Applied Psycholinguistics (2024), 1–29
doi:10.1017/S0142716424000298

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL ARTICLE

Honesty repeats itself: comparing manual and automated coding on the veracity cues total details and redundancy

Haneen Deeb[1], Aldert Vrij[1], Nicola Palena[2], Petra Hypšová[3], Gerges Dib[4], Sharon Leal[1] and Samantha Mann[1]

[1]Department of Psychology, University of Portsmouth, Portsmouth, Hampshire, UK, [2]Netherlands Institute for the Study of Crime and Law Enforcement, Amsterdam, Netherlands, [3]Department of Psychology, Palacký University, Olomouc, Czech Republic and [4]Amazon.com, Inc., Seattle, Washington, USA
Corresponding author: Haneen Deeb; Email: haneen.deeb@port.ac.uk

(Received 11 November 2023; revised 16 July 2024; accepted 22 July 2024)

Abstract

Lie detection research comparing manual and automated coding of linguistic cues is limited. In Experiment 1, we attempted to extend this line of research by directly comparing the veracity differences in manual coding and two coding software programs (Text Inspector and Linguistic Inquiry and Word Count [LIWC]) on the linguistic cue "total details" across eight published datasets. Mixed model analyses revealed that LIWC showed larger veracity differences in total details than Text Inspector and manual coding. Follow-up classification analyses showed that both automated coding and manual coding could accurately classify honest and false accounts. In Experiment 2, we examined if LIWC's sensitivity to veracity differences was the result of honest accounts including more redundant (repeated) words than false accounts as LIWC—but not Text Inspector or manual coding—accounts for redundancy. Our prediction was supported, and the most redundant words were function words. The results implicated that automated coding can detect veracity differences in total details and redundancy, but it is not necessarily better than manual coding at accurately classifying honest and false accounts.

Keywords: Automated coding; lie detection; linguistics; manual coding; redundancy; total details

Honesty repeats itself: comparing manual and automated coding on the veracity cues total details and redundancy

In lie detection experiments, honest and false accounts are commonly compared for verbal cues by analyzing data that is manually coded by humans (Chan & Bull, 2014; Deeb et al., 2017; Leal et al., 2010). Verbal cues are indicators that are based on the content of speech (Vrij, Granhag, et al., 2022). Recently, researchers started using automation to code data (Dzindolet & Pierce, 2005; Feldman et al., 2017; Newman et al., 2003), but only a few experiments directly compared manual and automated coding (Bogaard, van der Mark, et al., 2019; Jupe et al., 2018; Kleinberg et al., 2017; Kleinberg, Warmelink et al., 2018; Schutte et al., 2021; Vrij et al., 2007). Both manual (Nahari, 2018; O'Connell et al., 2023) and automated (Bond et al., 2017; van der Zee et al., 2022) coding can distinguish honest from false accounts. While manual coding is dominant in lie detection research, it is often criticized for being subjective (Kleinberg & Verschuere, 2021). Hence, automated coding was recommended as an objective alternative (Tomas et al., 2022).

In the present paper, we report two experiments. In Experiment 1, we compared manual coding with two coding software programs (Linguistic Inquiry and Word Count [LIWC] and Text Inspector) to test which method performs better at discriminating between honest and false accounts based on the verbal cue "total details." Experiment 2 was developed based on the findings of Experiment 1. LIWC detected differences in total details to a greater extent than Text Inspector and manual coding in Experiment 1, so in Experiment 2, we examined whether this was caused by the fact that LIWC—but not Text Inspector or manual coding—counts redundant (repeated) words in the same text. We also looked at the type of words that were most redundant.

Theoretical approaches to total details as a verbal veracity cue

Several theoretical approaches have been posited for explaining why truth tellers typically report more details than lie tellers. We elaborate below on the information management approach, the cognitive approach, and the Reality Monitoring (RM) approach. The information management approach postulates that both truth tellers and lie tellers attempt to control their verbal behavior during interviews, but they do this in a different manner (Granhag & Hartwig, 2008). Whereas truth tellers are generally forthcoming, lie tellers are more likely to focus on what information they will provide and on what information they will leave out (Hartwig et al., 2010). Lie tellers are more concerned than truth tellers about contradicting themselves (Deeb et al., 2024; Granhag & Strömwall, 1999; Strömwall & Willén, 2011), giving away information that may uncover their lie, and/or failing to remember information they previously reported (Toma & Hancock, 2012; Vrij et al., 2010). Thus, lie tellers provide "enough" information to appear honest while refraining from providing information that gives leads to their lie (Hines et al., 2010; McCornack, 1992).

According to the cognitive approach, lie telling is a cognitively engaging process (Vrij et al., 2011, 2017). Lie tellers have to suppress the truth, to think of and update their lies while responding to the interviewer, to control their verbal and nonverbal behavior, and to observe the interviewer's behavior to assess if they are providing convincing responses (DePaulo et al., 2003; Vrij, 2008). This exhausts the cognitive resources of lie tellers and makes them provide simple and short accounts (Vrij, Granhag, et al., 2022).

According to the RM approach (Johnson & Raye, 1981), experienced events contain more perceptual and contextual (spatial and temporal) information than imagined events. Truth tellers report about experienced events, whereas lie tellers report at least partially imagined events (Leins et al., 2013; Verigin et al., 2019). Hence, truth tellers' accounts should be richer in details than lie tellers' accounts (Amado et al., 2016; Bogaard, Colwell, et al., 2019; DePaulo et al., 2003; Sporer & Sharman, 2006). Research has shown that truth tellers not only provide more perceptual and contextual details than lie tellers (Harvey et al., 2017; Oberlader et al., 2016) but also other types of details such as person, location, action, temporal, and object (PLATO) details (Deeb, Vrij, Leal, & Mann, 2021; Deeb et al., 2022c; Leal, Vrij, Deeb, et al., 2018).

Overall, the theories implicate that truth tellers report more details than lie tellers which was corroborated in previous meta-analyses (Amado et al., 2016; Gancedo et al., 2021). However, not all researchers reached this conclusion (e.g., Vrij et al., 2018). Some factors may moderate the differences between truth tellers and lie tellers on total details, including culture or country in which the data is collected (Western vs. non-Western; Leal, Vrij, Vernham, et al., 2019), interview protocol (experimental vs. control; Bogaard et al., 2020), metric used (Schutte et al., 2021), coding scheme (Nahari, 2023), and coding method (manual vs automated; Kleinberg, van Toolen, et al., 2018).

Manual coding in lie detection research

In the present research, we were particularly interested in comparing honest and false accounts derived from interviews in mock forensic settings. The common experimental procedure in these types of experiments is that truth tellers honestly discuss activities they performed (e.g., Leins et al., 2012) or a video footage (e.g., Leal et al., 2023) they watched, whereas lie tellers make up details about these activities or video footage. The interviewees' responses are then transcribed and manually coded by human raters for specific verbal cues. The raters would either count the number of verbal cue(s) in the transcripts or rate on a scale the extent to which they think the cue(s) emerged. While the counting method is considered to be more objective than the rating method, it is also more labor-intensive, particularly if the rater examines more than one verbal cue (Nahari, 2016).

For any manual coding method, at least two raters are needed to measure inter-rater reliability. In many instances, inter-rater reliability is not high enough due to the subjectivity in coding between human raters (Tomas et al., 2022). Differences between raters frequently lead to a replicability problem in which subsequent research cannot replicate the original results (Kleinberg, Warmelink, et al., 2018). Thus, some researchers started recommending the use of automated coding as an alternative (Kleinberg & Verschuere, 2021; Plotkina et al., 2020; Tomas et al., 2022).

Automated coding in lie detection research

Automated coding is usually conducted via software programs that analyze transcribed interviews through a linguistic approach (Bond et al., 2017; Hauch et al., 2015). This can be done in different ways, including by providing a count of specific lexical categories (Pennebaker et al., 2015) or by deriving sentence specificity (Kleinberg, Mozes, et al., 2018). Unlike manual coding, automated coding is a faster way of coding large amounts of text. Also, automated coding has been recommended as an alternative to protect against human biases and subjective ratings (Tomas et al., 2022).

Automated coding seems to distinguish honest from false accounts well on different verbal cues (Newman et al., 2003). It has been suggested that its classification accuracy rate is approximately 70% and similar to manual coding (Bond & Lee, 2005; Mbaziira & Jones, 2016; Tomas et al., 2022). However, one meta-analysis (Hauch et al., 2015) that compared honest and false accounts when coding software programs were used showed that the effect sizes were small. The authors believed that the small effect sizes could be due to the inability of the software to consider the semantic context, a limitation of automated coding. Further, the authors found that different coding software programs yielded different results, depending on how the software was devised to code verbal cues. Thus, similar to manual coding, coding software programs can produce different conclusions.

Experiments comparing automated and manual coding revealed conflicting results. Whereas some experiments showed that manual coding outperformed automated coding (Vrij et al., 2007), others found the opposite effect (Plotkina et al., 2020). One experiment (Schutte et al. 2021) that compared automated and manual coding in four different datasets did not find any significant differences between the two coding methods when the same metrics were analyzed. The metrics were either frequency scores (raw number of specific details within an account) or percentage scores (number of specific details compared to the total number of details within an account). Overall, researchers have suggested that the contradictory results across experiments may be moderated by different factors, including manual coding schemes, coding software programs, and metric analyses (Kleinberg et al., 2017, Schutte et al., 2021; Vrij et al., 2007).

Experiment 1

In Experiment 1, we aimed to compare manual and automated coding using recently created datasets (see Table 1). We were particularly interested in total details as a veracity cue, because meta-analyses have shown that this cue had a larger magnitude of effect compared to most other tested verbal cues (Amado et al., 2016; DePaulo et al., 2003). Although this cue has been widely examined, its diagnosticity has yet to be determined when manual and automated coding are directly compared across multiple datasets. We were specifically interested in comparing the manual coding of total details to (a) a software program (LIWC) that is widely researched but that significantly differs from manual coding when coding total details and (b) another software program (Text Inspector) that has not been previously tested but that codes total details similar to manual coding. The automated coding and analyses were carried out in October 2023.

Manual coding of the transcripts

In the original experiments, manual coding schemes differentiated between total details and total word count. Only informative details were coded in a single transcript and the total number of informative details was computed toward the

https://doi.org/10.1017/S0142716424000298 Published online by Cambridge University Press

Table 1. Summary of the datasets and conditions included in the present paper

Experimental procedure	Present paper
Deeb et al. (2020): A total of 243 participants in the United Kingdom (UK) reported either a truthful or a false out of the ordinary memorable event. Participants were interviewed three times, immediately (T1), after one week (T2), and after two weeks (T3). They were exposed to the Model Statement (MS) either at T1 only, at T2 only, at T1 and T2, or not at all (control condition). All participants provided a free recall at T3. All three interviews involved only one free recall question. The experiment was conducted face-to-face. *Dependent variables:* Total details, core details, peripheral details, complications, common knowledge details (CKD), self-handicapping strategies (SHS). *Inter-rater reliability for total details:* Intra-class correlation coefficient (ICC) = 0.74	Analyses were conducted for the T1 free recall question on data from participants who were not exposed to the MS at T1 (i.e., control and T2 conditions). n = 122, of which 61 were lie tellers.
Deeb, Vrij, Leal, & Burkhardt (2021): A total of 243 participants in the UK reported either a truthful or a false out of the ordinary memorable event. Participants were interviewed three times, immediately (T1), after one week (T2), and after two weeks (T3). They were asked to sketch and narrate either at T1 only, at T2 only, at T1 and T2, or not at all (control condition). All participants provided a free recall at T3. All three interviews involved only one free recall question. The experiment was conducted face-to-face. *Dependent variables:* Total details, core details, peripheral details, complications, CKD, SHS, plausibility. *Inter-rater reliability for total details:* ICC = 0.70	Analyses were conducted for the T1 free recall question on data from participants who were not asked to sketch and narrate at T1 (i.e., control and T2 conditions). n = 123, of which 61 were lie tellers.
Deeb, Vrij, Leal, & Mann (2021): A total of 175 participants in the UK viewed a video footage. Participants were then interviewed three times, immediately (T1), after one week (T2), and after two weeks (T3). Two-thirds of the participants were asked to sketch and narrate at T1 and were exposed to the MS at T2 (sketch-MS condition) or vice versa (MS-sketch condition) and provided a free verbal recall at T3. One-third of the participants were asked for a verbal free recall across the three interviews (control condition). All three interviews involved only one free recall question. The experiment was conducted online. *Dependent variables:* Total details, PLATO details, complications, CKD, SHS. *Inter-rater reliability for total details:* ICC = 0.96	Analyses were conducted for the T1 free recall question on data from participants in the control condition. n = 61, of which 31 were lie tellers.

(*Continued*)

Table 1. (*Continued*)

Experimental procedure	Present paper
Deeb et al. (2022a): A total of 112 participants in the UK carried out a mission and were then instructed to lie or to tell the truth about it in an interview. Participants were asked two questions. In response to the first question, they provided a verbal free recall of the mission. For the second question, they were shown a detailed or a less detailed map and were asked to provide another verbal recall while sketching on the map. The experiment was conducted face-to-face. *Dependent variables:* PLATO details, complications. *Inter-rater reliability for total details:* ICC = 0.82	Analyses were conducted for the first free recall question on data from all participants. $N = 112$, of which 56 were lie tellers.
Deeb et al. (2022b): A total of 211 participants in the UK carried out a mission and were then instructed to lie or to tell the truth about it in an interview. Participants were asked two questions. In response to the first question, they provided a verbal free recall of the mission. For the second question, they either provided another verbal free recall or sketched (and narrated) on a blank sheet of paper or on a map. This was a face-to-face experiment, but the interview was conducted online. *Dependent variables:* Total details, PLATO details, complications, verifiable sources. *Inter-rater reliability for total details:* ICC = 0.79	Analyses were conducted for the first free recall question on data from all participants. $N = 211$, of which 106 were lie tellers.
Vrij et al. (2020): A total of 201 participants in Lebanon ($n = 56$), Mexico ($n = 65$), and South Korea ($n = 80$) were asked to report a truthful or a false trip they have made. To test countermeasures, prior to the interview participants were allocated to read information about the working of the MS, of details types (complications, CDK, SHS), or of MS+details types, or were not assigned to read any information (control condition). Participants were interviewed in their native language. They were first asked two free recall questions about (1) how they planned the trip and (2) everything they did during the trip. They were then exposed to the MS and asked to respond again to the same two questions. The experiment was conducted face-to-face. *Dependent variables:* Total details, complications, CDK, SHS, plausibility. *Inter-rater reliability for total details:* ICC = 0.72	Analyses were conducted for the first two free recall questions on data from the control condition. $n = 50$, of which 26 were lie tellers. 15 Lebanese, 15 Mexican, 15 South Korean.

(*Continued*)

Table 1. (*Continued*)

Experimental procedure	Present paper
Vrij, Leal, Deeb, Castro Campos, et al. (2022, Experiment 1): A total of 209 participants in Lebanon ($n = 60$), Mexico ($n = 70$), and South Korea ($n = 79$) were asked to report a truthful or a false trip they have made. To test countermeasures, prior to the interview participants were allocated to read information about the working of the MS, of details types (complications, CDK, SHS), or of MS+details types, or were not assigned to read any information (control condition). All participants were interviewed in their native language via an interpreter. They were first asked one free recall question to discuss their trip. Then, they were exposed to the MS and asked the same free recall question about their trip. The experiment was conducted face-to-face in Lebanon and South Korea but online in Mexico. *Dependent variables:* Total details, complications, CDK, SHS, plausibility. *Inter-rater reliability for total details:* ICC = 0.92	Analyses were conducted for the first free recall question on data from the control condition. $n = 52$, of which 24 were lie tellers. 18 Lebanese, 14 Mexican, 20 South Korean. 38 face-to-face and 14 online interviews.
Vrij, Leal, Deeb, Castro Campos, et al. (2022, Experiment 2): A total of 221 participants in Lebanon ($n = 71$), Mexico ($n = 70$), and South Korea ($n = 80$) were asked to report a truthful or a false trip they have made. The presence (or absence) of an interpreter and reading (or not) of information about detail types (complications, CDK, SHS) were manipulated. Thus, participants were interviewed in their native language through a native interviewer or through an interpreter. Participants were asked five questions about (1) what they did during their trip, (2), their accommodation, (3) planning places to visit, (4) planning accommodation and transport, and (5) verifiable details they can provide. The experiment was conducted online. *Dependent variables:* Total details, complications, CDK, SHS, plausibility. *Inter-rater reliability for total details:* ICC = 0.95	Analyses were conducted for the first free recall question (i.e., what they did during their trip) on data from the control conditions (i.e., no interpreter, no countermeasures). $n = 56$, of which 28 were lie tellers. 20 Lebanese, 16 Mexican, 20 South Korean.

total details score. For example, "It was a side road, I do not remember what the road was called. Uhh but apart from that there was a train station . . . " includes 24 words, but only the four informative details (underlined) were coded. Across all experiments, redundant words were not coded within a single transcript. That is, if an interviewee mentioned they "walked on the beach and then walked home," *walked* would be coded only once as it is a repetition and contains no new information the second time it is mentioned.

The data was manually coded using either an RM coding scheme (Deeb et al., 2020; Deeb, Vrij, Leal, & Burkhardt, 2021; Vrij, Leal, Deeb, Castro Campos, et al., 2022; Vrij et al., 2020), or a PLATO coding scheme (Deeb et al., 2022a, 2022b; Deeb, Vrij, Leal, & Mann, 2021). Thus, we included coding scheme as a covariate in the analyses.

The human raters were either one of the authors who had years of experience in coding verbal cues, research assistants who had previously coded verbal veracity cues, or research assistants who had no prior experience with coding cues but were trained. Training always occurred over several sessions. The experienced rater provided the trainee rater with one or more transcripts to code. The experienced rater then provided the trainee rater with feedback for each coded transcript. Afterward, the trainee rater was given another set of transcripts to code. This continued until the rater was able to code the transcripts independently. For all experiments, one rater coded all the transcripts and a second rater coded 15% to 29% of the total number of transcripts for inter-rater reliability purposes.

The inter-rater reliability scores achieved in each dataset are presented in Table 1. Reliability is considered poor for intra-class correlation coefficients (ICCs) less than .40, fair for coefficients between .40 and .59, good for coefficients between .60 and .74, and excellent for coefficients between .75 and 1 (Hallgren, 2012). The average ICC coefficient across datasets was excellent for total details (ICC $= 0.83$).

Automated coding of the transcripts

In Experiment 1, we used two software programs for analyzing the data. The first is the Linguistic Inquiry and Word Count software program which is widely used in the lie detection research field (Hauch et al., 2015). The second is the Text Inspector software program that to our knowledge was not utilized previously in lie detection research. The inspection of two software programs allowed us to examine potential differences between the programs.

Linguistic Inquiry and Word Count (LIWC) software program

LIWC is a linguistic tool that is psychologically based, analyzing texts for different parts of speech (e.g., pronouns, verbs), psychological constructs (e.g., affect, cognition), and other output variables that constitute more than 90 categories (Pennebaker et al., 2015). Words in analyzed texts are compared to LIWC's dictionary of linguistic and psychological words and categorized in the corresponding one or more categories (if they fit under more than one category). LIWC is regularly updated and revised based on empirical evidence (e.g., validated emotion rated scales) and other sources (e.g., word extraction software, social media platforms). Its internal consistency as reported by Pennebaker et al. (2015) is $\alpha = 0.69$.

LIWC was developed in 1993 (Francis, 1993) to examine language and expression within the context of health psychology. In 1996, the software was validated using groups of judges who evaluated the extent to which the dictionary of 2000+ words or word stems fit in different categories (Chung & Pennebaker, 2007; Pennebaker et al., 2015). Since then, LIWC has been translated into more than 16 languages. The software has been used by many psychologists, and it has been employed in different cultures and in different areas, including personality psychology, clinical psychology, and lie detection (Addawood et al., 2019; Newman et al., 2003; Pennebaker & Graybeal, 2001; Tausczik & Pennebaker, 2010).

Unlike manual coding, LIWC does not account for unique words but codes all words regardless of whether or not they are redundant in text. Also, LIWC provides percentage scores except for the total number of words, words per sentence, dictionary words, and punctuations which are presented as frequency scores. In Experiment 1, we used the academic license of LIWC2015 v1.6 and examined LIWC's total number of words which has the same metric (frequency score) as manual coding and Text Inspector.

LIWC's founders used it as a lie detection tool and demonstrated its success (Newman et al., 2003) which encouraged further lie detection research to utilize the software. It is now the major coding software program tested in lie detection research (e.g., Forsyth & Anglim, 2020; Taylor et al., 2017). LIWC can differentiate honest from false accounts based on several parts of speech and constructs (Dzindolet & Pierce, 2004; Markowitz & Griffin, 2020). However, the diagnosticity of LIWC's total number of words has yet to be determined. Some researchers found that it was diagnostic with honest accounts including more words than false accounts (Hirschberg et al., 2005; Toma & Hancock, 2012), whereas others showed the reverse pattern such that false accounts included more words than honest accounts (Bond et al., 2017; van der Zee et al., 2022). Still, other researchers did not find any significant differences between honest and false accounts on this cue (Bogaard, van der Mark, et al., 2019; Jupe et al., 2018; Masip et al., 2012).

Text Inspector software program

Text Inspector is an online language analysis tool that was developed in 2011 to analyze texts for lexical diversity, lexical complexity, and language proficiency (Bax et al., 2019; Weblingua, 2022). In its current form, it can analyze texts for 63 different parts of speech such as articles, verbs, and pronouns and provides the corresponding statistics as frequency scores. Text Inspector is based on empirical evidence in applied linguistics. Since its inception, it has been tested in over 180 countries (Weblingua, 2022). Unlike LIWC, it has not been tested on diverse samples but mostly on student samples. Nonetheless, its data is representative and it has been shown to accurately determine student proficiency levels similar to standardized linguistic tests (Rodríguez, 2023). It is regularly updated in line with emerging empirical evidence, and it has scored reliability rates up to 98% (Arslan & Eraslan, 2019; Gayed et al., 2022). According to Text Inspector's official website (https://textinspector.com/help/statistics-readability/), the tool is reliable for texts that are longer than 100 words.

Text Inspector is an easy-to-use software that was not tested previously by lie detection researchers. We decided to specifically use it because it can code unique (nonredundant) details similar to manual coding. Text Inspector is also a good alternative to the widely used LIWC as it codes words differently which allowed us to understand lie detection differences between the two coding software programs. The full version of Text Inspector was used for the coding and analyses.

Hypotheses

In line with the majority of previous research, we expected honest accounts to include more total details than false accounts. As the literature shows conflicting results concerning which coding method performs better at lie detection, we did not posit any hypotheses concerning the veracity × coding method interaction effect.

Method

We set several criteria for the inclusion of datasets in our analyses. First, the interviews should have been conducted with only one interviewee. Second, the interviews should have been about a past event as reporting about future events may yield different veracity effects (Sooniste et al., 2013). Third, the interviews should have included a verbal free recall question at the outset as only this question was used for the analyses to remove the effects of the interview protocol manipulation (see below for more details). Fourth, the relevant paper should have been peer-reviewed and published so that the coded data was readily available for analyses and relevant information on the experiments is accessible for interested readers. Fifth, we were interested in recent data, so only articles published after 2020 were selected. Sixth, the data should have already been manually coded for total details.

All datasets created by the first author and datasets from non-WEIRD samples created by the second author were included in the analyses if they met the above criteria. Including data from non-WEIRD countries in our analyses is an advantage over previous research in which automated and manual coding were compared on transcripts from WEIRD (Western, Educated, Industrialised, Rich, Democratic) samples. There has been an emerging call by researchers in the lie detection field in specific—and in the psychology field in general—to conduct more research in non-WEIRD countries as the majority of psychological research is conducted in WEIRD countries (Denault et al., 2022; Henrich et al., 2010; Vrij et al., 2023). Different cultures use different communication modes, and this difference is significant between WEIRD and non-WEIRD countries (Liu, 2016). As deception is a communication mode, verbal veracity cues may also differ across countries and cultures (Leal, Vrij, Vernham, et al., 2018, 2019; Taylor et al., 2015) which makes it important to cross-culturally examine manually and automatically coded verbal veracity cues. We thus included the country where the data was collected (i.e., sample's culture) as a covariate in the analyses.

A total of seven papers were selected for the analyses. One of the papers (Vrij, Leal, Deeb, Castro Campos, et al., 2022) included two experiments, so the total number of datasets that were analyzed was 8. The total sample analyzed included 787 interviewees. A description of the experiments and the data used is presented in

https://doi.org/10.1017/S0142716424000298 Published online by Cambridge University Press

Table 1. All experiments involved a face-to-face or an online oral interview. In two face-to-face experiments (Deeb et al., 2020; Deeb, Vrij, Leal, & Burkhardt, 2021), participants were asked to report a true or a false out-of-the-ordinary memorable event. In two other face-to-face experiments, participants reported truthfully or falsely a mission they completed in a face-to-face (Deeb et al., 2022a) or online interview (Deeb et al., 2022b). In one online experiment (Deeb, Vrij, Leal, & Mann, 2021), participants reported truthfully or falsely about a video they watched. In the remaining experiments (Vrij, Leal, Deeb, Castro Campos, et al., 2022; Vrij et al., 2020) in which some of the data was collected online, participants described truthfully or falsely a city trip they had made while or while not talking through an interpreter. The experiments by Vrij et al. (2020) and Vrij, Leal, Deeb, Castro Campos, et al. (2022) were the only experiments that were ran with non-WEIRD samples, namely in Lebanon, Mexico, and South Korea. Given that some of the data was collected via an online interview and/or via an interpreter, we added interview modality and interpreter presence as covariates in our analyses.

The original experiments tested different interview protocols (e.g., Model Statement interview technique, sketching and narrating interview technique) and compared them with a control condition which was a verbal free recall in all experiments (see Table 1 for the experimental interview protocol conditions). To be included in the present analyses, participants should have been asked for a free recall at the outset of the interview and should have not been subjected to the experimental interview condition. These exclusions minimized the confounding effects of experimental procedures and manipulations. For all analyses, we used the first free recall question which asked participants to discuss everything they did (or viewed), except for Vrij et al. (2020) for which we used the first two open-ended questions because participants were asked about their plans for the trip they made in the first question and to discuss everything they did in the second question.

The original datasets were cleaned from fillers (e.g., uhm, err), references to participants' behaviors (e.g., pausing, smiling), and interviewer's speech as these were irrelevant to the topic of investigation and/or to the coded cues. In the South Korean transcripts of the Vrij et al. experiments, the transcriber added pronouns to the transcripts to explain what the participants were saying because pronouns do not exist in South Korean language (Liu, 2016). We kept the pronouns to ensure that we can compare these transcripts with transcripts from other datasets.

The datasets are publicly shared as noted in the original papers or can be obtained from the original authors. The datasets were coded by various raters, and all were derived from the same lab. None of the datasets was previously analyzed using coding software programs.

Results

We ran a mixed effects model to account for clustering in our data as the data is nested in different datasets (Tate & Pitush, 2007; West et al., 2006). Our model included veracity (honest, false), coding method (manual, LIWC, Text Inspector), and their interaction as fixed factors. The intercepts of participants and datasets were entered as random factors and also as cluster factors. The coding scheme (RM

Table 2. Fixed effects parameters for total details as a function of veracity and coding method

	Estimate [95% CI]	*t*	*p*
Truth teller—Lie teller	74.73 [52.25, 97.21]	6.51	<.001
Text Inspector—Manual	67.54 [48.81, 86.27]	7.07	<.001
LIWC—Manual	325.12 [306.39, 343.85]	34.02	<.001
Truth tellers—Lie tellers* Text Inspector - Manual	2.19 [-35.27, 39.65]	0.11	.909
Truth tellers—Lie tellers* LIWC - Manual	104.52 [67.06, 141.99]	5.47	<.001

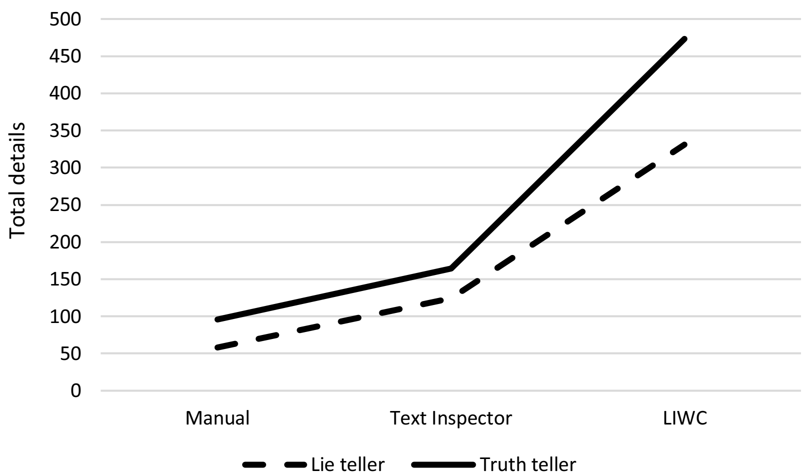

Figure 1. Simple effects for total details as a function of veracity and coding method.

and PLATO), country (Lebanon, Mexico, South Korea, United Kingdom), interview modality (face-to-face, online), and interpreter presence (present, not present) were treated as covariates. We carried out simple contrasts to compare the coding methods. The analysis was conducted using Jamovi 2.3.18 software and Gamlj package (Gallucci, 2019).

The variance and the ICC of the random intercepts showed variability in the data, so a mixed model analysis could be carried out on the data. The mixed effects model explained 52.4% of the variance (R^2 conditional) and showed significant effects of veracity, $F(1, 773.32) = 42.44$, $p < .001$, coding method, $F(2, 1570) = 644.55$, $p < .001$, and veracity × coding method, $F(2, 1570) = 19.53$, $p < .001$.

The parameter estimates of the fixed factors are shown in Table 2 (also see Figure 1 for an illustration). In line with our hypothesis, honest accounts included significantly more total details than false accounts. Both LIWC and Text Inspector coding resulted in more total details than manual coding, but the estimates and *t*-values were larger for LIWC coding.

For the interaction effect, veracity differences emerged for LIWC coding, but not for Text Inspector coding, compared to manual coding. To dissect this finding, we ran

https://doi.org/10.1017/S0142716424000298 Published online by Cambridge University Press

Table 3. Simple effects for total details as a function of veracity and coding method

	Truth tellers		Lie tellers		Estimate [95% CI]	t	p
	M (*SD*)	95% CI	*M* (*SD*)	95% CI			
Manual	95.84 (63.44)	89.55, 102.12	58.14 (39.88)	54.19, 62.10	39.2 [7.94, 70.4]	2.46	.014
Text Inspector	164.47 (85.84)	155.97, 172.98	124.59 (84.73)	116.19, 132.99	41.3 [10.13, 72.6]	2.60	.009
LIWC	473.22 (395.49)	434.05, 512.39	331.00 (369.42)	294.37, 367.64	143.7 [112.47, 174.9]	9.03	<.001

simple effects. As shown in Table 3, all the coding methods could differentiate honest from false accounts, but LIWC coding showed the largest estimate and significance. The estimates and *t*-values for manual coding and Text Inspector were almost similar.

To test the model's classification accuracy, we experimented with three classification methods: linear discriminant analysis, XGBoost, and random forests. We decided to use statistical and machine learning classification methods to understand if classification accuracy differs according to the learning model. We trained a different model for each coding method to ensure a fair comparison of the classification capability of each method and to remove any information leakage between the different methods if we train a joint model. We used 10-fold cross-validation to evaluate classification accuracy.

We ran separate analyses with and without the covariates (coding scheme, country, interview modality, and interpreter presence). We set veracity as the grouping variable and total details as the independent variable. The results are shown in Table 4. The analyses with and without the covariates showed similar results. The average accuracy rate across all three classification methods was highest for manual coding followed by Text Inspector and LIWC. Among the classification methods, the differences in accuracy were small, but the random forest classifier showed the best results (64.4%–65.3%).

We evaluated the feature importance of the model using the random forest classifier which showed the highest accuracy. We trained a random forest using all the data (total details, coding method, coding scheme, country, interview modality, and interpreter presence) for each coding method separately and calculated the decrease in impurity within each decision tree. Table 5 shows that for all coding methods, total details was the most important feature followed by country.

Discussion

We predicted that honest accounts will include more total details than false accounts, and this prediction was supported. The finding aligns with previous research showing that total details is a valid veracity cue and that a larger number of details is evident in honest accounts (Amado et al., 2016; Colwell et al., 2007). The mixed effects analysis also revealed that among all three coding methods, LIWC showed the highest estimates and differences between honest and false accounts.

https://doi.org/10.1017/S0142716424000298 Published online by Cambridge University Press

Table 4. Classification accuracy for each coding method based on total details using linear discriminant analysis, XGBoost classifier, and random forest classifier

Classification method	Lie accuracy	Truth accuracy	Average accuracy
Analyses with the covariates			
LIWC			
Discriminant analysis	71.4%	51.1%	61.2%
XGBoost classifier	59.7%	63.2%	61.5%
Random forest classifier	59.4%	71.3%	65.4%
Text Inspector			
Discriminant analysis	70.4%	54.9%	62.6%
XGBoost classifier	60.2%	63.4%	61.9%
Random forest classifier	55.9%	72.8%	64.4%
Manual			
Discriminant analysis	72.7%	56.4%	64.5%
XGBoost classifier	63.8%	59.9%	61.9%
Random forest classifier	80.9%	48.0%	64.4%
Analyses without the covariates			
LIWC			
Discriminant analysis	74.5%	47.1%	60.8%
XGBoost classifier	58.5%	60.9%	59.7%
Random forest classifier	59.2%	70.0%	64.7%
Text Inspector			
Discriminant analysis	70.1%	53.2%	61.7%
XGBoost classifier	63.3%	58.3%	60.9%
Random forest classifier	55.6%	74.0%	64.9%
Manual			
Discriminant analysis	75.9%	54.1%	64.9%
XGBoost classifier	66.5%	50.5%	58.4%
Random forest classifier	82.5%	48.3%	65.3%

Note: The covariates are coding scheme, country, interview modality, and interpreter presence.

Text Inspector and manual coding showed comparable performance. We further found that all three coding methods could significantly classify truth tellers and lie tellers with very small differences between them. The LIWC classification accuracy rates were the lowest (62.7% in the analysis with the covariates) and those of manual coding were the highest (63.6% in the analysis with the covariates), but the three coding methods showed similar lie detection performance. Further, the feature importance model revealed that total details contributed substantially to the model

Table 5. Feature importance of the model in Experiment 1

Total details	Country	Interview modality	Interpreter presence	Coding scheme
LIWC				
0.70	0.10	0.07	0.07	0.06
Text Inspector				
0.69	0.13	0.08	0.06	0.04
Manual				
0.68	0.11	0.07	0.06	0.08

compared to the covariates (country, interview modality, interpreter presence, and coding scheme) which corresponds with previous research showing total details to be a diagnostic verbal cue.

We also found that LIWC coding yielded the highest number of total details. This finding can be expected given that LIWC codes all words in an account (including redundant words), whereas Text Inspector and manual coding involve the coding of only unique (non-redundant) words. The overall findings suggest that researchers should consider the software program used when comparing manual and automated coding, but there is no coding method that can always be superior to the other.

Experiment 2

LIWC resulted in more pronounced veracity differences than Text Inspector and manual coding in Experiment 1. A main difference between LIWC and the other two coding methods is that LIWC counted redundant details, whereas the other two methods only counted nonredundant (unique) details. It could thus be that honest accounts included more redundant words than false accounts and that LIWC picked up this difference (although this difference did not seem to enhance lie detection accuracy). However, the data in Experiment 1 cannot inform us on whether honest accounts were more redundant, so we put this question to test in Experiment 2.

In previous research, redundancy was either examined under the construct of lexical diversity (i.e., unique words in text; e.g., Fuller et al., 2013) or as a cue on its own (e.g., Chen et al., 2020). The findings were generally inconsistent: Compared to false accounts, honest accounts were either more redundant (Burgoon, 2018), less redundant (Davis et al., 2005; DePaulo et al., 2003; Hauch et al., 2015; Mbaziira & Jones, 2016; Zhou et al., 2003), or equally redundant (Chen et al., 2020; Dunbar et al., 2023; Duran et al., 2010; Zhou, 2005). One potential reason for this inconsistency is that redundancy was measured differently across experiments. Whereas some researchers computed it as the ratio of unique details to total details (Burgoon, 2018; Dunbar et al., 2023), other researchers computed it as the ratio of total function words to total sentences (Zhou et al., 2003), the total number of redundant consecutive words (Chen et al., 2020), the total number of redundant nouns (Duran et al., 2010), the total number of redundant words in adjacent

https://doi.org/10.1017/S0142716424000298 Published online by Cambridge University Press

sentences (Davis et al., 2005), or the redundant words or phrases that are part of non-ah speech disturbances (DePaulo et al., 2003). This suggests some measurements were liberal (e.g., measuring all redundant components in text), whereas others were more restrictive (e.g., measuring only adjacent text for redundancy).

The inconsistent pattern of the findings can be explained by different theoretical frameworks. False accounts can become less redundant than honest accounts when lie tellers have prepared for their account (Dunbar et al., 2023) or are given time to interact with the interviewer (Zhou et al., 2004). That would help lie tellers to report longer accounts than they would have otherwise done which increases diversity in their accounts. In contrast, truth tellers who take their credibility for granted do not usually prepare for the interview and would not intentionally plan a diverse account as lie tellers do (Chan & Bull, 2014; Granhag & Hartwig, 2008; Vrij et al., 2010).

Another argument for the inconsistent pattern of the findings is that false accounts can become more redundant than honest accounts, because lie tellers prefer to keep their accounts simple (Vrij et al., 2010; Vrij, Granhag, et al., 2022), so they tend to repeat information rather than add new information (Alison et al., 2014; Deeb et al., 2024). Further, in an interview, lie tellers produce information on a follow-along basis as the account develops because they fabricate rather than retrieve information from memory (Duran et al., 2010). Hence, there is less possibility of producing new information and thus lie tellers default to a more redundant account. It can also be argued that lie tellers do not have the creativity to improvise a text with diverse wording (Vrij et al., 2021). In contrast, truth tellers can demonstrate more lexical diversity, because they have experienced the event and information is retrieved from memory at the global level so new information is continuously developed (Duran et al., 2010). Thus, truth tellers can be more specific in their accounts by including more perceptual and contextual details (Masip et al., 2005) without having to use redundant words.

Given these conflicting theoretical explanations, we tested redundancy in Experiment 2. Based on the findings from Experiment 1, we expected honest accounts to be more redundant than false accounts. We also explored which types of words are the most redundant. If, for example, content words (i.e., core structures of a sentence such as nouns and verbs) are particularly redundant, then speakers may be more focused on the content (semantics) of the message. However, if function words (e.g., conjunctions, prepositions) are particularly redundant, the focus would mostly be on the grammatical structure (syntax) of the message.

Method

The same eight datasets as in Experiment 1 were used in Experiment 2. SpaCy software program (https://spacy.io/) was employed to count redundant words in text. SpaCY is a library for the Python programming language that analyses texts based on pretrained language pipelines. A SpaCy pipeline has multiple components which utilize a base artificial intelligence model for natural language processing (NLP) tasks such as part of speech tagging, named entity recognition, and lemmatization. The analyses were carried out with the SpaCy English language transformer pipeline based on the RoBERTa model, which is an enhanced model of the original BERT language model (Liu et al., 2019).

The software tokenized each transcript into separate words. Punctuation marks and spaces were skipped. For each token, the software searched for the word lemma based on a set of rules and the word's part of speech and dependency. Lemmatization is the process of reducing words to their normalized form by grouping together different inflected forms of the same word (Khyani et al., 2021; Plisson et al., 2004). For example, in the context of going somewhere, "going" and "went" would both be lemmatized to "go" and allocated to the same group. Thus, where these three words are mentioned by the same participant in a single transcript, the software would count them as three redundant words.

Results

To account for the length of each transcript (see Schutte et al., 2021), we computed a redundancy ratio score by dividing the total number of redundant words by the total number of words in each transcript. A mixed effects model revealed that there was no variability in the data, so we conducted a one-way univariate analysis of variance with veracity (honest, false) as factor, redundancy ratio as dependent variable, and datasets (all eight), country (Lebanon, Mexico, South Korea, United Kingdom), interview modality (face-to-face, online), and interpreter presence (present, not present) as covariates. A significant effect of veracity emerged, $F(1, 781) = 30.58$, $p < .001$, $\eta 2 = .04$ (see Figure 2). Honest accounts ($M = 0.68$, $SD = 0.09$, 95% CI [0.67, 0.69]) were more redundant than false accounts ($M = 0.64$, $SD = 0.10$, 95% CI [0.63, 0.65]), $d = 0.42$ (95% CI [0.28, 0.56]). This result supported our hypothesis.

To explore which types of words were most redundant across participants, we further scrutinized the dataset. There were 3,233 redundant words with the most redundant word repeated 14,511 times. A total of 2,999 words were repeated less than 100 times, 200 words were repeated more than 100 times, and 34 words were repeated more than 1,000 times. We decided to analyze the 34 words that were repeated more than 1,000 times for two reasons. First, to have a better understanding of what types of words were most redundant, we needed to limit the number of interpreted words, and 34 words seemed enough for this purpose. Second, there was a significant gap in the times that these words were repeated. The least redundant word among these 34 words was repeated 1,146 times versus 14,511 times for the most redundant word. It thus made sense to include these 34 words rather than the words that were repeated less than 1,000 times as including the latter would further increase this gap.

We extracted the 34 most redundant words and subjected them to a t-test with veracity as factor. To control for multiple comparisons, we applied a strict p-value of less than .001 (two-sided). The redundant words that yielded significant differences between honest and false accounts are shown in Table 6. These redundant words are generally function words.

To explore if looking at the verbal cue redundancy enhances lie detection, we performed a discriminant analysis, XGBoost, and random forests with veracity as

Table 6. *T*-test results for redundant words that significantly differentiated truth tellers and lie tellers

Redundant word	Truth tellers		Lie tellers			
	M (*SD*)	95% CI	*M* (*SD*)	95% CI	*t*	*d* [95% CI]
And	19.24 (20.21)	17.24, 21.24	12.08 (14.46)	10.64, 13.51	5.72	0.41 [0.27, 0.55]
At	2.02 (2.69)	1.75, 2.28	1.20 (2.54)	0.95, 1.45	4.36	0.31 [0.17, 0.45]
Back	1.78 (2.11)	1.57, 1.99	1.13 (1.76)	0.96, 1.31	4.63	0.33 [0.19, 0.48]
But	2.62 (5.43)	2.08, 3.15	1.43 (2.63)	1.17, 1.69	3.89	0.28 [0.14, 0.42]
I	20.85 (20.04)	18.87, 22.84	13.65 (24.95)	11.17, 16.12	4.47	0.32 [0.18, 0.46]
In	3.90 (5.26)	3.38, 4.42	2.84 (3.55)	2.49, 3.19	3.32	0.24 [0.10, 0.38]
On	3.18 (3.39)	2.84, 3.51	1.67 (2.43)	1.43, 1.91	7.18	0.51 [0.37, 0.65]
She	4.37 (12.86)	3.10, 5.64	1.47 (4.82)	1.00, 1.95	4.19	0.30 [0.16, 0.44]
So	5.54 (7.87)	4.76, 6.31	3.88 (5.82)	3.30, 4.46	3.36	0.24 [0.10, 0.38]
Take	2.07 (2.36)	1.84, 2.30	1.03 (1.89)	0.85, 1.22	6.81	0.49 [0.34, 0.63]
The	23.40 (20.73)	21.35, 25.46	13.46 (16.77)	11.80, 15.12	7.40	0.53 [0.38, 0.67]
To	12.28 (13.46)	10.94, 13.61	8.93 (11.08)	7.83, 10.03	3.81	0.27 [0.13, 0.41]
Walk	2.11 (3.51)	1.76, 2.46	1.17 (2.09)	0.96, 1.38	4.58	0.33 [0.18, 0.47]
With	1.90 (2.40)	1.66, 2.14	1.07 (1.84)	0.89, 1.26	5.42	0.39 [0.25, 0.53]

Note: For all redundant words, $p < .001$.

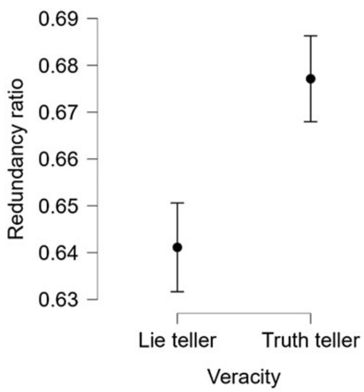

Figure 2. Means of the redundancy ratio as a function of veracity.

the grouping variable and redundancy ratio as the independent variable. We ran separate analyses with and without the covariates (country, interview modality, and interpreter presence). We used 10-fold cross-validation to evaluate classification accuracy. The results are shown in Table 7. The analyses with and without the covariates showed similar results. The average accuracy rate was highest for the discriminant analysis (59%) and lowest for the XGBoost classifier (51.4%-52.2%).

https://doi.org/10.1017/S0142716424000298 Published online by Cambridge University Press

Table 7. Classification accuracy based on redundancy ratio using linear discriminant analysis, XGBoost classifier, and random forest classifier

Classification method	Lie accuracy	Truth accuracy	Average accuracy
Analyses with the covariates			
Discriminant analysis	56.4%	60.9%	58.7%
XGBoost classifier	53.7%	50.8%	52.2%
Random forest classifier	54.1%	59.3%	56.8%
Analyses without the covariates			
Discriminant analysis	56.9%	61.6%	59.3%
XGBoost classifier	54.7%	48.2%	51.4%
Random forest classifier	57.7%	56.3%	57.0%

Note: The covariates are country, interview modality, and interpreter presence.

We evaluated the feature importance of the model using all the data (redundancy ratio, country, interview modality, and interpreter presence). Redundancy ratio was the most important feature (0.780) followed by country (0.107), interview modality (0.057), and interpreter presence (0.056).

Discussion

In line with our hypothesis, honest accounts were more redundant than false accounts, and this difference can be detected with approximately 59% accuracy. This finding suggests that accounting for redundancy in manual and automated coding can enhance detecting honest and false accounts. The most frequent redundant words were function words that do not have imperative lexical meaning. Function words are relevant to the grammatical structure of an account that make it look coherent (Afroz et al., 2012). Examples of function words include conjunctions (and, but), determiners (a, the), prepositions (on, at), personal pronouns (they, she), and modal verbs (should, might). While function words are not related to content, they are considered closely linked to psychological and social processes, including deceptive communication as they are less evident in false accounts than in honest accounts (Chung & Pennebaker, 2007).

There are several explanations for our findings. First, truth tellers provide more coherent, logical, and chronological accounts than lie tellers (Vrij, 2005). It thus makes sense that truth tellers would repeat function words to make the message flow and content more comprehensive to the interviewer (Zhou et al., 2004). In contrast, lie tellers produce more ambiguous and vague accounts (DePaulo et al., 2003) so that their messages do not appear clear and/or coherent. This should result in fewer redundant function words that link sentences together.

Second, in the original experiments, lie tellers self-reported preparing for the interviews more than truth tellers. While lie tellers kept their accounts shorter and simpler as we found in Experiment 1, preparing for the interview may have enhanced their lexical diversity and helped them avoid appearing redundant

(Dunbar et al., 2023). In contrast, truth tellers did not prepare for the interviews but reported from memory (Hartwig et al., 2007; Vrij et al., 2009). Truth tellers' focus was thus on maintaining a continuous flow of communication for establishing a coherent account which requires the inclusion of redundant language (Amado et al., 2016; Zhou et al., 2004).

Third, Zhou et al. (2004) argued that lie tellers can become less redundant when they are given time to interact with the receiver of their message. Given that our datasets were collected via interviews which are considered rich media venues (Zhou, 2005), the level of interaction was high which may have allowed lie tellers to produce a more diverse account than they would in less rich mediums (e.g., emails; Zhou et al., 2003).

General discussion

In Experiment 1, we showed that honest accounts were more detailed than false accounts. This finding is consistent with previous meta-analyses that found that compared to other verbal cues, total details was the most diagnostic cue with the largest magnitude (Amado et al., 2016; DePaulo et al., 2003; Gancedo et al., 2021). Given that we reached the same conclusion using a large number of participants, this increases our confidence in total details as a diagnostic cue (Lakens & Evers, 2014).

The theoretical approaches to deception, namely the information management approach, the cognitive approach, and the RM approach, can shed light unto this finding. Lie telling is a cognitively demanding task given that lie tellers typically have to fabricate at least some of the reported information (Vrij, 2008). At the same time, lie tellers want to appear cooperative and convincing so that their lie is believed (Granhag et al., 2015). This leads to lie tellers calculating what information they should disclose and what information they should avoid reporting. They thus strive to provide accounts that are simple to reduce cognitive demands and to avoid potentially contradicting themselves (Deeb et al., 2024; Vrij et al., 2017). Ultimately, lie tellers provide fewer details than truth tellers.

We found that the automated coding of total details can capture differences between truth tellers and lie tellers to a larger extent than manual coding, although this may depend on the software program used. More specifically, LIWC seemed to detect these differences more than Text Inspector. This finding aligns with previous research demonstrating that various software programs perform differently which can affect lie detection (Ceballos Delgado et al., 2021; Kleinberg et al., 2017). In Experiment 2, we showed that LIWC's performance may have been driven by its analysis of redundant words in text which makes it more sensitive for detecting differences between honest and false accounts. These findings suggest that it may be possible for software programs other than LIWC that also account for redundancy to effectively differentiate honest and false accounts. This question can be put to test by future research.

While LIWC showed larger veracity differences in reported details than Text Inspector and manual coding, the classification accuracy rate of the three coding methods were similar and all in the 60%–65% range. We expect the manual coding classification rate to increase up to 75% when human judges are asked to specifically look for total details. We base this prediction on a previous meta-analysis which has shown that when judges are asked to look for specific verbal cues that received empirical evidence, lie detection accuracy increases (Mac Giolla & Luke, 2021).

In Experiment 2, honest accounts were more redundant than false accounts and the veracity groups could be accurately classified via automated coding based on this verbal cue. While the direction of the differences between honest and false accounts on redundancy contradicts some previous research (DePaulo et al., 2003; Mbaziira & Jones, 2016; Zhou et al., 2003), it is consistent with other findings (Burgoon, 2018; Zhou et al., 2004). The nonconverging findings are likely the result of the redundancy cue being measured differently across experiments. In the present research, we used a simplified measure of redundancy that takes into account repeated lemmatizations in text and that accounts for the length of the account by calculating a percentage score (see Schutte et al., 2021).

Also unlike previous research on redundancy, all included datasets were rich media venues (oral interviews) that were conducted in a forensic context. Thus, at least in forensic interview contexts where free recalls are requested, redundancy may be diagnostic of truth telling. However, we have to be cautious when interpreting these findings as the effect size was medium. While such an effect size is practically significant for lie detection, a large effect size is usually preferred because it would make the veracity differences very noticeable to the naked eye (Cohen, 1992). These results can be ameliorated if interviewers look at multiple verbal cues at the same time (Deeb et al., 2024; Hartwig & Bond, 2011; Vrij, Hartwig, et al., 2019). Based on the present results, and as total details and redundancy could accurately classify honest and false accounts above chance levels, interviewers can look at both verbal cues to enhance lie detection accuracy.

We specifically found that the most redundant words were function words. While function words constitute less than 0.04% of the English vocabulary, they account for half of the words used in daily communications (Chung & Pennebaker, 2007). It is reassuring to find that words referring to syntax can be diagnostic similar to words referring to semantics (Afroz et al., 2012; Newman et al., 2003). However, we cannot infer from the data why function words were repeated more than content words. Future research can examine this through the use of metacognitive questions, whereby truth tellers and lie tellers are asked how they think about and use function words in their preparations and in their actual accounts.

Limitations and future research implications

We analyzed data from free recall (control) questions only. That meant that we only used passive interview protocols to reach our conclusions (Vrij, 2008). We did not code how honest and deceptive language changes as a function of asking different questions. We wanted to standardize the analyses and see how truth tellers and lie tellers respond to questions in a neutral context (i.e., when the interviewer is not actively asking questions that would increase differences between truth tellers and lie tellers). In the original experiments, the experimental interview questions (e.g., Model Statement; sketching while narrating) yielded more significant differences between honest and false accounts than the free recall question.

https://doi.org/10.1017/S0142716424000298 Published online by Cambridge University Press

We would thus expect more veracity differences to emerge when the interview protocol is manipulated. Future research can compare manual and automated coding on passive (free recall) versus active (experimental) questions.

Relevant to the above, our analyses are based on responses to one free recall question. In real life, interviews are usually longer and involve more specific questions (Griffiths & Milne, 2006; Oxburgh et al., 2010). There are also instances where a suspect may refuse to respond to questions (Moston et al., 1992). Thus, our results cannot generalize to all contexts and are limited to free recalls. We encourage researchers to compare manual and automated coding on other types of questions such as probing questions (Hartwig et al., 2011).

The analyses were limited to eight datasets collected in the same lab. Our research questions can be tested on more datasets by different labs and also in different countries. While our research involved the recruitment of participants from non-WEIRD countries, it is fundamental to recruit participants from different cultures as that may yield different results (Leal, Vrij, Vernham, et al., 2019; Taylor et al., 2017). Further, automated coding has yet to be tested on datasets in real-life forensic interviews where stakes are usually higher and may differ from stakes in laboratory settings. Whereas some research suggests that higher stakes affect differences in honest and false accounts (ten Brinke & Porter, 2013), a meta-analysis showed null effects (Hartwig & Bond, 2014). It is worth examining if and how suspects would change their language when they know that an automated system will be used to assess their accounts.

We compared manual and automated coding on one veracity cue (total details) in Experiment 1. Other cues that reflect richness within an account (e.g., person details, location details) can also be assessed. LIWC does not code these details in the same manner as manual coding. For example, pronouns, names, and people descriptions are coded as person details in PLATO manual coding schemes, but LIWC has different categories related to people (e.g., pronouns, social processes, body parts, etc.). The coding process would become subjective if the researcher has to decide on which LIWC categories to include under "person details" in the analysis. Other sophisticated software programs may be more appropriate for coding these details. For example, SpaCy can code "person" entities in a manner that is comparable to manual coding and can also account for redundant and non-redundant entities. When a software program already has a specific entity (category), researchers from different labs can use that same entity which creates a more standardized coding scheme across experiments and allows for a more proper comparison between outputs (Nahari & Vrij, 2015). Such software programs may also result in a higher accuracy rate than the more commonly used LIWC (Duran et al., 2010; Kleinberg, Mozes, et al., 2018; Kleinberg et al., 2017).

We further encourage the testing of other stylometric features. A major advantage of automated coding is that it allows for more sophisticated coding (e.g., by examining patterns in language or by coding multiple cues simultaneously) that humans are not capable of doing (Chung & Pennebaker, 2007; Hauch et al., 2015). Future research can look at features that were not widely examined in automated lie detection research but that have shown promising results, including sentence structure (Dykstra et al., 2022), average sentence and word length (Afroz et al., 2012; Zhou et al., 2004), and word concreteness (Kleinberg et al., 2019).

We found that automated coding can differentiate honest and false accounts on the verbal cues total details and redundancy. While we reported the advantages of automated coding and while we acknowledge that many advancements have been incorporated on coding software programs to enhance lie detection, automated coding has its own limitations (Tomas et al., 2022). First, although automated coding can examine content to a certain extent (such as words with similar meanings), it cannot accurately capture the context of an account such as its plausibility and predictability which may explain the conflicting results between different software programs (Hauch et al., 2015; Mann et al., 2023). Second, it cannot differentiate words used in different contexts (Chung & Pennebaker, 2007). For example, the word "lie" has different meanings in "She is lying to me" versus "She is lying on the floor." Third, while automated coding is more objective than manual coding, it is still subjective as different software programs include different libraries and dictionaries which varies their lie detection accuracy. Fourth, while automated coding has been recommended as an objective alternative to manual coding, it can still be biased as it was originally developed by humans and the output often requires human interpretation which is often bias- and error-prone (Jupe & Keatley, 2020; Kassin et al., 2013). Fifth, overreliance on automation can lead to erroneous decision-making. In applied forensic settings, interviewers may start basing their decisions solely on the automation output rather than on the overall evidence they have acquired which may lead to guilty suspects being judged as innocent or vice versa (Kleinberg & Verschuere, 2021; Tomas et al., 2022). Sixth, automated coding cannot be used in all contexts and at all times. For example, patrol officers who interview people in the field or on the spot do not have access to computerized venues, Also, in combat and military contexts, automated coding software may not be accessible.

Conclusions

The replicability crisis has taken its toll on the psychology field, so it is important to standardize procedures that yield robust and replicable results (Pashler & Wagenmakers, 2012; Tomas et al., 2022). For lie detection research, coding is a very important aspect of assessing accounts, and the subjectivity in coding which in many cases yields low inter-rater reliability scores is an obstacle for replicable results. Thus, automated coding has been suggested as a solution to this problem while at the same time allowing for a faster assessment of accounts than manual coding.

In the present research, we showed that automated software programs can indeed detect differences between honest and false free recalls on total details and redundant details, but the extent to which these differences are captured varies depending on the program used. In addition, automated coding performance was similar to manual coding when classifying truths and lies, at least in the tested context. The overall results thus implicate that both manual and automated coding could be implemented for lie detection purposes. Where time resources are limited, technology that automatically transcribes an interviewee's free recall, coupled with automated coding of total details and redundancy, can be used.

Replication package. This article was co-authored by Gerges Dib in his personal capacity. The views expressed in this article are his own and do not necessarily reflect those of Amazon.com, Inc.

The material, data, and analyses are available in the repository of the University of Portsmouth at https://doi.org/10.17029/73bf0f42-b599-4c36-81b7-0c87befb795f.

Competing interests. The authors declare none.

References

Addawood, A., Badawy, A., Lerman, K., & Ferrara, E. (2019). Linguistic cues to deception: Identifying political trolls on social media. Proceedings of the International AAAI Conference on Web and Social Media, 13, 15-25. https://doi.org/10.1609/icwsm.v13i01.3205

Afroz, S., Brennan, M., & Greenstadt, R. (2012, May). Detecting hoaxes, frauds, and deception in writing style online. In 2012 IEEE Symposium on Security and Privacy (pp. 461-475). https://10.1109/SP.2012.34

Alison, L., Alison, E., Noone, G., Elntib, S., Waring, S., & Christiansen, P. (2014). The efficacy of rapport-based techniques for minimizing counter-interrogation tactics amongst a field sample of terrorists. *Psychology, Public Policy, and Law*, **20**(4), 421–430.

Amado, B. G., Arce, R., Farina, F., & Vilarino, M. (2016). Criteria-Based Content Analysis (CBCA) reality criteria in adults: a meta-analytic review. *International Journal of Clinical and Health Psychology*, **16**(2), 201–210.

Arslan, A., & Erslan, A. (2019). Lexical analysis of a textbook based on the EVP. *International Journal of Languages' Education and Teaching*, **7**(1), 1–12.

Bax, S., Nakatsuhara, F., & Waller, D. (2019). Researching L2 writers' use of meta-discourse markers at intermediate and advanced levels. *System*, **83**, 79–95.

Bogaard, G., Colwell, K., & Crans, S. (2019). Using the reality interview improves the accuracy of the criteria-based content analysis and reality monitoring. *Applied Cognitive Psychology*, **33**(6), 1018–1031.

Bogaard, G., Meijer, E. H., & Van der Plas, I. (2020). A model statement does not enhance the verifiability approach. *Applied Cognitive Psychology*, **34**(1), 96–105.

Bogaard, G., van der Mark, J., & Meijer, E. H. (2019). Detecting false intentions using unanticipated questions. *Plos one*, **14**(12), e0226257.

Bond, G. D., Holman, R. D., Eggert, J. A. L., Speller, L. F., Garcia, O. N., Mejia, S. C.; Mcinnes, K. W., Ceniceros, E. C., & Rustige, R. (2017). 'Lyin' Ted', 'Crooked Hillary', and 'Deceptive Donald': language of lies in the 2016 US Presidential debates. *Applied Cognitive Psychology*, **31**(6), 668–677.

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.

Burgoon, J. K. (2018). Predicting veracity from linguistic indicators. *Journal of Language and Social Psychology*, **37**(6), 603–631.

Ceballos Delgado, A. A., Glisson, W. B., Shashidhar, N., McDonald, J. T., Grispos, G., & Benton, R. (2021). Detecting deception using machine learning. Proceedings of the 54th Hawaii International Conference on System Sciences. https://shsu-ir.tdl.org/items/39924527-455d-4016-b0ba-569d3080c948

Chan, S., & Bull, R. (2014). The effect of co-offender planning on verbal deception. *Psychiatry, Psychology and Law*, **21**(3), 457–464.

Chen, X., Ita Levitan, S., Levine, M., Mandic, M., & Hirschberg, J. (2020). Acoustic-prosodic and lexical cues to deception and trust: deciphering how people detect lies. *Transactions of the Association for Computational Linguistics*, **8**, 199–214.

Chung, C., & Pennebaker, J. (2007). The psychological functions of function words. In K. Fiedler (Ed.), *Social communication* (pp. 343–359). Psychology Press.

Cohen, J. (1992). Quantitative methods in psychology: a power primer. *Psychological Bulletin*, **112**(1),155–159.

Colwell, K., Hiscock-Anisman, C. K., Memon, A., Taylor, L., & Prewett, J. (2007). Assessment Criteria Indicative of Deception (ACID): an integrated system of investigative interviewing and detecting deception. *Journal of Investigative Psychology and Offender Profiling*, **4**(3), 167–180.

Davis, M., Markus, K. A., Walters, S. B., Vorus, N., & Connors, B. (2005). Behavioral cues to deception vs. topic incriminating potential in criminal confessions. *Law and Human Behavior*, **29**(6), 683–704.

Deeb, H., Vrij, A., Hope, L., Mann, S., Granhag, P. A., & Lancaster, G. L. (2017). Suspects' consistency in statements concerning two events when different question formats are used. *Journal of Investigative Psychology and Offender Profiling*, **14**(1), 74–87.

Deeb, H., Vrij, A., & Leal, S. (2020). The effects of a Model Statement on information elicitation and deception detection in multiple interviews. *Acta Psychologica*, **207**, 103080.

Deeb, H., Vrij, A., Leal, S., & Burkhardt, J. (2021). The effects of sketching while narrating on information elicitation and deception detection in multiple interviews. *Acta Psychologica*, **213**, 10236.

Deeb, H., Vrij, A., Leal, S., Fallon, M., Mann, S., Luther, K., & Granhag, P. A. (2022a). Mapping details to elicit information and cues to deceit: the effects of map richness. *The European Journal of Psychology Applied to the Legal Context*, **14**(1), 11–19.

Deeb, H., Vrij, A., Leal, S., Fallon, M., Mann, S., Luther, K., & Granhag, P. A. (2022b). Sketching routes to elicit information and cues to deceit. *Applied Cognitive Psychology*, **36**(5), 20149–1059.

Deeb, H., Vrij, A., Leal, S., Giorgianni, D., Hypšová, P., & Mann, S. (2024). Exposing suspects to their sketches in repeated interviews to elicit information and veracity cues. *The European Journal of Psychology Applied to the Legal Context*, **16**(1), 1–15.

Deeb, H., Vrij, A., Leal, S., & Mann, S. (2021). Combining the Model Statement and the sketching while narrating interview techniques to elicit information and detect lies in multiple interviews. *Applied Cognitive Psychology*, **35**(6), 1478–1491.

Deeb, H., Vrij, A., Leal, S., Mann, S., & Burkhardt, J. (2022c). The Model Sketch for enhancing lie detection and eliciting information. *Brain Sciences*, **12**(9), 1180.

Denault, V., Talwar, V., Plusquellec, P., & Larivière, V. (2022). On deception and lying: an overview of over 100 years of social science research. *Applied Cognitive Psychology*, **36**(4), 805–819.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, **129**(1), 74–118.

Dunbar, N. E., Burgoon, J. K., Chen, X., Wang, X., Ge, S., Huang, Q., & Nunamaker, J. (2023). Detecting ulterior motives from verbal cues in group deliberations. *Frontiers in Psychology*, **14**, 1166225.

Duran, N. D., Hall, C., McCarthy, P. M., & McNamara, D. S. (2010). The linguistic correlates of conversational deception: comparing natural language processing technologies. *Applied Psycholinguistics*, **31**(3), 439–462.

Dykstra, V. W., Lyon, T. D., & Evans, A. D. (2022). Maltreated and non-maltreated children's truthful and dishonest reports: linguistic and syntactic differences. *Frontiers in Psychology*, **13**, 1025419.

Dzindolet, M. T., & Pierce, L. G. (2004, September). A computerized text analysis can detect deception. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 48(3), 635–639. SAGE Publications.

Dzindolet, M. T., & Pierce, L. G. (2005, September). Using a linguistic analysis tool to detect deception. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 49(3), 563–567. SAGE Publications.

Feldman, G., Lian, H., Kosinski, M., & Stillwell, D. (2017). Frankly, we do give a damn: the relationship between profanity and honesty. *Social Psychological and Personality Science*, **8**(7), 816–826.

Forsyth, L., & Anglim, J. (2020). Using text analysis software to detect deception in written short-answer questions in employee selection. *International Journal of Selection and Assessment*, **28**(3), 236–246.

Fuller, C. M., Biros, D. P., Burgoon, J., & Nunamaker, J. (2013). An examination and validation of linguistic constructs for studying high-stakes deception. *Group Decision and Negotiation*, **22**, 117–134.

Gallucci, M. (2019). GAMLj: General analyses for linear models. [Jamovi module]. https://gamlj.github.io/

Gancedo, Y., Fariña, F., Seijo, D., Vilariño, M., & Arce, R. (2021). Reality monitoring: a meta-analytical review for forensic practice. *The European Journal of Psychology Applied to Legal Context*, **13**(2), 99–110.

Gayed, J. M., Carlon, M. K. J., Oriola, A. M., & Cross, J. S. (2022). Exploring an AI-based writing Assistant's impact on English language learners. *Computers and Education: Artificial Intelligence*, **3**, 100055.

Granhag, P. A., & Hartwig, M. (2008). A new theoretical perspective on deception detection: on the psychology of instrumental mind-reading. *Psychology, Crime & Law*, **14**(3), 189–200.

Granhag, P. A., & Strömwall, L. A. (1999). Repeated interrogations – Stretching the deception detection paradigm. *Expert Evidence*, **7**, 163–174.

Granhag, P. A., Vrij, A., & Verschuere, B. (Eds.) (2015). *Detecting deception: Current challenges and cognitive approaches*. Chichester, United Kingdom: John Wiley & Sons.

Griffiths, A., & Milne, R. (2006) Will it all end in tiers? Police interviews with suspects in Britain. In T. A. Williamson (Ed.), *Investigative interviewing: rights, research, regulation* (pp. 167–189). Willan.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, **8**(1), 23–34.

Hartwig, M., & Bond, C. F., Jr. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin*, **137**(4), 643–659.

Hartwig, M., & Bond Jr, C. F. (2014). Lie detection from multiple cues: a meta-analysis. *Applied Cognitive Psychology*, **28**(5), 661–676.

Hartwig, M., Granhag, P. A., & Strömwall, L. (2007). Guilty and innocent suspects' strategies during police interrogations. *Psychology, Crime, & Law*, **13**(2), 213–227.

Hartwig, M., Granhag, P. A., Stromwall, L., Wolf, A. G., Vrij, A., & Hjelmsäter, E. R. A. (2011). Detecting deception in suspects: verbal cues as a function of interview strategy. *Psychology, Crime & Law*, **17**(7), 643–656.

Hartwig, M., Granhag, P. A., Stromwall, L. A., & Doering, N. (2010). Impression and information management: on the strategic self-regulation of innocent and guilty suspects. *The Open Criminology Journal*, **3**(1), 10–16.

Harvey, A. C., Vrij, A., Leal, S., Hope, L., & Mann, S. (2017). Deception and decay: verbal lie detection as a function of delay and encoding quality. *Journal of Applied Research in Memory and Cognition*, **6**(3), 306–318.

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.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, **33**(2-3), 61–83.

Hines, A., CoIweII, K., Hiscock-Anisman, C., Garrett, E., Ansarra, R., & Montalvo, L. (2010). Impression management strategies of deceivers and honest reporters in an investigative interview. *European Journal of Psychology Applied to Legal Context*, **2**(1), 73–90.

Hirschberg, J. B., Benus, S., Brenier, J. M., Enos, F., Friedman, S., Gilman, S., Girand, C., Graciarena, M., Kathol, A., Michaelis, L., Pellom, B, Stolcke, A., & Shriberg, E. (2005). Distinguishing deceptive from non-deceptive speech. Proceedings of Eurospeech'05. Columbia University Libraries. https://doi.org/10.7916/D8697C06

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, **88**(1), 67–85.

Jupe, L. M., & Keatley, D. A. (2020). Airport artificial intelligence can detect deception: or am i lying?. *Security Journal*, **33**, 622–635.

Jupe, L. M., Vrij, A., Leal, S., & Nahari, G. (2018). Are you for real? Exploring language use and unexpected process questions within the detection of identity deception. *Applied Cognitive Psychology*, **32**(5), 622–634.

Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition*, **2**(1), 42–52.

Khyani, D., Siddhartha, B. S., Niveditha, N. M., & Divya, B. M. (2021). An interpretation of lemmatization and stemming in natural language processing. *Journal of University of Shanghai for Science and Technology*, **22**(10), 350–357.

Kleinberg, B., Mozes, M., Arntz, A., & Verschuere, B. (2018). Using named entities for computer-automated verbal deception detection. *Journal of Forensic Sciences*, **63**(3), 714–723.

Kleinberg, B., Nahari, G., Arntz, A, & Verschuere, B. (2017). An investigation on the detectability of deceptive intent about flying through verbal deception detection. *Collabra: Psychology*, **3**(1), 21.

Kleinberg, B., van der Toolen, Y., Vrij, A., Arntz, A., & Verschuere, B. (2018). Automated verbal credibility assessment of intentions: the model account technique and predictive modeling. *Applied Cognitive Psychology*, **32**(3), 354–366.

Kleinberg, B., van der Vegt, I., & Arntz, A. (2019). Detecting deceptive communication through linguistic concreteness. *PsyArXiv Preprints*. https://doi.org/10.31234/osf.io/p3qjh

Kleinberg, B., & Verschuere, B. (2021). How humans impair automated deception detection performance. *Acta Psychologica*, **213**, 103250.

Kleinberg, B., Warmelink, L., Arntz, A., & Verschuere, B. (2018). The first direct replication on using verbal credibility assessment for the detection of deceptive intentions. *Applied Cognitive Psychology*, **32**(5), 592–599.

Lakens, D., & Evers, E. R. (2014). Sailing from the seas of chaos into the corridor of stability: practical recommendations to increase the informational value of studies. *Perspectives on Psychological Science*, **9**(3), 278–292.

Leal, S., Vrij, A., Deeb, H., Burkhardt, J., Dabrowna, O., & Fisher, R. P. (2023). Verbal cues to deceit when lying through omitting information: examining the effect of a Model Statement interview protocol. *European Journal of Psychology Applied to Legal Context*, **15**(1), 1–8.

Leal, S., Vrij, A., Deeb, H., & Jupe, L. (2018). Using the Model Statement to elicit verbal differences between truth tellers and liars: the benefit of examining core and peripheral details. *Journal of Applied Research in Memory and Cognition*, **7**(4), 610–617.

Leal, S., Vrij, A., Mann, S., & Fisher, R. P. (2010). Detecting true and false opinions: the Devil's Advocate approach as a lie detection aid. *Acta Psychologica*, **134**(3), 323–329.

Leal, S., Vrij, A., Vernham, Z., Dalton, G., Jupe, L., Harvey, A., & Nahari, G. (2018). Cross-cultural verbal deception. *Legal and Criminological Psychology*, **23**(2), 192–213.

Leal, S., Vrij, A., Vernham, Z., Dalton, G., Jupe, L. M., Nahari, G., & Rozmann, N. (2019). Using the model statement to elicit verbal differences between truth tellers and liars amongst Arab interviewees: A partial replication of Leal, Vrij, Deeb, and Jupe (2018). *Applied Cognitive Psychology*, **33**(6), 1008–1017.

Leins, D. A., Fisher, R. P., & Ross, S. J. (2013). Exploring liars' strategies for creating deceptive reports. *Legal and Criminological Psychology*, **18**(1), 141–151.

Leins, D. A., Fisher, R. P., & Vrij, A. (2012). Drawing on liars' lack of cognitive flexibility: detecting deception through varying report modes. *Applied Cognitive Psychology*, **26**(4), 601–607.

Liu, M. (2016). Verbal communication styles and culture. In *Oxford Research Encyclopedia of Communication*. Oxford University Press.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.

Mac Giolla, E., & Luke, T. J. (2021). Does the cognitive approach to lie detection improve the accuracy of human observers? *Applied Cognitive Psychology*, **35**(2), 385–392.

Mann, S., Vrij, A., Deeb, H., & Leal, S. (2023). All mouth and trousers? Use of the Devil's Advocate questioning protocol to determine authenticity of opinions about protester actions. *Psychiatry, Psychology and Law*. Advance online publication. https://doi.org/10.1080/13218719.2023.2242433

Markowitz, D. M., & Griffin, D. J. (2020). When context matters: how false, truthful, and genre-related communication styles are revealed in language. *Psychology, Crime & Law*, **26**(3), 287–310.

Masip, J., Bethencourt, M., Lucas, G., Segundo, M. S. S., & Herrero, C. (2012). Deception detection from written accounts. *Scandinavian Journal of Psychology*, **53**(2), 103–111.

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.

Mbaziira, A., & Jones, J. (2016). A text-based deception detection model for cybercrime. International Conference on Technology and Management. https://www.researchgate.net/profile/Alex-Mbaziira/publication/307594168_A_Text-based_Deception_Detection_Model_for_Cybercrime/links/5851b1be08aef7d030a19101/A-Text-based-Deception-Detection-Model-for-Cybercrime.pdf

McCornack, S. A. (1992). Information manipulation theory. *Communications Monographs*, **59**(1), 1–16.

Moston, S., Stephenson, G. M., & Williamson, T. M. (1992). The effects of case characteristics on suspect behaviour during police questioning. *The British Journal of Criminology*, **32**(1), 23–40.

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.

Nahari, G. (2018). Reality monitoring in the forensic context: digging deeper into the speech of liars. *Journal of Applied Research in Memory and Cognition*, **7**(3), 432–440.

Nahari, G. (2023). Meta-research perspectives on verbal lie detection. *Brain Sciences*, **13**(3), 392.

Nahari, G., & Vrij, A. (2015). Systematic errors (biases) in applying verbal lie detection tools: richness in detail as a test case. *Crime Psychology Review*, **1**(1), 98–107.

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.

O'Connell, F., Carter, C., Taylor, P., Vernham, Z., & Warmelink, L. (2023). The effect of individual differences in episodic future thought on perceived credibility. *Psychology, Crime & Law*. Advance online publication. https://doi.org/10.1080/1068316X.2023.2226293

Oberlader, V. A., Naefgen, C., Koppehele-Gossel, J., Quinten, L., Banse, R., & Schmidt, A. F. (2016). Validity of content-based techniques to distinguish true and fabricated statements: a meta-analysis. *Law and Human Behavior*, **40**(4), 440–457.

Oxburgh, G. E., Myklebust, T., & Grant, T. (2010). The question of question types in police interviews: a review of the literature from a psychological and linguistic perspective. *International Journal of Speech, Language & the Law*, **17**(1), 45–66.

Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: a crisis of confidence?. *Perspectives on Psychological Science*, **7**(6), 528–530.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015 [Texas Scholar Works]. University of Texas Libraries. http://hdl.handle.net/2152/31333

Pennebaker, J. W., & Graybeal, A. (2001). Patterns of natural language use: Disclosure, personality, and social integration. *Current Directions in Psychological Science*, **10**(3), 90-93.

Plisson, J., Lavrac, N., & Mladenic, D. (2004, October). A rule based approach to word lemmatization. In Proceedings of IS (Vol. 3, pp. 83-86). https://ailab.ijs.si/dunja/SiKDD2004/Papers/Pillson-Lematization.pdf

Plotkina, D., Munzel, A., & Pallud, J. (2020). Illusions of truth—Experimental insights into human and algorithmic detections of fake online reviews. *Journal of Business Research*, **109**, 511–523. https://doi.org/10.1016/j.jbusres.2018.12.009

Rodríguez, M. A. (2023). Text Inspector corpus linguistics tool on trial: Checking accuracy for students' writings assessment [Student dissertation, Valladolid University]. https://uvadoc.uva.es/bitstream/handle/10324/64431/TFG_F_2023_145.pdf?sequence = 1&isAllowed = y

Schutte, M., Bogaard, G., Mac Giolla, E., Warmelink, L., & Kleinberg, B. (2021). Man versus Machine: Comparing manual with LIWC coding of perceptual and contextual details for verbal lie detection. PsyarXiv Preprint. https://psyarxiv.com/cth58/

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.

Sporer, S. L., & Sharman, S. J. (2006). Should I believe this? Reality monitoring of accounts of self-experienced and invented recent and distant autobiographical events. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, **20**(6), 837–854.

Strömwall, L. A., & Willén, R. M. (2011). Inside criminal minds: offenders' strategies when lying. *Journal of Investigative Psychology and Offender Profiling*, **8**(3), 271–281.

Tate, R. L., & Pituch, K. A. (2007). Multivariate hierarchical linear modeling in randomized field experiments. *The Journal of Experimental Education*, **75**(4), 317–337.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, **29**(1), 24–54.

Taylor, P. J., Larner, S., Conchie, S. M., & Menacere, T. (2017). Culture moderates changes in linguistic self-presentation and detail provision when deceiving others. *Royal Society Open Science*, **4**(6), 170128.

Taylor, P. J., Larner, S., Conchie, S. M., & van der Zee, S. (2015). Cross-cultural deception detection. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Detecting deception: Current challenges and cognitive approaches* (pp. 175–201). John Wiley & Sons.

ten Brinke, L., & Porter, S. (2013). Discovering deceit: Applying laboratory and field research in the search for truthful and deceptive behavior. In B.S. Cooper, D. Griesel, & M. Ternes (Eds.), *Applied issues in investigative interviewing, eyewitness memory, and credibility assessment* (pp. 221–237). Springer.

Toma, C. L., & Hancock, J. T. (2012). What lies beneath: the linguistic traces of deception in online dating profiles. *Journal of Communication*, **62**(1), 78–97.

Tomas, F., Dodier, O., & Demarchi, S. (2022). Computational measures of deceptive language: prospects and issues. *Frontiers in Communication*, **7**, 792378.

van der Zee, S., Poppe, R., Havrileck, A., & Baillon, A. (2022). A personal model of trumpery: linguistic deception detection in a real-world high-stakes setting. *Psychological Science*, **33**(1), 3–17.

Verigin, B. L., Meijer, E. H., Bogaard, G., & Vrij, A. (2019). Lie prevalence, lie characteristics and strategies of self-reported good liars. *PloS one*, **14**(12), e0225566.

Vrij, A. (2005). Criteria-based content analysis: a qualitative review of the first 37 studies. *Psychology, Public Policy, and Law*, **11**(1), 3–41.

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. John Wiley & Sons.

Vrij, A., Fisher, R. P., & Blank, H. (2017). A cognitive approach to lie detection: a meta-analysis. *Legal and Criminological Psychology*, **22**(1), 1–21.

Vrij, A., Granhag, P. A., Ashkenazi, T., Ganis, G., Leal, S., & Fisher, R. P. (2022). Verbal lie detection: its past, present and future. *Brain Sciences*, **12**(12), 1644.

Vrij, A., Granhag, P. A., Mann, S., & Leal, S. (2011). Outsmarting the liars: toward a cognitive lie detection approach. *Current Directions in Psychological Science*, **20**(1), 28–32.

Vrij, A., Hartwig, M., & Granhag, P. A. (2019). Reading lies: nonverbal communication and deception. *Annual Review of Psychology*, **70**, 295–317.

Vrij, A., Leal, S., Deeb, H., Castro Campos, C., Fisher, R. F., Mann, S., Jo, E., & Alami, N. (2022). The effect of using countermeasures in interpreter-absent and interpreter-present interviews. *The European Journal of Psychology Applied to Legal Context*, **14**(2), 53–72.

Vrij, A., Leal, S., Deeb, H., & Fisher, R. P. (2023). Verbal veracity indicators and the efficacy of countermeasures in three non-WEIRD populations. *Journal of Psychology & Behavior Research*, **5**(1), 34–63.

Vrij, A., Leal, S., Fisher, R. P., Mann, S., Deeb, H., Jo, E., Campos, C. C., Hamzeh, S. (2020). The efficacy of using countermeasures in a Model Account interview. *The European Journal of Psychology Applied to Legal Context*, **12**(1), 23–34.

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**, 159–166.

Vrij, A., Leal, S., Jupe, L., & Harvey, A. (2018). Within-subjects verbal lie detection measures: a comparison between total detail and proportion of complications. *Legal and Criminological Psychology*, **23**(2), 265–279.

Vrij, A., Mann, S., Kristen, S., & Fisher, R. P. (2007). Cues to deception and ability to detect lies as a function of police interview styles. *Law and Human Behavior*, **31**(5), 499–518.

Vrij, A., Mann, S., Leal, S., & Granhag, P. A. (2010). Getting into the minds of pairs of liars and truth tellers: an examination of their strategies. *The Open Criminology Journal*, **3**, 17–22.

Vrij, A., Palena, N., Leal, S., & Caso, L. (2021). The relationship between complications, common knowledge details and self-handicapping strategies and veracity: a meta-analysis. *European Journal of Psychology Applied to Legal Context*, **13**(2), 55–77.

Weblingua (2022). Text Inspector [software]. https://textinspector.com/help/who-are-we

West, B. T., Welch, K. B., & Galecki, A. T. (2006). *Linear mixed models: a practical guide using statistical software*. Chapman and Hall/CRC.

Zhou, L. (2005). An empirical investigation of deception behavior in instant messaging. *IEEE Transactions on Professional Communication*, **48**(2), 147–160.

Zhou, L., Burgoon, J. K., Twitchell, D. P., Qin, T., & Nunamaker Jr, J. F. (2004). A comparison of classification methods for predicting deception in computer-mediated communication. *Journal of Management Information Systems*, **20**(4), 139–166.

Zhou, L., Twitchell, D. P., Qin, T., Burgoon, J. K., & Nunamaker, J. F. (2003, January). An exploratory study into deception detection in text-based computer-mediated communication. In the Proceedings of the 36th Annual Hawaii International Conference on System Sciences. https://ieeexplore.ieee.org/abstract/document/1173793?casa_token=bk0BwJQmmcoAAAAA:mv5amH6frU8ztr7M7Blq2jUh6HR74_mLNPvr6NDu73tsLX9KGMUsHyDGX_KtUx9_oBLQlb0Em0VteeM

Cite this article: Deeb, H., Vrij, A., Palena, N., Hypšová, P., Dib, G., Leal, S., & Mann, S. (2024). Honesty repeats itself: comparing manual and automated coding on the veracity cues total details and redundancy. *Applied Psycholinguistics*. https://doi.org/10.1017/S0142716424000298