UNIVERSITY of CALIFORNIA PRESS | Collabra

Kleinberg, B., et al. (2017). An Investigation on the Detectability of Deceptive Intent about Flying through Verbal Deception Detection. *Collabra: Psychology,* 3(1): 21, DOI: https://doi.org/10.1525/collabra.80

ORIGINAL RESEARCH REPORT

An Investigation on the Detectability of Deceptive Intent about Flying through Verbal Deception Detection

Bennett Kleinberg*, Galit Nahari†, Arnoud Arntz* and Bruno Verschuere*

Background: Academic research on deception detection has largely focused on the detection of past events. For many applied purposes, however, the detection of false reports about someone's intention merits attention. Based on the verbal deception detection paradigm, we explored whether true statements on intentions were more detailed and more specific than false statements on intentions, particularly when instructed to be as specific as possible.
Method: Participants (n = 222) lied or told the truth about their upcoming travel plans either providing 'as much information as possible' (standard instructions) or being 'as specific as possible' (i.e., mentioning times, locations, places; specific instructions), resulting in four conditions (truthful vs. deceptive intention by standard vs. specific instructions). We collected data via a custom-made web app and performed automated verbal content analysis of participants' written answers.
Findings: We did not find a significant difference in the specificity of participants' statements. The instruction to be as specific as possible promoted more specific information but did not help to discern honest from deceptive flying intentions.
Conclusion: The experiment reported here attempted to demonstrate automated verbal deception detection of intentions. The difficulty in capturing genuine intentions, and the non-intrusive, non-interactive questioning approach might explain the null findings and raise questions for further research. We conclude with suggestions for a novel framework on semi-interactive information elicitation.

Keywords: deception detection; deceptive intentions; Reality Monitoring; airport screening; information protocol

With an increased demand for security systems like airport border control, researchers and practitioners alike have identified the need for applications to detect deception on a large scale (Honts & Hartwig, 2014; Vrij, Granhag, & Porter, 2010). For example, the context of airport border control excludes many tools used in deception research due to their limited applicability. With developments towards more seamless passenger flows and minimal passenger-security personnel interaction, an ideal deception detection system would be implementable at stages even before passengers arrive at the airport (e.g., a filter system during online check-in processes, Kleinberg, Arntz, & Verschuere, in press). A promising paradigm that might be applicable in settings that require the testing of a vast number of people is verbal deception detection. However, because the majority of verbal deception detection studies were conducted on the verbal content of face-to-face interviews, a key challenge is the transition towards large-scale applicable methods. This paper reports an attempt to apply verbal deception detection tools on a large-scale in an airport security context.

Verbal deception detection

The idea to use the verbal content as an indicator of deception is rooted in the Undeutsch Hypothesis (1967, 1982) stating that truthful statements differ from false declarations in quality and content because the process through which the particular statement is produced is different (Fornaciari & Poesio, 2013). One difference that emerges from that framework is that truthful statements contain more contextual embeddings (i.e., references to persons, events, locations) than deceptive ones (Köhnken, 2004). Research by Johnson and Raye (1981; Masip, Sporer, Garrido, & Herrero, 2005) has specified further that the source of one's memory determines how a remembered event is recalled. Genuine memories are obtained through sensory experiences whereas non-genuine memories are constructed through cognitive operations. Therefore, the content of these memories should differ so that descriptions of genuine memories should be richer in

* University of Amsterdam, NL
† Bar-Ilan University, IL
Corresponding author: Bennett Kleinberg (b.a.r.kleinberg@uva.nl)

sensory experiences (e.g., perceptual, spatial, temporal information), whereas non-genuine memories should contain more references to cognitive operations (Johnson & Raye, 1981; Masip et al., 2005). Reality Monitoring (RM) is a theoretical and analytical framework that incorporates this idea. Parallel to genuine and non-genuine memories, truthful statements are expected to be richer in detail compared to false statements (also labeled as Interpersonal Reality Monitoring, see Johnson, Bush, & Mitchell, 1998; Nahari & Vrij, 2014). Especially the amount of temporal, spatial and perceptual detail has been found to be higher in truth-tellers' statements than in liars' (Masip et al., 2005; Vrij, 2008).

The recently introduced the Verifiability Approach (VA, Nahari, Vrij, & Fisher, 2014a, 2014b) suggests that there might be an additional dimension to the number of details, namely the verifiability of details. The VA exploits the strategies used by liars to provide a believable, false account. During an interview, the liar faces the dilemma between being inclined to describe an event in sufficient detail to sound convincing and at the same time avoiding information that could potentially be verified (Nahari et al., 2014a). For example, an answer like *'I spoke to my friend James in the Vondelpark'* might be a detail that could theoretically be investigated further by the interviewer (e.g., by consulting James and asking for confirmation), whereas *'I spoke to someone in the park'* would not count as a verifiable detail. A series of studies showed that the number of verifiable details discriminates liars from truth-tellers (Harvey, Vrij, Nahari, & Ludwig, 2016; Nahari et al., 2014a). For example, the number of verifiable details was higher in truthful insurance claims than in deceptive ones when both liars and truth-tellers were instructed to mention as much verifiable information as possible (Harvey et al., 2016). The working definition of verifiable information includes any activity that i) has been done with an identifiable person, ii) has been witnessed by an identifiable person, or iii) has been documented or recorded through technology (e.g., CCTV, email, social networks, see Nahari et al., 2014a).

Taken together, these findings of verbal deception detection (contextual embeddings, the richness of detail, verifiable details) suggest that, in general, deceptive statements are less specific than truthful statements. This paper set out to test that assumption for deceptive intentions.

Detecting deceptive intentions

Academic deception research focused primarily on the detection of deception about recent events. For many practical purposes in law enforcement and intelligence services, it is becoming increasingly important to detect people with potentially malicious intent to prevent crimes from happening (Vrij & Granhag, 2012). For example, in border control settings, it might be more important to determine what someone is planning to do upon entering a country rather than learning what they have been doing before coming to border control. Recently, researchers have begun shifting the temporal dimension of verbal deception research paradigms towards that of intentions (Mac Giolla, Granhag, & Liu-Jönsson, 2013; Sooniste, Granhag, Knieps, & Vrij, 2013; Sooniste, Granhag, Strömwall, & Vrij, 2015).

For example, Sooniste et al. (2013, based on Granhag & Knieps, 2011) instructed half of their participants to plan and enact an innocent activity in a shopping mall (i.e., buy two gifts). The other half was told to prepare and enact a mock-crime (i.e., placing a USB stick illegally in a shop in the same shopping mall). Those instructed to plan the mock-crime were also told to develop a cover story similar to that of the innocent participants. Before any of the participants enacted their assigned task in the shopping mall, they were intercepted and interviewed about their intended behavior in the mall. Before the interview, liars were told to hide their true intentions, so that each interviewee tried to convince the interviewer of having planned the innocent activity. Most importantly, during the interview, questions about the planning phase and the intentions were asked, reasoning that the former was less expected and hence more diagnostic than the latter. They found that truth tellers' answers to planning-related questions were rated as more detailed than those of liars, whereas there was no such difference for intentions-related questions.

In a related study, airport passengers were asked in a quasi-experimental setup to either lie or tell the truth about their upcoming trip (Vrij, Granhag, Mann, & Leal, 2011). The authors found that truth tellers' answers were rated as more plausible than liars' answers but did not differ on the perceived amount of detail (for studies that do show such a difference in the richness of detail on intentions, see Sooniste et al., 2015; Warmelink et al., 2012; Warmelink, Vrij, Mann, Leal, et al., 2013). These findings were recently complemented with the Verifiability Approach (Jupe, Leal, Vrij, & Nahari, 2017). Truth-tellers mentioned slightly more verifiable details than liars about their upcoming trip (Cohen's $d = 0.28$).

The emerging academic literature on the detection of deceptive intention suggests that the verbal approach is promising. An important technique in verbal deception detection that might help increase verbal differences is that of exploiting differences in liars' and truth-tellers' preparedness.

Asking unanticipated questions

Because liars better prepare for an interview than truth-tellers (Granhag & Hartwig, 2014), the interviewer can ask unexpected questions to exploit the liars' preparedness. For example, a liar may prepare for questions like '*Where have you been yesterday?*' but not for questions like '*What did the spatial arrangement look like in the cafe?*' (Vrij et al., 2009). There is evidence that asking unanticipated questions is beneficial to deception detection regarding past events (e.g., mock crimes, Shaw et al., 2013; Vrij et al., 2009) and regarding lies about someone's occupation (Warmelink, Vrij, Mann, Leal, et al., 2013). The effectiveness of question expectedness on the detection of deceptive intentions is less clear. When participants lied or told the truth about their travel plans (Warmelink et al., 2012), the expected higher richness of detail for truth telling was indeed only found for unexpected questions (i.e., "*How are you going to travel to your destination?*"), but

not for general questions (i.e., "*What is the main purpose of your trip?*"; where liars were, in fact, more detailed). In another experiment, truth-tellers were to find material in a library whereas liars intended to steal the material but prepare a believable cover story (Fenn, McGuire, Langben, & Blandón-Gitlin, 2015). During the interview, they were asked to either tell about their activities in chronological order, or in the reverse order (i.e., going backward in time). The unexpected reverse order question appeared detrimental for deception detection accuracy: truth-tellers were more likely to be misclassified as liars. We, therefore, explore how question expectedness impacts upon the detection of intentions.

The current investigation

The aim of the present study is to examine whether we can identify people lying about their intentions of traveling by airplane. To work towards potentially large-scale applicable methods of deception detection, we built an online platform where we asked questions about people's upcoming flight plans. We stayed close to the questions asked in previously successful studies on verbal deception detection about intentions (Sooniste et al., 2013; Warmelink et al., 2012; Warmelink, Vrij, Mann, Leal, et al., 2013). In addition to being able to collect data on a large scale, another requirement for implementable tools is an automated analytical framework (Kleinberg et al., in press). Here, we aimed to address this by analyzing verbal content both computer-automated and complemented by human coding.

We instructed participants to either tell the truth or lie, and they were subsequently asked ten questions about their next or most recent flight. Because previous studies highlight the importance of asking the right questions, we used unexpected (e.g., transportation-related) questions in addition to expected (e.g., general) questions (Warmelink et al., 2012; Warmelink, Vrij, Mann, Leal, et al., 2013). Furthermore, there are indications that informing both liars and truth tellers about the quality of the expected information in truthful answers may benefit deception detection (Harvey et al., 2016; Nahari et al., 2014b, but see Nahari & Pazuelo, 2015). We, therefore, instructed half of the interviewees to provide highly specific information (e.g., names of persons or locations, dates) instead of merely asking to provide as much information as possible. Asking for highly specific information may pose an additional difficulty to liars because they wish to avoid providing damaging information, whereas truth-tellers could quickly recall specific, potentially verifiable information (Nahari et al. 2014b).

Hypotheses

Our primary hypothesis in this study was motivated by findings from deception research on both past events (e.g., Nahari et al., 2014a) and intentions (e.g., Vrij et al., 2011; Warmelink, Vrij, Mann, Leal, et al., 2013): Truth-tellers' accounts contain more detailed information than those of liars ("*richness of detail hypothesis*"). Similar to the information protocol procedure (Nahari et al., 2014b), we further hypothesized that truth tellers could provide more specific information than liars if they are explicitly told to do so ("*information protocol hypothesis*"). For exploratory purposes, we were interested in investigating *i)* human coded variables for differentiating truthful from deceptive statements that may be harder to automatize (e.g., plausibility); *ii)* how the question type affected the richness of detail, *iii)* how richness of detail differed between past events and future intentions, and *iv)* how the temporal immediacy of flight plans moderated the effect of richness of detail.

Method

This experiment was approved by the IRB of the University of Amsterdam (dossier #2016-CP-7230). All materials, raw and aggregated data, as well as the source code to the experimental task, are available via https://osf.io/knhz4/.

Participants

We aimed to collect data from 518 participants based on *a priori* power analysis for the 2 (Veracity: truthful vs. deceptive) by 2 (Information Protocol: standard vs specific) interaction for an effect size of Cohen's $f = 0.25$, power = .95, alpha = .05.[1] We opened spots for participation on the online platform *crowdflower.com* until this number of participants was reached. Of the initial sample of 518 participants, there were no data for nine participants, and we further excluded data of participants whose IP address has been registered more than once, resulting in an additional 94 exclusions (see Kleinberg & Verschuere, 2015; Verschuere, Kleinberg, & Theocharidou, 2015). The relatively high exclusion number for duplicate IP addresses might be due to the block-wise data collection that made double participation possible. From the remaining sample (n = 415), we excluded participants who were outliers (larger than 2.5 SDs above the mean) on the number of weeks until their flight and the number of times having visited the destination before (n = 33) and those who indicated to not have provided genuine information at the beginning (n = 28), resulting in a final sample of 354 participants.[2] Participants were randomly allocated to the truthful or deceptive condition and further to the standard or specific information protocol condition. Our quasi-experimental manipulation of participants who were or were not flying in the next three months resulted in two groups (flyers who reported about their future flight and non-flyers who reported about their past flight). The focus of this investigation is on those who report about their future flight.[3] Of those who were going to fly (n = 222),[4] 109 participants were in the truthful condition (standard: n = 49, M_{age} = 32.51, SD = 9.10, 32.65% female; specific: n = 60, M_{age} = 35.78, SD = 9.35, 41.94% female), and 113 were in the deceptive condition (standard: n = 52, M_{age} = 33.90, SD = 10.21, 34.54% female; specific: n = 61, M_{age} = 32.25, SD = 7.61, 34.43% female).

Materials

Experimental task

The experimental task was presented in a custom-made web app programmed in HTML and JavaScript. Participants needed an Internet connection and a standard web browser to do the task on their computer. To ensure additional control over the experimental task, we disabled

the translation function, the copy-and-paste function, tested whether the input was provided where necessary, whether text input was real English, and whether the required minimum length of answers was provided. If not, the participants were informed about this via a pop-up and alerted that this could result in invalidation of their participation.

Computer-automated analysis: Richness of detail

Many studies that adopted a computer-automated approach to verbal deception detection have used the Linguistic Inquiry and Word Count software (LIWC, Mihalcea & Strapparava, 2009; Pennebaker, Boyd, Jordan, & Blackburn, 2015). Text statements processed with LIWC return proportions of word categories occurring in the text. Each word category is intended to model different psycholinguistic variables such as the category 'affect' which is used to model emotional processes. Underlying each category are extensive dictionaries of words against which the words in the statements are analyzed. LIWC has successfully been employed in multiple contexts (Ott, Cardie, & Hancock, 2013; Pérez-Rosas & Mihalcea, 2014) and was shown to be acceptable for modeling human-coded RM annotation (Bond & Lee, 2005). In the current investigation, we used the LIWC word categories "percept," "space" and "time," to model perceptual, spatial and temporal details, respectively. For each participant, we summed the three categories across all ten answers to derive the dependent variable richness of detail.

Sentence specificity: Speciteller

Motivated by the observation that two sentences can contain the same propositional meaning but convey that information with different degrees of specificity, Li and Nenkova (2015) developed speciteller. Speciteller is a python-implemented machine learning-based classifier giving the specificity of a sentence ranging from 0 (lowest) to 1 (highest). Five independent annotators judged a sample of 885 sentences from the Wall Street Journal, New York Times, and Associated Press. The annotators determined that 54.58% of the sample be specific sentences which were then used to build a classifier with shallow surface features (e.g., the number of words, the estimated number of named entities) and dictionary features (e.g., subjective words, concreteness). Using machine learning (supervised logistic regression, semi-supervised and co-training classification), they derived a final classifier model they released as open-source software under the name speciteller. We used the speciteller tool to calculate the average sentence specificity per statement as a dependent variable.

Information specificity: Named Entity Recognition

We operationalize information specificity as the number of named entities recognized by the SpaCy python natural language processing tool (Honnibal, 2016). Named entity recognition is a sub-field of computational linguistics focused on the extraction of information from text. The information is extracted in so-called named entities that refer to specific information such as persons, places or dates. In general, the approach to developing a named entity recognition (NER) system is to define grammar-based rules, regular expressions, and machine learning classification to identify entities in text automatically. In this investigation, we extract named entities in the categories persons, nationalities or religious groups, facilities, organizations, geopolitical entities, locations, products, events, works of art, law documents, languages, dates, times, percentages, quantities, ordinals, and cardinals (Honnibal, 2016; Kleinberg, Nahari, & Verschuere, 2016). We obtained the dependent variable information specificity by summing all named entities per statement divided by the number of words (see also Kleinberg, Mozes, Arntz, & Verschuere, 2017).

Human coding

In addition to automated report coding, two students coded the statements of those participants that were flying in no more than four weeks (n = 110). Coders were presented with the entire account of the participant (i.e., their answers to the ten questions), and asked to rate the entire statement. All variables were scored on a 7-point Likert scale (1 = very low; 7 = very high, see Sooniste et al., 2013). Definitions were adopted from Vrij (2015) and MacGiolla et al. (2013). The coders were trained in rating the statements on richness of detail ("the inclusion of specific descriptions of place, time, persons, objects and events in the statement"), plausibility ("the coherency of the statement in terms of not containing logical inconsistencies or contradictions and the degree to which the message seems plausible, likely, or believable"), complications ("the reporting of either an unforeseen interruption or difficulty, or spontaneous termination of the event"), occurrence of how-utterances ("concrete descriptions of activities"), occurrence of why-utterances ("first, wider motivations/reasons why someone planned an activity; second, motivations/reasons for doing something in a certain way"), and truthfulness. Both coders received a 2.5-hour training session on statements of non-flying participants and a subset (n = 16) of the selected statements which were excluded from the analysis. Ninety-one statements (48 truthful, 43 deceptive), 31 statements were coded by both coders (ICCs: plausibility = 0.67, richness of detail = 0.85, how-utterances = 0.36,[5] why-utterances = 0.86, complications = 0.71, truthfulness = 0.82). The remaining 60 statements were randomly distributed between the two coders.

Procedure

The experimental task was advertised on crowdflower.com as a survey about people's flying behavior. Upon accessing the custom-made web app via a link provided in the task description (tinyurl.com/jny6p9w) participants were introduced to the task and told that serious participation was necessary and would be rewarded with the chance of winning a $100, Amazon.com voucher.[6] After giving informed consent, on the next page, all participants were asked whether they would be flying in the next three months (answer options: "yes", "no", "not sure yet"). If they indicated that they would fly in the next three months, the next page asked the following flight-related questions; i)

how many weeks this flight was away, ii) what the purpose of this flight was (pre-defined selection menu, e.g., "work"; see Appendix), iii) what the final destination of their trip was (e.g., "London", and iv) how often they had visited that place before. For all pages where any input was required, participants could only proceed after providing the required information. Participants were either instructed to lie or to tell the truth about their flight. Those who were in the truthful condition were told to provide honest answers about their trip to (say) "London for work." Those who were allocated to the deceptive condition were assigned a new destination (e.g., "Madrid") and a new purpose (e.g., "holiday") and were told to pretend they are planning to fly to this new destination with the new purpose.

Also, in both conditions, we told participants that they should either provide as much information as possible on the next ten questions about their flight or that they should provide as specific information as possible (e.g., names, locations, dates). The instructions were repeated in bullet points on the next page, and all participants had 30 seconds to prepare for the upcoming questions. In total, all participants answered eleven questions including one test question ("Please describe your task for this experiment") to help participants become acquainted with the task (**Table 1**). The remaining ten questions were identical to all participants whereby the destination and purpose were filled in according to the participants' experimental condition and their assigned destination/purpose pair. The questions were selected to reflect the structure and content of related studies (i.e., asking questions on the core event – Question 2 and 3; on the planning and preparation – Question 4 and 5; and on the transportation – Question 8 and 9, see Sooniste et al., 2013; Warmelink et al., 2012; Warmelink, Vrij, Mann, Leal, et al., 2013), and to take into account meta-analytical findings showing that emotion-related questions (Question 6 and 7) are a useful to elicit truth-lie differences (Hauch, Blandón-Gitlin, Masip, & Sporer, 2015). We supplemented these eight questions with two questions that we reasoned to be uniquely related to properly planned intended actions (Question 10) and should be rather unexpected (Question 11). During all questions, below the actual wording of the questions, the instructions regarding the veracity and information protocol manipulation were repeated (e.g., "Remember: please lie about your original trip by giving very specific information (persons, locations, times, etc.) about a trip to Madrid for a holiday". Questions were presented one at a time in identical order.

After typing in the answers to these questions, we asked for demographic variables (age, gender, education, country of origin, native language) and asked for each question, how expected they found it on a Likert scale from 1 (not expected at all) to 10 (absolutely expected). Also, we asked how motivated they were and had them rate their language proficiency as well as doing two language assessment tasks which were part of another study and are not reported here. Those participants who indicated at the beginning that they were not flying in the next three months proceeded through the same task but answered all flight-related questions about their most recent past flight. The truthful/deceptive manipulation was adjusted accordingly (i.e., answer truthfully or lie about the last past trip). The wording of the questions changed automatically.

At the end of the task, as a control check, participants were asked whether they provided accurate information at the beginning of the task regarding their upcoming or past flight (answer options: "yes", "no"). Participants were then debriefed and could provide their email address for the draw on the $100, voucher. The task took approx. 15 min.

Experimental manipulations

There were two experimental manipulations as well as one quasi-experimental manipulation in this study. First, we manipulated the veracity of people's answers by allocating them to either the truthful or the deceptive condition. If participants indicated that the purpose of their upcoming flight would be returning home, they responded to questions about their past trip. Those in the truthful condition (for both past and upcoming trip) were asked questions about the self-reported destination and purpose whereas those in the deceptive condition were allocated a different destination/purpose pair. This allocation ensured that neither the purpose nor the destination for liars matched the original one. We further attempted to apply a semi-yoked matched design by randomly allocating a destination/purpose pair that genuine flyers reported in pilot studies. Second, we manipulated the information protocol by changing the additional instructions to answer the questions. Those in the standard information protocol condition were told to provide as much information as possible, whereas those in the specific information protocol condition were told to provide as much specific information as possible. The latter also received examples of what specific information was (names, times, locations, etc.). Third, the quasi-experimental manipulation was the temporal focus of flying (past flight or upcoming flight) which was self-reported by participants.

Analysis plan

Although the full design of this study was 2 (Temporal focus: future vs. past flight, between-subjects) by 2 (Veracity: truthful vs. deceptive, between-subjects) by 2 (Information Protocol: standard vs. specific, between-subjects), as reported above, the primary aim of the analysis were participants who had future flight plans (i.e., intentions). Therefore, for the main hypotheses tested, the particular design was 2 (Veracity: truthful vs. deceptive, between-subjects) by 2 (Information Protocol: standard vs. specific, between-subjects). As the dependent variable, we tested richness of detail, average sentence specificity, and information specificity in the written answers.

Also, in exploratory analyses, we provide human coding of verbal content variables of a subset of statements. For exploratory analyses, we included an additional factor into the analysis, namely Question type (general vs. planning

Table 1: Questions asked in the experimental task (verbatim).

#	Question	Rationale/ reference	Minimum length (characters)	Observed length (*M*, *SD*)	Example of answer[7]
1	"This is a test question and a check whether you understood the instructions. Please briefly state your task in this experiment."	Control question	15	71.08 (34.26)	"To accurately provide information on my trip to London to visit family. I plan to fly to London to visit family and friends. I will also be traveling to Brighton and maybe Southampton."
2	"What is the main purpose of your flight to [DESTINATION]?"	General question (Warmelink, Vrij, Mann, Leal, et al., 2013)	50	133.15 (95.03)	"The main purpose is to visit family in London. I will also be going to Brighton."
3	"Who will you meet in [DESTINATION] and for which reason?"	General question	50	110.91 (70.93)	"I will be visiting family that live in London. I will visit some friends as well."
4	"Please describe in which order you did the planning for your trip to [DESTINATION]. What was first, what second, and what last?"	Planning question (Warmelink, Vrij, Mann, Leal, et al., 2013)	50	169.95 (103.67)	"The first thing I had to do was check the flights to London. The second was to book the flights according to my schedule."
5	"What was the hardest to plan?"	Planning question (Warmelink, Vrij, Mann, Leal, et al., 2013)	50	111.04 (60.15)	"The hardest thing to plan was booking a hotel. There are so many hotels with so many reviews. It was difficult to choose one and pick the location."
6	"What is the most pleasant event you expect to happen during your trip?"	Emotion-related question (Hauch et al., 2015)	50	105.75 (50.33)	"The most pleasant event that I expect to happen during my trip is to see family that I haven't seen in a couple of years."
7	"What is the most unpleasant event you expect to happen during your trip?"	Emotion-related question (Hauch et al., 2015)	50	104.67 (47.88)	"The most unpleasant event will likely be the travelling part. I will be departing at 6am, so it is likely to be an early morning."
8	"If you have to wait during your journey, for example in the airport or changing train stations, what will you do while you're waiting?"	Transportation question (Warmelink et al., 2012)	10	92.78 (47.47)	"While waiting on my journey, I will likely be on my phone or laptop."
9	"How will you get from the airport to your accommodation?"	Transportation question (Warmelink et al., 2012)	10	74.00 (46.53)	"I will travel from the airport to my accommodation via rental car."
10	"What is the first thing you will do when you arrive at your final destination?"	Other specific question	50	94.03 (36.80)	"The first thing I will do when I arrive will be to look for a Starbucks."
11	"What is the first thing you will do when you return home from your trip to [DESTINATION]?"	Other specific questions	50	95.12 (34.38)	"The first thing I will do when I return home is unpack and shower."

Note: The tense of the question was adjusted when the questions pertained to the most recent past flight.

vs. emotion-related vs. transportation vs. other specific). All analyses were conducted with an alpha level of .05.

Results

Descriptive statistics

Table 2 shows descriptive statistics for the final sample.

Manipulation check

There was no difference in the distribution of participants who failed the control question between the flyers in the truthful and deceptive condition, $X^2(1) = 0.07$, $p = .795$, Cramer's $V = 0.05$. A one-way ANOVA on the question expectedness revealed that expectedness differed across

Table 2: Descriptive statistics for participants who reported upon their upcoming flight (means and standard deviation in parentheses).

	Standard		Specific	
	Truthful	**Deceptive**	**Truthful**	**Deceptive**
Final *n*	49	52	60	61
Weeks until flight	5.12 (3.64)	7.08 (7.06)	6.07 (5.48)	5.85 (7.73)
Times visited before	4.24 (5.02)	4.82 (6.89)	4.73 (5.84)	4.70 (7.23)
Motivation	8.00 (1.99)	8.12 (1.83)	7.81 (1.87)	8.23 (1.53)
Failed control question (%)	3.92 (19.60)	3.70 (19.06)	4.76 (21.47)	1.61 (12.70)
Number of words*	195.18 (66.26)	224.33 (112.04)	210.22 (75.52)	214.75 (80.54)
Expectedness general questions	6.84 (2.10)	7.53 (1.72)	7.08 (2.33)	7.57 (1.90)
Expectedness planning questions	5.94 (2.30)	6.10 (2.29)	6.37 (2.30)	6.70 (2.41)
Expectedness emotion-related questions	5.46 (2.47)	6.65 (1.93)	6.24 (1.90)	7.07 (2.25)
Expectedness transportation questions	6.22 (2.22)	6.62 (1.96)	6.43 (2.14)	6.95 (1.85)
Expectedness other specific questions	6.21 (2.25)	6.41 (2.09)	6.26 (2.14)	6.78 (2.26)
Richness of detail (LIWC)	13.93 (3.30)	14.21 (3.75)	14.98 (3.94)	14.96 (3.14)
Average sentence specificity*100	57.41 (32.48)	53.43 (32.60)	54.00 (30.60)	56.20 (32.43)
Named entity-based information specificity*10^4	97.99 (92.31)	118.47 (106.09)	144.87 (115.78)	149.46 (123.12)

Note: *There were no main effects or an interaction between Veracity and Information Protocol for the number of words, all *p*'s > .143.

Question type, F (4, 880) = 13.87, $p < .001$, f = 0.13. Follow-up tests indicated that the general questions were perceived as more expected (M = 7.26, SD = 2.04) than questions of all other types ($M_{collapsed}$ = 6.24, $SD_{collapsed}$ = 1.80, *p*s > .05, see **Table 2**).

Main analysis

For richness of detail, the 2 (Veracity: truthful vs deceptive) by 2 (Information protocol: standard vs specific) between-subjects ANOVA revealed that there was no significant main effect of Veracity, F (1, 218) = 0.07, p = .787, f = 0.02, nor for Information protocol, F (1, 218) = 3.57, p = .060, f = 0.13. This main effect of Information protocol suggests a trend that those who received the instruction to provide specific information (M = 14.97, SD = 3.54) provided more detailed information than those with standard instructions (M = 14.07, SD = 3.53). The interaction between Veracity and Information protocol was not significant, F (1, 218) = 1.00, p = .754, f = 0.07 (**Table 2**).

For information specificity, there was no significant main effect of Veracity, F (1, 218) = 0.70, p = 0.40, f = 0.06, and no significant Veracity by Information protocol interaction, F (1, 218) = 0.28, p = .596, f = 0.04. However, the main effect of Information protocol was significant, F (1, 218) = 6.78, p = .010, f = 0.18, suggesting that those instructed to provide specific information (M = 1.47, SD = 1.19) did in fact provide more information than those with standard instructions (M = 1.09, SD = 1.00).

For sentence specificity, there was no significant main effect of Veracity, F (1, 218) = 0.04, p = .836, f = 0.04, or Information protocol (specific: M = 0.55, SD = 0.31; standard: M = 0.55, SD = 0.32), F (1, 218) = 0.01, p = .941, f = 0.01. The interaction between Veracity and Information protocol not significant either, F (1, 218) = 0.51, p = .475, f = 0.05.

Exploratory analyses

Question expectedness

When we collapsed the question types into expected (i.e., the 2 general questions) versus unexpected (i.e., the 2 planning, 2 emotion-related, 2 transport, and 2 'other' questions), the 2 (Veracity: truthful vs deceptive) by 2 (Question expectedness: expected vs unexpected) ANOVA on the richness of detail revealed only a significant main effect of Question expectedness, F (1, 220) = 10.09, p = .002, f = 0.21. Unexpected questions (M = 14.56, SD = 9.11) resulted in more detailed answers than expected questions (M = 12.93, SD = 8.72). Likewise for information specificity: only a significant main effect of Question expectedness emerged, F (1, 220) = 87.24, $p < .001$, f = 0.63, which revealed that expected questions (M = 23.55, SD = 31.62) elicited a higher information specificity than unexpected ones (M = 8.60, SD = 19.32). For sentence specificity, the same pattern emerged. The significant main effect of Question expectedness, F (1, 220) = 37.91, $p < .001$, f = 0.42, showed that the sentence specificity of the answers was higher for expected

Table 3: Means (*SD*s) of the human-coded variables per Veracity and Information Protocol.

	Standard		Specific	
	Truthful	**Deceptive**	**Truthful**	**Deceptive**
n	24	14	24	29
Richness of detail	1.38 (0.65)	1.93 (1.00)	2.67 (1.88)	2.90 (1.76)
Plausibility	5.33 (1.55)	5.57 (1.83)	5.50 (1.50)	5.24 (1.38)
Complications	2.20 (1.38)	2.29 (1.33)	2.08 (1.59)	2.48 (1.09)
Why-utterances	2.88 (1.39)	2.50 (1.29)	3.42 (1.59)	3.21 (1.42)
How-utterances	3.33 (1.09)	3.36 (1.08)	3.79 (1.41)	3.62 (1.32)
Truthfulness	3.71 (1.57)	4.71 (2.16)	4.83 (1.86)	4.62 (1.76)

Table 4: The richness of detail, average sentence specificity and information specificity per Question type.

	General question		Planning question		Emotion-related question		Transportation question		Other specific question	
	T	**D**	**T**	**D**	**T**	**D**	**T**	**D**	**T**	**D**
Richness of detail (LIWC)	12.92 (9.24)	12.94 (8.21)	16.52 (8.54)	16.36 (7.91)	14.48 (9.85)	14.87 (9.60)	12.47 (10.17)	11.28 (8.94)	14.05 (8.26)	15.62 (8.21)
Average sentence specificity*100	14.91 (21.47)	16.46 (22.73)	15.75 (24.42)	17.07 (24.35)	9.26 (17.98)	9.14 (18.97)	7.54 (15.87)	8.87 (19.56)	6.91 (15.41)	8.01 (17.44)
Named entity-based information specificity*100	22.53 (32.60)	24.54 (30.68)	13.56 (27.76)	13.96 (20.31)	6.79 (14.81)	9.07 (20.01)	5.35 (15.88)	5.90 (18.76)	6.75 (17.88)	7.38 (13.96)

Note: T = truthful; D = deceptive.

(*M* = 15.70, *SD* = 22.11) than for unexpected questions (*M* = 10.33, *SD* = 19.83).

Human coded variables

Human coders blind to the experimental conditions and hypotheses scored a subset (n = 91)[8] of statements (i.e., those who fly within no more than four weeks) on plausibility, complications, richness of detail, why-utterances, and truthfulness on 7-point Likert scales. For each statement that was coded by the two coders, we used an odd-even split to determine which scoring to use for the analysis. We conducted 2 (Veracity: truthful vs. deceptive) by 2 (Information Protocol: standard vs. specific) between-subjects ANOVAs on each of the five variables (**Table 3**). There was a significant main effect of Information Protocol for richness of detail, $F(1, 87) = 12.32$, $p < .001$, $f = 0.37$;[9] and for why-utterances, $F(1, 87) = 3.97$, $p = .050$, $f = 0.21$.[10] Statements were rated as more detailed and containing more why-utterances when the instructions for participants were to provide specific information. There were, however, no effects of Veracity, Fs < 1.

Question type

To test how the Question type affected the richness of detail of the answers, we added Question type as within-subjects factor and conducted a 2 (Veracity: truthful vs deceptive) by 5 (Question type: general, planning, emotion-related, transport, other) ANOVA on the LIWC-scored richness of detail. There was no significant main effect of Veracity, F (1, 220) = 0.07, p = .789, f = 0.02, and no significant Veracity*Question type interaction, F (4, 880) = 1.49, p = .203, f = 0.04. The main effect of Question type was significant, $F(4, 880) = 18.95$, $p < .001$, f = 0.14. **Table 4** shows the means (SDs) per Question type and follow-up contrasts between the different question types. For the average sentence specificity, the same pattern emerged with only a significant main effect of Question type, F (4, 1408) = 40.11, $p < .001$, f = 0.20; as well as for information specificity with a significant main effect of Question type, $F(4, 880) = 48.49$, $p < .001$, $f = 0.22$.

Past events versus future intentions

We examined exploratory whether the temporal dimension of the flight moderated the richness of detail of participants' answers and potentially the effect of Veracity. A 2 (Veracity: truthful vs deceptive) by 2 (Temporality: past flight vs. upcoming flight) between-subjects ANOVA on the LIWC-based richness of detail revealed only a significant main effect of Temporality, F (1, 350) = 9.80, p = .002, f = 0.17. Answers about past flights regardless of Veracity contained more detailed information (M = 15.80, SD = 3.57) than answers about upcoming flights (M = 14.56, SD = 3.55). There was no such effect for average sentence specificity or information specificity.

Temporal immediacy of intentions

To test whether there was a relation between the immediacy of flying (i.e., how long away in the future/past the flight was) and the richness of detail, we included the number of weeks until/after the flight in the ANOVA model. There was no significant main effect of or interaction with the number of weeks, all *p*s > .05.

Discussion

In this study, we examined whether computer-automated verbal content analysis could differentiate between participants who provided truthful or deceptive statements about their upcoming flight. To address challenges of large-scale applicability, we tried to adopt an online data collection process and a computer-automated analytical approach with natural language processing tools to model the richness of detail of statements. Our core hypothesis was that truthful statements contained more detailed information than false statements and that this might be moderated by the instructions given to participants.

The data did not support the hypothesis that truthful statements contain more detailed information than false statements which is in contrast to some previous intentions studies (Sooniste et al., 2015; Warmelink, Vrij, Mann, & Granhag, 2013, but see Sooniste et al., 2013). Those studies found that truthful statements tended to be richer in detail than false statements. In our data, none of the dependent variables indicated a significant main effect of the veracity of the answers given. Our results showed a trend in support of the hypothesis that promoting specific answers resulted indeed in slightly more detailed and specific answers than giving standard instructions. These findings corroborate the information protocol hypothesis (Nahari et al., 2014b): promoting specific answers did seem to result in more specific answers, although not to the effect of eliciting differences between truthful and deceptive answers. In exploratory analyses, human judgments of the statements corroborate the finding that the information protocol manipulation facilitated the elicitation of information. However, the gain in information was found not to be conducive to identifying deceptive and truthful statements. Likewise, regarding the types of questions asked, unexpected questions resulted in more information than expected ones but did not facilitate the detection of deceptive or truthful reviews. The information gain due to unexpected questions was found for the named entity and the sentence specificity operationalization but not for the LIWC composite variable of "richness of detail". These contradictory findings would need further corroboration, but one explanation could be that the LIWC is less suitable for modeling the richness of detail than named entities or sentence specificity. A comparative analysis of these three operationalizations indeed showed that the LIWC richness of detail was less appropriate for modeling the theoretical lines of verbal deception theory than the other two (Kleinberg et al., 2017).

In several ways, the results from this study are not in line with previous studies on the detection of false intentions (Mac Giolla et al., 2013; Vrij et al., 2011). We will first discuss limitations related to the experimental design and data collection and then elaborate on those related to the theory, data analysis and operationalization of constructs.

Experimental design and data collection

There are some lessons learned from the current study. First, our setting was non-interactive whereas previous studies within the verbal deception paradigm on intentions used face-to-face interview settings (e.g., Jupe et al., 2017; Sooniste et al., 2013; Warmelink, Vrij, Mann, & Granhag, 2013, but see Bogaard, Meijer, Vrij, & Merckelbach, 2016). Our data collection procedure may have affected the statements in two ways. That the participants were merely filling in forms in our study implies that the interviewing process was passive (i.e., without an interviewer as conversation counterpart) and anonymous rather than actively engaging the interviewees. Moreover, the flow of the questions was pre-scripted and non-dynamic. Such a static interviewing precluded the possibility of asking follow-up questions or providing clarifications.

Second, contrary to the vast majority of studies on verbal deception detection (see Vrij Fisher, & Blank, 2015), in the current experiment there was no interviewer present and hence no time pressure for the interviewee to reply. From a theoretical perspective, it is possible that the assumption that additional "cognitive load" makes lying harder than telling the truth (Vrij et al., 2015; Zuckerman, DePaulo, & Rosenthal, 1981) might be moderated by the temporal immediacy of the answers. For example, in a face-to-face interview, an interviewee may be inclined to respond rather quickly to avoid any irregularities in the conversation. On the contrary, if there is no interviewer, there is no time pressure, and therefore, it may seem irrelevant to the interviewee how long they take to reply. Although there was no difference in the response time (see Appendix), future research could shed light on the effect of time pressure in typed statements.

Third, a critical assumption made by us, inspired by previous studies (Sooniste et al., 2013; Warmelink et al., 2012), was that planning and transportation questions, in particular, would be perceived as unexpected. The unexpectedness should have put truth tellers in an advantage of being able to report on their genuine trip freely. Our data suggest that this assumption was only partially met: participants did indicate that the general questions were more expected than all others, but there was no difference between the questions of the remaining four topics. In Sooniste et al.'s (2013) study, the general (intention-related) question was perceived as less difficult than the planning-related questions. Those general questions did not result in any truth-lie differences in richness of detail, whereas questions on the planning phase did. Interestingly, part of these findings from Sooniste et al. (2013) can be found in the current experiment as well: no differences emerged for 'general questions', although in contrast to Sooniste et al. (2013) we did not observe the differences for the planning-related questions either. The nuances in the interplay between question expectedness and difficulty would be an interesting avenue for future research on intentions.

For example, it remains unclear how planning-questions differ from transportation-questions. Similarly, it has been suggested that the perceived difficulty of questions might moderate their effectiveness (Fenn et al., 2015). If an unanticipated question is equally difficult for a truth-teller as for a liar (e.g., because no concrete plans exist yet about a future event), this might put truth-tellers at risk of providing an unbelievable answer. For future studies on question expectedness, it will be worthwhile looking at the perceived difficulty as well as the perceived expectedness of questions.

Fourth, the information protocol manipulation we used could have worked in two opposite directions. By instructing participants in the standard information protocol condition to provide *as much detail as possible*, it is imaginable that this gave the participants, especially the liars, a hint that richness of detail is a cue of interest. The information protocol has been shown to work with instructing participants to provide as much *verifiable* information as possible (Nahari et al., 2014b). However, research by Nahari and Pazuelo (2015) suggests that the information protocol pointing towards detailed (rather than verifiable) information might diminish truth-lie differences. In the current study, the beneficial effect of the 'specific' instructions on the detectability of the statements' veracity could have been canceled out by the detrimental effect of the 'as detailed as possible' instructions. Although the instruction to provide *as detailed answers as possible* has been used as interviewing tool in other studies (Sooniste et al., 2013), further research should try to adopt novel ingredients for verbal deception detection like the model statement technique (Leal, Vrij, Warmelink, Vernham, & Fisher, 2015) for the detection of intention.

Fifth, our participants often did not have immediate and specific intentions yet, as was the case in most previous studies (Sooniste et al., 2013, 2015). On average, the flight upon which participants reported was five to seven weeks away, whereas in previous intentions studies participants had direct plans to implement their intention on the spot (Sooniste et al., 2013; Suchotzki et al., 2013; Warmelink, Vrij, Mann, & Granhag, 2013). The lack of implementation intentions might have put both liars and truth tellers in the difficult situation when asked to provide information regarding a flight that they had not yet planned in sufficient detail (Fenn et al., 2015). One way to address this limitation is in-vivo studies directly at the airport which allows for direct intentions.

Verbal deception detection and deceptive intentions

The current investigation was based on findings from verbal deception theory, in particular, the Reality Monitoring framework and the Undeutsch hypothesis. While the findings reported here might be related to the methodology and design of the experiment (see this discussion section), another explanation might stem from the theory. A small set of experiments reported successful applications of verbal deception detection for intentions (e.g., Sooniste et al., 2013; Warmelink et al., 2012) whereas others found no truth-lie differences in the verbal content (Fenn et al., 2015). The vast body of evidence for the verbal approach on *past* events motivated the theoretical angle of the current study. However, it could be that the detection of intentions represents a boundary condition for the classic verbal deception detection approach. For example, a core assumption is that experiencing an event leaves a memory trace which leads to richer accounts of that event. Not yet experienced activities do not meet that assumption. Sooniste et al. (2013) highlight the role of asking questions about the planning, that is the part of an intention that allows participants to talk about the past. The current experiment incorporated that finding to no beneficial effect on the detection of deception.

More research is needed to map out potential refinements of verbal deception theory for intentions as well as for the development of novel approaches. It is imaginable that for truth-lie differences to emerge the findings of the verifiability approach (Nahari et al., 2014a) could be extended. Parallel to past events, one can argue that genuinely (truthfully) intended actions often entail detailed planning which is accompanied by, for instance, making a car reservation, booking a hotel, or arranging visiting a friend. Liars would be expected to provide fewer checkable information (e.g., contact details of the friend, details on the hotel booking) than truth-tellers due to a risk of being unmasked. Preliminary findings suggest that there is a role for the verifiability notion for the detection of intentions (Jupe et al., 2017) and it is worthwhile exploring that line of inquiry to understand differences in verbal content regarding deceptive and truthful intentions.

Data analysis and operationalization

In our operationalization procedure, it merits attention that we adhered predominantly to fully computer-automated scoring of verbal content. The quantification of qualitative measures (e.g., plausibility) remains a key challenge for social sciences and computational disciplines alike, but there is evidence that this is feasible (Bachenko, Fitzpatrick, & Schonwetter, 2008; Bond & Lee, 2005; for a review see Fitzpatrick, Bachenko, & Fornaciari, 2015). Moreover, the operationalizations used in the current study (esp. speciteller and named entities) were found to discriminate truthful from deceptive statements elsewhere (Kleinberg et al., 2017). Skepticism towards automated text analysis has been voiced elsewhere for context-sensitive scoring tools like Reality Monitoring (Vrij, 2008). The argument is that human coders are more attentive to context-dependency than lexicon approaches like the LIWC (but see Mihalcea & Strapparava, 2009; Newman, Pennebaker, Berry, & Richards, 2003). Although the automated analysis of verbal statements is necessary for quick and scalable applications of the verbal deception detection method, the manual human annotation might offer valuable insights. Human-scored verbal content variables did not reveal differences between truthful and deceptive statements in the current experiment. Moderate to high intra-class coefficients (ICCs 0.67 – 0.86, except for how-utterances) suggest that a poor reliability was not the

cause for these null-findings. An alternative explanation is that the coding procedure was not validly measuring the constructs of, for example, plausibility and richness of detail. There are indications that a frequency count method (i.e., counting the occurrences of details) is better suited than a scaling method as employed here (Nahari, 2016). In spite of the close adherence to the procedure of a related experiment that did find significant truth-lie differences (Sooniste et al., 2013, but see Warmelink, Vrij, Mann, & Granhag, 2013; Warmelink et al., 2012), it might be interesting for further research to test how the frequency count vs. scale method affects deception detection accuracy for intentions.

Future research

Although all of the limitations mentioned above merit the attention of future research on deceptive intentions, we believe an essential requirement for applied purposes is that of large-scale applicability. Research efforts could be directed towards hybrid approaches consisting of elements from remote data collection methods, question expectedness, and verbal deception detection cues. For example, rather than providing participants a form to be filled in, one could develop an instant-messaging framework that asks a set of pre-tested questions to be answered in a semi-interactive online conversation (e.g., Derrick, Meservy, Jenkins, Burgoon, & Nunamaker, 2013; Zhou, 2005). Such a framework would ideally *i)* allow for active information elicitation through interviewee-interviewer interaction, *ii)* give the interviewee a feeling of non-anonymity and accountability through interaction with an interviewer, *iii)* provide higher information gain (i.e., shorter replies with more information), *iv)* facilitate quick interview procedures, and *v)* lay the foundation for automated chatbot-like systems that would be a step towards large-scale applicability of verbal deception detection.

Conclusion

The reported experiment was an attempt to investigate deceptive intentions using a remote data collection procedure and an automated analytical procedure. Participants' truthful or deceptive statements about their upcoming flight did not reveal differences in the verbal content. The study of future behavior may need implementation intentions to be able to determine their veracity. Moreover, future research on large-scale verbal deception detection approaches might want to explore novel paths towards active information elicitation processes, and strategic question approaches at scale.

Notes

[1] Note that power calculations indicate a sample size of 210 for the given parameters (i.e., 53 per cell in a 2 × 2 between-subjects design). We deliberately added 20% to this number for potential drop-out (*n* per cell: 64) in this online study (Kleinberg & Verschuere, 2015, 2016) and multiplied this number by eight given that the overall design was 2 × 2 × 2 between-subjects.

[2] Post-hoc power calculations showed that the final sample size of 354 was indeed sufficient to pick up an effect of $f = 0.25$; achieved power: 0.99.

[3] For the scope of this paper, we focus mainly on the data for participants with upcoming travel plans and use the data for the participants with "past" travel activities only in the exploratory analysis.

[4] Achieved power for $f = 0.25$ and $n = 222$: 0.96.

[5] Due to the low ICC of how-utterances, we did not include this variable further into the analysis.

[6] Of all participants that provided their email address for this, one was chosen at random and was awarded the voucher.

[7] ed The locations and times are anonymized here. The examples are taken from a participant in the truthful condition.

[8] Of the total 110 statements, 16 were randomly chosen for the training of the coders and therefore excluded from the analysis.

[9] The main effect of Information Protocol, $F(1, 87) = 1.48$, $p = .227$, $f = 0.13$; as well as the Veracity*Information Protocol interaction, $F(1, 87) = 0.25$, $p = .616$, $f = 0.05$, were non-significant.

[10] The main effect of Information Protocol, $F(1, 87) = 0.87$, $p = .354$, $f = 0.10$; as well as the Veracity*Information Protocol interaction, $F(1, 87) = 0.07$, $p = .783$, $f = 0.03$, were non-significant.

Additional File

The additional file for this article can be found as follows:

- **Appendix 1.** For further exploratory analyses, we added response time, the number of deletions, and gaps between key-press events as dependent variables. DOI: https://doi.org/10.1525/collabra.80.s1

Competing Interests

The authors have no competing interests to declare.

References

Bachenko, J., Fitzpatrick, E., & **Schonwetter, M.** (2008). Verification and implementation of language-based deception indicators in civil and criminal narratives. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*: 41–48. Association for Computational Linguistics. Retrieved from: http://dl.acm.org/citation.cfm?id=1599087. DOI: https://doi.org/10.3115/1599081.1599087

Bogaard, G., Meijer, E. H., Vrij, A., & **Merckelbach, H.** (2016). Scientific Content Analysis (SCAN) Cannot Distinguish Between Truthful and Fabricated Accounts of a Negative Event. *Frontiers in Psychology*, *7*. DOI: https://doi.org/10.3389/fpsyg.2016.00243

Bond, G. D., & **Lee, A. Y.** (2005). Language of lies in prison: Linguistic classification of prisoners' truthful and deceptive natural language. *Applied Cognitive Psychology*, *19*(3): 313–329. DOI: https://doi.org/10.1002/acp.1087

Derrick, D. C., Meservy, T. O., Jenkins, J. L., Burgoon, J. K., & **Nunamaker, J. F.** (2013). Detecting Deceptive Chat-Based Communication Using Typing Behavior and Message Cues. *ACM Transactions on Management Information Systems*, *4*(2): 1–21. DOI: https://doi.org/10.1145/2499962.2499967

Fenn, E., McGuire, M., Langben, S., & **Blandón-Gitlin, I.** (2015). A reverse order interview does not aid deception detection regarding intentions. *Frontiers in Psychology*, *6*. Retrieved from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4553365/. DOI: https://doi.org/10.3389/fpsyg.2015.01298

Fitzpatrick, E., Bachenko, J., & **Fornaciari, T.** (2015). *Automatic detection of verbal deception*, *8*. Morgan & Claypool Publishers. Retrieved from: http://www.morganclaypool.com/doi/abs/10.2200/S00656ED1V01Y201507HLT029. DOI: https://doi.org/10.2200/S00656ED1V01Y201507HLT029

Fornaciari, T., & **Poesio, M.** (2013). Automatic deception detection in Italian court cases. *Artificial Intelligence and Law*, *21*(3): 303–340. DOI: https://doi.org/10.1007/s10506-013-9140-4

Granhag, P. A., & **Hartwig, M.** (2014). The Strategic Use of Evidence Technique: A Conceptual Overview. In: Granhag, P.-A., Vrij, A., & Verschuere, B. (Eds.). *Detecting Deception*, 231–251. Chichester, UK: John Wiley & Sons, Ltd. DOI: https://doi.org/10.1002/9781118510001.ch10

Granhag, P. A., & **Knieps, M.** (2011). Episodic future thought: Illuminating the trademarks of forming true and false intentions. *Applied Cognitive Psychology*, *25*(2): 274–280. DOI: https://doi.org/10.1002/acp.1674

Harvey, A. C., Vrij, A., Nahari, G., & **Ludwig, K.** (2016). Applying the Verifiability Approach to insurance claims settings: Exploring the effect of the information protocol. *Legal and Criminological Psychology*, n/a-n/a. DOI: https://doi.org/10.1111/lcrp.12092

Hauch, V., Blandón-Gitlin, I., Masip, J., & **Sporer, S. L.** (2015). Are computers effective lie detectors? A meta-analysis of linguistic cues to deception. *Personality and Social Psychology Review*, *19*(4): 307–342. DOI: https://doi.org/10.1177/1088868314556539

Honnibal, M. (2016). SpaCy (Version 1.3.0). Retrieved from: https://spacy.io/.

Honts, C., & **Hartwig, M.** (2014). Credibility assessment at portals. In: Raskin, D. C., Honts, C., & Kircher, J. (Eds.). *Credibility assessment: Scientific research and applications*, 37–62. Academic Press. DOI: https://doi.org/10.1016/B978-0-12-394433-7.00002-6

Johnson, M. K., Bush, J. G., & **Mitchell, K. J.** (1998). Interpersonal reality monitoring: Judging the sources of other people's memories. *Social Cognition*, *16*(2): 199–224. DOI: https://doi.org/10.1521/soco.1998.16.2.199

Johnson, M. K., & **Raye, C. L.** (1981). Reality monitoring. *Psychological Review*, *88*(1): 67. DOI: https://doi.org/10.1037/0033-295X.88.1.67

Jupe, L. M., Leal, S., Vrij, A., & **Nahari, G.** (2017). Applying the Verifiability Approach in an international airport setting. *Psychology, Crime & Law*, (just-accepted), 1–29. DOI: https://doi.org/10.1080/1068316X.2017.1327584

Kleinberg, B., Arntz, A., & **Verschuere, B.** (in press). Detecting deceptive intentions: Possibilities for large-scale applications. In: Docan-Morgan, T. (Ed.). *The Handbook of Deceptive Communication*.

Kleinberg, B., Mozes, M., Arntz, A., & **Verschuere, B.** (2017). Using named entities for computer-automated verbal deception detection. *Journal of Forensic Sciences.* DOI: https://doi.org/10.1111/1556-4029.13645

Kleinberg, B., Nahari, G., & **Verschuere, B.** (2016). Using the verifiability of details as a test of deception: A conceptual framework for the automation of the verifiability approach. In *Proceedings of NAACL-HLT*, 18–25. Retrieved from: http://www.anthology.aclweb.org/W/W16/W16-0803.pdf. DOI: https://doi.org/10.18653/v1/W16-0803

Kleinberg, B., & **Verschuere, B.** (2015). Memory Detection 2.0: The First Web-Based Memory Detection Test. *PLOS ONE*, *10*(4): e0118715. DOI: https://doi.org/10.1371/journal.pone.0118715

Kleinberg, B., & **Verschuere, B.** (2016). The role of motivation to avoid detection in reaction time-based concealed information detection. *Journal of Applied Research in Memory and Cognition*, *5*(1): 43–51. DOI: https://doi.org/10.1016/j.jarmac.2015.11.004

Köhnken, G. (2004). Statement validity analysis and the 'detection of the truth.' *The Detection of Deception in Forensic Contexts*, 41–63. DOI: https://doi.org/10.1017/CBO9780511490071.003

Leal, S., Vrij, A., Warmelink, L., Vernham, Z., & **Fisher, R. P.** (2015). You cannot hide your telephone lies: Providing a model statement as an aid to detect deception in insurance telephone calls. *Legal and Criminological Psychology*, *20*(1): 129–146. DOI: https://doi.org/10.1111/lcrp.12017

Li, J. J., & **Nenkova, A.** (2015). Fast and Accurate Prediction of Sentence Specificity. In: *AAAI*, 2281–2287. Retrieved from: https://pdfs.semanticscholar.org/69f5/a7032605a88e7bed7bf0c9c2218c5e3f2512.pdf.

Mac Giolla, E., Granhag, P. A., & **Liu-Jönsson, M.** (2013). Markers of good planning behavior as a cue for separating true and false intent: Good planning behavior and true and false intent. *PsyCh Journal*, *2*(3): 183–189. DOI: https://doi.org/10.1002/pchj.36

Masip, J., Sporer, S. L., Garrido, E., & **Herrero, C.** (2005). The detection of deception with the reality monitoring approach: a review of the empirical evidence. *Psychology, Crime & Law*, *11*(1), 99–122. DOI: https://doi.org/10.1080/10683160410001726356

Mihalcea, R., & **Strapparava, C.** (2009). The lie detector: Explorations in the automatic recognition of deceptive language. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, 309–312. Association for Computational Linguistics. Retrieved from: http://dl.acm.org/citation.cfm?id=1667679. DOI: https://doi.org/10.3115/1667583.1667679

Nahari, G. (2016). When the long road is the shortcut: a comparison between two coding methods for content-based lie-detection tools. *Psychology, Crime & Law*, *22*(10): 1000–1014. DOI: https://doi.org/10.1080/1068316X.2016.1207770

Nahari, G., & **Pazuelo, M.** (2015). Telling a convincing story: Richness in detail as a function of gender and information. *Journal of Applied Research in Memory*

and Cognition, *4*(4): 363–367. DOI: https://doi.org/10.1016/j.jarmac.2015.08.005

Nahari, G., & **Vrij, A.** (2014). Are you as good as me at telling a story? Individual differences in interpersonal reality monitoring. *Psychology, Crime & Law*, *20*(6): 573–583. DOI: https://doi.org/10.1080/1068316X.2013.793771

Nahari, G., Vrij, A., & **Fisher, R. P.** (2014a). Exploiting liars' verbal strategies by examining the verifiability of details. *Legal and Criminological Psychology*, *19*(2): 227–239. DOI: https://doi.org/10.1111/j.2044-8333.2012.02069.x

Nahari, G., Vrij, A., & **Fisher, R. P.** (2014b). The Verifiability Approach: Countermeasures Facilitate its Ability to Discriminate Between Truths and Lies: The verifiability approach and countermeasures. *Applied Cognitive Psychology*, *28*(1): 122–128. DOI: https://doi.org/10.1002/acp.2974

Newman, M. L., Pennebaker, J. W., Berry, D. S., & **Richards, J. M.** (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, *29*(5): 665–675. DOI: https://doi.org/10.1177/0146167203029005010

Ott, M., Cardie, C., & **Hancock, J. T.** (2013). Negative Deceptive Opinion Spam. In *HLT-NAACL*, 497–501. Retrieved from: http://www.aclweb.org/website/old_anthology/N/N13/N13-1.pdf#page=535.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & **Blackburn, K.** (2015). *The development and psychometric properties of LIWC2015*. Retrieved from: https://repositories.lib.utexas.edu/handle/2152/31333.

Pérez-Rosas, V., & **Mihalcea, R.** (2014). Cross-cultural Deception Detection. In: *ACL (2)*, 440–445. Retrieved from: http://www.anthology.aclweb.org/P/P14/P14-2072.pdf. DOI: https://doi.org/10.3115/v1/P14-2072

Shaw, D. J., Vrij, A., Leal, S., Mann, S., Hillman, J., Granhag, P. A., & **Fisher, R. P.** (2013). Expect the unexpected? Variations in question type elicit cues to deception in joint interviewer contexts. *Applied Cognitive Psychology*, *27*(3): 336–343. DOI: https://doi.org/10.1002/acp.2911

Sooniste, T., Granhag, P. A., Knieps, M., & **Vrij, A.** (2013). True and false intentions: asking about the past to detect lies about the future. *Psychology, Crime & Law*, *19*(8): 673–685. DOI: https://doi.org/10.1080/1068316X.2013.793333

Sooniste, T., Granhag, P. A., Strömwall, L. A., & **Vrij, A.** (2015). Statements about true and false intentions: Using the Cognitive Interview to magnify the differences. *Scandinavian Journal of Psychology*, *56*(4): 371–378. DOI: https://doi.org/10.1111/sjop.12216

Verschuere, B., Kleinberg, B., & **Theocharidou, K.** (2015). RT-based memory detection: Item saliency effects in the single-probe and the multiple-probe protocol. *Journal of Applied Research in Memory and Cognition*, *4*(1): 59–65. DOI: https://doi.org/10.1016/j.jarmac.2015.01.001

Vrij, A. (2008). Reality Monitoring. In *Detecting lies and deceit: Pitfalls and opportunities* (2nd ed.). John Wiley & Sons. Retrieved from: https://books.google.nl/books?hl=en&lr=&id=20pg76wmAucC&oi=fnd&pg=PR7&dq=Detecting+Lies+and+Deceit:+Pitfalls+and+Opportunities.&ots=weuu8nJljc&sig=DUiBNG-H7a9vNuPnNQiKsYGwtqI.

Vrij, A. (2015). Verbal Lie Detection tools: Statement validity analysis, reality monitoring and scientific content analysis. In *Detecting deception: Current challenges and cognitive approaches*, (1st ed.), pp. 3–35. John Wiley & Sons, Ltd. Retrieved from: https://books.google.nl/books?hl=en&lr=&id=4brlBQAAQBAJ&oi=fnd&pg=RA1-PA3&dq=Verbal+Lie+Detection+tools:+Statement+validity+analysis,+reality+monitoring+and+scientific+content+analysis&ots=4sFTBKx24S&sig=5lA5qnbszpbpaGcYokvw8n37ekw.

Vrij, A., Fisher, R. P., & **Blank, H.** (2015). A cognitive approach to lie detection: A meta-analysis. *Legal and Criminological Psychology*, *22*(1): 1–21. DOI: https://doi.org/10.1111/lcrp.12088

Vrij, A., & **Granhag, P. A.** (2012). Eliciting cues to deception and truth: What matters are the questions asked. *Journal of Applied Research in Memory and Cognition*, *1*(2): 110–117. DOI: https://doi.org/10.1016/j.jarmac.2012.02.004

Vrij, A., Granhag, P. A., Mann, S., & **Leal, S.** (2011). Lying about flying: the first experiment to detect false intent. *Psychology, Crime & Law*, *17*(7): 611–620. DOI: https://doi.org/10.1080/10683160903418213

Vrij, A., Granhag, P. A., & **Porter, S.** (2010). Pitfalls and Opportunities in Nonverbal and Verbal Lie Detection. *Psychological Science in the Public Interest*, *11*(3): 89–121. DOI: https://doi.org/10.1177/1529100610390861

Vrij, A., Leal, S., Granhag, P. A., Mann, S., Fisher, R. P., Hillman, J., & **Sperry, K.** (2009). Outsmarting the liars: The benefit of asking unanticipated questions. *Law and Human Behavior*, *33*(2): 159–166. DOI: https://doi.org/10.1007/s10979-008-9143-y

Warmelink, L., Vrij, A., Mann, S., & **Granhag, P. A.** (2013). Spatial and Temporal Details in Intentions: A Cue to Detecting Deception: Spatial and temporal details in lie detection. *Applied Cognitive Psychology*, *27*(1): 101–106. DOI: https://doi.org/10.1002/acp.2878

Warmelink, L., Vrij, A., Mann, S., Jundi, S., & **Granhag, P. A.** (2012). The effect of question expectedness and experience on lying about intentions. *Acta Psychologica*, *141*(2): 178–183. DOI: https://doi.org/10.1016/j.actpsy.2012.07.011

Warmelink, L., Vrij, A., Mann, S., Leal, S., & **Poletiek, F. H.** (2013). The Effects of Unexpected Questions on Detecting Familiar and Unfamiliar Lies. *Psychiatry, Psychology and Law*, *20*(1): 29–35. DOI: https://doi.org/10.1080/13218719.2011.619058

Zhou, L. (2005). An Empirical Investigation of Deception Behavior in Instant Messaging. *IEEE Transactions on Professional Communication*, *48*(2): 147–160. DOI: https://doi.org/10.1109/TPC.2005.849652

Zuckerman, M., DePaulo, B. M., & **Rosenthal, R.** (1981). Verbal and nonverbal communication of deception. *Advances in Experimental Social Psychology*, *14*: 1–59. DOI: https://doi.org/10.1016/S0065-2601(08)60369-X

Peer review comments
The author(s) of this paper chose the Open Review option, and the peer review comments are available at: http://doi.org/10.1525/collabra.80.pr

How to cite this article: Kleinberg, B., Nahari, G., Arntz, A. and Verschuere, B. (2017). An Investigation on the Detectability of Deceptive Intent about Flying through Verbal Deception Detection. *Collabra: Psychology,* 3(1): 21, DOI: https://doi.org/10.1525/collabra.80

Submitted: 06 February 2017 **Accepted:** 28 August 2017 **Published:** 26 September 2017

UNIVERSITY of CALIFORNIA PRESS | Collabra

Collabra: Psychology is a peer-reviewed open access journal published by University of California Press.

OPEN ACCESS