentally damaged, stolen, or lost. At least three working days passed between being informed about the phone call and the actual phone call taking place so that the participants could prepare themselves. The claim call took place at the agreed date and time, and the insurance callers called each mock claimant at their home, work or mobile phone, which is normal practice in this insurance company. The insurance callers, thinking that they were dealing with real claimants (see above), carried out their normal interview phone calls and their normal procedure after the phone calls. The insurance callers asked: 'OK, just so I can understand, I am going to need you to take me back to the day of the incident, and tell me in as much detail as possible everything that happened from before the incident through to your contacting the insurance company'. On completing the claim call the insurance callers completed their normal procedure. The insurance callers filled out their coding sheet (confidential information) and made their veracity judgements, by using a dichotomous yes/no scale and also a 7-point Likert scale ranging from 1 (*definitely truth telling*) to 7 (*definitely lying*). The insurance callers normally do not make veracity judgements, but those were added to the callers' coding sheet for the duration of the experiment. They also wrote their correspondence letter to the claimant. The letters written to the mock claimants were removed by the line manager from the post bag without the insurance callers noticing this. Recorded experimental telephone call data and the corresponding insurance callers' veracity ratings were collected by the experimenter for analysis. The telephone calls were then transcribed to allow for CBCA analysis.

Before the telephone call the mock claimants completed a questionnaire in which they were asked to indicate their preparation for the interview in terms of three measures: 'thorough', 'sufficient', and 'good'. The answers were given on 7-point Likert scales ranging from 1 (*not at all*) to 7 (*very much so*). After the telephone call the mock claimants were thanked, debriefed, and given £10 cash for their efforts.

Ethics

Ethical BPS guidelines for research with human participants state that the deception of participants should be avoided wherever possible. In this study, insurance callers were deceived as they received mock calls. To ascertain how insurance callers detect lies and truths within their normal working context it was vital that they should *not* be aware that they would receive mock calls. We believe that the level of deception we introduced was acceptable for the following reasons: (a) insurance callers' individual

accuracy rates were not given to their employers; (b) as part of their contract all insurance callers had already agreed that their conversations will be recorded and can be listened to by their management; and (c) all insurance callers were aware that their veracity judgements would be checked according to claim outcomes. Thus, similar to shop assistants, they work in an environment where they can expect to be observed publicly.

CBCA coding

The transcribed answers to the open-ended question were coded by two experienced CBCA raters, who were blind to the veracity status of the transcripts. They coded 15 criteria, see Table 1. Four CBCA criteria were left out: Criterion 10 *accurately reported details misunderstood*, criterion 13 *attributions of perpetrator's mental state*, criterion 18 *pardoning the perpetrator*, and criterion 19 *details characteristic of the offence*. They were left out as we felt that they were not appropriate for insurance claim interviews. Omitting CBCA criteria that are considered inappropriate for the particular experimental setting is common practice in CBCA research (Vrij, **2005**, **2008**). The coders counted the frequency of occurrence of visual details and contextual embeddings. For the remaining variables, the coders counted whether the criterion was absent or present. One coder noticed several instances of unusual details (N = 8) and superfluous details (N = 8), whereas the other coder did not see any instances of such details. Those disagreements were relegated to a third experienced CBCA rater. We report only the total CBCA score, because total CBCA scores obtain better results than the individual criteria in terms of reliability and (veracity) diagnosticity (Vrij, **2005**, **2008**). We calculated the intra class coefficient (ICC) between the two CBCA scores based on the scores of the two individual raters. Before calculating the total CBCA scores for each of the raters, we dichotomized their quantity of detail and contextual embedding criteria (half of the participants with the lowest score on quantity of detail received the score '0', and the remaining half received the score '1', and this procedure was repeated for the contextual embedding scores). For the unusual details and superfluous details we used the third rater's scores. The ICC for the two raters' total CBCA scores was high and sufficient (r = .91, p = .000). The two raters' CBCA scores were averaged and this average CBCA score is reported. Table 2 shows the results for the individual CBCA criteria when p < .10.

Table 2. Univariate Criteria-Based Content Analysis results in Study 1

	Truth			Lie			F	p	η^2	d
	M	*SD*	CI	*M*	*SD*	CI				
5. Descriptions of interactions	.11	.27	−.07, .28	.33	.46	.17, .50	3.61	.07	.09	.60
11. Related external associations	.32	.48	.16, .48	.05	.15	−.11, .29	5.98	.02	.14	.86
12. Accounts of subjective mental state	.47	.49	.27, .68	.21	.41	.02, .41	3.39	.07	.08	.58
15. Admitting lack of memory	.16	.34	.05, .27	.03	.11	−.08, .13	3.01	.09	.07	.58

	Truth			Lie			F	p	η^2	d
	M	*SD*	CI	*M*	*SD*	CI				
17. Self-deprecation	.13	.33	.03, .24	.00	.00	−.10, .10	3.42	.07	.08	.39

Plausibility coding

Five coders (researchers), also blind to the veracity status of the transcripts and the model statement manipulation, each rated the plausibility of all transcripts on a 7-point Likert scale ranging from 1 (*not at all*) to 7 (*very much*). Plausibility was defined as 'Could this incident have happened as described'? and 'Could an honest interviewee remember the amount of detail as described'? A reliability analysis showed good reliability amongst the five raters (Cronbach's $\alpha = .71$). The scores of the five raters were averaged and this averaged score was used in the analyses.

Results

Pre-interview questionnaire

Participants judged their preparation for the interview as thorough ($M = 5.43$, $SD = 1.1$), sufficient ($M = 5.65$, $SD = 0.9$), and good ($M = 5.59$, $SD = 0.9$). No differences emerged between truth tellers and liars regarding these aspects (all F's < 1.07, all p's > .31). In terms of percentages, 87% of the participants scored 5 or higher on each of the three 7-point Likert scales.

Insurance callers' accuracy rates

The insurance callers showed a truth bias and believed 66% ($N = 27$) of the participants. Their ability to detect deceit was poor. Their overall accuracy rate was 50%, which is the percentage that could be expected by chance. They judged 68% of the truth tellers correctly, but only 33% of the liars. Regarding the Likert scale results, truth tellers: $M = 3.50$, $SD = 1.9$, 95% CI (2.77, 4.32), and liars: $M = 3.15$, $SD = 1.5$, 95% CI (2.37, 3.93), were perceived as equally suspicious, $F(1, 36) = .39$, $p = .54$, $\eta^2 = .01$, $d = .21$. The mean scores show that, if anything, truth tellers were perceived to be slightly more suspicious than liars, and, as the means score were close to '4', that the insurance callers were not confident in their judgements. These scores demonstrate the difficulty insurance callers faced when discriminating between truth tellers and liars.

Number of words in response to the open-ended question

When asked to describe what had happened in as much detail as possible, participants used on average $M = 142.03$ words ($SD = 82.20$) to report the incident. No difference was found between truth tellers: $M = 143.94$, $SD = 82.90$, 95% CI (105.28, 182.61), and liars: $M = 140.28$, $SD = 83.57$, 95% CI (103.51, 177.06), $F(1, 38) = .02$, $p = .89$, $\eta^2 = .00$, $d = .04$.

CBCA scores

Go To

Truth tellers': M = 4.00, SD = 2.0, 95% CI (3.12, 4.88), CBCA scores were minimally higher than liars' CBCA scores: M = 3.36, SD = 1.7, 95% CI (2.52, 4.19), but this difference was not significant, $F(1, 38) = 1.15$, p = .29, η^2 = .03, d = .34.

Plausibility

Truth tellers' statements: M = 4.21, SD = 1.00, 95% CI (3.73, 4.69), and liars' statements: M = 4.31, SD = 1.1, 95% CI (3.86, 4.77), were seen as equally plausible by the five judges, $F(1, 38) = .10$, p = .75, η^2 = .00, d = .10.

Discussion

In Study 1, genuine and fraudulent mock insurance claims were filed into the existing claim assessment system of a large multinational insurance fraud detection company. Following the normal procedure of this insurance company, the mock claimants were then called by the insurance company operators (insurance callers), who were unaware that they were talking to mock claimants. The study revealed two important findings. First, insurance callers performed only at chance level (50% accuracy rate) and could not discriminate between truth tellers and liars on the Likert scale. This should come as no surprise to deception researchers, as lie detection based on audiotapes is a notoriously difficult task (Bond & DePaulo, **2006**). In that respect, we have no reasons to believe that other insurance companies that assess claims via telephone interviews will fare better than the company that participated in Study 1. Second, CBCA scores did not differ significantly between truth tellers and liars in Study 1. This is unusual as CBCA typically does discriminate successfully between truth tellers and liars. However, in typical CBCA research the statements are considerably longer than the statements we assessed in this study. CBCA has more opportunity to discriminate between truth tellers and liars in longer statements.

Our selection of participants was governed by two competing forces: ensuring ground truth versus having a representative sample of participants. In order for us to ensure ground truth, we had to have personal knowledge of the participants, however, that obviously precludes random assignment. We opted for ensuring 100% ground truth, in part because we felt that our sample was likely to generate responses that were similar to a true random selection from the population. Or, at least, we cannot think of a plausible theoretical reason why our sample would be very different from a true random sample. Had we opted to maximize representativeness, and chosen a true random sample, we would have guaranteed not having 100% ground truth. Hence, we opted to ensure knowing ground truth.

To lengthen truth tellers' statements, we provided participants with a model of a detailed answer (734 words) in Study 2. Increasing the length of the participants' answers would create the opportunity for differences between truth tellers and liars to emerge, as we explained in the Introduction.

STUDY 2

Method

Go To

Participants

The 83 participants (22 men and 61 women) were undergraduate students and university staff. Their average age was M = 25.31 years (SD = 8.4). A total of 26 participants reported accidental damage, 44 theft, and the remaining 13 reported loss.

The model interview

In the experimental condition, prior to answering the open-ended question to discuss their incident in detail, participants were asked to listen to a 734 words audiotape in which a person gives a detailed account of attending a motor racing event. We did not want to give participants an idea what to say during the interview (hence, the unrelated event of motor racing), but wanted to give them an idea about what a detailed account entails.

The model statement audiotape was a recall of a witness (son of the first author) describing his experiences when attending motor racing for 1 day. This was a spontaneous, unscripted, recall of the event, and the only instruction the witness received was to be as detailed as possible. The witness was aware that this recall would be used as an example for others. We did not give any guidance about what types of detail to include and about what to say. Below is one, 125 word, paragraph of the interview. A transcript of the full account is available from the first author.

'So I was walking down the middle of the grid but I was advised to go to one side as the cars were coming in, so's not to get run over! I remember looking up at the grandstand on my right and there were a fair few people stood there watching the grid. I walked down towards position 11 which is where Tom was located. I remember looking quite closely at some of the pit girls which you would as a man! So I walked past, I was on my own at this point and I got out my phone to do a little bit of filming as it's not every day that you get to be on the grid at a formula 2 race'!

Procedure

Participants were recruited via advertising posters and an announcement on the university Intranet. The advert explained that the study investigated what cues to deceit may be present in insurance claims, and that we were looking for mock truthful and deceptive claimants.

Truthful claimants were required to be people who have recently had some genuine incident that they could have claimed for. Examples of incidents include the following: an accident such as a dropped laptop/TV; broken mobile; spillage on carpet etc.; a theft such as a stolen phone, purse, or laptop; or a motor claim such as a car theft or accident. The amount of money the incident would (or did cost) should range £150–£9,000 (from $232 to $13,950). The advert further stated that, ideally, the incident would have taken place within the last 4 months or so, but the important thing is that the person could remember and truthfully answer questions about the details of what happened. A £5 cash award was offered for participation. Those who volunteered to participate were approached by the experimenter, who checked that the claim would be covered by the insurance policy and that the person actually remembered what had happened. All participants said that they remembered the incident as it was

quite emotional to them. The accident took place up to a few months ago for 16 participants, but longer than a few months ago for the remaining 25 participants. Those two groups did not differ from each other in terms of CBCA scores, $F(1, 39) = .01$, $p = .92$, $\eta^2 = .000$, $d = .04$, and plausibility scores, $F(1, 39) = .00$, $p = .96$, $\eta^2 = .00$, $d = .02$, and this time element is therefore ignored in the further analyses.

To recruit liars, the advert also mentioned that those who did not have experienced theft, loss or accidental damage could contact the experimenter to participate as liars. Those liars were asked to think of a made up incident they were happy to talk about, and to contact the experimenter again when they had thought of an incident. The experimenter checked again whether this claim would be covered by the insurance policy.

The experimenter arranged a date and time to meet with the participants. Truth tellers were instructed to answer the questions truthfully during the interview, whereas liars were instructed to answer the questions in such a way that the interviewer would think that an item had actually been accidentally damaged, lost or stolen. A minimum of three working days passed before the interview took place to give the participant time to prepare for the interview. Participants then arrived at the agreed time in the psychology department where they were greeted by the experimenter. The experimenter took them to a cubicle where the telephone call would be received. The experimenter informed the female mock insurance caller (a PhD student blind to the veracity status of the participants) that the participant was ready to be interviewed and the insurance caller subsequently called the participant. The participant was aware that this was a simulated insurance claim call and that the insurance caller was an experimenter rather than someone working for an insurance company. After greeting the participant and recording some background information the mock insurance caller said in the no-model response (control) condition ($N = 43$, 21 truth tellers and 22 liars): 'OK, just so I can understand, I am going to need you to take me back to the day of the incident, and tell me in as much detail as possible everything that happened from before the incident through to you contacting the insurance company'.

In the model statement condition ($N = 40$, 20 truth tellers and 20 liars) the interviewer said: 'OK, just so I can understand, I am going to need you to take me back to the day of the incident, and tell me in as much detail as possible everything that happened from before the incident through to you contacting the insurance company. I know that sometimes people are not sure just *how* much detail to include. In order to give you an idea of what I am looking for I'd like to play you an example of what we consider a detailed answer'.

We then played the 734-word model response audiotape. After the participant listened to the audiotape the insurance caller said: 'OK, I know that wasn't too relevant to your claim but hopefully you have an idea of the amount of detail it takes for us to get a clear rounded idea of an event! Could you now please tell me in as much detail as possible everything that happened from before the incident through to you contacting the insurance company'? After the initial response the insurance caller thanked the participant for his/her cooperation and ended the conversation. After the phone the participants were thanked, debriefed, and received £5 cash for their efforts.

CBCA coding

One experienced CBCA coder and a newly trained CBCA coder coded all the transcripts independently from each other. They were blind to the veracity status of the transcripts as well as the model response

manipulation. The trainee first received background reading about CBCA and a list of definitions of the criteria presented in Table 1. The experienced CBCA coder and trainee then coded a few transcripts together (a sample of Study 1). After that the trainee coded a few transcripts on her own and received feedback from the experienced CBCA coder. After this the actual coding took place. The coders coded the same criteria as were used in Study 1.

Before calculating the total CBCA scores for each of the raters, we dichotomized their quantity of detail and contextual embedding criteria. Half of the participants with the lowest score on quantity of detail received the score '0', and the remaining half received the score '1', and this procedure was repeated for the contextual embedding scores. The ICC for the CBCA total scores for the two raters was very high and sufficient, $r = .91$. We averaged the two CBCA scores and used that average score in the analyses. Table 3 shows the results for the individual CBCA criteria when $p < .10$.

Table 3. Univariate Criteria-Based Content Analysis results in Study 2, scores of expert coder

	Truth			Lie			*F*	*p*	η^2	*d*
	M	*SD*	CI	*M*	*SD*	CI				
5. Descriptions of interactions	.46	.50	.32, .61	.21	.41	.07, .36	6.04	.02	.07	.55
6. Reproduction of conversation	.12	.33	.05, .19	.00	.00	−.07, .07	5.79	.02	.07	.37
7. Unexpected complications during the incident	.37	.49	.23, .50	.17	.38	.03, .30	4.35	.04	.05	.46
8. Unusual details	.12	.33	.05, .19	.00	.00	−.07, .07	5.69	.02	.07	.36
14. Spontaneous corrections	.12	.33	.04, .20	.02	.15	−.05, .10	3.02	.09	.04	.42
15. Admitting lack of memory	.29	.46	.17, .41	.10	.30	−.02, .21	5.41	.02	.06	.50

Plausibility coding

Go To

Five coders (students), also blind to the veracity status of the transcripts and the model statement manipulation, each rated the plausibility of all transcripts on a 7-point Likert scale ranging from 1 (*not at all*) to 7 (*very much*). Plausibility was defined as 'Could this incident have happened as described'? and 'Could an honest interviewee remember the amount of detail as described'? A reliability analysis showed good reliability amongst the five raters (Cronbach's $\alpha = .71$). The scores of the five raters were averaged and this averaged score was used in the analyses.

Results

Number of words

On average the participants spoke $M = 204.3$ words ($SD = 155.8$). A 2 (veracity) × 2 (model statement) analysis of variance (ANOVA) showed that Veracity had no impact on the number of words spoken, $F(1, 79) = 2.97$, $p = .09$, $\eta^2 = .04$, $d = .36$, but that Model Statement had, $F(1, 79) = 23.00$, $p < .001$, $\eta^2 = .23$, $d = 1.05$. Participants who had listened to the model statement: $M = 279.70$, $SD = 155.9$, 95% CI (236.45, 322.95), said more than twice as much as participants who did not listen to the statement: $M = 134.14$, $SD = 120.0$, 95% CI (93.15, 176.60). The Veracity × Model Statement interaction effect was not significant, $F(1, 79) = .14$, $p = .71$, $\eta^2 = .00$.

CBCA scores

A 2 (veracity) × 2 (model statement) ANOVA with the total CBCA score as dependent variable showed significant main effects for Veracity, $F(1, 79) = 10.66$, $p = .002$, $\eta^2 = .12$, $d = .76$ and Model Statement, $F(1, 79) = 16.19$, $p = .000$, $\eta^2 = .17$, $d = .86$. Truth tellers obtained higher CBCA scores: $M = 4.63$, $SD = 2.6$, 95% CI (4.03, 5.29), than liars: $M = 3.17$, $SD = 1.8$, 95% CI (2.59, 3.82), and the participants who had been exposed to the model statement obtained higher CBCA scores: ($M = 4.83$, $SD = 2.2$, 95% CI (4.20, 5.46), than the participants who had not been exposed to the statement: $M = 3.02$, $SD = 2.0$, 95% CI (2.43, 3.66). The Veracity × Model Statement interaction effect was not significant, $F(1, 79) = .01$, $p = .91$, $\eta^2 = .00$.

Plausibility

A 2 (veracity) × 2 (model statement) ANOVA with plausibility as dependent variable showed significant main effects for Veracity, $F(1, 79) = 21.05$, $p = .000$, $\eta^2 = .21$, $d = 1.97$ and Model Statement, $F(1, 79) = 4.92$, $p = .029$, $\eta^2 = .06$, $d = .45$, and a significant Veracity × Model Statement interaction effect, $F(1, 79) = 4.33$, $p = .041$, $\eta^2 = .05$. Truth tellers: $M = 5.40$, $SD = 1.0$, 95% CI (4.22, 4.79), sounded more plausible than liars: $M = 3.59$, $SD = 0.9$, 95% CI (3.31, 3.87), and participants who had been exposed to the model statement sounded more plausible: $M = 4.27$, $SD = 1.1$, 95% CI (3.98, 4.56), than those who were not exposed to that statement: $M = 3.82$, $SD = 0.9$, 95% CI (3.55, 4.10). The interaction effect revealed that in the model statement condition truth tellers: $M = 4.94$, $SD = 0.9$, 95% CI (4.53, 5.34), sounded more plausible than liars: $M = 3.60$, $SD = 0.9$, 95% CI (3.19, 4.01), $F(1, 38) = 21.48$, $p = .000$, $\eta^2 = .36$, $d = 1.24$, whereas the difference in plausibility

between truth tellers: M = 4.07, SD = 1.0, 95% CI (3.68, 4.47), and liars: M = 3.57, SD = 0.8, 95% CI (3.19, 3.97), was not significant for participants in the control condition, $F(1, 41) = 3.26$, p = .08, η^2 = .07, d = .55. The increase in plausibility between the model statement condition and control condition was large and significant for truth tellers (M = 4.07 vs. M = 4.94); $F(1, 39) = 8.23$, p = .007, η^2 = .17, d = .92, and small and non-significant for liars (M = 3.57 vs. M = 3.60); $F(1, 40) = .01$, p = .92, η^2 = .00, d = .03. The interaction effect is provided in Figure 1.

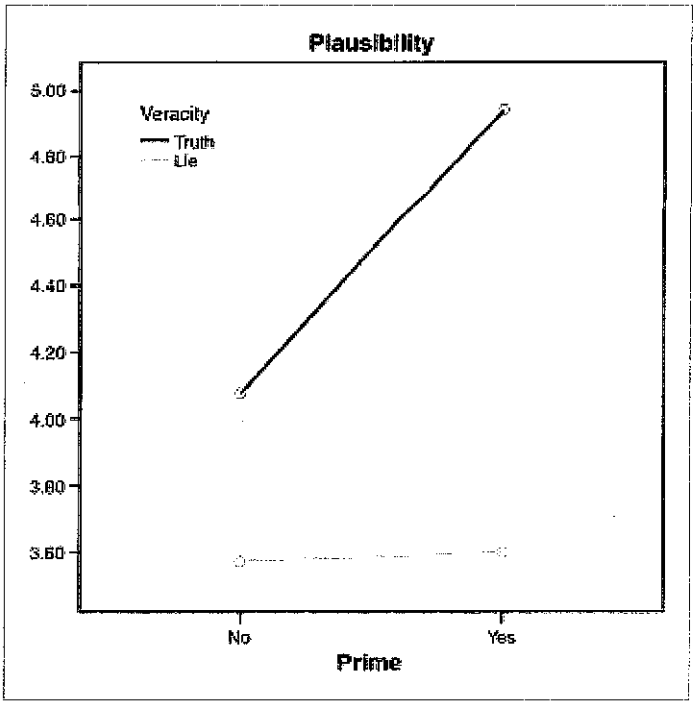

Figure 1.

Open in figure viewer

Plausibility as a function of veracity and model statement.

Discrimination between truth tellers and liars

A discriminant analysis was conducted to examine the extent to which the CBCA and plausibility scores enabled discrimination between truth tellers and liars. The objective group-belonging was the classifying variable and CBCA and plausibility were the two predictor variables. The stepwise method was used and separate analyses were carried out for the model statements and control conditions. In the control condition only the CBCA factor emerged in the function, $\chi^2(1) = 6.05$, Wilk's Lambda = .86, p = .048. The function correctly identified 72.7% of liars but only 52.4% of truth tellers, resulting in a total accuracy of 62.5%. In the model statement condition only the plausibility factor emerged in the function, $\chi^2(1) = 16.80$, Wilk's Lambda = .64, p = .000. The function correctly identified 75.0% of liars and 85.0% of truth tellers, resulting in a total accuracy of 80%.

Discussion

The number-of-words analysis showed a similar problem in the control condition as in Study 1: Participants said very little (this time 134 words). Being exposed to a model statement increased this substantially to 279 words, and the increase was similar for truth tellers and in liars. The results further showed that being exposed to a model statement also increased the truth tellers' and liars' CBCA scores to a similar extent. However, being exposed to a model statement had a differential effect on truth tellers and liars in terms of plausibility: Truth tellers' statements sounded significantly more plausible than liars' statements, but only in the model statement condition. In other words, truth tellers benefited more from the model statement than truth tellers. Being exposed to a model statement resulted in truth tellers and liars adding a similar number of words to their stories which contained similar detail as measured with CBCA. However, these additional details sounded more plausible in truth tellers. Apparently, adding words and details to a story is one thing; doing that in a plausible way is another issue.

The absence of a Veracity × Model Statement is worth discussing. The findings of this experiment resemble CBCA coaching studies in which it was found that as a result of coaching both truth tellers and liars included more CBCA criteria in their statements (Coaching main effect) whereas the Veracity × Coaching interaction effect was not significant (Vrij, Akehurst, Soukara, & Bull, **2002**, **2004**). In CBCA coaching studies, participants are informed about specific CBCA criteria and are instructed to include such criteria in their statement. In this experiment participants were not informed about specific criteria, but just listened to a detailed story. This apparently results in similar effects as coaching, but is far easier to implement, as no training needs to be provided to participants.

We examined plausibility as it is an efficient way to judge the quality of an account. However, we realize that plausibility is a subjective cue, and someone may argue that it is difficult to teach someone how to code it. To counteract this argument, in both Studies 1 and 2 we used five raters to judge the plausibility of the accounts, and across both studies eight different raters were used. Their scores were consistent as indicated by a satisfactory Cronbach's α rating (.71) in both studies. This implies that plausibility coding was not an idiosyncratic activity, at least not in this study.[1]

The model statement intervention fits well in the rapidly growing 'interviewing to detect deception' literature (Granhag & Vrij, **2010**), and can be considered a new method in the outsmarting the liars approach (Vrij, Granhag, Mann, & Leal, **2011**; Vrij, Granhag, & Porter, **2010**). The core of that approach is that cues to deceit can be elicited by imposing tasks on interviewees that liars find particularly difficult to cope with. Suggested and experimentally tested approaches to impose cognitive load include asking participants to recall their story in reverse order (Vrij *et al*., **2008**) and to recall while maintaining eye contact with the interviewer (Vrij, Mann, Leal, & Fisher, **2010**). An alternative way of imposing cognitive load on liars is to ensure that in a given interview setting truth tellers will provide more information. Talkative truth tellers raise the standard for liars, who also need to become more talkative to match truth tellers. Liars may find it too cognitively difficult to add as many details as truth tellers, or if they do add a sufficient number of details the additional information may be of lesser quality or sound less plausible. The model statement intervention is one way to make truth tellers more talkative, but there are other ways, such as introducing a silent second interviewer who is supportive throughout (Mann *et al*., **2013**).

In addition to eliciting cues to deceit, exposing participants to a model statement has several additional advantages. First, it is very easy to apply as no interview skill is required. The only activity required

from the interviewer is to switch on an audiotape. In that respect, exposing interviewees to a model statement is easier to apply in real life than other ways to increase talkativeness suggested in the investigative interviewing literature. For example, a well-known procedure is the narrative elaboration procedure (Bowen & Howie, **2002**; Brown & Pipe, **2003a,b**; Camparo, Wagner, & Saywitz, **2001**; Dorado & Saywitz, **2001**; Saywitz & Snyder, **1996** Saywitz, Snyder, & Lamphear, **1996**). In that procedure, investigators show interviewees (typically children) reminder cue cards with story grammar categories (e.g., participants, setting, actions, conversations). Second, a model statement intervention is not related to any specific interview protocol, and it could therefore be applied to every interview setting where detailed answers from interviewees are desirable (e.g., Fisher & Geiselman, **1992**). This is not just the case in insurance claim interviews, but also in police interviews, airport interviews, and security screening interviews, and many other types of information-gathering interview. Third, exposure to a model statement leads to extra information (from both truth tellers and liars) which gets to the core of interviewing: To obtain as much information from an interviewee as possible (Fisher, **2010**). Fourth, the fact that exposure to a model statement leads to extra information increases the chance of verbal cues to deception to occur as such cues are more likely to occur in longer than in shorter statements (Vrij *et al*., **2007**). Fifth, exposure to a model statement is a much more user-friendly way of imposing cognitive load. A disadvantage of other imposing cognitive load techniques introduced in the literature so far, such as asking to recall in reverse order or to maintain eye contact with the interviewer during the recall, is that investigators may feel uncomfortable to apply it in real life (Vrij, Leal, Mann, & Fisher, **2012**).

Conclusion

We carried out a field study and a laboratory experiment regarding lying and truth telling in insurance claims. The field study showed that, in a highly ecologically valid setting, insurance company interviewers could not distinguish truth tellers from liars (50%), but the findings also showed that their task was particularly challenging. The interviewees' answers to the open-ended question were short, which gave little opportunity for differences between truth tellers and liars to emerge. The laboratory experiment showed a way to make lie detection easier. Exposing interviewees to an audiotape of a detail account of an unrelated event led to an equal amount of additional information in both truth tellers and liars as measured via CBCA, but truth tellers' additional information sounded more plausible than liars' additional information. This change in plausibility adds another cue to discriminate between the two groups. Exposure to a model statement fits well in the growing literature on interviewing to deception and outsmarting the liars.

Note

1 In addition, unreported analyses revealed a similar pattern among the raters. In Study 1, none of the individual raters could successfully discriminate between truth tellers and liars. In Study 2, in the model statement condition all five raters successfully discriminated between truth tellers and liars, and all five raters found truth tellers significantly more plausible than liars. In contrast, in the control condition, only one rater could successfully discriminate between truth tellers and liars. This is again a consistent pattern of results that suggests that plausibility coding was not an idiosyncratic exercise.

Go To

Related content

Citing Literature

WILEY

Browse Publications

Browse by Subject

Resources

Help & Support

Cookies & Privacy

Terms of Service

About Us

Wiley Job Network

Advertisers & Agents

Powered by Wiley Online Library Copyright © 1999 - 2017 John Wiley & Sons, Inc. All Rights Reserved