Received: 13 April 2020 | Revised: 14 September 2020 | Accepted: 14 September 2020
DOI: 10.1002/acp.3743

RESEARCH ARTICLE

WILEY

Keyboard dynamics discrepancies between baseline and deceptive eyewitness narratives

Frédéric Tomas[1] | Ioannis Tsimperidis[2] | Samuel Demarchi[1] | Farid El Massioui[1]

[1]Human and Artificial Cognitions Laboratory, Department of Psychology, University Paris 8, Saint-Denis, France

[2]Department of Electrical and Computer Engineering, Democritus University of Thrace, Komotini, Greece

Correspondence
Frédéric Tomas, Département de Psychologie, Laboratoire Cognitions Humaine et Artificielle (CHArt), Université de Paris 8, 2 Rue de la Liberté, 93526 Saint-Denis, France.
Email: tomas.phd@gmail.com

Summary

Information manipulation and cognitive load imposition make the production of deceptive narratives difficult. But little is known about the production of deception, and how its mechanisms may help distinguish truthful from deceitful narratives. This study focuses on the measurement of keystroke dynamics while typing truthful and deceptive eyewitness testimonies after a baseline assessment. While typing their narrative, some participants would undergo an auditory cognitive load. Results show that liars typed their story slower, and in less time than the truthful participants when compared to their respective baselines. The imposition of the auditory cognitive load showed adverse results, enhancing the amount of keystrokes and the time necessary to type the narrative. Classification shows better results for deceptive narrative when no auditory cognitive load is imposed. These results are discussed in terms of expanding current models to include the cognition of linguistic production and writing strategies.

KEYWORDS

baseline, cognitive load, deception detection, keylogger, written report

1 | INTRODUCTION

Several studies have shown that focusing on linguistic cues is more efficient when it comes to detecting deception, although effect sizes remain low (DePaulo et al., 2003; Hauch, Blandon-Gitlin, Masip, & Sporer, 2015). Methods have been elaborated for the linguistic analysis of deceit: Their accuracy ranges from above-average to notable 85% and higher classification rates (Masip, Bethencourt, Lucas, Segundo, & Herrero, 2012; Mihalcea & Strapparava, 2009; Ott, Choi, Cardie, & Hancock, 2011). Computerized means of detecting deceit via linguistic cues allow for the extraction and analysis of many cues, some of which varying in regards to a truthfulness factor (Hauch et al., 2015). But the examination of linguistic deception detection via natural language processing methods show a recurrent pattern: Most studies use a dictionary-based approach such as the Linguistic Inquiry and Word Count (e.g., Bond & Lee, 2005; Newman, Pennebaker, Berry, & Richards, 2003). Yet, computer-assisted analyses may encompass other ways of scanning the linguistic production of deception, such as behavioral keystroke dynamics.

2 | KEYSTROKE DYNAMICS AND DECEPTION DETECTION

Biometric and behavioral cues of deception detection have recently been studied, especially in the domain of identity deception (e.g., Monaro et al., 2018; Monaro, Fugazza, Gamberini, & Sartori, 2017; Monaro, Gamberini, & Sartori, 2017a, 2017b; Sartori, Zangrossi, & Monaro, 2018). These behavioral approaches rely on two types of measurements: mouse dynamics and keystroke dynamics. Mouse dynamics allow researchers to measure reactions times, velocity, accelerations, or average and prototypical trajectories when answering questions on a screen. A study focusing on the yes–no response of identity thieves to unexpected questions showed that these data combined with machine-learning classifications allow to detect deceptive identities correctly in more than 90% of the cases (e.g., Monaro, Gamberini, & Sartori, 2017b). Keystroke dynamics are behavioral biometric measures relying on typing behavior. In brief, the capture of data (i.e., typing behavioral cues) allows to extract unique features (e.g., how one relies on the deletion key) and use these

 | © 2020 John Wiley & Sons Ltd

features for mapping (i.e., the comparison and matching of the unique features against a large amount of data) and classification purposes (Bhatt & Santhanam, 2013). In other words, when considering keystroke dynamics, what matters is not what one types but how it is typed (for a review, see Crawford, 2010). Two different types of measures may be considered in keystroke dynamics: temporal and nontemporal measurements (Tsimperidis, Arampatzis, & Karakos, 2018). Temporal features encompass keystroke duration (i.e., how long a key is pressed) and bigram latency (i.e., the time between two consecutive keystrokes). Nontemporal features are commonly including total amount of keystrokes, or corrections (amount of *Deletes* or *Backspace* keystrokes used).

Keystroke patterns have been used for various purposes. A study found that keystroke dynamics, even with scarce data, helped recognize participants with an accuracy rate of 73% (Rybnik, Tabedzki, & Saeed, 2008). Zahid, Shahzad, Khayam, and Farooq (2009) worked on identification matters related to smartphones, and showed that the combination of keystroke dynamics during Personal Identification Number typing, and algorithmic classifications reduced error rates to 2%. Finally, keystroke patterns have also been studied to determine the gender of a computer user. In a study, participants were asked to record the everyday use of their computer while knowing that their keystroke patterns were registered (e.g., Tsimperidis et al., 2018). This method lead to interesting results: Relying on a *radial basis function basis network* algorithm, gender accuracy reached 95.6%.

To our knowledge, only a few studies have focused on the deception of detection by analyzing typing behavior. Derrick, Meservy, Jenkins, Burgoon, and Nunamaker Jr (2013) used a synchronous communication chatbot that would ask questions, and require a spontaneous truthful or a deceptive answer. This study examining synchronous computer-mediated-communication highlighted that deception implied more edits and deletions, and a longer reaction time than when truthful responses were typing. Another study focused on the differentiation between truthful and deceptive online opinions of a restaurant, gay marriage, and gun control (Banerjee, Feng, Kang, & Choi, 2014). By combining timespan measurements, bag of words (i.e., a statistical model where words are considered independently of their order), editing behaviors, and using support vector machine classifiers, Banerjee et al. (2014) showed that the classifications were overall better when taking all three measurements into account. Accuracy rose up to 94.3%. A few studies applied keystroke dynamics to detect deceitful identities (Monaro, Spolaor, et al., 2017; Monaro et al., 2018; Sartori et al., 2018). These studies demonstrated that liars made more conceptual errors in their answers than truth-tellers, have slower reaction times when answering questions, and were slower to answer than truth-tellers. Accuracy rate reached 93.75% (Monaro et al., 2018). Other similar experiments show accuracy rates above 90% (Monaro et al., 2019; Monaro, Spolaor, et al., 2017).

However, to our knowledge, no study has investigated keyboard dynamics in the case of deceptive episodic allegations. This is an important topic, as many agencies allow for the complainant to file a report online. This means that a potentially malicious report may be filed, and thus be costly for the agency. We thus focused in this study on the keyboard dynamics of truthful and deceptive eyewitness reports, and the imposition of an additional cognitive load during the typing of the account.

3 | INFORMATION MANIPULATION, DECEPTION DETECTION, AND AUDITORY COGNITIVE LOAD

Information manipulation theory allows to conceptualize the production of deceptive narratives (Markowitz, 2020; McCornack, Morrison, Paik, Wisner, & Zhu, 2014). It relies on the Cooperative Principle and Grice's maxims (Grice, 1989). A message emitted toward the receiver is supposed to provide all the necessary information necessary to its understanding (Cooperative Principle; Grice, 1989). If one is being cooperative (and by extension, honest), the information emitted should be sufficient (Quantity maxim), should provide supporting evidence for one's claims (Quality maxim), be relevant to the context of utterance (Relation maxim), and be clearly uttered (Manner maxim; Markowitz, 2020). Covert violation of these maxims is then considered as deception. For instance, providing more information than one has in one's possession, or not enough information, can be considered as a violation of the Quantity maxim. These covert manipulations of the available information imply an increase of cognitive load for the deceiver (McCornack et al., 2014). Deception is a cognitively complex task because one has to inhibit the truth, build up a new plausible alternative reality and, simultaneously, monitor their own behavior and the reactions of the addressee for feedback as most liars do not take their credibility for granted (Gombos, 2006; Vrij, Fisher, & Blank, 2017; Walczyk, Harris, Duck, & Mulay, 2014). Moreover, the liar has to remember what has been said beforehand in order not contradict previous accounts (Driskell & Driskell, 2019). This would be the reason why small differences between truthful and deceptive narratives are observable.

The imposition of an additional cognitive load should help emphasize discrepancies between truthful and deceptive statements (Blandón-Gitlin, Fenn, Masip, & Yoo, 2014; Walczyk, Igou, Dixon, & Tcholakian, 2013). This supplementary cognitive load increases the complexity of the deceptive act, and thus accentuates differences between deceptive and truthful narratives. The imposition of cognitive load may for instance entail (a) asking the potential liar to narrate his/her account from end to beginning (i.e., the reverse-order technique), (b) asking unanticipated question which cause the potential liar to react fast in order not lose his/her credibility credit, (c) encouraging for further details, as liars tend to provide less information than truth-tellers, or (d) the imposition of a simultaneous second task to perform while narrating one's account (Vrij et al., 2017).

One of the propositions explaining the effects of an additional cognitive load on deceptive narrative relies on Baddeley's working memory model (e.g., Baddeley, 2012) and provides insights as to how cognitive load is being manipulated. It argues that the central executive part of the working memory model responsible of the overall control of action could undergo heavy perturbations in dual tasks. This is in line with Wickens's multiple resource theory (Driskell & Driskell, 2019; Wickens, 2008). This theory posits that our attentional resources are spread according to four dimensions. Each of these dimensions is divided in two levels. These dimensions are: stages (i.e., perceptual and cognitive activities vs. response), perceptual modalities (i.e., visual vs. auditory, where cross-modal is easier to treat than intra-modal), visual channels (i.e., resources dedicated to focal vs. ambient vision), and processing codes (i.e., verbal vs. spatial). Two tasks to occur simultaneously are considered as more demanding if they require the same level of a dimension. For instance, focusing on a visual stimulus while someone else speaks in the nearby environment (both perceptual, intra-modal tasks in during the memorization stage) should be more difficult than focusing on a visual stimulus while pressing on a button when the lights go off.

Sporer (2016) encourages researchers to focus on task demand and working memory overload by using auditory stimuli. From a psycholinguistic viewpoint, a phenomenon like the *irrelevant speech effect* could provide evidence-based background and framework for the auditory stimuli testing. The irrelevant speech effect is a linguistic perturbation that has been explored for decades (e.g., Baddeley & Salame, 1986). Its deleterious effect has been noticed during encoding (e.g., Knez & Hygge, 2002), serial recall (e.g., Baddeley & Salame, 1986; Tremblay, Nicholls, Alford, & Jones, 2000), or episodic recall (e.g., Enmarker, 2004). It has been found to be independent of the one's age (e.g., Van Gerven, Meijer, Vermeeren, Vuurman, & Jolles, 2007), or whether one pays attention or not to the distraction noise (e.g., Elliott & Briganti, 2012). Some factors seem to make the irrelevant speech effect more efficient, such as the fact that meaningful words have a more disrupting effect than nonwords (e.g., Oswald, Tremblay, & Jones, 2000). Negatively valenced words also seem to have an important depleting impact on memory and performance (e.g., Buchner, Mehl, Rothermund, & Wentura, 2006), and so would also the specific foreknowledge of the stimulus that will be presented (Röer, Bell, & Buchner, 2015).

Interestingly, Baddeley's working memory model also provides explanations for the irrelevant speech effect. The irrelevant speech effect would have a saturating effect on the phonological loop, one of the subsystems that guides verbal structuring and articulation (e.g., Baddeley & Salame, 1986). Studies show that the cognitive overload created by the irrelevant speech effect can be important. Large effect sizes have been reported, including in episodic memory contexts (e.g., calculated Hedge's g = 1.34 for Enmarker, 2004). This cognitive load has importance from a practical viewpoint. For instance, medical students who would be distracted by an auditory stimulus would have more trouble distinguishing the right from the left (McKinley, Dempster, & Gormley, 2015). Similarly, when someone is writing, the wording and translation of abstract concepts into words would be impeded, which may lead to more errors and a certain delay in answering written questions (e.g., Hayes & Chenoweth, 2006; Levy & Marek, 1999).

To determine the effect of the irrelevant speech effect, keyboard analytics may be of interest. When undergoing various types of cognitive loads, such as time pressure, extensive text copying, or language unfamiliarity in text copying, participants type generally less keystrokes per second, and tend to pause more often as their fluency drops (e.g., Vizer, Zhou, & Sears, 2009). This is coherent with the cognitive load paradigm in deception detection research: Liars generally need more time to react, type, or answer questions than truth-tellers (e.g., Suchotzki, Verschuere, Van Bockstaele, Ben-Shakhar, & Crombez, 2017). Pauses in typing behavior have been determined to appear for two reasons: high cognitive demand and impression management (Schilperoord, 2002). Both explanations for pauses in typing behavior fit with the concepts mentioned by Gombos (2006) and Vrij et al. (2017) to justify the cognitive complexity of deception, namely the decrease in cognitive capacity while one prevents the truth from erupting, and the observation of one's own behavior to appear as credible as possible. Consequently, there seems to be a tendency to slowing down the process of writing while typing (i.e., less keystrokes per second, cognitive pauses) and editing (i.e., the amount of *Backspaces* and *Delete* keystrokes, impression management pauses). Moreover, there seems to be a difference in typing behavior according to the strength of the cognitive load. Recent research has highlighted that text revision (i.e., the amount of *Delete* and *Backspace* strokes) allowed to distinguish highly loaded tasks from lightly loaded tasks (e.g., Brizan et al., 2015).

The influence of the auditory cognitive load on the production of deceptive messages has not been studied yet. Some studies focused on the impact of the irrelevant speech effect of typing behavior while transcribing text. A study by Levy and Marek (1999) showed no effect of the auditory cognitive load on one's typing behavior, concluding to a lack of influence of verbal working memory perturbation on mechanic typing behavior. But more recent work showed that the saturation of the phonological loop could have an impact on typing behavior (Chenoweth & Hayes, 2003; Hayes & Chenoweth, 2006; Sörqvist, Nöstl, & Halin, 2012). In studies by Chenoweth and Hayes, typing rate was reduced (i.e., less keystrokes per second), and errors increased as the participants were actively saturating their phonological loop by repeating the word 'tap' regularly (i.e., articulatory suppression). Their argument displayed as an explanation for their result differences with Levy and Marek (1999) consists in the strength of the cognitive load perturbation: articulatory suppression disrupts verbal working memory more severely than the irrelevant speech effect. However, there is evidence that the hearing of discourse while typing is a disturbing phenomenon: the irrelevant speech effect was shown to decrease the amount of words, characters, sentences, writing fluency (i.e, characters/time), and propositions, while increasing the amount of errors and the length of pauses (Sörqvist et al., 2012).

4 | RESEARCH QUESTIONS

4.1 | Hypotheses regarding the presence of deceit on keystroke dynamics

Based on the aforementioned literature, we hypothesize the following phenomena. Relying on the cognitive load approach to deception and the literature regarding cognitive load and keystroke dynamics, we expect one's global typing behavior to be impeded by the presence of deception when compared to the baseline (H_1).

At a univariate level, when compared to a baseline measure, we hypothesize that the deceptive narrative would be typed slower than their truthful counterparts because the cognitive load induce by deception would impede the writer's cognitive processes and require more time to structure the narrative (e.g., Suchotzki et al., 2017).

We also hypothesize, based on the observation that deceptive narratives are generally shorter than their truthful counterparts (Hauch et al., 2015), that liars would use less keystrokes than the truthful participants. We thus computed the amount of keystrokes comprised between the first stroke and the last one, included.

Liars also tend to perfect their narratives before delivering these as they do not take their credibility for granted, and thus deploy impression management techniques (e.g., Brizan et al., 2015; DePaulo et al., 2003; Sporer, 2016). Moreover, the load imposed by the deceitful act may lead them to make more mistakes and thus to correct their narratives more often (e.g., Vizer et al., 2009). We thus hypothesize that deceptive writers would correct themselves more by using the *Backspace* and the *Delete* keys.

Moreover, as the cognitive load induced by the act of lying generally involves more thinking and more control over actions, we supposed that deceptive texts would contain longer pauses (i.e., the average pause length above 0.5 s, based on Vizer et al., 2009), and a higher pause rate (i.e., the amount of pauses divided by the amount of keystrokes, based on Vizer et al., 2009).

Finally, as far as the time taken to write one's narrative, two hypotheses may be contrasted: either deceptive narratives take more time, due to the inherent cognitive load of deception when compared to truth-telling, or truthful narratives take more time to write, as they contain more information and more details.[1] The timing necessary to write one's narrative will thus be considered as an exploratory hypothesis. A precise description of the dependent variables is available in Data S1.

4.2 | Hypotheses regarding the presence of the irrelevant speech effect on keystroke dynamics

We further hypothesized that the imposition of a cognitive load (in this case, an auditory cognitive load defined as an irrelevant speech effect) would deplete one's ability to type the required narratives. At a univariate level, based on the cognitive load hypothesis and the literature regarding the impact of an auditory cognitive load on typing dynamics (e.g., Sörqvist et al., 2012), we expect that narratives written while undergoing the imposition of a auditory stimulus would, when compared to their baselines, be typed slower, be written with less keystrokes, be composed of more Backspace and Delete keystrokes, need more time to be typed, contain longer pauses, and a higher pause rate (H_2).

4.3 | Interaction hypotheses

We expect an interaction between the presence of deception, and the presence of cognitive load, as foreseen by current models (H_3, based on Sporer, 2016). In other words, one's typing abilities would be depleted with the presence of an auditory cognitive load, and even more so when the narrative is deceptive. We thus foresee that narratives written while undergoing the imposition of a auditory stimulus would, when compared to their baselines, be typed slower, be written with less keystrokes, be composed of more *Backspace* and *Delete* keystrokes, need more time to be typed, contain longer pauses, and a higher pause rate, and that these effect would be emphasized if the narratives considered are deceptive.

4.4 | Classification hypotheses

Classification models will be tested to determine which fits best to the data gathered in this study, and to what extent it is able to classify correctly the collected narratives. We hypothesize that deceptive narratives will be more accurately classified than truthful ones, based on the baseline hypothesis (Feeley, deTurck, & Young, 1995). Deception detection tasks based on common (but wrong) beliefs or random guessing generally classify data accurately around the chance level (i.e., 54% according to Aamodt & Custer, 2006). Our hypotheses and measurements being based on the current scientific literature, we expect our classification for deceptive and truthful narratives results to be above this 54% overall accuracy threshold, with a higher accuracy score for deceptive narrative.

In terms of cognitive load, we expect results to show higher classification accuracy for narratives in the presence of cognitive load than in its absence. More precisely, we expect to reach the same levels of classification accuracy as highlighted by a recent meta-analysis on the topic (i.e., around 70%; Vrij et al., 2017).

The rest of the article is structured as follows. We describe how the data was acquired, and how the keystroke dynamics variables were extracted. We then present the results for the four hypotheses and show how keyboard dynamics may be helpful to classify deceitful and truthful narratives by relying on different algorithms such as support vector machine (SVM), multilayer perceptron (MLP), radial basis function network (RBFN), and simple logistic (SL). We then discuss and conclude with an overview for further research, and how it may apply to the needs of the field.

5 | METHOD

5.1 | Participants

Eighty French-speaking adult participants (34 female) took part in the experiment. Age ranged from 18 to 59 years, with an average of M = 26.34 years old (SD = 8.02 years). Seventy-three participants

(91.25%) declared to be right-handed. Seventy-seven participants were pursuing or had a Bachelor degree (96.25%), the three others (3.75%) mentioning a higher degree. All of them reported using the computer 1–2 hr a day. One of the participants failed to be recorded, which brings the amount of participants to $N = 79$. We computed a sample size estimation with GPower for special effects and interaction for a MANOVA, with Cohen's $f^2 = .15$ (i.e., a medium effect size according to Cohen, 1988), $\alpha = .05$ and power at 80%. It indicated that 68 participants were necessary for this study. As more participants were available, and in order to avoid cherry-picking, we included all participants' responses in our study.

5.2 | Design

The experiment utilized a 2 (*Truthfulness*: truthful vs. deceitful) × 2 (*Cognitive load*: present vs. absent) between-subjects design. Prior to the experimental condition, all participants were all asked to type about their day with as many details as possible from the moment that they opened their eyes until the moment they arrived in the laboratory. This would constitute their baseline. Participants were afterwards randomly assigned to one of the four conditions.

Regarding the truthfulness treatment, truthful people had to tell what they had seen in a video that they were showed, and to provide as many details as possible. Deceptive people were asked to remember as many details from a summary of the event, and to inflate the story in order to create the false belief that they had really witnessed the event. Similar instances of this summary protocol were used in other studies (e.g., Logue, Book, Frosina, Huizinga, & Amos, 2015).

For the cognitive load treatment, during their report of the event, participants wore a headset. If cognitive load was to be present, an excerpt of the 2012 French presidential election debate would be heard in the headphones. If no cognitive load was expected, no sound would be heard.

Keystroke measurements included the average count of characters per second (i.e., the average amount of characters typed in 1 s), the average count of keystrokes (i.e., the counting of all the pressures and releases on the keyboard), the count of *Delete* keystrokes, the count of *Backspace* keystrokes, the total elapsed time required to type one's narrative, the average pause length (i.e., the average time that the keyboard remained idle for more than 0.5 s), and the pause rate (i.e., the amount of pauses as defined previously divided by the count of keystrokes) to determine if truthfulness and/or cognitive load had an effect on them.

5.3 | Procedure

Participants were approached by email lists available at the university, and in class. They were informed that the experiment concerned the assessment of eyewitness testimony and lasted around 30 min. Participants entered the laboratory and were asked to read and sign the informed consent form and the information form before entering the room of the experiment. Meanwhile, the experimenter went into the room, and launched the keylogger. The experimenter entered all the sociodemographic data in the software. After saving the data, the software opens a blank page in a text editor. Participants were then brought into the room. The experimenter then explained that the participant should write down everything that they remember of their day, starting the moment they woke up to the moment of arrival in the laboratory. After collecting this baseline, the participants were sat in another part of the room for their next task.

Truthful participants would watch a video of an event lasting 90 s. This event described an encounter between a young man and a young woman. After discussing a while, the young woman would receive a text on her phone, and say that she had to leave. The volume was adjustable to suit the participant's hearing (and to later serve as a landmark for the auditory cognitive load). For the Deceptive condition, a summary of the video was given. It was brief enough for them to know what happened without having many details. They had 90 s to remember as much from the summary as possible. Participants in the deceitful condition were told after this that the girl went missing, and that they had to imagine that the authorities had a reward for anyone who would have information about her disappearance.

All participants were then told that the girl was missing, that everything they could remember would be useful, and that they would have to produce a detailed account of the event. They were also told that the police is looking for her, and that a reward would be given to anyone providing new information. In the deceptive condition, participants were clearly told to inflate the data that they were given so that if a police officer were to read their account, they would consider it as credible. All participants then passed a nonverbal distraction task (i.e., a maze solving task) during 3 min. This task was pre-tested to be certain that it was not too easy ($N = 5$, $M = 13$ min 58 s, $SD = 5$ min 0 s). After this distraction task, all participants given two instructions. The first one concerned the irrelevant speech effect. Participants were given a headset to wear. Some of them (i.e., those who underwent cognitive load) would hear an excerpt of the 2012 French presidential election debate and if so, should ignore it as much as possible. The excerpt was 10 min long, and put in a repeating loop to be certain that the irrelevant speech effect would always be ongoing. The participants who did not undergo auditory cognitive load would hear nothing in the headset. Second, participants were asked to narrate with as many details as possible the event. Finally, the participants were debriefed and free to ask any question they wanted.

The next step consisted in the extraction of the features from the log files. To do so, the second author developed a free program entitled 'ISqueezeU'.[2] This software scans the raw data from IRecU and extracts the desired features. The data was analyzed with JASP for statistical inferences and the computation of the Bayes Factors.

5.4 | Data normalization

The goal of this study is to determine the typing behaviors of liars and truthful persons compared to their baseline. We subtracted the results from the keylogger for the either truthful or deceitful narrative each participant produced from the baseline. For instance, this means that if a participant had 6,455 keystrokes during his/her baseline, and 6,435 keystrokes during his/her narrative, the positive difference between both (i.e., 20 keystrokes) shows that the baseline had more keystrokes than the treatment narrative. A negative result shows that the narrative was written with more keystrokes than the baseline. This linear data modification was applied on all the extracted cues: For each observed variables, the keystroke results from the eyewitness testimony were subtracted from the ones of the baseline.

5.5 | Data classification

Classifications were ran on Weka, a well-known and well-documented open source software, containing dozens of machine-learning algorithms.

Results will be analyzed in terms of accuracy, *F*-score, and AUC. The *F*-score is a combined measurement of precision and recall that favors the algorithms with higher sensitivity (i.e., single-class effectiveness of a classification model). The area under the curve (AUC) measurement focuses on the area under the receiving operating characteristics (ROC) curve. The ROC curve plots in a two-dimensional space the true positive rate in the *Y* axis, and the true negative rate on the *X* axis. The AUC measures the area under this curve, and compares it to a straight diagonal with a value of .5, dividing the two-dimensional space in two, representing chance level. In other words, the faster and the higher the AUC of a classifier is above the diagonal, the less its results are considered as chance.

6 | RESULTS

6.1 | Preliminary checks

A MANOVA was ran to examine the effect of the participant's sex on the observed variables. No effect was found, $F(7,69) = 1.06$, $p = .40$, Wilk's $\Lambda = 0.89$. Concerning the age of the participants, IRecU stores them by age categories. No effect of the age category could was observed, $F(14,136) = 0.88$, $p = .59$, Wilk's $\Lambda = 0.82$.

6.2 | Hypothesis testing

A MANOVA was conducted with truthfulness (truthful vs. deceitful) and cognitive load (present vs. absent) as the between-subject factors. At the multivariate level, a main effect for truthfulness was found, $F(7,69) = 3.12$, $p = .006$. This implies an overall effect of the truthfulness factor on keystroke measurements, in line with H_1. A main effect of was found for cognitive load, $F(7, 69) = 1.60$, $p = .15$, and did not yield support to H_2. No interaction was observable between truthfulness × cognitive load, $F(7,69) < 1$, ns = .89. No support was thus found H_3.

A one-way ANOVA showed a main effect was found for truthfulness on the difference in keystrokes per second between baseline and narratives, $F(1,77) = 6.36$, $p = .01$, $\eta^2_p = .08$, $BF_{10} = 3.51$. This implies that, compared to their baseline, liars typed their narratives slower ($M = 0.14$, $SD = 0.69$, 95% CI [−0.08; −0.36]) than truth-tellers ($M = -0.19$, $SD = 0.48$, 95% CI [−0.36; −0.04]). These results bring support to H_1. Truthfulness was also found to influence the difference in total elapsed time between baseline and narrative, $F(1,77) = 5.58$, $p = .03$, $\eta^2_p = .06$, $BF_{10} = 2.15$. Deceitful narratives, when compared to their baseline, were typed in less time ($M = -88{,}110.65$, $SD = 225{,}720.21$, 95% CI [−160,299.47, −15,921.83]) than their truthful counterparts ($M = 23{,}527.18$, $SD = 208{,}530.30$, 95% CI [−44,070.52, 91,124.88]), bringing support to H_1. No support was found for the other observed variables, all *p*s > .05, as shown in Table 1.

Regarding the impact of cognitive load on the dependent variables, a one-way ANOVA showed that cognitive load had an effect on the difference in the count of keystrokes between baseline and narrative, $F(1,77) = 4.69$, $p = .03$, $\eta^2_p = .06$, $BF_{10} = 1.73$. The presence of cognitive load seemed to affect the count of keystrokes contrary to what was expected from H_2: Compared to their baseline, participants undergoing auditory cognitive load typed more ($M = 1.74$, $SD = 556.77$, CI 95% [−178.74, 182.23]) than when no cognitive load was present ($M = 312.10$, $SD = 706.27$, CI 95% [86.22, 537.98]). A Pearson correlation was run in comparison with the Word Count function from the linguistic inquiry and word count software (Pennebaker, Francis, & Booth, 2001). It showed a strong correlation between the amount of keystrokes typed and the word count, $r = .96$, $p < .001$, 95% CI [1.72; 2.17]. These results challenge H_2, in which we hypothesized that

TABLE 1 ANOVA results for the effect of truthfulness on keystroke dynamics

Variables	F	p	η^2_p	M_{true} (SD_{true})	M_{deceit} (SD_{deceit})
Keystrokes per second	6.38	.01**	.08	−0.19 (0.48)	0.14 (0.69)
Count of keystrokes	3.31	.07	.04	25.8 (532)	289 (734)
Count of *Backspace* keystrokes	1.61	.21	.02	−15.3 (62.7)	6.58 (87.9)
Count of *Delete* keystrokes	1.00	.32	.01	2.97 (17.9)	0.125 (1.88)
Total time	5.58	.03*	.06	23,527 (208,530)	−88,111 (225,720)
Average pause length	0.08	.78	.00	45.7 (353)	77.6 (627)
Pause rate	0.62	.43	0.01	−0.01 (0.02)	−0.02 (0.13)

*$p < .05$.
**$p < .01$.

cognitive load would lower the amount of keystrokes in deceptive narratives. Cognitive load was also found to influence the difference in average typing time between baseline and narrative, $F(1,77) = 5.25$, $p = .02$, $\eta^2_p = .06$, $BF_{10} = 2.19$. The presence of auditory cognitive load was found to increase the time duration of the narratives relative to their baseline (*M* = 23,740.79, *SD* = 184,952.26, 95% CI [−36,213.79, 83,695.38]) as compared to when no cognitive load was heard (*M* = −88,318.93, *SD* = 244,829.50, 95% CI [−166,619.20, −10,018.65]). No other effect was statistically significant for the other variables observed in this study, all *ps* > .05 as shown in Table 2.

We then used classification techniques to determine whether the participants' report of the event when compared to their baseline, could be accurately identified in terms of truthfulness and presence of cognitive load. Therefore, four different classification experiments were performed, one for each different condition, examining whether the way users typed differed from their baseline. The seven keystroke dynamics features that were extracted were used for the classification as independent variables. The four treatments in this study (i.e., authentic/deceptive with/without cognitive load) were considered as the dependent variables for this machine-learning experiment.

For this purpose, classifiers such as neural networks, decision trees, Bayesian classifiers, were selected based on their high accuracy as reported in the literature in similar studies of keystroke dynamics (e.g., Tsimperidis et al., 2018). Those who had the best results were support vector machine (SVM), multilayer perceptron (MLP), radial basis function network (RBFN), and simple logistic (SL), and only their results will be presented. The term 'best results' refers to the performance of the classifier in terms of accuracy (acc.), F-measure (F1), and area under the ROC curve (AUC), as they are explained in Tsimperidis et al. (2018). Classification results are displayed in Table 3.

TABLE 2 ANOVA results for the effect of cognitive load on keystroke dynamics

Variables	*F*	*p*	η^2_p	M_{withCL} (SD_{withCL})	$M_{withoutCL}$ ($SD_{withoutCL}$)
Keystrokes per second	0.3	.87	.00	−0.01 (0.49)	−0.03 (0.723)
Count of keystrokes	4.69	.03*	.06	1.74 (557)	312 (706)
Count of *Backspace* keystrokes	1.49	.22	.02	−14.9 (65.4)	6.17 (86.1)
Count of *Delete* keystrokes	0.78	.38	.01	0.26 (3.13)	2.77 (17.16)
Total elapsed time	5.25	.02*	.06	23,741 (184,952)	−88,319 (244,830)
Average pause length	1.36	.25	.02	−5.34 (364)	127 (614)
Pause rate	0.73	.39	0.01	0.00 (0.02)	−0.02 (0.13)

$^*p < .05$.

TABLE 3 Best performance of classifiers for truthfulness and deception detection

Conditions	Classifiers	Acc. (%)	95% CI (*LL*–*UL*)	F1	AUC
Truthful with cognitive load	SVM	53.9	38.3–69.5	.537	.537
	MLP	51.3	35.6–67.0	.506	.437
	RBFN	43.6	27.9–59.3	.425	.459
	SL	61.5	46.2–76.8	.612	.584
Truthful without cognitive load	SVM	57.5	42.2–72.8	.568	.575
	MLP	57.5	42.2–72.8	.573	.473
	RBFN	57.5	42.2–72.8	.562	.177
	SL	57.5	42.2–72.8	.562	.643
Deception with cognitive load	SVM	52.5	37.0–68.0	.525	.525
	MLP	47.5	32.0–63.0	.475	.477
	RBFN	47.5	32.0–63.0	.472	.414
	SL	55.0	39.6–70.4	.550	.598
Deception without cognitive load	SVM	67.5	53.0–82.0	.665	.675
	MLP	62.5	47.5–77.5	.625	.533
	RBFN	70.0	55.8–84.2	.699	.633
	SL	62.5	47.5–77.5	.625	.658

Abbreviations: Acc., accuracy for correct classification; AUC, area under the curve; CI, confidence interval; F1, *F*-score; LL, lower limit; MLP, multilayer perceptron; RBFN, radial basis function network; SL, simple logistic; SVM, support vector machine; UL, upper limit.

For each classifier, a hyperparameter tuning process was followed, in order to find those values that lead to the best performance. Moreover, to assess the performance of each classifier fairly, the 10-fold cross-validation method was used. This divides the data into 10 disjoint parts, uses 9 of them for training and the remaining one for testing, in a round-robin fashion (Ramezan, Warner, & Maxwell, 2019).

The best performances for truthful narratives with cognitive load was achieved for SVM with a polynomial kernel and a *C* parameter, representing the margin, of 2000.0. MLP reached best performance while having learning rate (*L*) equal to 0.9 and momentum (*M*) equal to 0.8. RBFN achieved best performance by using 10 clusters for *K*-means and a 0.1 minimum *SD* for the clusters. SL achieved best performance with 500 as the maximum number of iterations for LogitBoost (*M*), 100 as the last iteration of LogitBoost if no new error minimum has been reached (*H*), and 65% for weight trimming (*W*).

In the case of truthful narratives without cognitive load, the best performances were achieved by the SVM with RBFKernel (radial basis function kernel) and $C = 0.60$, by MLP with $L = 0.50$ and $M = 0.20$, by RBFN with three clusters and $SD = 0.50$, and by *SL* with $M = 5{,}000$, $H = 50$, and $W = 40\%$.

In the case of deceptive narratives with cognitive load, the parameters for best classification rates for SVM were the use of a RBFKernel and $C = 0.6$; $L = 0.5$ and M = 0.5 for MLP; 10 clusters and $SD = 0.10$ for RBFN; and M = 500, $H = 50$, and $W = 30\%$ for SL.

Finally, in the case of deceptive report without cognitive load, the best results for SVM were achieved with a polynomial kernel, and $C = 11$; for MLP with $L = 0.30$ and $M = 0.50$; for RBFN with two clusters and $SD = 2.60$; and for SL with $M = 500$, $H = 400$, and $W = 60\%$.

Accuracy results shows that the SL classifier reaches the best classification rates for truthful narrative with and without cognitive load imposition, with 61.5 and 57.5% namely. The SL classifier also shows best classification rates for deceptive narrative with cognitive load imposition with 55%. The RBFN classifier shows the best classification rate for deceptive narrative without cognitive load imposition with 70% accuracy. This brings partial support to our classification hypothesis, as deception was indeed better detected in the absence of cognitive load, while its presence allows for better classification of truthful narratives.

Considering the *F*-score, the SL classifier performed better than the other classifiers when cognitive load was imposed during truthful and deceptive narratives (i.e., 0.612 and 0.550 namely). In the absence of additional cognitive load, the MLP classifier performed better than the other classification model at classifying correctly truthful narratives, with an *F*-score of 0.573, whereas the RBFN classifier reached an *F*-score of 0.6999 for the classification of deceptive narratives written in the absence of cognitive load.

Results in this study show that for truthfulness with and without cognitive load, and deception with cognitive load, the SL classifier performs better than the other classification models, with AUC scores of namely .58, .64, and .60. Regarding the classification of deceptive narratives without cognitive load imposition, the SVM classifier performed better than the other classifiers, reaching .68 accuracy.

7 | DISCUSSION AND CONCLUSION

The goal of this study was to determine whether deception detection could be operated in eyewitness testimonies by means of keyboard biometrics. We investigated whether deception induced differences in keyboard measurements when compared to a typed baseline. The goal of the baseline was to account for typing variability among participants, and allowed to take idiosyncratic measurements and typing familiarity into account. We also imposed an irrelevant speech effect in order to magnify the differences in keyboard measurements based on the cognitive load paradigm. To our knowledge, this is the first study to investigate keystroke dynamics in potentially deceptive eyewitness testimonies with the inclusion of a cognitive load stimulus.

We found that the truthfulness of the participants had an impact on their keystroke measurements. Deception involved typing slower than the baseline, whereas truthful people typed faster. Moreover, deceitful narratives were shown to be typed in less time. This provides support to the cognitive load theory: As lying is more complex than telling the truth, we can expect people to type slower and to want to stop writing faster than truthful eyewitnesses as a strategy to cope with cognitive load (Schilperoord, 2002; Vrij et al., 2017).

However, the induction of the irrelevant speech effect suggests that the cognitive load theory might more nuanced than what is claimed regarding its capacity at distinguishing liars from truth-tellers (Sporer, 2016; Vrij et al., 2017). In our study, the imposition of cognitive load was shown to increase the amount of keystrokes in narratives when compared to a baseline. In other words, when participants underwent the irrelevant speech effect, they typed more without correcting more, which contradicts the predictions of the deception detection cognitive load theory (Sporer, 2016; Vrij et al., 2017), the irrelevant speech effect (Enmarker, 2004) and Wickens's (2008) model. This brings support to the suggestion that cognitive load is not monolithic, and that it should be investigated with more interest.

Finally, classification results show that, when compared to the baseline, deception is better detected in the absence of cognitive load imposition. The radial basis function network classifier shows that it is possible to detect deception while relying on keyboard dynamics with 70% accuracy, and a *F*-score of 70%. These results are above the 54% chance level established in the literature on deception detection (Aamodt & Custer, 2006). However, when cognitive load is introduced in order to increase classification rates, it has the adverse effect of decreasing correct classification to levels barely at chance level. This situation contradicts the findings of the meta-analysis ran by Vrij et al. (2017), and is far from the 70% claimed in their paper.

To explain these results, we intend to provide two hypotheses regarding cognitive load to be tested in the future. The first hypothesis concerns the ambivalence of the cognitive load stimuli studied in deception detection research. Most studies rely on the imposition of stimuli that have a beneficial effect for the truth teller, and a negative effect for the deceivers. For instance, the use of the reverse-order technique has an ambivalent effect: It favors the truthful eyewitness by allowing him/her to remember more details and provide new information (e.g., Fisher & Geiselman, 2010), where as it impedes the structuration and production of deceptive narratives (e.g., Gombos, 2006). Similarly, the invitation to pursue one's narrative after being done is another ambivalent cognitive load stimulus: It has been supported by experimental research as an interesting technique to help the truthful person give more details (e.g., Bogaard, Meijer, & Vrij, 2014). Conversely, it creates a burden for the deceiver who cannot provide new information, and thus undergoes important cognitive load (Vrij et al., 2017).

The cognitive load in this study does not favor the production of a truthful narrative: The auditory cognitive load is a burden for both the truthful and the deceptive participant. For this reason, we hypothesize that the auditory stimulus creates noise, and prevents algorithms from relying on keystroke dynamics. We thus suggest further research in the nature of auditory cognitive load by manipulating the auditory stimulus (language understandability, emotional valence, etc.). The examination of the auditory cognitive load might give insights on the theories currently established in deception detection research.

These suggestions for further research should include the linguistic models of written production. If one relies on Sporer (2016), one might want to extend the model by adding Kellogg's working memory model of written production (Kellogg, Whiteford, Turner, Cahill, & Mertens, 2013). This model deepens the potential hypotheses from Sporer (2016), as it combines Baddeley's model (2012) with a temporally distributed model of the linguistic written production. This might allow researchers to enhance their knowledge of word production, and thus provide a more fine-grained analysis of how deception alters its mechanisms. In this study, for instance, the irrelevant speech effect was supposed to use resources simultaneously to other linguistic productions. However, according the literature regarding the cognition of writing, the irrelevant speech effect would only act on the translation of concepts into words, which should impact the timing of the narrative and keystrokes in a very particular way.

The second, nonexclusive, hypothesis concerns the strategies at stake when confronted to cognitive load. The cognitive load theory has been under scrutiny for a few years, as its mechanisms are to be explored cautiously (e.g.,Blandón-Gitlin et al., 2014; Gombos, 2006). A model relying on working memory principles was established, which we tested in this study (Sporer, 2016). Although our results contradict some hypotheses extracted from the model, we do not argue against the model, but rather for an inclusion of writing production as mentioned above, and writing strategies one may deploy when writing with high cognitive load. In other words, we argue that one has to focus on the coping mechanisms of participants, and to a greater extent, eyewitnesses when undergoing cognitive load. The differences regarding the impact of the auditory cognitive load, although significant, remained small, and the *SD* measurements show important differences as to how one reacts in the presence of the irrelevant speech effect. We argue that sensitivity to the irrelevant speech effect is to be investigated in order to better understand the strategies at the level of individual self-regulation and self-control, and thus the coping mechanisms against cognitive load imposition (Blandón-Gitlin et al., 2014; Gombos, 2006). Moreover, factors such as motivation to provide the expected effort and working memory capacity to achieve the task independently from the auditory distraction should also be taken into account (Carver & Scheier, 2012).

More research on keystroke dynamics applied to eyewitness narratives and deception detection is necessary. For instance, the evaluation of idiosyncratic knowledge and use of computers, and how this impacts one's ability to type truthful and deceptive accounts, would be crucial. Indeed, one may expect skilled typers to experience less cognitive load than typers who are not familiar with a computer (e.g., Doubé & Beh, 2012). We argue that factors such as age, familiarity with keyboard typing, and education should be further investigated in the future. Moreover, the accuracy of memories should also be investigated. Studies on false memories show that narratives may be very vivid and clear despite being inaccurate (e.g., Lampinen, Ryals, & Smith, 2008). The vividness of these memories might also influence the typing prosody, and pass for truthful narratives while being absolutely inaccurate. This might have important practical implications if one were to rely on keystroke dynamics to assess credibility.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

ETHICS STATEMENT

All procedures performed in studies involving human participants were in accordance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

ENDNOTES

[1] We wish to thank the reviewer who made this remark, and allowed us to improve the quality of our manuscript.

[2] The software can requested and hand-coded on demand from Ioannis Tsimperidis.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available at https://doi.org/10.17605/OSF.IO/M3RQ4.

ORCID

Ioannis Tsimperidis https://orcid.org/0000-0003-0682-1750
Samuel Demarchi https://orcid.org/0000-0003-1985-4754

REFERENCES

Aamodt, M. G., & Custer, H. (2006). Who can best catch a liar? *Forensic Examiner*, *15*(1), 6–11.

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1–29.

Baddeley, A., & Salame, P. (1986). The unattended speech effect: Perception or memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(4), 525–529.

Banerjee, R., Feng, S., Kang, J. S., & Choi, Y. (2014). Keystroke patterns as prosody in digital writings: A case study with deceptive reviews and essays. In A. Moschitti, B. Pang, & W. Daelemans (Eds.), *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1469–1473). Doha, Qatar: Association for Computational Linguistics.

Bhatt, S., & Santhanam, T. Salem, Tamilnadu, India, IEE (2013). Keystroke dynamics for biometric authentication—A survey. *2013 international conference on pattern recognition, informatics and mobile engineering* (pp. 17–23). https://doi.org/10.1109/ICPRIME.2013.6496441.

Blandón-Gitlin, I., Fenn, E., Masip, J., & Yoo, A. H. (2014). Cognitive-load approaches to detect deception: Searching for cognitive mechanisms. *Trends in Cognitive Sciences*, *18*(9), 441–444.

Bogaard, G., Meijer, E. H., & Vrij, A. (2014). Using an example statement increases information but does not increase accuracy of CBCA, RM, and SCAN. *Journal of Investigative Psychology and Offender Profiling*, *11* (2), 151–163.

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. https://doi.org/10.1002/acp.1087

Brizan, D. G., Goodkind, A., Koch, P., Balagani, K., Phoha, V. V., & Rosenberg, A. (2015). Utilizing linguistically enhanced keystroke dynamics to predict typist cognition and demographics. *International Journal of Human-Computer Studies*, *82*, 57–68.

Buchner, A., Mehl, B., Rothermund, K., & Wentura, D. (2006). Artificially induced valence of distractor words increases the effects of irrelevant speech on serial recall. *Memory & Cognition*, *34*(5), 1055–1062.

Carver, C. S., & Scheier, M. F. (2012). *Attention and self-regulation: A control-theory approach to human behavior*. New York, NY, USA: Springer Science & Business Media.

Chenoweth, N. A., & Hayes, J. R. (2003). The inner voice in writing. *Written Communication*, *20*(1), 99–118.

Cohen, J. E. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Crawford, H. (2010). Keystroke dynamics: Characteristics and opportunities. In *2010 eighth annual international conference on privacy, security and trust (PST 2010)* (pp. 205–212). New York, NY, USA: Institute of Electrical and Electronics Engineers. Retrieved from 10.1109/PST.2010.5593258

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. https://doi.org/10.1037/0033-2909.129.1.74

Derrick, D. C., Meservy, T. O., Jenkins, J. L., Burgoon, J. K., & Nunamaker, J. F., Jr. (2013). Detecting deceptive chat-based communication using typing behavior and message cues. *ACM Transactions on Management Information Systems (TMIS)*, *4*(2), 1–21.

Doubé, W., & Beh, J. (2012). Typing over autocomplete: Cognitive load in website use by older adults. In V. Farrell, G. Farrell, C. Chua, W. Huang, R. Vasa, & C. Woodward (Eds.), *Proceedings of the 24th Australian computer–human interaction conference* (pp. 97–106). New York, NY: Association for Computer Machinery.

Driskell, T., & Driskell, J. E. (2019). Got theory? Multitasking, cognitive load, and deception. In T. Docan-Morgan (Ed.), *The Palgrave handbook of deceptive communication* (pp. 145–165). Cham, Switzerland: Springer International Publishing AG.

Elliott, E. M., & Briganti, A. M. (2012). Investigating the role of attentional resources in the irrelevant speech effect. *Acta Psychologica*, *140*(1), 64–74.

Enmarker, I. (2004). The effects of meaningful irrelevant speech and road traffic noise on teachers' attention, episodic and semantic memory. *Scandinavian Journal of Psychology*, *45*(5), 393–405.

Feeley, T. H., deTurck, M. A., & Young, M. J. (1995). Baseline familiarity in lie detection. *Communication Research Reports*, *12*(2), 160–169.

Fisher, R. P., & Geiselman, R. E. (2010). The cognitive interview method of conducting police interviews: Eliciting extensive information and promoting therapeutic jurisprudence. *International Journal of Law and Psychiatry*, *33*(5–6), 321–328.

Gombos, V. A. (2006). The cognition of deception: The role of executive processes in producing lies. *Genetic, Social, and General Psychology Monographs*, *132*(3), 197–214.

Grice, P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Hauch, V., Blandon-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.

Hayes, J. R., & Chenoweth, N. A. (2006). Is working memory involved in the transcribing and editing of texts? *Written Communication*, *23*(2), 135–149.

Kellogg, R. T., Whiteford, A. P., Turner, C. E., Cahill, M., & Mertens, A. (2013). Working memory in written composition: An evaluation of the 1996 model. *Journal of Writing Research*, *5*(2), 159–190.

Knez, I., & Hygge, S. (2002). Irrelevant speech and indoor lighting: Effects on cognitive performance and self-reported affect. *Applied Cognitive Psychology*, *16*(6), 709–718.

Lampinen, J. M., Ryals, D. B., & Smith, K. (2008). Compelling untruths: The effect of retention interval on content borrowing and vivid false memories. *Memory*, *16*(2), 149–156.

Levy, C. M., & Marek, P. (1999). Testing components of Kellogg's multicomponent model of working memory in writing: The role of the phonological loop. In M. Torrance & G. C. Jeffery (Eds.), *The cognitive demands of writing: Processing capacity and working memory in text production* (pp. 25–41). Amsterdam, The Netherlands: Amsterdam University Press.

Logue, M., Book, A., Frosina, P., Huizinga, T., & Amos, S. (2015). Using reality monitoring to improve deception detection in the context of the cognitive interview for suspects. *Law and Human Behavior*, *39*(4), 360–367.

Markowitz, D. M. (2020). The deception faucet: A metaphor to conceptualize deception and its detection. *New Ideas in Psychology*, *59*, 100816. https://doi.org/10.1016/j.newideapsych.2020.100816

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. https://doi.org/10.1111/j.1467-9450.2011.00931.x

McCornack, S. A., Morrison, K., Paik, J. E., Wisner, A. M., & Zhu, X. (2014). Information manipulation theory 2: A propositional theory of deceptive discourse production. *Journal of Language and Social Psychology*, *33*(4), 348–377.

McKinley, J., Dempster, M., & Gormley, G. J. (2015). 'Sorry, I meant the patient's left side': Impact of distraction on left–right discrimination. *Medical Education*, *49*(4), 427–435.

Mihalcea, R., & Strapparava, C. (2009). The lie detector: Explorations in the automatic recognition of deceptive language. In *Proceedings of the ACL-IJCNLP 2009 conference short papers* (pp. 309–312). Suntec, Singapore: Association for Computational Linguistics.

Monaro, M., Businaro, M., Spolaor, R., Li, Q. Q., Conti, M., Gamberini, L., & Sartori, G. (2019). The online identity detection via keyboard dynamics. In K. Arai, R. Bhatia, & S. Kapoor (Eds.), *Proceedings of the future technologies conference 2018* (pp. 342–357). Cham, Switzerland: Springer.

Monaro, M., Fugazza, F. I., Gamberini, L., & Sartori, G. (2017). How human–mouse interaction can accurately detect faked responses about identity. In L. Gamberini, A. Spagnolli, G. Jacucci, B. Blankertz, & J. Freeman (Eds.), *Symbiotic interaction. Symbiotic 2016. Lecture notes in computer science* (Vol. 9961, pp. 115–124). Cham, Switzerland: Springer.

Monaro, M., Galante, C., Spolaor, R., Li, Q. Q., Gamberini, L., Conti, M., & Sartori, G. (2018). Covert lie detection using keyboard dynamics. *Scientific Reports*, *8*(1), 1976–1985. https://doi.org/10.1038/s41598-018-20462-6

Monaro, M., Gamberini, L., & Sartori, G. (2017a). Identity verification using a kinematic memory detection technique. In K. Hale & K. Stanney

(Eds.), *Advances in neuroergonomics and cognitive engineering* (pp. 123–132). Cham, Switzerland: Springer.

Monaro, M., Gamberini, L., & Sartori, G. (2017b). The detection of faked identity using unexpected questions and mouse dynamics. *PLoS One*, *12*(5), e0177851. https://doi.org/10.1371/journal.pone.0177851

Monaro, M., Spolaor, R., Li, Q., Conti, M., Gamberini, L., & Sartori, G. (2017). Type me the truth!: Detecting deceitful users via keystroke dynamics. In *Proceedings of the 12th international conference on availability, reliability and security*. 60 (pp. 1–6)New York, NY: ACM.

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.

Oswald, C. J., Tremblay, S., & Jones, D. M. (2000). Disruption of comprehension by the meaning of irrelevant sound. *Memory*, *8*(5), 345–350.

Ott, M., Choi, Y., Cardie, C., & Hancock, J. T. (2011). Finding deceptive opinion spam by any stretch of the imagination. In D. Lin (Ed.), *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies* (Vol. 1, pp. 309–319). Stroudsburg, PA: Association for Computational Linguistics.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiryand word count: LIWC2001*. Mahwah, NJ: Erlbaum.

Ramezan, C. A., Warner, T. A., & Maxwell, A. E. (2019). Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification. *Remote Sensing*, *11*(2), 185.

Röer, J. P., Bell, R., & Buchner, A. (2015). Specific foreknowledge reduces auditory distraction by irrelevant speech. *Journal of Experimental Psychology*, *41*(3), 377–391.

Rybnik, M., Tabedzki, M., & Saeed, K. New York, NY, USA: IEE (2008). A keystroke dynamics based system for user identification. In *2008 7th computer information systems and industrial management applications* (pp. 225–230).

Sartori, G., Zangrossi, A., & Monaro, M. (2018). Deception detection with behavioral methods. In J. P. Rosenfeld (Ed.), *Detecting concealed information and deception* (pp. 215–241). San Diego, CA: Academic Press.

Schilperoord, J. (2002). On the cognitive status of pauses in discourse production. In T. Olive & M. Levy (Eds.), *Studies in writing: Volume 10: Contemporary tools and techniques for studying writing* (pp. 61–87). Doordrecht, The Netherlands: Springer.

Sörqvist, P., Nöstl, A., & Halin, N. (2012). Disruption of writing processes by the semanticity of background speech. *Scandinavian Journal of Psychology*, *53*(2), 97–102.

Sporer, S. L. (2016). Deception and cognitive load: Expanding our horizon with a working memory model. *Frontiers in Psychology*, *7*, 1–12. https://www.frontiersin.org/articles/10.3389/fpsyg.2016.00420/full

Suchotzki, K., Verschuere, B., Van Bockstaele, B., Ben-Shakhar, G., & Crombez, G. (2017). Lying takes time: A meta-analysis on reaction time measures of deception. *Psychological Bulletin*, *143*(4), 428–453. https://doi.org/10.1037/bul0000087

Tremblay, S., Nicholls, A. P., Alford, D., & Jones, D. M. (2000). The irrelevant sound effect: Does speech play a special role? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1750–1754.

Tsimperidis, I., Arampatzis, A., & Karakos, A. (2018). Keystroke dynamics features for gender recognition. *Digital Investigation*, *24*, 4–10.

Van Gerven, P. W., Meijer, W. A., Vermeeren, A., Vuurman, E. F., & Jolles, J. (2007). The irrelevant speech effect and the level of interference in aging. *Experimental Aging Research*, *33*(3), 323–339.

Vizer, L. M., Zhou, L., & Sears, A. (2009). Automated stress detection using keystroke and linguistic features: An exploratory study. *International Journal of Human-Computer Studies*, *67*(10), 870–886.

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.

Walczyk, J., Igou, F., Dixon, L., & Tcholakian, T. (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* 1–14. . Retrieved from. https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00014/full

Walczyk, J. J., Harris, L. L., Duck, T. K., & Mulay, D. (2014). A social-cognitive framework for understanding serious lies: Activation-decision-construction-action theory. *New Ideas in Psychology*, *34*, 22–36.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, *50*(3), 449–455.

Zahid, S., Shahzad, M., Khayam, S. A., & Farooq, M.. Heidelberg, Germany: Springer-Verlag; (2009). Keystroke-based user identification on smart phones. In *International workshop on recent advances in intrusion detection* (pp. 224–243).

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Tomas F, Tsimperidis I, Demarchi S, El Massioui F. Keyboard dynamics discrepancies between baseline and deceptive eyewitness narratives. *Appl Cognit Psychol*. 2021;35:112–122. https://doi.org/10.1002/acp.3743