Applied Cognitive Psychology

WILEY

RESEARCH ARTICLE OPEN ACCESS

Reaction Time as an Actual and a Perceived Cue to Deception Under Cognitive Load

Evan Brennan | Keith A. Hutchison

Montana State University, Bozeman, Montana, USA

Correspondence: Keith A. Hutchison (khutch@montana.edu)

Received: 31 October 2024 | **Revised:** 14 May 2025 | **Accepted:** 18 May 2025

Funding: The authors received no specific funding for this work.

Keywords: cognitive load | deception | reaction time

ABSTRACT

The predictive validity of reaction time as an actual (objective) and a perceived cue to deception was tested in two experiments differing in question presentation methodology. Participants were video recorded while giving truthful and dishonest verbal responses to autobiographical questions under high and low cognitive load, and coders later viewed the recordings to detect their responses. We hypothesized that lie reaction times (RTs) would be significantly longer than truthful RTs and that longer RTs would be associated with differential lie and truth detection accuracy. We did not make any predictions regarding cognitive load, considering the current literature has produced mixed results. Our hypotheses were supported by the data. Results of our load manipulation differed between Experiment 1 and Experiment 2, leaving us questioning its validity. We provide many suggestions for future research regarding experimental methodologies measuring deception and cognitive load.

1 | Introduction

Decades of deception research demonstrates that there are no universally valid individual cues to detect deception (Hartwig and Bond 2011; Vrij 2008) and that humans fare no better than chance accuracy at discriminating truthful from dishonest messages (Bond and DePaulo 2006). However, recent research on the cognitive processes involved in deception has found promising evidence for the use of reaction time (RT) measures as an indicator of deceit (Suchotzki et al. 2017). We define RT as synonymous with response latency, which is the time elapsed from the end of a question being presented until the initiation of a response (DePaulo et al. 2003). Lies are generally associated with longer RTs than truthful responses, which may indicate deception to be a more cognitively demanding activity. Walczyk et al. (2003) demonstrated this to be at least partially due to the construction (Experiment 1) and decision-making (Experiment 2) components that theoretically underlie the process of lie generation, finding lies to take reliably longer to initiate than truthful responses in both experiments (see Walczyk et al. 2003, 2014 for a detailed description of the decision and construction components involved in lie generation). Furthermore, paradigms like the Concealed Information Test (formally the Guilty Knowledge Test) in which participants must only respond with simple yes or no verbal or behavioral responses to old or new stimuli show evidence of longer RTs to stimuli participants deceptively conceal from examiners (Seymour et al. 2000), which suggests a higher cognitive load experienced when deceiving driven by necessarily suppressing truthful information (Vrij 2015).

Because deception is theorized as more cognitively demanding than telling the truth (Gombos 2006; Zuckerman et al. 1981), a common experimental approach to assess differences between liars and truth tellers is to increase cognitive load upon individuals answering questions, with the goal of exacerbating observable cues in liars (Blandón-Gitlin et al. 2014; Vrij 2015).

Evidence is mixed as to whether increasing cognitive load disproportionately affects liars' RTs (Verschuere et al. 2018) and further research is needed to examine this relationship.

Despite evidence suggesting that RT is an objective cue to deception, there is little evidence that human lie detectors reliably use RTs to detect deception. It is an empirical question whether longer RTs observed in experimental settings are associated with lie judgments made by detectors. Furthermore, the influence of cognitive load on RTs and deception detection is unexplored.

1.1 | Reaction Time as an Actual Cue to Deception

RTs were originally considered unreliable deception cues in early meta-analyses (DePaulo et al. 2003; Zuckerman et al. 1981). However, studies included in these meta-analyses used a broad range of paradigms and imprecise measures for assessing RTs (e.g., stopwatches; Suchotzki et al. 2017). Moreover, DePaulo et al. (2003) found that the RT effect size between unplanned and planned deceptive responses was significant ($d = 0.20$), highlighting planning as an important moderator of deceptive RTs. This suggests that RTs may be effective deception cues under certain conditions, such as when lies are told spontaneously relative to when they are rehearsed.

More recent meta-analyses provide promising evidence that RTs are an objective deception cue. For example, Sporer and Schwandt (2006) assessed paraverbal deception cues, a category that includes RT, message duration, total words produced, pauses, speech errors, speech rate, repetitions, and speech pitch, finding that lie RTs were significantly longer than truth RTs across five out of their six moderator variable analyses. They concluded that longer lie RTs reflect cognitive difficulty rather than any experienced emotions or arousal, finding the largest effect size between studies varying in preparation time. Further support for RTs as a deception cue was demonstrated in Suchotzki et al. (2017), whose meta-analysis accounted for limitations presented in DePaulo et al. (2003) by using computer measures to record RTs, instructing participants to answer each question immediately after it was asked, and averaging RTs across a large number of trials (~20/condition). The authors found a large RT effect ($d = 1.049$) even after accounting for publication bias, such that lies took reliably longer to initiate than truthful responses.

1.2 | Reaction Time as a Perceived Cue to Deception

There is evidence that RTs are used by human lie detectors when detecting deception from meta-analyses (Hartwig and Bond 2011; Zuckerman et al. 1981) and some older research. For example, Harrison et al. (1978) found long RTs were more often judged as lies, regardless of whether they preceded actual lie or truth responses. This suggests that long RTs are associated with an increased probability of correctly classifying lie responses, but also with a decreased probability of correctly classifying truth responses.

Other studies manipulated audio recordings by inserting pauses between questions and answers to test whether RT length would affect detection accuracy. Baskett and Freedle (1974) inserted nine different lengths of RTs [from very short (0.07 s) to very long (6.07 s)] into audio recordings of participants' stating whether certain adjectives were self-descriptive and then had detectors judge each responses' veracity. They found that either very short or very long RTs were significantly associated with the probability of responses being judged as lies, but that the intermediate range of RTs had no effect on detector judgments. Similarly, Kraut (1978, Experiment 2) inserted either a 1 or a 7 s pause into audio recordings in which a confederate either confessed or denied using marijuana in an interview. Detectors were more likely to judge the long pause denial responses as lies, but the long pause admission responses as truthful, as the former was seen as self-serving and the latter as self-damaging. Both studies suggest RTs may influence human lie detector accuracy, and the latter mentioned study suggests a combined effect of long RTs and the content of the lie on detection accuracy.

Finally, how may RTs be associated with deception detection in real-world settings? Experimental evidence shows that RTs can be used in questioning procedures like the concealed information test to measure differences in the time taken to respond to questions when individuals are concealing information relative to questions in which no information is concealed (Seymour et al. 2000). This suggests that under controlled questioning procedures, RTs may be used to discriminate individuals who are concealing information from those who are not concealing information, with information concealment (i.e., truth suppression) considered to be a necessary component of lying (Vrij 2015). Additionally, RT may serve as a substitute measure for pauses in deception detection, which Vrij and Mann (2001) found to reliably discriminate deceptive from truthful messages for a criminal suspect when interrogated for murder. Long pauses in that study were associated with having to think hard. In two follow-up studies analyzing high-stakes lies told by samples of 16 and 14 individuals, respectively, having to think hard was associated with pauses occurring during lie responses (Mann et al. 2002) and better lie detection accuracy for human lie detectors (Mann and Vrij 2006). Thus, the ability to detect lies in real-world settings may be affected by how long it takes someone to construct an answer either following a question or after they have begun speaking, suggesting lie detection may be affected by the length of RTs preceding answers.

1.3 | The Cognitive Load Approach (CLA)

Given that lying is theorized to impose an intrinsic cognitive load upon the deceiver (Walczyk et al. 2013; Zuckerman et al. 1981), researchers have attempted to exacerbate any observable deception cues by further increasing respondents' cognitive load, using what is called the cognitive load approach (CLA) (Vrij et al. 2008). The CLA proposes that increasing cognitive load should disproportionately affect lie responses more than truth responses, because the increased cognitive effort imposed by the load should interfere with the already cognitively effortful act of lying. Experimentally, cognitive load has been increased by having respondents maintain eye contact with the interviewer, tell their stories in reverse chronological order (Vrij et al. 2017), answer unexpected questions (Walczyk et al. 2013), or perform demanding secondary tasks (Vrij et al. 2006). These manipulations are effective at extracting cues exhibited by liars more so than truth-tellers during structured interviews (Vrij 2015); however, they are less often examined using experimental short-answer paradigms purposed to assess the cognitive processes that facilitate lie generation. Some studies have successfully increased lie RTs using cognitive load manipulations (Visu-Petra et al. 2013; Williams et al. 2013), but overall results are mixed regarding the effects of increasing cognitive load on lie and truth RTs (Verschuere et al. 2018). More research is needed to verify whether lie RTs can be reliably increased in experimental settings.

Because longer RTs when lying than telling the truth are suggestive of additional cognitive effort needed to lie (Sporer and Schwandt 2006; Walczyk et al. 2003), and because the CLA proposes that lying should be more discernable under high cognitive load than low cognitive load, it is possible that increasing cognitive load will be associated with increased RTs when lying and subsequently, increased human lie detector accuracy. We base this speculation on findings from Maldonado et al. (2018), in which participants answered questions truthfully and dishonestly under high and low cognitive load while either being simultaneously detected by another participant (Experiment 1a) or audio–video recorded and detected later by observers of the recordings (Experiment 1b). Before the deception task, participants completed two complex span tasks to assess working memory capacity, to assess the association between the cognitive ability of liars and the ability of observers to correctly detect their responses as truthful or dishonest. Maldonado et al. (2018) demonstrated that participants lower in working memory capacity were more detectable than higher working memory capacity participants under increased cognitive load. This possibly suggests that the complexity of lying under load was better handled by individuals higher in cognitive ability. Furthermore, those detecting participants in person were asked at the end of the experiment what they thought indicated that the other participant was lying, with ~75% of participants suggesting long RTs (e.g., "hesitation") as being associated with lie responses. Considering these findings, and because RTs were not measured in that study, our goal with the current study was to examine the relationship between RTs, cognitive load, and human lie detection. To our knowledge, no study has assessed this relationship, but we believe it is a worthwhile endeavor to improve our understanding of how cognitively difficult lies may be detected using RTs.

2 | Current Experiments and Hypotheses

We tested whether RT serves as both an actual and perceived cue to deception. We also tested whether imposing additional cognitive load upon participants would increase lie RTs, making them more discriminable from truth RTs, and whether this effect of cognitive load on RTs would influence observers' abilities to correctly detect participants' lie responses. Specifically, we conducted two experiments in which individuals gave spontaneous truthful and dishonest responses to test the following hypotheses:

H1. *RTs will be significantly longer preceding lie than truth responses.*

This hypothesis is based on the RT deception effect size reported in Suchotzki et al. (2017), as well as the finding that spontaneously delivered lies are associated with longer RTs than spontaneously delivered truthful messages (DePaulo et al. 2003; Sporer and Schwandt 2006).

H2. *Long lie RTs will be associated with increased lie detection accuracy and an increased probability of hits, in which detectors correctly classify lie responses as lie responses.*

This hypothesis is based on a majority of Maldonado et al.'s (2018, Experiment 1a) sample mentioning RTs as used in deception detection, the other studies in which RTs were reported by detectors (Harrison et al. 1978; Hartwig and Bond 2011; Zuckerman et al. 1981), and studies in which experimentally manipulated RTs affected detection ability (Baskett and Freedle 1974; Kraut 1978, Experiment 2).

H3. *Long truth RTs will be associated with decreased truth detection accuracy and an increased probability of false alarms, in which detectors incorrectly classify truth responses as lie responses.*

This hypothesis is based on Harrison et al.'s (1978) finding that both truth and lie responses preceded by long RTs were judged as lies, and the finding that appearing to think hard before responding is more often an objective deception cue that discriminates lies and truthful messages given by individuals being interrogated for serious crimes (Mann et al. 2002; Vrij and Mann 2001) and a perceived deception cue that improves lie detection accuracy (Mann and Vrij 2006).

Because evidence is mixed regarding the effects of increasing cognitive load on deceptive RTs (Verschuere et al. 2018) and because the relationship between cognitive load, RTs, and detectability has not been empirically examined, no hypotheses were made regarding our cognitive load manipulation. We performed exploratory analyses testing whether increasing cognitive load disproportionately increases lie RTs, which may subsequently affect lie detection.

3 | Experiment 1

Our first experiment was conducted similarly to Maldonado et al. (2018), but included RTs as the main dependent measure. Our procedure followed the suggestions of Suchotzki et al. (2017) by using precise methods to measure RTs and instructing participants to answer rapidly following questions. Additionally, our deception task did not allow participants to prepare their deceptive responses in advance to prevent planning from potentially affecting RTs. Another test battery was performed to test separate hypotheses that were not part of the current study, so it will not be mentioned further.[1,2]

3.1 | Method

3.1.1 | Participants

The overall response type effect on RTs is very large ($d = 1.049$; Suchotzki et al. 2017). However, we were interested in the interaction between response type and cognitive load. Because the evidence for this interaction effect is mixed (Verschuere et al. 2018), we estimated our sample size to be large enough to yield a small-to-moderate interaction effect size (Cohen's d between 0.2 and 0.5). A power analysis using G*Power version 3.197 estimating a Cohen's $d=0.35$ (which corresponds to an effect size $f=0.175$) yielded a sample size of 72 participants for a repeated measures ANOVA with power of 0.95. By the end of the semester, we had run 76 students (60% male) from Montana State University, who participated via an online SONA systems recruitment pool in which participants signed up for an hour-long session to receive partial course credit for an Introductory to Psychology course. This experiment was approved by the Montana State University Institutional Review Board (Protocol # 2023-915-EXPEDITED) and all participants signed a written informed consent form immediately upon arrival. We did not ask participants to report their ages, but this population consists mostly 18- to 20-year-old students. Only one participant was allowed to sign up per session. Our final sample included 75 participants after we eliminated data from one participant whose RTs were dramatically greater than the mean (all conditions >3.5 SD).

3.1.2 | Design and Procedure

The experiment included both Load (high and low) and Response Type (truth and lie) manipulated within-subjects, and RTs as the dependent variable. When participants arrived, they signed a consent form and then received a 64-item autobiographical questionnaire taken from Maldonado et al. (2018) that they were instructed to answer truthfully. The questions required simple 1- to 2-word short answers. Maldonado et al. chose questions with simple 1- to 2-word answers because they allowed for intrusion of a correct response that needed to be suppressed when lying, controlled for complexity of answers (Walczyk et al. 2013), and avoided repetition, which allowed us to examine RTs without establishing episodic stimulus–response contingencies between questions and responses.

3.1.2.1 | Deception Task. Research assistants, who were unaware of the experimental hypotheses, ran participants through the study. Specifically, although the research assistants knew the study included prompts for participants to give lie and truth responses, they were not told anything about the possible effects of cognitive load or anything regarding a hypothesized relation between RT and the different response types. Participants first read instructions on screen that described the upcoming task. For each of four memory load blocks, participants would see an onscreen 4×4 matrix (adapted from Heyman et al. 2015) for 4 s containing either four dots in a straight line for the low load condition or six dots scattered throughout the matrix for the high load condition. Participants were told to remember the location of the dots for later recall. After the matrix disappeared, participants were instructed to respond as quickly as possible, via microphone response, to eight pseudo-randomly picked questions from the 64-item questionnaire completed earlier. Participants saw each of the eight questions in the following sequence and all computer instructions were displayed in Courier New 18-point font. First, a response cue ("Truth" or "Lie") appeared centrally on screen for 4 s to indicate how participants should respond to the following question. After the cue, a screen appeared showing "Experimenter will read question" and the experimenter read the question aloud from a script and then immediately pressed the space bar, which triggered the E-prime serial response box to record RTs from the attached microphone during a screen that displayed "Please provide your answer." We chose to have experimenters press the space bar to trigger the voice key because questions varied greatly in length (from 5 words in "What is your dream car?" to 15 words in "If you could only have one thing, what would you bring onto a deserted island?"). As a result, the measurement error introduced by differential reading times across questions would have been much greater than any potential error created by trial-to-trial differences in experimenter's time to press the space bar. We further discuss possible measurement error concerns from this procedure in the results and discussion sections. After the response, a 1000 ms blank screen appeared followed by the response cue for the next trial. After answering all eight questions, a blank 4×4 matrix appeared onscreen, and participants filled in the corresponding blank matrix on a paper in front of them with the locations of dots held in memory. Participant performance on the cognitive load dot matrix task was determined by how many dots they recalled in their correct location within the matrix, receiving a point for each correct location.

Participants completed a total of four blocks of eight questions each, with two blocks under low load and two blocks under high load. There were four versions to counterbalance questions across load and response type. Participants' faces and upper bodies were audio–video recorded throughout the task, so these recordings could be used in the detection task.

3.1.2.2 | Detection Task. Three research assistants,[3] who did not assist with the deception task procedure and were also unaware of the experimental hypotheses, served as coders of the video recordings of the participants from that task. Specifically, alhtough they knew that participants had given lie and truth responses, they were not told of the cognitive load manipulation or that the participants's RTs to answer questions had been collected. The coders were given packets with the participants' number at the top of the page and question numbers listed beneath them with blank lines next to the question numbers. The coders were instructed to watch the video recordings of the participants and to indicate in the blank lines whether they believed the participants were responding with a truthful response (by writing a "T") or a lie response (by writing an "L") after each of their answers. No further instructions were given as part of this procedure to the coders, and they were only to make their judgments based on their own intuitive beliefs about deception.

3.2 | Results

3.2.1 | Data Scoring

Correct responses for the deception task were based on the answers marked by participants on the autobiographical questionnaire. Correct lie responses were any responses given that differed from the answer marked on the questionnaire when instructed to lie, and correct truth responses matched those from the questionnaire when instructed to tell the truth. Only these correct responses were included in all RT analyses, and questions that participants did not answer were omitted before analysis. This involved removing trials with microphone errors (3.8%) or when the participant did not answer or comply with instructions (6.8%). Next, because RT distributions tend to be positively skewed, outliers were removed per Van Selst and Jolicoeur's (1994) nonrecursive procedure. It removed an additional 1.1% of the correct RTs.

For the detection task, we examined the coders' abilities to discriminate deceptive from truthful responses using signal detection analysis (Macmillan and Creelman 2005) with d' (standardized hit score minus standardized false alarm divided by the square root of 2) and C (the negated sum of the standardized hit and false alarm scores divided by the square root of 2) as criterion variables. These formulas are based on two alternative forced choice designs (Macmillan and Creelman 2005, 271). We use MacMillan and Creelman's (Macmillan and Creelman 2005, 21) correction by changing 0 or 1 values to 0.05 and 0.95, respectively. All confidence intervals reported are 95% confidence intervals. In all analyses reported below, significant effects contain a two-tailed p value <0.05 and partial eta-square (η_p^2) as our measure of effect size.

3.2.2 | Cognitive Load Manipulation

Because participants each performed two blocks of questions under two levels of cognitive load, they were able to receive a total of 8 points for recall performance under low load (4 memorized dots per block) and 12 points for recall performance under high load (6 memorized dots per block). Each participant's proportion of accurate dot recall was determined by dividing their total number of correctly identified dot locations by the total possible score within each load condition. A paired samples t-test determined participants were significantly more accurate under low load than under high load ($\mu=0.95$ for low load; $\mu=0.81$ for high load; $t\,(74)=5.726$, $p<0.001$, CI [0.091, 0.189]), confirming that the high load condition was indeed significantly more difficult than the low load condition.

3.2.3 | ANOVA Results

RTs were analyzed using the general linear model with Load and Response Type as within-subject factors. Table 1 presents the mean trimmed RTs from all participants in each condition. Our first hypothesis was supported, such that RTs were longer for lie than for truth responses, as confirmed by a main effect of Response Type ($F(1,74)=42.52$, $p<0.001$, $\eta_p^2=0.37$). The main effect of Load was not significant ($F(1,74)=1.152$, $p=0.287$, $\eta_p^2=0.02$). However, the Response Type × Load interaction was also significant ($F(1,74)=6.147$, $p=0.015$, $\eta_p^2=0.08$). Simple effects tests demonstrated that lie RTs were longer than truth RTs under high load (353 ms) than under low load (181 ms), which is consistent with the CLA. Table 1 shows that truth RTs were non-significantly slower under low load than under high load, contributing to the larger RT difference in Response Type under high load.

TABLE 1 | Reaction time means and standard deviations (in ms) for Experiment 1.

	High load	Low load	Difference
Lie	1294 (591)	1266 (558)	28
Truth	941 (398)	1085 (725)	(−144)
Difference	353***	181**	

Note: ** indicates the difference is significant at the $p<0.01$ level. *** indicates the difference is significant at the $p<0.001$ level.

TABLE 2 | Correlations between coders when detecting responses under high load for Experiment 1.

	Coder 1	Coder 2	Coder 3
Coder 1	—		
Coder 2	0.255*	—	
Coder 3	0.288*	0.165	—

Note: * indicates the difference is significant at the $p<0.05$ level.

3.2.4 | Reaction Time Measurement Error?

One potential concern with our procedure is that having the experimenter trigger the voice key may add random error (i.e., noise) into the RT data, which could diminish the study's power to detect differences across conditions. Such noise would reduce the consistency (i.e., reliability) of participants' RTs for each condition. However, our obtained RT reliabilities were very high (0.76 for truth responses; 0.75 for lie responses). In fact, this value surpassed reliabilities from past studies using computer-triggered methods in word pronunciation or lexical decision studies (0.55–0.72, see Hutchison et al. 2008; Keuleers et al. 2012). Given this, we believe our procedure produced very limited noise in our dependent measure, allowing us to effectively detect the main effect of Response Type and the significant Response Type × Load interaction.

3.2.5 | Multiple Regression

To test our second and third hypotheses, first we correlated d' between our coders separately for participant responses given under high and low load, as shown in Tables 2 and 3, respectively. These analyses reveal mostly significant positive correlations between our coders when detecting participants' responses under high load, but not when detecting participants' responses under low load. Second, we regressed the average d' across the three coders on the trimmed RT values in separate regression analyses within their respective load conditions. For the high load condition, the regression model was significant [$r^2=0.094$, $F(2,74)=3.722$, $p=0.029$]. RTs for lie responses were a significant positive predictor of d' [$\beta=0.409$, $t\,(74)=2.269$, $p=0.026$]. As higher d' is associated with a higher hit rate, this means that longer lie RTs were associated with coders' abilities to correctly detect lie responses, supporting our second hypothesis. Also, supporting our third hypothesis, RTs for truthful responses were a significant negative predictor of d' [$\beta=-0.490$, $t\,(74)=2.719$, $p=0.008$], meaning longer truthful RTs were associated with more false alarms, impairing coders' truth detection accuracy. Partial regression plots depicting the relationships between d' and lie and truth RTs under high load are shown in Figures 1 and 2, respectively. In contrast to the high load condition, the regression model for the low load condition was not significant [$r^2 = 0.015$, $F(2,74) = 0.559$, $p = 0.574$]. Although the relationship between RTs and d' was in the same predicted direction as that found under high load, neither lie RTs nor truth RTs under low load significantly predicted d' [$\beta = 0.115$, $t(74) = 0.725$, $p = 0.471$ for lie RTs; $\beta = -0.167$, $t(74) = 1.057$, $p = 0.294$ for truth RTs].

TABLE 3 | Correlations between coders when detecting responses under low load for Experiment 1.

	Coder 1	Coder 2	Coder 3
Coder 1	—		
Coder 2	0.159	—	
Coder 3	0.149	0.157	—

We next ran the same set of regression analyses by regressing C on the same predictor variables. Neither regression model was significant [$r^2 = 0.059$, $F(2,74) = 2.251$, $p = 0.113$ for high load; $r^2 = 0.056$, $F(2, 74) = 2.118$, $p = 0.128$ for low load] suggesting that RTs were not associated with our coders' criterion used for detection under either load condition.

3.3 | Discussion

We found lie RTs to be significantly longer than truthful RTs and provide partial support for the CLA by showing a larger difference between lie and truth RTs under high load than under low load. Because truth RTs were numerically greater under low load ($M = 1085$, SD $= 725$) than high load ($M = 941$, SD $= 396$; $t(74) = 1.985$, $p = 0.051$), and lie RTs were unaffected

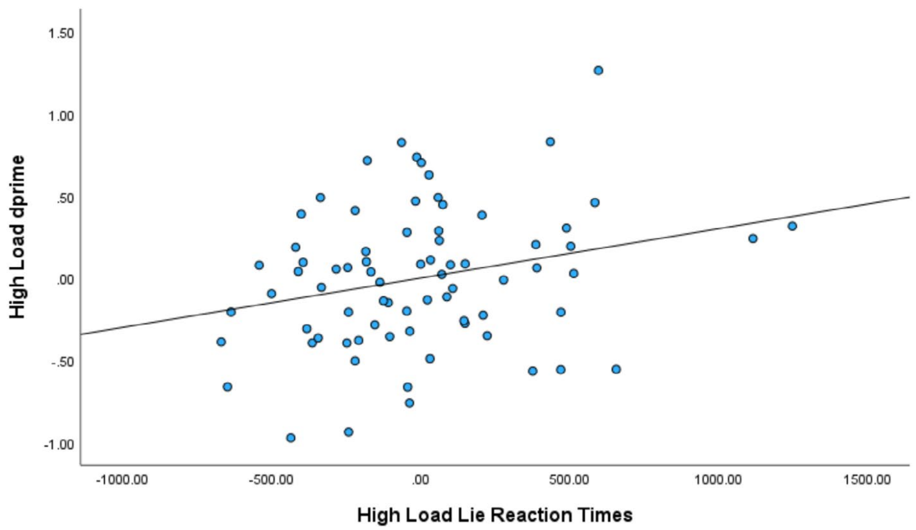

FIGURE 1 | Partial regression plot of participant detectability and lie reaction times under high load in Experiment 1.

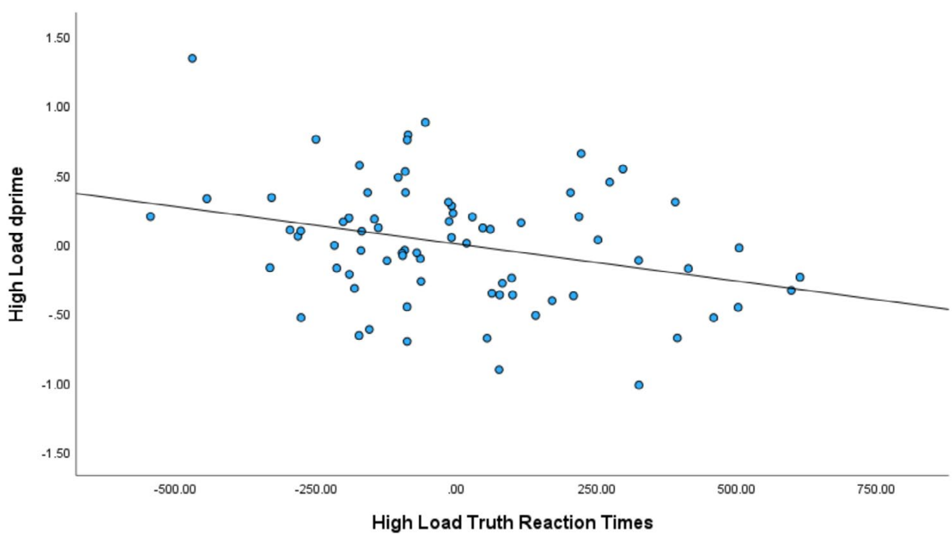

FIGURE 2 | Partial regression plot of participant detectability and truth reaction times under high load in Experiment 1.

by increased cognitive load, our results do not fully support the CLA. Our results also showed that participants' longer lie RTs under high load were associated with increased lie detection accuracy and longer truth RTs under high load were associated with decreased truth detection accuracy from our coders. These findings are consistent with previous research (Harrison et al. 1978) and support findings from Maldonado et al. (2018, Experiment 1a) in which most participants reported long RTs as a perceived deception cue used in detection. Because the results of our cognitive load manipulation are exploratory; however, they require replication to confirm that they are reliable.

Also, it is possible that allowing the experimenter in the deception task to trigger the voice key may have introduced error into the RT data. We used this procedure to adjust for differences in reading speed, so that all participants could respond at the same time once each question was asked. Our reliability analysis suggested very little noise was introduced in the RT data using this procedure. However, we ran a second study to replicate the results of our cognitive load manipulation and to address the potential of experimenter error using a more controlled questioning procedure for our deception task.

4 | Experiment 2

Our second experiment contained a larger more powerful sample of participants and a more controlled experimental procedure. We modified the questionnaire used in Maldonado et al. (2018) and Experiment 1 to prevent participants from being able to prepare their deceptive responses during question presentation. We used the same cognitive load manipulation to replicate the results from Experiment 1. In addition, we used a more specific participant packet for our coders' veracity judgments that had them indicate transparent lie responses and responses that may not fit under either the "lie" or "truth" categories (e.g., "did not answer" or unintelligible answers) along with truth and lie judgments. Finally, we had participants in the deception task report their ages in this study.

4.1 | Method

4.1.1 | Participants

One hundred ninety-two students from Montana State University participated via an online SONA systems recruitment pool in which participants signed up for an hour-long session to receive partial course credit for an Introductory to Psychology course. Participants in our sample ranged from 17 to 33 years old ($M = 19.42$) and consisted of 52% female and 43% male participants, with 4.8% and 2.4% missing data for gender and age, respectively. Only one participant was allowed to sign up per session. Our goal was to run a total of at least 150 participants to double the sample size from Experiment 1. At the end of the semester, we had 192 participants. However, our final sample included 155 participants, after eliminating data from 25 participants due to experimenter error (i.e., failing to turn on the microphone that measured participant's RTs), another nine participants who had < 65% usable RT data, and three more participants in which we did not have detectability data from all our coders. This remains more than double the sample size from Experiment 1. The experiment was approved by the Montana State University Institutional Review Board (Protocol # 2023-915-EXPEDITED) and all participants signed a written informed consent form immediately upon arrival.

4.1.2 | Design and Procedure

The same within subjects design was used as in Experiment 1. To allow for a more controlled test of RTs preceding the two response types, we modified the procedure from Experiment 1 such that questions were posed by the computer instead of the experimenter, with only the first word or couple of words being presented first, and the remaining words masked by "xxxx's." Participants were instructed to press the "SPACE" key to reveal each additional portion of the question. We rephrased questions such that many of them began with the same few words (e.g., "What is your favorite xxxx?") so participants would not know what was being asked until the final keypress when the full question was revealed, eliminating the possibility of them preparing a response in advance. After four successive keypresses, the full question was revealed and the microphone was automatically triggered for participants to respond, eliminating the possibility of experimenter error affecting RTs.

4.1.2.1 | Deception Task. As in Experiment 1, participants completed four blocks of questions, with eight questions per block, with the same dot matrix cognitive load manipulation differing between low and high load in each block of questions. The 4 × 4 matrix containing the load condition again preceded blocks of questions for 4 s, but the question prompt reading either "Lie" or "Truth" on screen was only presented for 2 s before each question screen. Once participants fully navigated through question screens, such that the full question was presented and the microphone was triggered, they had 6 s to respond before the 5000 ms intertrial interval preceding the next trial. After each block of questions, a blank 4 × 4 matrix was presented on screen, at which time participants filled in the matrix manually from memory in a packet provided to them by the experimenter. Illustrations of this deception task procedure are shown in Figures 3 and 4. Again, participants' faces and upper bodies were audio–video recorded throughout this task.

4.1.2.2 | Detection Task. Three research assistants (see endnote 3), who did not assist with the deception task and were blind to the hypotheses, served as coders judging the veracity of participants' recorded responses. Because the experimenter did not read the questions, like in Experiment 1, recordings were edited with question screens added so coders knew what question the participant was answering. For each participant who had usable RT data, a 5-s screen was edited into videos following the fourth "SPACE" bar keypress in each trial, displaying the full question that was being answered. Coders were given packets that showed them what questions were being answered, with the four response options of "Lie," "Truth," "Obvious Lie," or "Did Not Answer." The "Obvious Lie" option was added in these packets to assess how often participants gave agreeably unconvincing lie responses (e.g.,

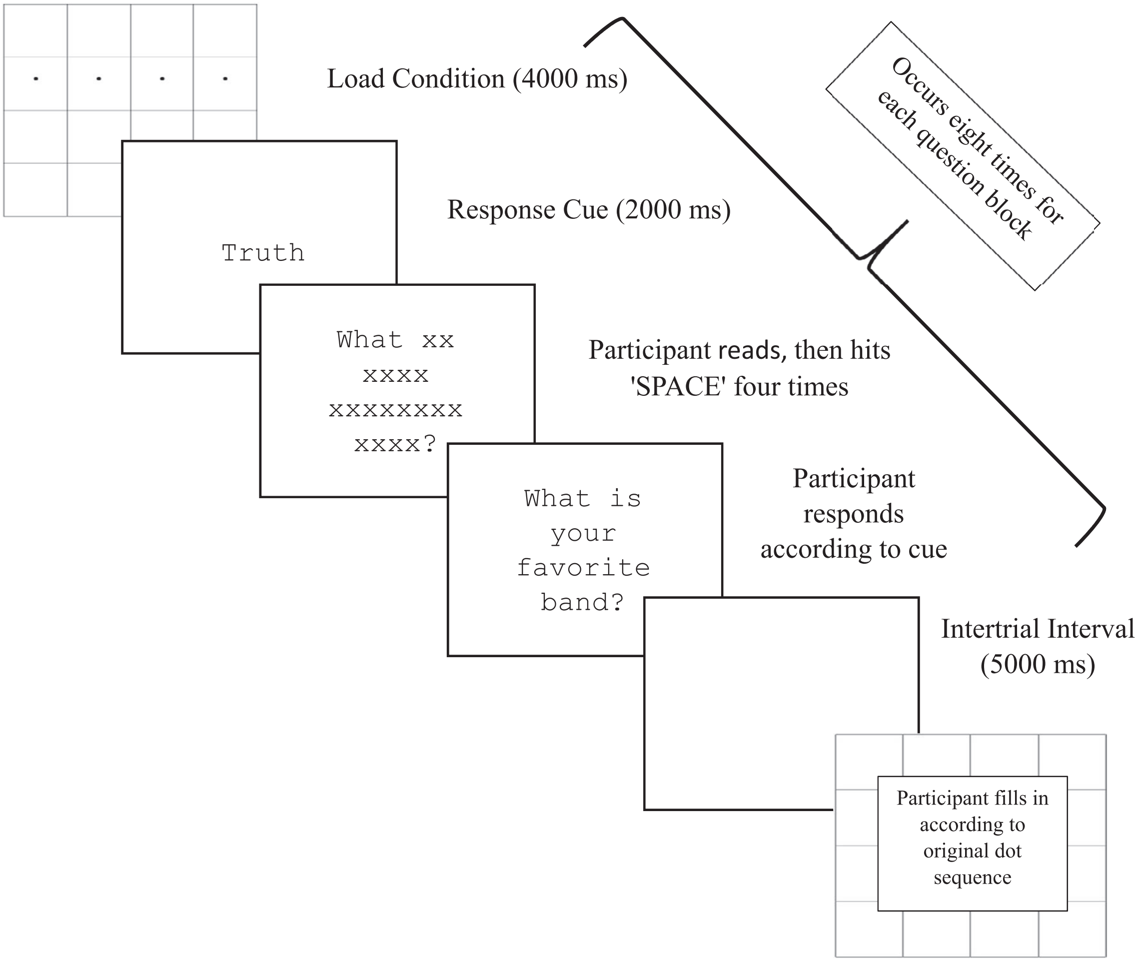

FIGURE 3 | Low load truth trial procedure example of the deception task in Experiment 2.

Saying, "1912" in response to the question, "What is your birth year?"). The "Did Not Answer" option was added so we could keep track of non-answers (i.e., answers that fit into neither the "Truth" nor "Lie" categories) that would not be included in the calculations of our criterion variables used in signal detection analysis. Coders were able to view participants' responses as much as needed to judge their responses and performed their coder duties at their own pace.

4.2 | Results

4.2.1 | Data Scoring

For the deception task, we used the same procedure as in Experiment 1 to score correct responses, with the only difference being that trials in which responses were not made within 6 s following the final keypress by participants were also coded as incorrect. Only correct responses were included in all RT analyses. We again removed trials with microphone errors (0.1%) or when the participant did not comply with instructions (17.9%). The same nonrecursive procedure (Van Selst and Jolicoeur's 1994) was used as in Experiment 1 to remove outliers, which removed an additional 2.8% of the RT data.

For the detection task, we used signal detection analysis (Macmillan and Creelman 2005) with d' and C as the criterion variables to determine the coder's detection abilities. A small percentage of responses were judged by our coders to be Obvious Lies (3.6%) and Did Not Answer (2.6%) responses. However, convergence between coders when making these judgments was rare, with only 0.2% of responses coded unanimously by judges as "Obvious Lie" and only 0.4% of responses coded unanimously by judges as "Did Not Answer." It is unclear why there was low agreement between coders when judging responses as "Did Not Answer." We were not able to perform any follow-up assessments of why there was low agreement between coders in the detection task when judging answers as "Did Not Answer." Nevertheless, because agreement on these two judgments across coders was so low, we used the same data scoring procedure for correct, incorrect, and unanswered responses in our detector task for coders as we did for our research assistants scoring the deception task, such that "Obvious Lies" were scored as "Lies" and "Did Not Answer" was scored according to whether the research assistants during the original deception task indicated that participants did not provide an answer to the question. If questions were marked as answered by research assistants, but not by coders, we omitted the coders' responses from their calculated scores for d' and C. All confidence intervals reported are 95% confidence intervals. All significant effects reported in the

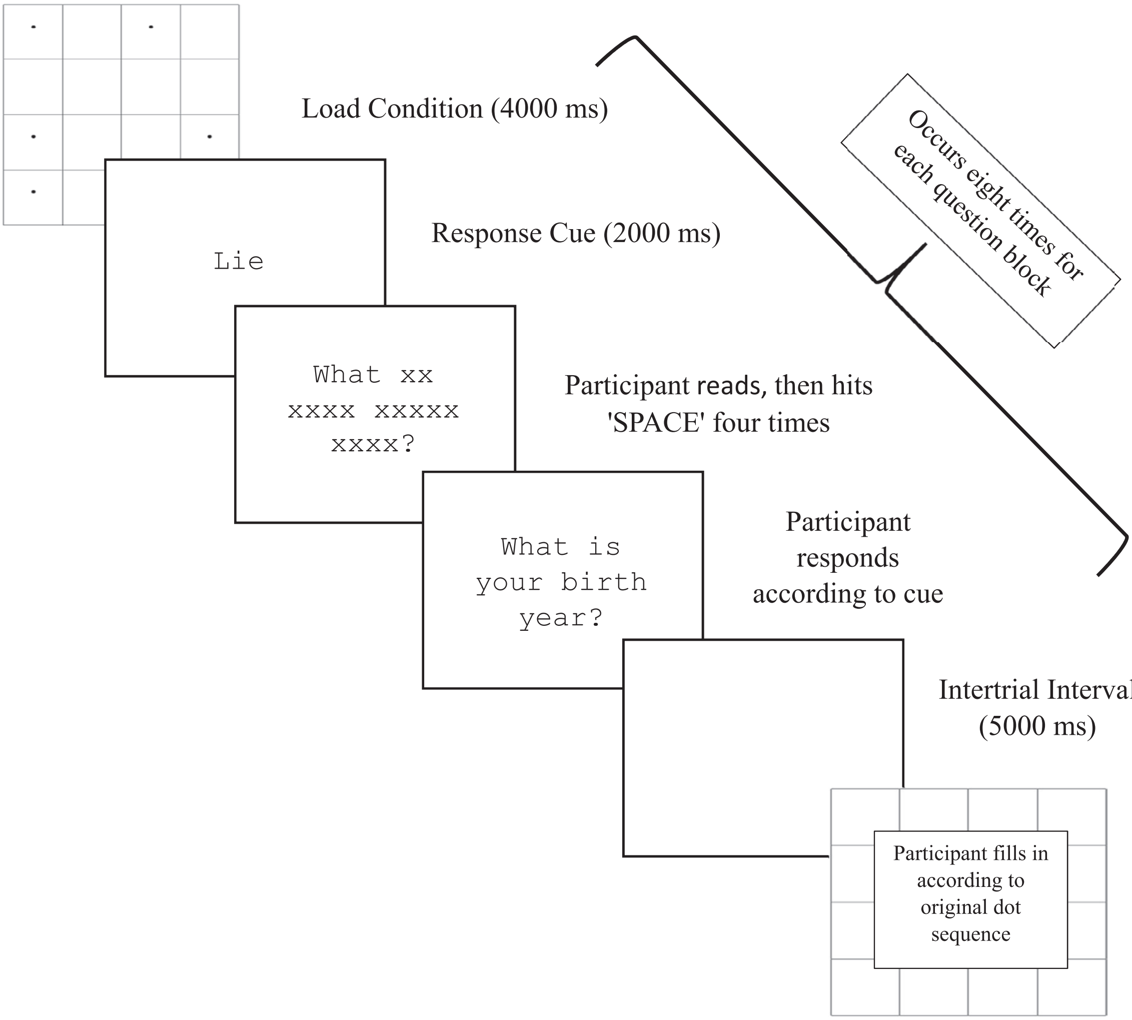

FIGURE 4 | High load lie trial procedure example of the deception task in Experiment 2.

analyses below contain a two-tailed p value <0.05 and partial eta-square (η_p^2) as our measure of effect size.

4.2.2 | Cognitive Load Manipulation

The dot matrix task was scored the same as in Experiment 1. A paired samples t-test determined that accuracy was again significantly higher under low load than high load ($\mu=0.95$ for low load; $\mu=0.78$ for high load; t (154) = 9.58, $p<0.001$, CI [0.131, 0.198]). Again, this shows that this task was significantly more difficult under high load than low load.

4.2.3 | ANOVA Results

RTs were analyzed using the general linear model with Load and Response Type as within-subject factors. We replicated the main effect of Response Type found in Experiment 1, such that RTs were longer for lie than for truthful responses ($F(1,154)=210.98$, $p<0.001$, $\eta_p^2=0.58$). However, neither the main effect of Load nor the Response Type × Load interaction were significant (all p's >0.342, all $\eta_p^2<0.007$). This relationship is depicted in Table 4.

4.2.4 | Multiple Regression

To test our second and third hypotheses, first we correlated d' between our coders separately for participant responses given under high and low load, as shown in Tables 5 and 6, respectively. These reveal significant positive correlations between our coders for both load conditions. Second, we regressed d' on the trimmed RT values in separate analyses within their respective load conditions. Significant regression models were found for both load conditions [$r^2=0.052$, $F(2,154)=4.15$, $p=0.018$, for high load; $r^2=0.076$, $F(2,154)=6.27$, $p=0.002$, for low load]. Supporting our second hypothesis, RTs for lie responses acted as a significant positive predictor of d' under both load conditions [$\beta=0.257$, $t\,(152)=2.65$,

TABLE 4 | Reaction time means and standard deviations (in ms) for Experiment 2.

	High load	Low load	Difference
Lie	2104 (542)	2089 (565)	15
Truth	1668 (471)	1634 (473)	34
Difference	436***	455***	

Note: *** indicates the difference is significant at the $p<0.001$ level.

$p = 0.009$ for high load; $\beta = 0.329$, $t\ (152) = 3.52$, $p < 0.001$ for low load], meaning longer lie RTs were associated with improved lie detection accuracy. Partial regression plots for lie RTs under high and low load are depicted in Figures 5 and 6, respectively. Finally, consistent with our third hypothesis, RTs for truthful responses acted as a significant negative predictor of d' under both load conditions [$\beta = -0.240$, $t\ (152) = 2.47$, $p = 0.015$, for high load; $\beta = -0.209$, $t\ (152) = 2.24$, $p = 0.027$ for low load], meaning longer truth RTs were associated with more false alarms. Partial regression plots for truth RTs under high and low load are depicted in Figures 7 and 8, respectively.

We next ran the same analyses by regressing C on the same predictor variables within their respective load conditions. The only significant model was when regressing C on RTs under low load [$r^2 = 0.058$, $F(2,154) = 4.66$, $p = 0.011$; the model under high load was non-significant, $p = 0.08$]. However, neither lie RTs [$\beta = -0.166$, $t\ (152) = 1.76$, $p = 0.081$] nor truth RTs [$\beta = -0.106$, $t\ (152) = 1.21$, $p = 0.264$] significantly predicted C. Again, this suggests that RTs are not associated with our coders' criterion for judging participants' responses under either load condition.

TABLE 5 | Correlations between coders when detecting responses under high load for Experiment 2.

	Coder 1	Coder 2	Coder 3
Coder 1	—		
Coder 2	0.232**	—	
Coder 3	0.213**	0.229**	—

Note: ** indicates the difference is significant at the $p < 0.01$ level.

TABLE 6 | Correlations between coders when detecting responses under low load for Experiment 2.

	Coder 1	Coder 2	Coder 3
Coder 1	—		
Coder 2	0.329***	—	
Coder 3	0.250**	0.221**	—

Note: ** indicates the difference is significant at the $p < 0.01$ level. *** indicates the difference is significant at the $p < 0.001$ level.

4.3 | Discussion

Using a higher-powered sample of participants and a different methodology, we again showed lie RTs were significantly longer than truth RTs but failed to show a similar interaction between response type and load. Furthermore, long lie and truth RTs were associated with participants' detectability regardless of the load condition. Our measure of lie detection accuracy (d') was significantly positively correlated between our coders in both load conditions. This differed from Experiment 1 in which d' was only correlated between our coders for responses given under high load. Better reliability among our coders may have been associated with approximately twice the number of detection judgments performed in the second experiment (see Levine et al.'s 2022 position on the number of deception judges and senders).

Concerning our load manipulation, dot recall performance was again significantly reduced under high load compared with low load, but this did not affect RTs when lying or participant detectability. Participants' dot memory accuracy did not differ for the high load condition between Experiment 1 and Experiment 2 ($t\ (232) = 0.871$, $p = 0.385$), suggesting our load manipulation was equally difficult across samples.

4.4 | General Discussion

Our findings suggest that RTs may be both an actual and a perceived cue to deception. Lie RTs were longer than truth RTs in both experiments, suggesting lying is more cognitively demanding than telling the truth (Walczyk et al. 2003), and responses with longer RTs were more often classified as lies, which was

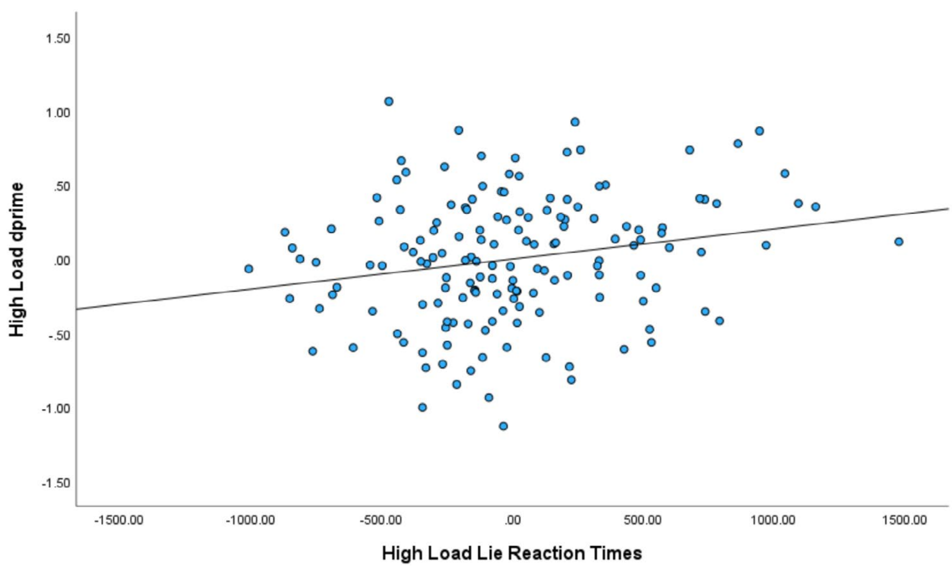

FIGURE 5 | Partial regression plot of participant detectability and lie reaction times under high load in Experiment 2.

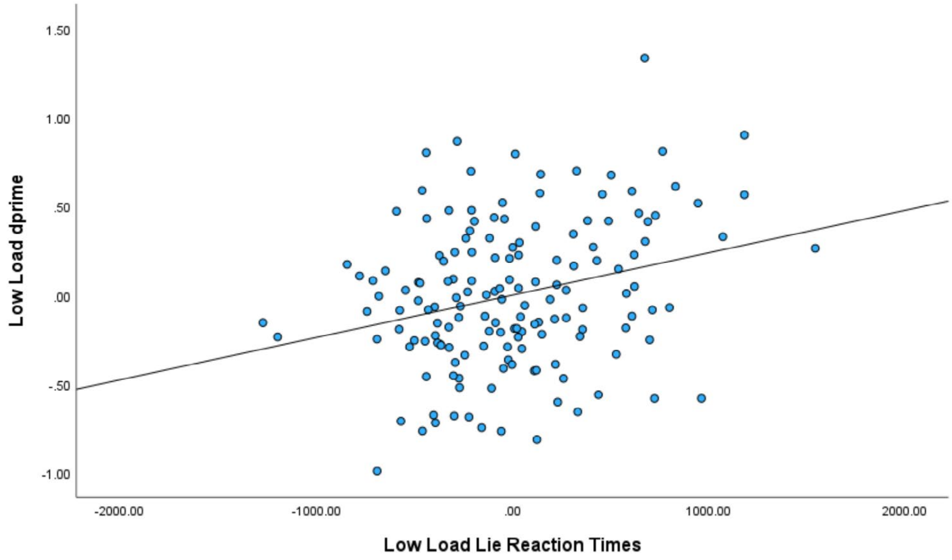

FIGURE 6 | Partial regression plot of participant detectability and lie reaction times under low load in Experiment 2.

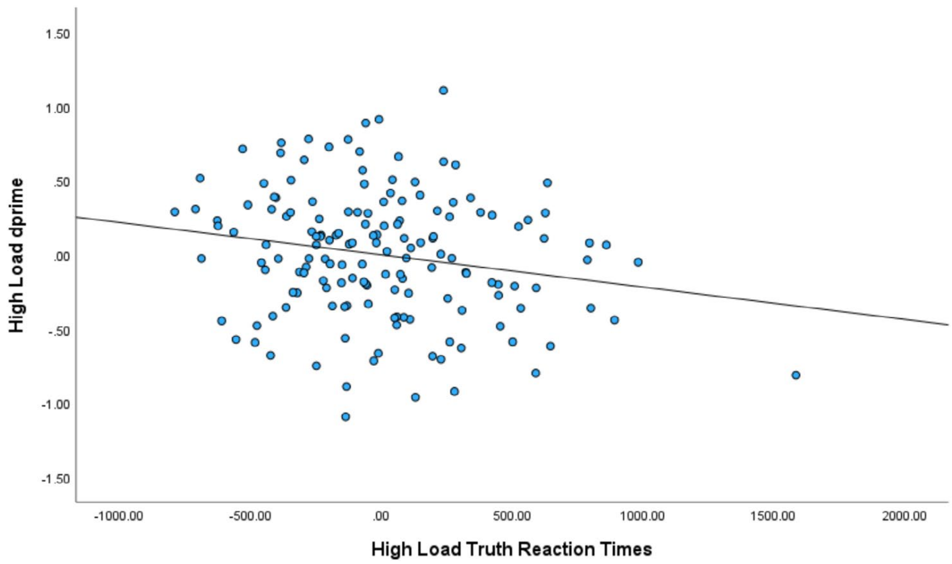

FIGURE 7 | Partial regression plot of participant detectability and truth reaction times under high load in Experiment 2.

associated with differences in detectability. We speculate whether question content may have affected RTs and discuss the results from our load manipulation and the CLA more broadly.

4.4.1 | Question Content

Our results suggest that RTs influence participants' detectability, as our coders were more likely to judge any response as a lie if it was associated with a long RT. This raises the question of what may be contributing to long RTs, to better understand when truthful responses may be judged as deceptive. The updated working memory model of deception (Sporer 2016) discusses how the concreteness of the question information affects retrieval from long-term memory. Borrowing principles from fuzzy trace theory (Reyna and Brainerd 1995), Sporer mentions that memories for autobiographical information that do not contain highly salient details are stored at a gist level of representation. This means that specific details are less well remembered and decay with the passage of time, but the general (gist) level of details is remembered in abstraction. The more often gist memory details are experienced and encoded into long-term memory, the easier their retrieval becomes when necessary.

We speculate that less well encoded information may have contributed to longer RTs when telling the truth in our experiments. This speculation is consistent with Walczyk et al.'s (2014) proposal that retrieval of truthful information in lie generation may require controlled attention when the information is infrequently or not recently accessed, implicating greater cognitive effort relative to retrieval of highly rehearsed truthful information. Our question list contained items that were about the past (e.g., "As a child, what was your favorite toy?"), questions pertaining to someone other than the participant (e.g., "What is the last name of your father's mother?"), or questions about

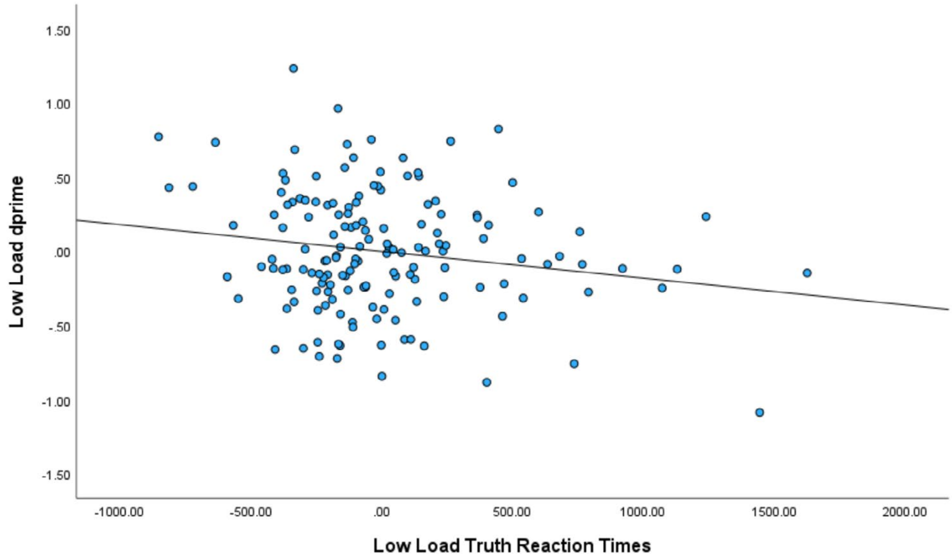

FIGURE 8 | Partial regression plot of participant detectability and truth reaction times under low load in Experiment 2.

hypotheticals or non-experienced events (e.g., "What would it be, if you could only bring one item onto a deserted island?"). Sporer (2016) mentions this latter example as demanding a response based on mental schemas and action scripts of events, being only an estimation of the event memory details given one's actual experience. It would be beneficial to test this speculation by controlling for the level of encoding of truthful information to verify its effects on RTs and detection accuracy.

4.4.2 | Cognitive Load

Dot recall performance in our cognitive load manipulation was significantly reduced in the high load condition relative to the low load condition, but participants' RTs when lying were not affected by increased cognitive load. These findings do not support the CLA but also do not directly replicate Verschuere et al.'s (2018) findings either. Counter to what would be predicted by the CLA, Verschuere et al. showed that increased cognitive load was associated with a larger RT deception effect under low load than under high load. We demonstrated an RT effect consistent with the CLA in Experiment 1, but did not replicate it in Experiment 2. Our results are more consistent with those of Verschuere et al. (2018), in that increasing cognitive load did not reliably increase the time taken to lie. Perhaps increasing cognitive load affects other deception cues (e.g., verbal cues; Evans et al. 2013) more than it does RTs.

It is also possible that increasing cognitive load must occur continuously with responding to affect RTs. The two studies from Verschuere et al. (2018) that showed an RT deception effect that aligned with the CLA used load manipulations that continuously increased the complexity of deceptive and honest responding. Williams et al. (2013, Experiment 3) increased the number of possible response alternatives when lying, and Visu-Petra et al. (2013) used continuous memory and set-shifting tasks integrated into question presentation. Furthermore, the effectiveness of common cognitive load manipulations (e.g., telling story in reverse order; Vrij et al. 2017) seems to rely on their ability to continuously interfere with responding. Supporting this notion, Debey et al. (2012) found that a goal neglect manipulation that occurred continuously with participants' deceptive responding produced an RT deception effect aligned with the CLA, but two ego depletion manipulations performed before responding did not produce a comparable RT deception effect.

It is unclear whether our load manipulation interfered with participants' responses. It is possible that participants answered questions without continuously updating their memory of the dot sequence and then were reminded of the sequence when the blank matrix was presented. This lack of continuous response interference may explain our results. Supporting this contention, Rowthorn (2016) used a similar cognitive load manipulation in which participants retained a digit sequence in memory while giving truthful and dishonest responses. At the end of question blocks, another digit sequence was presented for them to judge as the same or different from the original by one digit. They also failed to produce an RT deception effect aligned with the CLA and reasoned that participants may not have been consciously holding the sequence of digits in memory while responding. Thus, our results and those of Rowthorn (2016) suggest that increasing cognitive load probably needs to directly interfere with responding to effectively increase lie RTs.

4.4.3 | Limitations and Future Research Directions

Our small sample of coders for the detection task limits our ability to firmly conclude anything regarding our findings. While our methodology used only a few coders, each coder judged the veracity of a very large number of observations (~2400 for Experiment 1 and ~5000 for Experiment 2). It is often the case that deception detection experiments will use a much larger sample of coders who judge a much smaller number of observations (Bond and DePaulo 2006). Levine et al. (2022) argue that the number of senders of dishonest messages and the number of observations from each sender are more important for stabilization of accuracy scores than is the sample size of coders. The authors contend that there is more variability among senders than among coders (see Bond and DePaulo 2008) and that detection accuracy is more variable when derived from large samples of coders judging a small number of observations. However, Levine et al. (2022) also contend that a small number of coders is problematic as well. It is necessary to replicate our findings from these two experiments using many and larger samples of coders to verify their reliability when an adequate number of observations is assessed.

One other potential limitation is the generalizability of RT measures to lie detection in real world settings. For instance, when detecting real world lies, it may not be clear how to mark the end of a question to determine how long it takes someone to respond. A couple of measures can be taken to account for this potential limitation to better ensure that RTs represent valid and reliable measures of deceptive responding. First, baseline RTs should be assessed, so that they can be compared against RTs associated with answers to critical questions used to discriminate liars from truth-tellers. Second, it is important to gather many RT observations of questions that may be answered both truthfully and dishonestly by liars, which is often accomplished by repeating sets of questions multiple times (Vrij 2008). Incorporating such control measures could potentially improve the accuracy and validity of any observed truthful and dishonest RT differences in applied settings.

As is often the case in experimental deception research, our results are limited by the low ecological validity of our methodology. Lies told in these experiments are not tantamount to lies told in real-world situations; however, the speculation that question content affects RTs for truthful and dishonest responses is applicable outside our experimental context. Future research could use a similar questioning format, but with questions analyzed separately by the concreteness of memory details and how recently and/or frequently truthful information used to answer the questions was accessed. Additionally, real-world situations in which lying is commonplace, like criminal interrogations, are expected to differentially affect the mindsets of truthful and dishonest individuals (Granhag and Hartwig 2008), with liars often preparing their messages ahead of time (Inbau et al. 1999). Future research using paradigms like ours could test whether preparation moderates the effect of RTs on detectability.

Also, our participants were not incentivized or motivated in any way when carrying out the deception task procedures. Real world lies are often told by individuals motivated to avoid punishment, embarrassment, and/or reputation damage, which may affect the speed with which their lies are generated. Future experiments could motivate participants when lying to assess its effect on RTs and detectability by using experimental scenarios appearing more realistic to potential liars. For example, Dulaney (1982) found that participants produced faster lie RTs than truth RTs when threatened with "punishment" by their university after being questioned about cheating in an experiment. Future experiments could also use identity-relevant motivation techniques (DePaulo et al. 2003) to motivate participants, like convincing them that the ability to lie is related to admirable character traits like intelligence (Cutrow et al. 1972; Streeter et al. 1977) or future career success (Ekman and Friesen 1974; Hocking and Leathers 1980).

Finally, although our manipulation check showed that the high load condition was indeed more difficult than the low load condition, it is unclear whether this load manipulation interfered with participant responding in the deception task. As stated earlier, although the high load dot condition is more difficult, we do not know whether participants were continually exerting more effort trying to remember the dots in the high load condition during the deception task. Future studies could verify this using a post-experiment manipulation check that asks participants to self-report how difficult the deception and dot matrix tasks were under both high and low load and that asks them to report any strategies they used in remembering the dot sequence during responding. This may reveal whether our load manipulation directly interfered with responding in the deception task. Additionally, to verify whether our load manipulation effectively increases cognitive load for respondents during the deception task itself, it should be compared with other load manipulations that continuously interfered with responding and affected RTs (e.g., Debey et al. 2012; Visu-Petra et al. 2013).

5 | Conclusion

Our findings suggest that RTs may be both an actual and a perceived cue to deception. Lying took reliably longer than telling the truth, and longer RTs were more often judged as lies regardless of their veracity. These findings require replication with more and larger samples of coders to verify their reliability. Our exploratory results from our load manipulation suggest that increasing cognitive load may not increase the time taken to lie and may not aid in the discrimination of lie and truth responses. However, it remains an empirical question whether our load manipulation was effective at increasing cognitive load while participants were responding. Many future research directions are proposed based on our findings.

Author Contributions

Evan Brennan: conceptualization, investigation, writing – original draft, methodology, formal analysis, data curation, writing – review and editing, project administration. **Keith A. Hutchison:** conceptualization, investigation, writing – original draft, writing – review and editing, methodology, software, formal analysis, project administration, data curation, supervision, resources.

Acknowledgements

We would like to thank Chris Callahan, Tom Bart, Nicole Bobo, Cora Tollerson, Sage Heroux, and Hailey Kopischke for their work as coders of participants' responses. Also, we would like to thank Cara Stettler, Max Schneider, Taylor Merchant, Katy McCauley, Johanna Rydell, Ashley Pluhar, Bradley Buckland, Davia Whitt, Jamie Rush, Josiah Shirley, and Penelope Allen for their roles in gathering data as research assistants.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data for the present study are available at OSF via https://osf.io/d4gfw/?view_only=5918733e00cd4363854e4b97548ac06b.

Endnotes

[1] Participants also completed a cognitive test battery composed of three tasks (Stroop, antisaccade, and Reading Span) used to measure individual differences in attentional control. These measures are not discussed here because they were part of separate hypotheses testing whether attentional control moderated lie RTs and errors committed. We found high relative to low attentional control was associated with significantly fewer lie response errors in Experiment 2, but no relationship between attentional control and lie RTs. Overall, we felt this detracted from the main point of this manuscript.

[2] Although our sample of coders is small (for both experiments), it is based on comparable research (Maldonado et al. 2018, Experiment 2). Levine et al. (2022) contends that, "Extremely low numbers of either judges or senders need to be avoided" (p. 200), but also argue primarily that the number of total judgments [(the number of judges) × (the number of judgments)] is most critical to stable detection accuracy across judges and recommends at least 500 total judgments. Because we had 75 participants in Experiment 1 providing 32 responses each and 155 participants in Experiment 2 providing 32 responses each, we had ~7200 and 14,880 total judgments across Experiments 1 and 2, respectively, well above the number recommended by Levine et al. (2022).

[3] We performed the analyses using standard formulas for d' ($Z_{\text{Hits}} - Z_{\text{FalseAlarms}}$) and C [$1 - (Z_{\text{Hits}} - Z_{\text{FalseAlarms}})/2$] (Macmillan and Creelman 2005) as well and the results were not significantly different from those that we reported.

References

Baskett, G. D., and R. O. Freedle. 1974. "Aspects of Language Pragmatics and the Social Perception of Lying." *Journal of Psycholinguistic Research* 3, no. 2: 117–131. https://doi.org/10.1007/BF01067571.

Blandón-Gitlin, I., E. Fenn, J. Masip, and A. H. Yoo. 2014. "Cognitive-Load Approaches to Detect Deception: Searching for Cognitive Mechanisms." *Trends in Cognitive Science* 18, no. 9: 441–444. https://doi.org/10.1016/j.tics.2014.05.004.

Bond, C. F., and B. M. DePaulo. 2006. "Accuracy in Deception Judgments." *Personality and Social Psychology Review* 10, no. 3: 214–234. https://doi.org/10.1207/s15327957pspr1003_2.

Bond, C. F., and B. M. DePaulo. 2008. "Individual Differences in Judging Deception: Accuracy and Bias." *Psychological Bulletin* 134, no. 4: 477–492. https://doi.org/10.1037/0033-2909.134.4.477.

Cutrow, R. J., A. Parks, N. Lucas, and K. Thomas. 1972. "The Objective Use of Multiple Physiological Indices in the Detection of Deception." *Psychophysiology* 9, no. 6: 578–588. https://doi.org/10.1111/j.1469-8986.1972.tb00767.x.

Debey, E., B. Verschuere, and G. Crombez. 2012. "Lying and Executive Control: An Experimental Investigation Using Ego Depletion and Goal Neglect." *Acta Psychologica* 140, no. 2: 133–141. https://doi.org/10.1016//j.actpsy.2012.03.004.

DePaulo, B. M., J. J. Lindsay, B. E. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper. 2003. "Cues to Deception." *Psychological Bulletin* 129, no. 1: 74–118. https://doi.org/10.1037/0033-2909.129.1.74.

Dulaney, E. F. 1982. "Changes in Language Behavior as a Function of Veracity." *Human Communication Research* 9, no. 1: 75–82. https://doi.org/10.1111/j.1468-2958.1982.tb00684.x.

Ekman, P., and W. V. Friesen. 1974. "Detecting Deception From the Body or Face." *Journal of Personality and Social Psychology* 29, no. 3: 288–298. https://doi.org/10.1037/h0036006.

Evans, J. R., S. W. Michael, C. A. Meissner, and S. E. Brandon. 2013. "Validating a New Assessment Method for Deception Detection: Introducing a Psychologically Based Credibility Assessment Tool." *Journal of Applied Research in Memory and Cognition* 2, no. 1: 33–41. https://doi.org/10.1016/j.jarmac.2013.02.002.

Gombos, V. A. 2006. "The Cognition of Deception: The Role of Executive Processes in Producing Lies." *Genetic, Social, and General Psychology Monographs* 132, no. 3: 197–214. https://doi.org/10.3200/MONO.132.3.197-214.

Granhag, P. A., and M. Hartwig. 2008. "A New Theoretical Perspective on Deception Detection: On the Psychology of Instrumental Mind-Reading." *Psychology, Crime & Law* 14, no. 3: 189–200. https://doi.org/10.1080/106831607016-45181.

Harrison, A. A., M. Hwalek, D. F. Raney, and J. G. Fritz. 1978. "Cues to Deception in an Interview Situation." *Social Psychology* 41, no. 2: 156–161. https://doi.org/10.2307/3033575.

Hartwig, M., and C. M. Bond. 2011. "Why Do Lie-Catchers Fail? A Lens Model Meta-Analysis of Human Lie Judgments." *Psychological Bulletin* 137, no. 4: 643–659. https://doi.org/10.1037/a0023589.

Heyman, T., B. V. Rensbergen, G. Storms, K. A. Hutchison, and S. De Deyne. 2015. "The Influence of Working Memory Load on Semantic Priming." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 41, no. 3: 911–920. https://doi.org/10.1037/xlm0000050.

Hocking, J. E., and D. G. Leathers. 1980. "Nonverbal Indicators of Deception: A New Theoretical Perspective." *Communication Monographs* 47, no. 2: 119–131. https://doi.org/10.1080/03637758009376025.

Hutchison, K. A., D. A. Balota, M. Cortese, and J. M. Watson. 2008. "Predicting Semantic Priming at the Item-Level." *Quarterly Journal of Experimental Psychology* 61, no. 7: 1036–1066. https://doi.org/10.1080/17470210701438111.

Inbau, F. E., J. E. Reid, J. P. Buckley, and B. C. Jayne. 1999. *Criminal Interrogations and Confessions*. 4th ed. Aspen Publishers.

Keuleers, E., P. Lacey, K. Rastle, and M. Brysbaert. 2012. "The British Lexicon Project: Lexical Decision Data for 28,730 Monosyllabic and Disyllabic English Words." *Behavior Research Methods* 44, no. 1: 287–304. https://doi.org/10.3758/s13428-011-0118-4.

Kraut, R. 1978. "Verbal and Nonverbal Cues in the Perception of Lying." *Journal of Personality and Social Psychology* 36, no. 4: 380–391. https://doi.org/10.1037/0022-3514.36.4.380.

Levine, T. R., Y. Daiku, and J. Masip. 2022. "The Number of Senders and Total Judgments Matter More Than Sample Size in Deception–Detection Experiments." *Association for Psychological Science* 17, no. 1: 191–204. https://doi.org/10.1177/1745691621990369.

Macmillan, N. A., and C. D. Creelman. 2005. *Detection Theory: A User's Guide*. 2nd ed. Lawrence Erlbaum Associates Publishers.

Maldonado, T., F. M. Marchak, D. M. Anderson, and K. A. Hutchsion. 2018. "The Role of Working Memory Capacity and Cognitive Load in Producing Lies for Autobiographical Information." *Journal of Applied Research in Memory and Cognition* 7, no. 4: 574–586. https://doi.org/10.1016/j.jarmac.2018.05.007.

Mann, S., and A. Vrij. 2006. "Police Officers' Judgments of Veracity, Tenseness, Cognitive Load and Attempted Behavioural Control in Real-Life Police Interviews." *Psychology, Crime & Law* 12, no. 3: 307–319. https://doi.org/10.1080/10683160600558444.

Mann, S., A. Vrij, and R. Bull. 2002. "Suspects, Lies, and Videotape: An Analysis of Authentic High-Stakes Lies." *Law and Human Behavior* 26, no. 3: 365–376. https://doi.org/10.1023/A:1015332606792.

Reyna, V. F., and C. J. Brainerd. 1995. "Fuzzy-Trace Theory: An Interim Synthesis." *Learning and Individual Differences* 7, no. 1: 1–75. https://doi.org/10.1016/1041-6080(95)90031.

Rowthorn, H. 2016. "Detecting Lies From Autobiographical Memories: A Cognitive Load and Deadline Approach to Differentiating Fakers From Truth-Tellers in the Autobiographical Implicit Association Test." Master Thesis. https://doi.org/10.13140/RG.2.1.1001.6405.

Seymour, T. L., C. M. Seifert, M. G. Shafto, and A. L. Mosmann. 2000. "Using Response Time Measures to Assess "Guilty Knowledge"."

Journal of Applied Psychology 85, no. 1: 30–37. https://doi.org/10.1037/0024-9010.85.1.30.

Sporer, S. L. 2016. "Deception and Cognitive Load: Expanding Our Horizon With a Working Memory Model." *Frontiers in Psychology* 7, no. 420: 1–12. https://doi.org/10.3389/fpsyg.2016.00420.

Sporer, S. L., and B. Schwandt. 2006. "Paraverbal Indicators of Deception: A Meta-Analytic Synthesis." *Applied Cognitive Psychology* 20, no. 4: 421–446. https://doi.org/10.1002/acp1190.

Streeter, L. A., R. M. Krauss, V. Geller, C. Olson, and W. Apple. 1977. "Pitch Changes During Attempted Deception." *Journal of Personality and Social Psychology* 35, no. 5: 345–350. https://doi.org/10.1037/0022-3514.35.5.345.

Suchotzki, K., B. Verschuere, B. Van Bockstaele, G. Ben-Shakhar, and G. Crombez. 2017. "Lying Takes Time: A Meta-Analysis on Reaction Time Measures of Deception." *Psychological Bulletin* 143, no. 4: 428–453. https://doi.org/10.1037/bul0000087.

Van Selst, M., and P. Jolicoeur. 1994. "A Solution to the Effect of Sample Size on Outlier Elimination." *Quarterly Journal of Experimental Psychology* 47, no. 3: 631–650. https://doi.org/10.1080/14640749408401131.

Verschuere, B., N. C. Köbis, Y. Bereby-Meyer, D. Rand, and S. Shalvi. 2018. "Taxing the Brain to Uncover Lying? Meta-Analyzing the Effect of Imposing Cognitive Load on the Reaction-Time Costs of Lying." *Journal of Applied Research in Memory and Cognition* 7, no. 3: 462–469. https://doi.org/10.1016/j.jarmc.2018.04.005.

Visu-Petra, G., M. Varga, M. Miclea, and L. Visu-Petra. 2013. "When Interference Helps: Increasing Executive Load to Facilitate Deception Detection in the Concealed Information Test." *Frontiers in Psychology* 4, no. 146: 146–156. https://doi.org/10.3389/fpsyg.2013.00146.

Vrij, A. 2008. *Detecting Lies and Deceit: Pitfalls and Opportunities*. 2nd ed. Wiley.

Vrij, A. 2015. "A Cognitive Approach to Lie Detection." In *Detecting Deception: Current Challenges and Cognitive Approaches*, edited by P. A. Granhag, A. Vrij, and B. Verschuere, 1st ed., 205–229. Wiley Blackwell.

Vrij, A., R. Fisher, S. Mann, and S. Leal. 2006. "Detecting Deception by Manipulating Cognitive Load." *Trends in Cognitive Science* 10, no. 4: 141–142. https://doi.org/10.1016/j.tics.2006.02.003.

Vrij, A., R. Fisher, S. Mann, and S. Leal. 2008. "A Cognitive Load Approach to Lie Detection." *Journal of Investigative Psychology and Offender Profiling* 5, no. 1–2: 39–43. https://doi.org/10.1002/jip.82.

Vrij, A., R. P. Fisher, and H. Blank. 2017. "A Cognitive Approach to Lie Detection: A Meta-Analysis." *Legal and Criminological Psychology* 22, no. 1: 1–21. https://doi.org/10.1111/lcrp.12088.

Vrij, A., and S. Mann. 2001. "Telling and Detecting Lies in a High-Stake Situation: The Case of a Convicted Murderer." *Applied Cognitive Psychology* 15, no. 2: 187–203. https://doi.org/10.1002/1099-0720(200103/04)15:2<187LAID-ACP696>3.0CO12-A.

Walczyk, J. J., L. L. Harris, T. K. Duck, and D. Mulay. 2014. "A Social-Cognitive Framework for Understanding Serious Lies: Activation-Decision-Construction-Action Theory." *New Ideas in Psychology* 34: 22–36. https://doi.org/10.1016/j.newideapsych.2014.03.001.

Walczyk, J. J., F. P. Igou, A. P. Dixon, and T. Tcholakian. 2013. "Advancing Lie Detection by Inducing Cognitive Load on Liars: A Review of Relevant Theories and Techniques Guided by Lessons From Polygraph-Based Approaches." *Frontiers in Psychology* 4, no. 14: 14–27. https://doi.org/10.3389/fpsyg.2013.00014.

Walczyk, J. J., K. S. Roper, E. Seeman, and A. M. Humphrey. 2003. "Cognitive Mechanisms Underlying Lying to Questions: Response Time as a Cue to Deception." *Applied Cognitive Psychology* 17, no. 7: 755–774. https://doi.org/10.1002/acp.914.

Williams, E. J., L. A. Bott, J. Patrick, and M. B. Lewis. 2013. "Telling Lies: The Irrepressible Truth?" *PLoS One* 8, no. 4: 1–14. https://doi.org/10.1371/journal.pone.0060713.

Zuckerman, M., B. M. DePaulo, and R. Rosenthal. 1981. "Verbal and Nonverbal Communications of Deception." In *Advances in Experimental Social Psychology*, edited by L. Berkowitz, Vol. 14, 1–59. Academic Press.