Using the verifiability of details as a test of deception: A conceptual framework for the automation of the verifiability approach

Kleinberg, B.; Nahari, G.; Verschuere, B.

DOI
10.18653/v1/W16-0803

Publication date
2016

Document Version
Final published version

Published in
NAACL HLT 2016 : The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies

Using the verifiability of details as a test of deception: A conceptual framework for the automation of the verifiability approach

Bennett Kleinberg[1], Galit Nahari[2] and Bruno Verschuere[1]
[1] Department of Psychology, University of Amsterdam, The Netherlands.
[2] Department of Criminology, Bar-Ilan University, Ramat Gan, Israel.
{b.a.r.kleinberg,b.j.verschuere}@uva.nl galit.nahari@biu.ac.il

Abstract

The Verifiability Approach (VA) is a promising new approach for deception detection. It extends existing verbal credibility assessment tools by asking interviewees to provide statements rich in verifiable detail. Details that *i)* have been experienced with an identifiable person, *ii)* have been witnessed by an identifiable person, or *iii)* have been recorded through technology, are labelled as verifiable. With only minimal modifications of information-gathering interviews this approach has yielded remarkable classification accuracies. Currently, the VA relies on extensive manual annotation by human coders. Aiming to extend the VA's applicability, we present a work in progress on automated VA scoring. We provide a conceptual outline of two automation approaches: one being based on the *Linguistic Inquiry and Word Count* software and the other on rule-based shallow parsing and named entity recognition. Differences between both approaches and possible future steps for an automated VA are discussed.

1 Cognitive deception detection

Based on the rationale that the default setting in human communication is honesty (Levine, 2014; Verschuere & Shalvi, 2014), the cognitive approach to deception (e.g. Zuckerman et al., 1981) postulates that the act of lying requires extra mental effort compared to telling the truth (e.g. trying to fabricate a convincing lie; Vrij, 2014). The general idea of lying being correlated to increased mental effort has been corroborated by a great body of research, using self-reports, behavioral, autonomic, electrophysiological, and neural measures (Ganis et al., 2003; Verschuere et al., 2011). To further increase cognitive differences between lying and truth telling, the cognitive approach advocates applying minimal interventions in information-gathering interviewing situations that enlarge the differences between the truth tellers and liars (e.g. asking unexpected questions, asking to recall a story in reverse order; Vrij et al., 2015, Meissner et al., 2012). A recent meta-analysis indicates that cognitive techniques outperform standard interviews (Vrij et al., 2015). This body of work has also found the most reliable differences between truths and lies to be manifested in verbal rather than nonverbal behavior. Ormerod and Dando (2014), for instance, applied cognitive interviewing techniques on mock-passengers and found verbal detection methods to by far outperform its behavioral counterparts (e.g. spotting suspicious behavior). Also, objective judgments (i.e., algorithmic scoring such as discriminant analysis) outperformed human judgments (truths: 60% vs. 80%; lies: 64% vs. 73%, for human vs. objective judgments, respectively, Vrij et al., 2015). The superiority of objective criteria might be explained by the sheer amount of information for humans to take into account to derive a binary truth versus lie judgment (e.g. Rubin & Conroy, 2012).

Proceedings of NAACL-HLT 2016, pages 18–25,
San Diego, California, June 12-17, 2016.

The verbal content of statements seems to offer potential for cognition-based deception detection. Verbal deception detection offers multiple levels of analysis (e.g. overall content of statements, lexical analysis, syntactic analysis, see Fitzpatrick et al., 2015) and the most promising results of deception research fall under the umbrella term of verbal deception detection (e.g. Ott et al., 2011, 2013; Mihalcea et al., 2013; Harvey et al., 2016). Unlike other approaches, verbal deception detection is suitable for large-scale applications due to its potential for computer-automation. The cognitive approach to increasing the differences between liars and truth tellers provides a theoretical framework for a synthesis of computer-automated approaches and validated information-gathering interviewing techniques.

2 The Verifiability Approach

2.1 Rationale

Information gathering interviews typically ask the interviewee for a detailed account of the events (e.g. *"Describe in as much detail as possible what happened"*). Derived bottom-up from the liars' verbal strategies (Nahari et al., 2014a), the Verifiability Approach (VA) aims to further increase the differences between liars and truth tellers. The VA is based on three assumptions about the liars' dilemma in an interview:

(1) Liars are inclined to mention sufficient details to provide a convincing false account.
(2) Liars try to avoid mentioning those details that can potentially be verified by the interviewer.
(3) As solution to (1) and (2), the liars provide many *non-verifiable details*.

Working from this dilemma, Nahari et al. (2014a) developed a set of criteria deemed appropriate as an indication of the verifiability of a detail given in a statement. Specifically, a detail is categorized as verifiable if at least one of the three criteria below applies:

- The detail describes an activity with an identifiable person.
- The detail describes an activity that has been witnessed by an identifiable person.
- The detail describes an activity that may have been documented or recorded through technology.

Type of detail	**Example**
Verifiable detail	*"My husband and I parked the car outside our house and noticed our neighbour's daughter kissing her boyfriend good-bye on their doorstep."* [with an identifiable person] *"There were 3 people present who witnessed me use my phone before placing it in my bag at 22:00: Ben Kleinberg, Galit Nahari, and Bruno Verschuere."* (names changed) [witnessed by an identifiable person] *"My phone was last used at 12.30 am in which I sent a text message to my friend who was with me that evening."* [recorded through technology]
Non-verifiable detail	*"When I got back to my house I realised that my phone was no longer in my pocket, but had gone."* *"I spoke to a lady in the shop about the general chit chat of an airport, where she was going to and when her flight was."* *"A guy offered to buy me a drink after noticing it was my birthday."*

Table 1: Illustration of verifiable (criterion in square brackets) and non-verifiable details (verbatim from Vrij et al., 2016).

Table 1 shows verbatim examples for each category from the most recent VA study (Vrij et al., 2016).

2.2 Verifiable details and verifiable facts

Hancock et al. (2005) outline that liars use more details when the nature of the deception permits it (i.e., when the narrative *per se* is non-verifiable). For example, opinions are inherently non-verifiable fact

Figure 1: Number of details per category and condition in Nahari et al. (2014a).

scenarios. In contrast, event-based deception settings like mock-crimes are verifiable fact scenarios. In this sense, verifiable facts are interwoven with the ground truth of a deception study: established ground truth provides verifiable facts for the researcher regarding the narrative, and vice versa.

Interestingly, this important distinction between verifiable and non-verifiable facts made by Hancock et al. (2005, 2007), is relevant to the VA in two ways: First, the application of the VA is appropriate only when the scenario is based on theoretically verifiable facts (e.g., a crime). In its current state, the VA is less relevant for non-verifiable fact scenarios (e.g., opinions), as one cannot verify details that are not event-based. Second, Hancock et al. (2005) discuss how liars' verbosity could depend on the verifiability of the overall scenario. In situations where the verifiability or ground truth is difficult to establish (e.g. opinions), the liars may choose to include many details in their statement, whereas this strategy is expected to be counterproductive when the ground truth can be established. The VA extends this idea by actively challenging interviewees in different ways: liars and truth tellers are explicitly asked to provide verifiable details. This technique results in a disproportionally difficult task for the liars, whereas the truth teller can easily recall verifiable details from memory.

3 Experimental findings using the VA

In the initial experiment on the VA, Nahari et al. (2014a) modified a mock-crime procedure by instructing participants to do their normal daily business (e.g. drinking coffee, visiting a book shop) and return to the lab after 30 minutes. Upon returning to the lab, the participants were allocated to the truth-condition or the lie-condition. Those in the truth-condition were instructed to provide a truthful account of their activities in the previous 30 minutes, whereas those in the lie-condition were required to give an entirely false statement. The findings (Figure 1) show that truthful statements contained more overall details ($p < .05$) as a function of the number of verifiable details ($p < .001$). Moreover, the number of details translated to promising classification rates (Table 2). These general findings have been corroborated in several studies (e.g. Nahari & Vrij, 2014; Nahari & Vrij, 2015).

	F-measure true statements	**F-measure false statements**	**Accuracy**
No. of verifiable details	80.95	76.47	78.94
No. of all details	69.57	53.33	63.16
No. of verifiable details/no. of non-verifiable details	73.17	68.57	71.05

Table 2: F-measures and accuracies of the VA for three decision criteria (from Nahari et al., 2014a)

From a methodological point of view, two essential elements in this study are the annotation of details in statements and subsequently the annotation of these details as either verifiable or non-verifiable.

3.1 Annotation of details

In order to extract details from the statements written by the participants, the researchers adopted the Reality Monitoring approach (RM; Sporer, 2004). RM has gained popularity in deception research because it offers a theoretical framework about content differences of true and false statements (Johnson & Raye, 1981; Vrij, 2015). The underlying assumption of RM is that true experiences are obtained through perceptual processes whereas imagined (or false) experiences are obtained through cognitive operations. This in turn is thought to be reflected in, for example, the amount and type of detail when recalling an experience. Specifically, three

of the eight RM criteria are suitable for a verifiability approach (see Nahari et al., 2014a): spatial (e.g. locations or spatial arrangements), temporal (e.g. points in time or sequence of events), and perceptual details (e.g. all sensorial information like visual information and sounds), all which truth tellers are expected to produce more of.

In Nahari et al. (2014a) two independent coders were trained in coding example statements on the three detail criteria for 2.5 hours. Within each statement, the two coders manually annotated all details that were spatial, temporal or perceptual, with an inter-rater reliability of 78%, 77%, and 78%, for perceptual, spatial and temporal details respectively. If the two coders disagreed, a third trained coder made the final decision about the presence of a detail. Consider two examples from Table 1 as illustration:

a) *"My husband and I parked the car outside our house and noticed our neighbour's daughter kissing her boyfriend good bye on their doorstep."*

b) *"When I got back to my house I realised that my phone was no longer in my pocket, but had gone"*

These statements are annotated as:

a) *"*My husband [PERCEPTUAL] *and I* parked [PERCEPTUAL] *the car* outside our house [SPATIAL] *and* noticed our neighbour's daughter [PERCEPTUAL] kissing her boyfriend [PERCEPTUAL] *good-bye* on their doorstep [SPATIAL]*."*

b) *"*When [TEMPORAL] *I got back to* my house [SPATIAL] *I realised that my phone was no longer in my pocket, but had gone"*

Within this statement, all details that fit one of the three RM criteria are annotated. The next step in the annotation is to decide for each detail whether or not it can be deemed potentially verifiable.

3.2 Annotation of the verifiability of details

The annotated details were further classified as verifiable or non-verifiable (Table 1). Similar to the detail annotation, the same two independent coders made a judgment for each detail whether it fit at least one of the three verifiability criteria (see 2.1). The coders agreed on 87.95% of the detail verifiability (Nahari et al., 2014a). Each disagreement was referred to a third coder who made the final decision. After this final annotation, the overall number of details and of verifiable details was subjected to a discriminant analysis. Applied to the two examples, this phase would result in the following annotation:

a) *"*My husband [PERC.-VERIFIABLE] *and I* parked [PERC.-VERIFIABLE] *the car* outside our house [SPATIAL-VERIFIABLE] *and* noticed our neighbour's daughter [PERC.-VERIFIABLE] kissing her boyfriend [PERC.-VERIFIABLE] *good-bye* on their doorstep [SPATIAL-VERIFIABLE]*."*

b) *"*When [TEMP.-NONVERIFIABLE] *I got back to* my house [SPATIAL-NONVERIFIABLE] *I realised that my phone was no longer in my pocket, but had gone"*

3.3 The VA information protocol

Despite the promising initial results, a key challenge to the VA is that liars can embed their lies into mainly true events, e.g., their normal daily routine (Nahari et al., 2014b), or altogether within non-verifiable scenarios (Hancock et al., 2005). For example, when using the VA on participants' statements about false and true insurance claims, initial findings indicated that the VA does not benefit when the lies are embedded in mainly non-verifiable contexts. However, contrary to other content analysis tools (Nahari & Pazualo, 2015) the VA has been shown to allow for higher classification accuracy when the participants were aware of the working mechanisms of the tool. Harvey et al. (2016) manipulated the information liars and truth tellers received about the VA in the insurance claim setting: one group in each condition (truth vs. lie) was told to provide as much detail as possible whereas another group was informed that verifiable details are used as an indicator for truthfulness. The accuracy in the informed group (77.5%) was higher than that in the uninformed group (57.5%). This information protocol manipulation affected liars and truth tellers in

	Detail annotation	Verifiability annotation
Human coding	Manually	Manually
LIWC-based system	Word frequencies of 'perceptual processes', 'space', and 'time'	*Unlikely to be captured*
NER-based system	Shallow parsing of verb phrases	Presence of named entities as proxy for verifiability

Table 3: Conceptual comparison of VA automation approaches.

unequal ways (see 2.2) and it has now become standard procedure in VA research.

3.4 Towards large scale application of VA

Although the VA has yielded promising results in laboratory studies it is limited with regard to its large-scale applicability. Most importantly, the annotation of details and verifiability relies on manual coding making this procedure resource-intensive (e.g. Harvey et al., 2016). In other words, the manual coding of verbal criteria can be seen as a key impediment to large-scale investigations with the VA and potential applications. In the remainder of this paper, we report on a work in progress about the automation of the VA.

4 Related work on automated verbal deception detection

To put our work into perspective we briefly discuss three key studies on automated approaches to verbal deception detection. Zhou et al. (2004) conducted one of the early experimental attempts to automate the detection of deception in a computer-mediated communication eliciting lies in a group decision problem. Crucially, they found that computerized analysis added significantly to the identification of linguistic cues for the detection of deception in asynchronous settings. Mihalcea et al. (2013) answer another relevant question for the broader aim of this investigation. They showed that a data-driven machine learning approach achieved classification accuracies of up to 74% for low-stakes lies (see also Mihalcea & Strappavara, 2009). A study by Bachenko et al. (2008) complements this finding by addressing the critical issue of low-stakes with a linguistic analysis of genuine crime documents. On the one hand, they showed that a theory-based selection of cues can successfully be automated for linguistic annotation of texts, and on the other hand, they were able to develop a tagging system that discriminated true from false declaration within statements. The latter is of particular relevance for the problem of embedded lies.

5 Automating the VA

The main research question guiding this conceptual paper is whether we can automate the VA. As this is the first automated annotation approach of the VA, we will use the data of existing VA studies (e.g. Nahari et al., 2014a; Vrij et al., 2016) which will both be readily available and provide human scoring as baseline. We discuss two automation approaches that could both address the annotation of details but differ in their potential of annotating the verifiability. The first system is based on the *Linguistic Inquiry and Word Count* system (LIWC, Pennebaker et al., 2015) and the second system is a two-phasic annotation approach relying on named entity recognition.

5.1 LIWC-based automation

The LIWC system (Pennebaker et al., 2015) has been applied widely in psycholinguistic research (e.g. Bond & Lee, 2005; Hancock et al., 2007; Ramirez-Esparza et al., 2008). The software analyzes text statements and produces frequency tables of word categories that fit psychological processes (e.g. cognitive mechanism, affect). Bond and Lee (2005) applied LIWC to automate RM criteria in a sample of prisoners. We aim to adopt their approach to modelling the VA detail categories (perceptual, spatial, temporal) with the LIWC word categories 'perceptual processes' (e.g., "saw", "heard"), 'space' (e.g., "down", "under"), and 'time' (e.g., "before", "until"). We will use the frequency of these word categories as proxy for overall detail in existing statements from VA studies (Table 3). Whereas we expect the annotation of details to be feasible in this system (see Bond & Lee, 2015), we

argue that the verifiability is less likely to be captured. The current LIWC system does not provide a word category or output that we think is able to function as proxy for detail verifiability.

5.2 Shallow parsing and NER system

In the second automation approach, we aim to develop a system adopting essentially rule-based shallow parsing (Pradhan et al., 2004) of statement activities and details that will then be coupled with *named entity recognition* (NER; e.g. persons, locations, organizations; Weischedel et al., 2013; Honnibal, 2016; see Appendix A). Specifically, in this first automation attempt, we aim to use the existing VA data and perform part-of-speech tag rule-based shallow parsing to annotate verb phrases as proxy for activities.

As a second step, we add the verifiability annotation using NER. By extracting named entities, we hypothesize to be able to add significantly to the verifiability annotation as compared to the LIWC-based system. It could be possible that the presence of named entities comes close to fulfilling the actual verifiability criteria (i.e., presence of or witnessing by an identifiable person and reference to technology). For example, the statements "I saw a guy in a cafe" contains no named entity, whereas "I saw Dan in the Starbucks" contains two named entities (Dan = PERSON, Starbucks = ORGANIZATION). Note how in both cases, LIWC and the shallow parser would identify the same detail ("saw"), but only the NER-based system would also annotate the two additional named entities. We will investigate whether the NER-added information functions as a proxy for verifiability. We aim to use Cython-based *spaCy* (Honnibal, 2016) software for pre-processing and annotation. The tokenizer, POS-tagger and NER algorithms are trained on the OntoNotes5 corpus (Weischedel et al., 2013).

6 Outlook and conclusion

The conceptual approach to an automation of the VA is thought to eventually enable large-scale applications. Given the novelty of the VA, the approach to its automation is still in its infancy and requires multiple phases of development. For example, it could be that the LIWC-based system and the NER-approach are not mutually exclusive but that both complement each other. A successful automation of the VA could open up new directions of verbal deception research. First, it would enable researchers to conduct VA experiments on large sample sizes and corpora sizes efficiently. Second, an automated VA could have considerable impact on applied deception research. Especially in the area of crime prevention, the practitioners' focus is identifying the few persons out of a large population who may have false intentions. Inherent to such aims is the need for large-scale automated deception detection tools. Third, on a psycholinguistic level, the VA adds an interesting dimension to scoring mechanisms by explicitly looking at the verifiability of details. Future directions could, for example, involve modifications of the VA as a tool to identify propositions subject to ground truth checking or applying the VA to smaller units of analysis like single propositions instead of whole statements. The latter could also be a step towards detecting embedded lies.

In summary, the VA offers a promising framework for the detection of verbal deception and would benefit from automation. Two approaches were outlined, one based on the LIWC software tool and another based on named entity recognition algorithms. Successfully automated approaches of the VA could contribute to novel research paths and to further integration of cognitive deception detection and computational linguistics.

Acknowledgments

This work is supported by a grant from the Dutch Ministry of Security and Justice. We thank three critical and attentive reviewers and the editor for their invaluable input.

References

Bachenko, J., Fitzpatrick, E., & Schonwetter, M. (2008). Verification and implementation of language-based deception indicators in civil and criminal narratives. *Proceedings of the 22nd International Conference on Computational Linguistics*, 41-48.

Bond, G.B., & Lee, A.Y. (2005). Language of lies in prison: Linguistic classification of prisoners' truthful

and deceptive natural language. *Applied Cognitive Psychology, 19*, 313-329. doi:10.1002/acp.1087

Fitzpatrick, E., Bachenko, J., Fornaciari (2015). *Automatic detection of verbal deception*. Morgan & Claypool Publishers.

Ganis, G., Kosslyn, S.M., Stose, S., Thomposon, W.L., & Yurgelun-Todd, D.A. (2003). Neural correlates of different types of deception: An fMRI investigation. *Cerebral Cortex, 13*(8), 830-836. doi: 10.1093/cercor/13.8.830

Hancock, J.T., Curry, L.E., Goorha, S., & Woodworth, M. (2007). On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes, 45*(1), 1-23, doi: 10.1080/01638530701739181

Hancock, J.T., Curry, L., Goorha, S., Woodworth, M. (2005). Automated linguistic analysis of deceptive and truthful synchronous computer-mediated communication. *Proceedings of the 38th Hawaii International Conference on System Sciences*, 1-10.

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:10.1111/lcrp.12092

Honnibal, M. (2016). spaCy (Version 0.100.6) [Computer software]. Available from https://spacy.io/

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*, 67–85. Retrieved from http://www.apa.org/pubs/journals/rev/index.aspx

Levine, T. R. (2014). Truth-Default Theory (TDT): A theory of human deception and deception detection. *Journal of Language and Social Psychology, 33*(4), 378–392. doi:10.1177/0261927X14535916

Meissner, C.A., Redlich, A.D., Bhatt, S., & Brandon, S. (2012). Interview and interrogation methods and their effects on true and false confessions. Campbell Systematic Reviews, 13. Doi: 10.4073/csr.2012.13

Mihalcea, R., & Strapparava, C. (2009). The lie detector: Explorations in the automatic recognition of deceptive language. *In Proceedings of the Association for Computational Linguistics*, 309-312.

Mihalcea, R., Perez-Rosas, V., & Burzo, M. (2013). Automatic detection of deceit in verbal communication. *Proceedings of the 15th ACM on International conference on multimodal interaction*, 131-134.

Nahari, G., & Vrij, A. (2014). Can I borrow your alibi? The applicability of the verifiability approach to the case of an alibi witness. *Journal of Applied Research in Memory and Cognition*, *3*(2), 89–94. doi:10.1016/j.jarmac.2014.04.005

Nahari, G., & Vrij, A. (2015). Can someone fabricate verifiable details when planning in advance? It all depends on the crime scenario. *Psychology, Crime & Law*, *21*(10), 987–999. doi:10.1080/1068316X.2015.1077248

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: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. *Applied Cognitive Psychology*, *28*(1), 122–128. doi:10.1002/acp.2974

Ormerod, T. C., & Dando, C. J. (2014). Finding a Needle in a Haystack: Toward a Psychologically Informed Method for Aviation Security Screening. *Journal of Experimental Psychology: Genera*l, *144*(1), 76–84. doi:10.1037xge0000030

Ott, M., Cardie, C., & Hancock, J.T. (2013). Negative deceptive opinion spam. *Proceedings of NAACL-HLT 2013*, 497–501.

Ott, M., Choi, Y., Cardie, C., & Hancock, J.T. (2011). Finding deceptive opinion spam by any stretch of the imagination. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, 309-319.

Pennebaker, J.W., Booth, R.J., Boyd, R.L., & Francis, M.E. (2015). *Linguistic Inquiry and Word Count: LIWC2015*. Austin, TX: Pennebaker Conglomerates (www.liwc.net).

Pradhan, S., Ward, W., Hacioglu, K., Martin, J.H., & Jurafsky, D. (2004). Shallow semantic parsing using support vector machines. *Proceedings of the Human Language Technology - North American Association for Computational Linguistics*.

Ramirez-Esparza, N., Chung, C.K., Kacewicz, E., & Pennebaker, J.W. (2008). The psychology of word use in depression forums in English and in Spanish: Testing two text analytic approaches. *Association for the Advancement of Artificial Intelligence*, 102-108.

Rubin, V.L., & Conroy, N. (2012). Discerning truth from deception: Human judgments and automation efforts. *First Monday, 17*(3). doi: 10.5210/fm.v17i3.3933

Sporer, S. L. (2004). Reality monitoring and detection of deception. In P. A. Granhag & L. A. Stromwall (Eds.), *The detection of deception in forensic contexts* (pp. 64–102). Cambridge, UK: Cambridge University Press.

Verschuere, B., Ben-Shakhar, G., & Meijer, E. (2011). *Memory detection: theory and application of the Concealed Information Test*. Cambridge, U.K.: Cambridge University Press.

Verschuere, B., & Shalvi, S. (2014). The truth comes out naturally! Does it? *Journal of Language and Social Psychology, 33*, 417-423.

Vrij, A. (2014). Interviewing to detect deception. *European Psychologist, 19,* 184–195. doi:10.1027/1016-9040 a000201

Vrij, A. (2015). Verbal lie detection tools: Statement Validity Analysis, Reality Monitoring and Scientific Content Analysis. In: P.A. Granhag, A. Vrij, and B. Verschuere (Eds.), *Detecting Deception: Current Challenges and Cognitive Approaches*, John Wiley & Sons.

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

Vrij, A., Nahari, G., Isitt, R., & Leal, S. (2016). Using the verifiability lie detection approach in an insurance claim setting. *Journal of Investigative Psychology and Offender Profiling* (Early view). Doi: 10.1002/jip.1458

Weischedel, R., Palmer, M., Marcus, M., Hovy, E., Pradhan, S., Ramshaw, L., Xue, N., Taylor, A., Kaufman, J., Franchini, M., El-Bachouti, M., Belvin, R., & Houston, A. (2013). *OntoNotes Release 5.0* LDC2013T19. Web Download. Philadelphia: Linguistic Data Consortium, 2013. Retrieved from https://catalog.ldc.upenn.edu/LDC2013T19.

Zhou, L., Burgoon, J., Nunamaker, J., & Twitchell, D. (2004). Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communication. *Group Decision and Negotiation, 13*, 81-106.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. New York, NY: Academic Press.

Appendix A

Entities as recognized by spaCy NER algorithm (adopted from https://spacy.io/docs#annotation; most relevant categories in bold):

- **PERSON** (e.g. Bill Gates)
- NORP (nationalities or religious groups)
- **FACILITY** (e.g. Heathrow Airport)
- **ORG** (organization, e.g. Starbucks)
- **GPE** (countries, states, cities)
- **LOC** (locations other than GPE)
- PRODUCT (e.g. vehicles, food)
- **EVENT** (e.g. wars, sports events)
- WORK_OF_ART (title of books, songs)
- LAW (named documents made into laws)
- LANGUAGE (any named language)
- **DATE** (e.g. last week)
- **TIME** (e.g. for 2 hours)
- PERCENT (e.g. 85%)
- QUANTITY (e.g. 15kg, 42km)
- ORDINAL (e.g. first, second, 15th)
- CARDINAL (other numerals)